COVID-19 Genomics UK (COG-UK) Consortium Report - 12th March 2021

Impact of travel restrictions on importations to England from May-Sept 2020

This report includes information on the ongoing state of the research being carried out. It should not be considered formal or informal advice. The conclusions of the ongoing scientific studies may be subject to change as further evidence becomes available and as such any firm conclusions may be premature.

Executive Summary

● Genomic epidemiology is a powerful tool for tracking importation and transmission of SARS-CoV-2 and assessing the effectiveness of public health measures, as demonstrated by previous COG-UK analyses for Scotland and Wales.

● A new genomic epidemiology analysis of travel related cases of SARS-CoV-2 infection in England between the 27th May and 13th of September 2020 assessed the effectiveness of travel restrictions (closed travel-corridors and home quarantine vs open travel corridors and no quarantine) imposed on entering and leaving the country during this period.

● Travel restrictions were effective in reducing the number of contacts for imported index cases and the number of subsequent cases owing to onward transmission, although not completely eliminating either [Strong].

● Age had a significant effect on the number of contacts, with the 16-20 year old age group representing the greatest number of travel related cases. This effect was moderated by travel restrictions. [Moderate]

● Combined, COG-UK analyses for Scotland, Wales and Englend suggest that there was a substantial reduction in cases and circulating lineages associated with the first lockdown, that travel was associated with increased importation and transmission following the easing of lockdown, and that where imposed, travel restrictions were effective in reducing transmission events.

An integrated analysis of contact tracing and genomics to assess the efficacy of travel restrictions on SARS-CoV-2 introduction and transmission in England from June to September, 2020

*Dinesh Aggarwal[1,2,3,4], Andrew J. Page[5,^], Ulf Schaefer[2,^], George M. Savva[5], Richard Myers[2], Erik Volz[6], Nicholas Ellaby[2], Steven Platt[2], Natalie Groves[2], Eileen Gallaghar[2], Niamh M. Tumelty[7], Thanh Le Viet[5], Gareth J. Hughes[8], Cong Chen[2], Charlie Turner[2], Sophie Logan[9], Abbie Harrison[2], The COVID-19 Genomics UK (COG-UK) Consortium[10,#], Sharon J. Peacock[1,2,3,4], Meera Chand[2^^,] *Ewan M. Harrison[1,2,4,11^^]

Affiliations
1. University of Cambridge, Department of Medicine, Cambridge, UK.
2. Public Health England, 61 Colindale Ave, London, NW9 5EQ, UK.
3. Cambridge University Hospital NHS Foundation Trust, Cambridge, UK.
4. Wellcome Sanger Institute, Hinxton, Cambridge, UK.
5. Quadram Institute Bioscience, Norwich Research Park, Norwich, NR4 7UQ, UK.
6. Imperial College London, Department of Infectious Disease Epidemiology, London, UK.
7. University of Cambridge, Cambridge University Libraries, Cambridge, UK
8. Public Health England National Infections Service, Field Service, Leeds, UK.
9. Public Health England, National Infections Service, Field Service, Nottingham, UK.
10. https://www.cogconsortium.uk
11. University of Cambridge, Department of Public Health and Primary Care, Cambridge, UK.

[#] Full list of consortium names and affiliations are in the appendix
^ Contributed equally
^^ Contributed equally (joint senior authors)

* corresponding authors:
Dr Dinesh Aggarwal, University of Cambridge, Department of Medicine, Cambridge, UK. Email: dinesh.aggarwal@nhs.net

Dr Ewan Harrison, Wellcome Sanger Institute, Hinxton, Cambridge, UK. Email: eh6@sanger.ac.uk

Abstract

Background: Mitigation of SARS-CoV-2 transmission from international travel is a priority. Travellers from countries with travel restrictions (closed travel-corridors) were required to quarantine for 14 days over Summer 2020 in England. We describe the genomic epidemiology of travel-related cases in England and evaluate the effectiveness of this travel policy.

Methods: Between 27/05/2020 and 13/09/2020, probable travel-related SARS-CoV-2 cases and their contacts were identified and combined with UK SARS-CoV-2 sequencing data. The epidemiology and demographics of cases was identified, and the number of contacts per case modelled using negative binomial regression to estimate the effect of travel restriction, and any variation by age, sex and calendar date. Unique travel-related SARS-CoV-2 genomes in the COG-UK dataset were identified to estimate the effect travel restrictions on cluster size generated from these. The Polecat Clustering Tool was used to identify a travel-related SARS-CoV-2 cluster of infection.

Findings: 4,207 travel-related SARS-CoV-2 cases are identified. 51.2% (2155/4207) of cases reported travel to one of three countries; 21.0% (882) Greece, 16.3% (685) Croatia and 14.0% (589) Spain. Median number of contacts per case was 3 (IQR 1-5), and greatest for the 16-20 age-group (9.0, 95% C.I.=5.6-14.5), which saw the largest attenuation by travel restriction. Travel restriction was associated with a 40% (rate ratio=0.60, 95% C.I.=0.37-0.95) lower rate of contacts. 827/4207 (19.7%) of cases had high-quality SARS-CoV-2 genomes available. Fewer genomically-linked cases were observed for index cases related to countries with travel restrictions compared to cases from non-travel restriction countries (rate ratio=0.17, 95% C.I.=0.05-0.52). A large travel-related cluster dispersed across England is identified through genomics, confirmed with contact-tracing data.

Interpretation: This study demonstrates the efficacy of travel restriction policy in reducing the onward transmission of imported cases.

Funding: Wellcome Trust, Biotechnology and Biological Sciences Research Council, UK Research & Innovation, National Institute of Health Research, Wellcome Sanger Institute.

RESEARCH IN CONTEXT

Evidence before this study

We searched PubMed, medRxiv, bioRxiv, Web of Science and Scopus for the terms (COVID-19 OR SARS-COV-2) AND (imported or importation) AND (sequenc* OR genom* or WGS). We filtered the 55 articles identified through this search and rejected any that did not undertake SARS-CoV-2 sequencing as part of an epidemiological investigation for importation into a different country. The remaining 20 papers were reviewed in greater detail to understand the patterns of importation and the methods used in each case.

Added value of this study

This is the first published study on importations of SARS-CoV-2 into England using genomics. Plessis et al., (2021) used a predictive model to infer the number of importations in to the UK from all SARS-CoV-2 genomes generated before 26th June 2020. The current study assesses the period 27/05/2020 to 13/09/2020 and presents findings of case-reported travel linked to genomic data. Two unpublished reports exist for Wales and Scotland, although only examine a comparatively small number of importations.

Implications of all the available evidence

This large-scale study has a number of findings that are pertinent to public health and of global significance, not available from prior evidence to our knowledge. The study demonstrates travel restrictions, through the implementation of ‘travel-corridors’, are effective in reducing the number of contacts per case based on observational data. Age has a significant effect on the number of contacts and this can be mitigated with travel restrictions. Analysis of divergent clusters indicates travel restrictions can reduce the number of onwards cases following a travel-associated case. Analysis of divergent clusters can allow for importations to be identified from genomics, as subsequently evidenced by cluster characteristics derived from contact tracing. The majority of importations of SARS-CoV-2 in England over Summer 2020 were from coastal European countries. The highest number of cases and onward contacts were from Greece, which was largely exempt from self-isolation requirements (bar some islands in September at the end of the study period). Systematic monitoring of imported SARS-CoV-2 cases would help refine implementation of travel restrictions. Finally, along with multiple studies, this study highlights the use of genomics to monitor and track importations of SARS-CoV-2 mutations of interest; this will be of particular use as the repertoire of clinically relevant SARS-CoV-2 variants expand over time and globally.

Introduction

A new coronavirus related disease (COVID-19) was first reported in Wuhan, China (2) in December 2019, with the causative virus identified as a novel coronavirus SARS-CoV-2 (3). Since then, SARS-CoV-2 has been imported into virtually every country and region in the world. Understanding and tracking the sources of importations can give important information for policy makers, and for managing the pandemic, by informing policies aimed at reducing the further spread of virus.

Public health measures can help mitigate and suppress the spread of the virus, but the threat of importations will remain. The available brakes on imported SARS-CoV-2 cases include travel bans, quarantine measures, and testing of returning travellers. These can apply to all countries or targeted to high-risk countries, for variable durations, and with variable degree of enforcement. In England, travel restrictions were assigned on a country by country basis from 6 July 2020 entailing the use of 'travel-corridors'; travellers returning from countries that were on the travel restrictions list (with 'closed travel-corridors') (4) were required to quarantine for 14 days (reduced to 10 days on 15/12/20), or from the 15th December 2020, choose to quarantine for 5 days and then pay for a SARS-CoV-2 diagnostic test (Figure 1). This policy aims to limit onwards transmission of SARS-CoV-2, and as a secondary outcome possibly deter travel to those countries. Upon identification of an imported case, contact tracing and quarantine/self-isolation measures can limit onwards transmission. The CORSAIR study reported that 18.2% of individuals adhered to general SARS-CoV-2 self-isolation guidance recommended by Public Health England in the UK (5). The PHE Isolation Assurance Service however have identified up to 97% self-reported compliance with travel-specific self-isolation guidance (6). These data do not include countries exempt from quarantine, contact-tracing data or link to genomic data to evaluate travel-related clusters.

Studies from numerous countries have used genome sequencing to complement epidemiological investigations in order to characterise importations of SARS-CoV-2 (Supplementary Table 1). Primarily these are in-depth case reports on small datasets but demonstrate the utility of genomics combined with contact tracing. Genomic sequencing of returning travellers was useful in allowing for the first case of reinfection of SARS-CoV-2 in the world to be identified in Hong Kong (7) and identify a new variant of SARS-CoV-2 (B.1.177/20A.EU1, variant A222V) (8).

This study combines contact-tracing data from National Health Service (NHS) Test and Trace (T&T) for probable importation cases with genomic data made available through the

COVID-19 Genomics UK (COG-UK) consortium (9), which receives samples from NHS hospital diagnostic labs and mass community testing labs (UK Lighthouse labs network) across the UK. We aimed to characterise the known imported cases and the effectiveness of travel restrictions on onwards transmission.

A total of 4,207 SARS-CoV-2 positive importation cases were analysed, along with 18,856 contacts, of which 888 sequenced genomes were available for comparison to all UK genomic data (131,387 sequences from the UK and in the COG-UK dataset by 5 December 2020). The number of contacts reported by a case was used as an indicator of adherence to quarantining.

Methods and materials

Contact tracing and case identification

Contact-tracing data was obtained from T&T. All cases and contacts had a field for demographic data, but this was not always reported (Table 2 and Supplementary Table 3). 'Highly probable' travel-related cases were defined as individuals who reported international travel as an activity in the two days before symptom onset/testing. On 12/08/2020 the additional facility to report international travel in the seven days prior to symptom onset/testing became available, and also included in this study and defined as 'probable' travel-related cases.

Cases were asked to provide details of all contacts for activities in the 2 days prior to onset/testing up to completing the system which were gathered. If any contacts become cases they would then also be included in T&T data as a case separately, but if they did not report direct travel themselves, then they would not meet the definition for a travel-associated case.

Case identification from T&T data

Data included free-text destination city or country. A freetext country and city search with a custom python script on travel-related T&T was used to identify destination country. Results and remaining entries were manually checked and corrected (see Supplementary methods for more details).

Clinical samples, Genome sequencing and Quality Control

Clinical samples were collected passively as part of national SARS-CoV-2 testing. This included both community testing through lighthouse labs and testing through hospital

diagnostic labs. Samples were sequenced at one of seventeen COG-UK sequencing sites (Figure 1). The samples were prepared for sequencing using either the ARTIC (10) or veSeq (11) protocols, and were sequenced using Illumina or Oxford Nanopore platforms. All samples were uploaded to and processed through COVID-CLIMB pipelines (12,13). Genomes were aligned to the Wuhan Hu-1 reference genome (MN908947.3). Genomes which contained more than 10% missing data were excluded from further analysis to ensure high quality phylogenetic analysis.

Lineages and minor variants

Global and UK Lineages (14) were assigned to each genome using Pangolin (https://github.com/cov-lineages/pangolin) with analysis performed on COVID-CLIMB (13). Minor variants were pre-defined within the COG-UK database using type_variants (https://github.com/cov-ert/type_variants).

Identification of extinct and unique genomes

The 827 high-quality travel-related genomes were compared to the COG-UK dataset on 16/10/2020. Genomes were only compared to other genomes with the same UK lineage assigned by COG-UK, since we assume that no relatedness relevant to transmission exists between genomes of different UK lineages. A ‘unique’ genome in the community was deemed to be one that was known to be from a travel-related case and either: (1) A UK lineage that had not been sampled in the previous 4 weeks in the UK, (2) >3 SNPs distance to the closest relative in the COG-UK dataset.

Within the same UK lineage we identified those genomes sampled within 4 weeks prior to the genome of interest. We determined the minimum SNP distance between the sequence of interest and these genomes. ‘Unique’ genomes were compared to sequences that were generated in the COG-UK dataset within 2 and 4 weeks after their sampling date, to identify samples with the same UK lineage and within 2 SNPs. These would represent onward transmission or further introductions of similar genomes. The analysis was run with an in-house custom Python script developed by US and RM. Further detail in supplementary methods.

Identification of a travel-related SARS-CoV-2 cluster

We used the Polecat clustering tool (https://cog-uk.github.io/polecat) to systematically identify outliers in COG-UK genomic dataset and link to contact-tracing data.

Statistical analysis

All models were estimated using the glmmTMB package (version 1.0.1) (15) with marginal means and effects calculated using the emmeans package (1.5.2-1) (16) for R (version 3.5.1) (17). Figures were generated using R (version 4.0.2) and Microsoft Excel (version 1908). The number of contacts per case was modelled using negative binomial regression analysis, to estimate the effect of travel restriction, and whether this varied by age-group, sex of the index case and calendar date. Travel destination and ethnic group were included as covariates (as random effects). A similar approach was taken when estimating the effect of travel restriction on genomic cluster size.

Role of the funding source

The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

From 17/03/20 – 04/07/20 the Foreign & Commonwealth Office advised against all non-essential travel worldwide (18). From the 04/07/20 – 01/02/21 travel corridors to countries deemed to be low risk for COVID-19 disease (subject to assessment and change) were established in which returning travellers were no-longer required to quarantine. Persons returning from countries outside this list (except for exemptions e.g. specific employment) were required to quarantine. We sought to both gauge the impact of this policy and to attempt to quantify the numbers of onward transmissions using genomic epidemiology.

Between 27/05/2020 and 13/09/2020, using contact-tracing data for cases who have tested positive for SARS-CoV-2, we identified 4,207 travel-related cases (Figure 1). Supplementary tables 2 and 3 show the case characteristics.

Travel to European countries accounted for 85.9% (3612/4207) of cases, of which 51.2% (2155/4207) had visited one of Greece (21.0%, 882/4207), Croatia (16.3%, 685/4207) and Spain (14.0%,589/4207) (Figure 2 and Table 2). For 284 cases the country of travel was unclear or unknown. Travel restrictions were first eased on 03/07/2020; 2.9% of travel-related cases were recorded before this date. For the countries associated with the highest numbers of imports, the duration of the peak of imported cases differs, with variable association with changes in travel restriction policy (Figure 3). Geographically variations in imported cases across England were apparent, with the greatest number in Greater London (28.6%, 1205/4207) (Figure 2, Supplementary Figures 1 and 2, and Supplementary Table 3).

The median number of reported contacts per travel-associated case was 3 (IQR 1-5), with a maximum of 172. Overall, travel restriction reduced the number of contacts per case by 40% (rate ratio (R.R.)=0.60, 95% C.I.=0.37-0.95). The mean number of contacts (adjusting and averaging over all over covariates) was 5.85 when no travel restriction was in place and 3.50 when there was. The effect of travel restriction varied significantly with age-group and over time (Supplementary Table 4 and Figure 4). The number of contacts per case was greatest for the 16-20 age-group without travel restriction with a marginal mean of 9.0 (95% C.I.=5.6-14.5) but with restriction reduced to 4.7 (95% C.I.=3.9-5.7), and similar to other age-groups. After adjusting for all other covariates the numbers of contacts per imported case was roughly half in September compared to May, June and July, whether or not a travel restriction was in place.

Transmission patterns identified by analysis of traveller SARS-CoV-2 genomes

We next sought to quantify onward transmission from an imported case using genomics. High-quality sequencing data was available for 827/4207 (19.7%) of cases (Figure 1) and demographics of the sequenced cases was broadly similar to the entire travel-related cohort (Supplementary Table 3).

186/827 (22.4%) imported cases had viral lineages that were sufficiently unique in the COG-UK dataset to monitor onward spread. Of these, 146/186 isolates had not been sampled in the entire UK dataset in the 4 weeks prior and 40/186 isolates were >3 SNPs to their closest matching sequence in the UK dataset.

To compare the effect of travel restrictions on the subsequent spread of likely imported cases (excluding 18/186 cases before 14/07/2020 to ensure the dates of cases with and without a travel restriction overlapped), the entire COG-UK dataset was interrogated to identify isolates within 2 SNPs of these distinct imported cases during the period 0-2 and 0-4 weeks following the importation case. The number of subsequent cases detected during the four weeks since the unique index case increased from a mean of 1.2 new cases where a travel restriction was in place to 11.3 cases where there was not. The proportions of cases leading to a subsequent newly detected case (e.g. likely transmission), and the number of new cases where at least one is detected are shown in Figure 5. Overall, 56/168 of genomes from cases that were genetically unique were detected in subsequent cases (up to four weeks later). Among cases diagnosed after returning from a country where a travel restriction was in place, 25% of (20/81) were detected in later cases (up to four weeks later), rising to 41% (29/71) when cases were imported from a country without a travel restriction. Destination country for 16 index cases was unknown. There was a high variation in the

number of subsequent cases matching each genome (range 1 to 210, IQR 0-4) with a small number of imported cases corresponding to large numbers of subsequent cases (Figure 5).

There was some evidence that imported cases with higher numbers of contacts gave rise to more cases in the subsequent month (Figure 5). Although the number of cases with any subsequent matching genome was not affected by the number of contacts, the average number of subsequent cases detected was substantially higher when the index case had five or more contacts (mean=10.8) compared to none (mean=2.8) or 1 to 4 contacts (mean=1.7).

To estimate the effect of travel restriction on spread, considering possible confounding effects of calendar date and the mediating effect of reported contacts of the index case, and to test the statistical significance of observed effects, a series of negative binomial regression models were fitted (Figure 5 and Supplementary Table 5). In the four weeks following the index case, fewer genomically-linked cases were reported when the index case was imported from a country with travel restrictions compared to cases from a non-travel restriction country (R.R.=0.17, 95% CI=0.05-0.52). When the number of contacts of the index case was included in the model, this rate ratio was attenuated slightly toward 1 (R.R.=0.25; 0.08-0.81) suggesting a limited mediating effect of the number of contacts of the index case. The effect of contacts was still seen, but the rate of subsequent cases (over four weeks) with the same genome was 4.0 (1.1-15.1) times higher for index cases with five or more reported contacts compared to those with none.

Genomic identification of a large imported cluster

The Polecat Clustering tool (https://cog-uk.github.io/polecat) was used to analyse genomes in UK data on 14 September 2020. An outlier cluster was observed (Supplementary Figure 3). This cluster (UK1897) was associated with high diversity with a long stem length compared to samples from the UK, suggesting that this lineage evolved outside the UK. The geographic distribution of this lineage is demonstrated in Supplementary Figure 4, likely representing multiple importations across the UK (Supplementary Figure 4). This cluster contained the D614G mutation but no others associated with increased transmission. The root of the cluster was associated with a Swiss phylotype when linked to data in GISAID. During the course of the study period (04/08/2020 to 14/09/2020) there were 304 genomes. These were linked to 238 individuals, of whom 159 could be linked to a contact-tracing record. 143/159 had contact-tracing information indicating international travel or not. 72/143 (50.3%) individuals were linked to international travel and associated with ten,

dispersed European countries (four individuals had travel to more than one European Country) and most commonly Croatia (35/72, 48.6%) (Supplementary Figure 5). A further 4 cases were identified as contacts of individuals who had reported travel to mainland Europe. There is a trend towards an increased proportion of cases that do not report travel over time, and possibly representing dispersion and onwards transmission locally of this lineage (Supplementary Figure 6).

Lineage diversity of imported SARS-CoV-2 cases

The 827 imported genomes reflected 238 UK lineages (see Supplementary Materials), of which 214 were seen less than 5 times (142 singletons) and 24 were seen 5 or more times (Supplementary Table 6). The most commonly observed were UK5 (152 genomes, 18.4%) and UK1897 (73 genomes, 8.8%). There were 39 global lineages within the genomes. The most commonly observed lineages were B.1.1 (159 genomes, 19.2%) and B.1.177 (128 genomes, 15.5%) (Supplementary Tables 7 and 8). Further, potentially functionally important mutations were identified (Supplementary Table 9 and Supplementary Figure 8): D614G, 824/827 (99.6%) cases; N439K, 65/827 (7.86%) of cases; A222V, 131/827 (15.84%) of cases. ΔH69/V70 was identified in 53 cases associated with lineage B.1.258. We evaluated the introduction of A222V (B.1.177) over time, demonstrating a clear epidemiological link to Spain through contact tracing (Supplementary Figure 9). By the end of the study period, this variant was introduced from 16 separate countries indicating dispersion across Europe (Supplementary Figures 10). The mutations co-occur, with the proportion of cases represented by these combinations varying over time (Supplementary Figure 11).

Discussion

We demonstrate, through the analysis of both contact-tracing data and the use of genomics, that travel restrictions (use of travel corridors) reduced the detected linked cases of SARS-COV-2. From 27/05/2020 to 13/09/2020, 85.9%% of importations were from European countries with three countries accounting for 51.2% of all imported cases. Along with travel restriction, age was a significant determinant of onwards contacts, and this effect was mitigated with closing travel corridors. After a period of national lockdown, systematic monitoring of imported genomes can identify sequences that are sufficiently unique and provide utility for monitoring of onwards transmission.

Whilst the study period covers nearly 5 months, the importations are concentrated after the implementation of travel corridors. The peaks for imports for each country occur at different

times and with different epidemic curves. For the most common destination, barring Spain, imported cases appear to reduce after country-specific travel restrictions. Importations from Greece came at the end of August and continued into September, with the steepest of all curves. No travel restrictions were imposed on Greece during this time period and it was the source of greatest imported SARS-CoV-2 cases during this study period. This highlights the need for active surveillance of imported cases of SARS-CoV-2 for the introduction of travel corridors in a timely manner. London accounts for 15.4% of the population in England (19), but had 28.6% of the imports, possibly reflecting a younger age demographic. The overall R0 remained largely similar to other parts of the country during the study period potentially indicating imports are unlikely to have had a substantial impact on onward infection rates (20). Other explanations may include a small possible effect of higher seropositivity rates in London (17.5%, 27/4/2020, (21)) from the first wave of SARS-CoV-2 infections in England seen and a potential lower detection rate in London.

The number of onwards contacts are significantly reduced by the introduction of travel restrictions. Age is also a significant determinant of onwards contacts, with the 16-20 year old age-group representing the greatest number of travel-related cases and onwards contacts. This identifies an opportunity to direct public health awareness campaigns to younger travellers, with the intention to promote behaviours that will reduce the risk of SARS-CoV-2 acquisition and enhance compliance with quarantine on return to the UK.

The use of genomic sequencing, specifically after a period of national lockdown, allowed identification of a cohort of unique genomes that could be monitored for cluster growth. The cluster size for genomes that were related to a country without travel restrictions was significantly higher than those related to countries under travel restriction guidance. Further, when comparing the number of genomes in a cluster to the number of contacts that their respective cases reported, there was a trend towards a positive correlation suggesting self-isolation is effective. The total effect of travel restrictions was not explained by forward contacts alone and it is a possible that a reduction in the absolute number of individuals travelling to countries with travel restrictions also contributes to this.

The Polecat Clustering Tool highlighted a cluster that developed largely through travel to Croatia. Programmatic analysis of genomics data can therefore identify putative importation clusters. Integration with contact-tracing information was vital for the true picture of the sources of introduction and the subsequent spread, due to the SARS-CoV-2 sequencing bias observed globally. In this instance an introduced lineage was associated with wide-spread dispersal and onward transmission during a period when England had limited social

distancing measures. The lineage, B.1.160, associated with this cluster is not associated with increased transmissibility but this study highlights a supplementary method for the detection and monitoring of expanding imported clusters and could prove particularly useful for the investigation of introduced variants of concern.

Our study is subject to multiple limitations. The COG-UK dataset has a limited sequencing coverage across England and cluster sizes detected will under-estimate absolute numbers (Supplementary Figure 12). The dates of country-specific travel restriction guidance was aligned with the date of travel-related case sampling, the earliest date reliably available. The effect of this should not however markedly affect results or conclusions; the period of travel restrictions are long and the effect size seen is large and therefore this discrepancy is unlikely to account for the significant difference observed. Further, most countries accounting for the imported SARS-CoV-2 cases went into a period of a travel restriction over the study period; by using a date later than the date of return from travel, we are more likely to over-represent contacts for countries under 'travel-restriction' guidance. Our study evaluates a period of time following a national lockdown and the associated reduced travel would likely exaggerate the diversity of genomes when compared with the COG-UK dataset. Outcomes such as travel and the number of contacts are self-reported by cases which will have inherent biases. Finally, there will be an artificial reduction in cases at the end of the study period when accounting for case incubation period, testing and report, with data provided 3 days after study close.

Conclusions

We present an integrated epidemiological and genomic evaluation of the largest dataset of confirmed SARS-CoV-2 imported cases into the UK (or any other country) to our knowledge. We demonstrate the efficacy of closing 'travel-corridors' in reducing onward transmission of imported cases, and highlight the importance for targeted public health campaigns to reduce SARS-CoV-2 importations and onwards transmission. Our data demonstrates how routine genomic monitoring of travel-related cases could be used to refine travel restrictions and the genomic diversity of the SARS-CoV-2 import cases.

Funding

DA is a Clinical PhD Fellow and gratefully supported by the Wellcome Trust [Grant number: 222903/Z/21/Z]. EMH is supported by a UK Research and Innovation (UKRI) Fellowship: MR/S00291X/1. AJP, TLV, GMS gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); their research was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project BBS/E/F/000PR10352, also Quadram Institute Bioscience BBSRC funded Core Capability Grant (project number BB/CCG1860/1). The COVID-19 Genomics UK (COG-UK) Consortium is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author contributions

All authors read the manuscript and consented to its publication.
DA led the study.
DA, AJP, GS wrote the first draft of the manuscript.
All authors contributed to revision of the first draft of the manuscript.
DA, US, AJP, RM, NE undertook data analysis.
GMS provided statistical guidance and analysis.
DA, SP, NG, EG contributed to data curation.
GJH, CC, CT, SL, AH provided oversight over data acquisition and data definitions.
TLV wrote scripts to perform analysis.
NMT undertook the literature search.
DA, US, RM, EV, MC, EMH contributed to study design
MC, EMH, SJP conceived the study and provided overall leadership.
DA, MC and SJP provided clinical oversight.

Conflicts of interests and disclosures

None declared.

Ethics

This study was conducted as part of surveillance for COVID-19 infections under the auspices of Section 251 of the NHS Act 2006 and/or Regulation 3 of The Health Service (Control of Patient Information) Regulations 2002. They therefore did not require individual

patient consent or ethical approval. Public Health England affiliated authors had access to identifiable patient data. Other authors only had access to anonymised or summerised data. The COG-UK study protocol was approved by the Public Health England Research Ethics Governance Group (reference: R&D NR0195).

Acknowledgements

We thank members of the COVID-19 Genomics Consortium UK and Test and Trace contact tracers or their contributions to generating the data used in this study. We thank Sarah Mitchell from the Department of Plant Sciences, University of Cambridge, for direction on statistical analysis.

Data availability

Assembled/consensus genomes are available from GISAID (22) subject to minimum quality control criteria. Raw reads are available from European Nucleotide Archive (ENA) (23). All genomes, phylogenetic trees, basic metadata are available from the COG-UK consortium website (https://www.cogconsortium.uk/data). For confidentiality reasons, extended metadata (24) is not publicly available, however some may be available upon request from Public Health England.

References

1. Plessis L du, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science. 2021 Feb 12;371(6530):708–12.

2. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020 Feb 15;395(10223):497–506.

3. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020 Feb 20;382(8):727–33.

4. GOV UK. Coronavirus (COVID-19): travel corridors [Internet]. GOV.UK. 2020 [cited 2020 Nov 19]. Available from: https://www.gov.uk/guidance/coronavirus-covid-19-travel-corridors

5. Smith LE, Potts HW, Amlôt R, Fear NT, Michie S, Rubin GJ. Adherence to the test, trace and isolate system: results from a time series of 21 nationally representative surveys in the UK (the COVID-19 Rapid Survey of Adherence to Interventions and Responses [CORSAIR] study). medRxiv. 2020 Sep 18;2020.09.15.20191957.

6. Border Force. Data on health measures at the UK border [Internet]. GOV.UK. 2020 [cited 2021 Feb 19]. Available from: https://www.gov.uk/government/publications/data-on-health-measures-at-the-uk-border/data-on-health-measures-at-the-uk-border

7. To KK-W, Hung IF-N, Ip JD, Chu AW-H, Chan W-M, Tam AR, et al. COVID-19 re-infection by a phylogenetically distinct SARS-coronavirus-2 strain confirmed by whole genome sequencing. Clin Infect Dis [Internet]. 2020 Aug 25 [cited 2020 Sep 8]; Available from: https://academic.oup.com/cid/advance-article/doi/10.1093/cid/ciaa1275/5897019

8. Hodcroft EB, Zuber M, Nadeau S, Comas I, Candelas FG, Consortium S-S, et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. medRxiv. 2020 Oct 28;2020.10.25.20219063.

9. COG-UK. An integrated national scale SARS-CoV-2 genomic surveillance network. Lancet Microbe [Internet]. 2020 Jun 2 [cited 2020 Jun 9];0(0). Available from: https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(20)30054-9/abstract

10. Quick J. nCoV-2019 sequencing protocol v2. 2020 Apr 9 [cited 2020 Apr 25]; Available from: https://www.protocols.io/view/ncov-2019-sequencing-protocol-v2-bdp7i5rn

11. Bonsall D, Golubchik T, Cesare M de, Limbada M, Kosloff B, MacIntyre-Cockett G, et al. A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. J Clin Microbiol [Internet]. 2020 Sep 22 [cited 2021 Mar 10];58(10). Available from: https://jcm.asm.org/content/58/10/e00382-20

12. Nicholls SM, Poplawski R, Bull MJ, Underwood A, Chapman M, Abu-Dahab K, et al. MAJORA: Continuous integration supporting decentralised sequencing for SARS-CoV-2 genomic surveillance. bioRxiv. 2020 Oct 7;2020.10.06.328328.

13. Connor TR, Loman NJ, Thompson S, Smith A, Southgate J, Poplawski R, et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. Microb Genomics. 2016;2(9):e000086.

14. Rambaut A, Holmes EC, O'Toole Á, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020 Nov;5(11):1403–7.

15. Brooks ME, Kristensen K, Van Benthem KJ, Magnusson A, Berg CW, Nielsen A, et al. glmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modeling. R J. 2017;9(2):378–400.

16. Lenth R, Singmann H, Love J, Buerkner P, Herve M. Emmeans: Estimated marginal means, aka least-squares means. R Package Version. 2018;1(1):3.

17. Team RC. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna. WwwR-Proj. 2018;

18. Foreign & Commonwealth Office. Travel advice: coronavirus (COVID-19) [Internet]. GOV.UK. 2021 [cited 2021 Feb 19]. Available from: https://www.gov.uk/guidance/travel-advice-novel-coronavirus

19. Office for National Statistics, National Records of Scotland, Northern Ireland Statistics and Research Agency. 2011 Census aggegate data (Data downloaded: 1 June 2016) [Internet]. UK Data Service; 2016. Available from: https://beta.ukdataservice.ac.uk/datacatalogue/studies/study?id=7427

20. DHSC, SAGE. The R value and growth rate in the UK [Internet]. GOV.UK. 2021 [cited 2021 Feb 19]. Available from: https://www.gov.uk/guidance/the-r-number-in-the-uk

21. GOV UK. Sero-surveillance of COVID-19 [Internet]. GOV.UK. 2020 [cited 2020 Nov 19]. Available from: https://www.gov.uk/government/publications/national-covid-19-surveillance-reports/sero-surveillance-of-covid-19

22. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance. 2017 Mar 30;22(13):30494.

23. Cochrane G, Karsch-Mizrachi I, Takagi T, Sequence Database Collaboration IN. The International Nucleotide Sequence Database Collaboration. Nucleic Acids Res. 2016 Jan 4;44(D1):D48–50.

24. Griffiths EJ, Timme RE, Page AJ, Alikhan N-F, Fornika D, Maguire F, et al. The PHA4GE SARS-CoV-2 Contextual Data Specification for Open Genomic Epidemiology. 2020 Aug 9 [cited 2020 Sep 3]; Available from: https://www.preprints.org/manuscript/202008.0220/v1

25. Au CH, Chan WS, Lam HY, Ho DN, Lam SYM, Zee JST, et al. Genome Sequences of SARS-CoV-2 Strains Detected in Hong Kong. Microbiol Resour Announc. 2020 Jul;9(31):e00697-20.

26. Böhmer MM, Buchholz U, Corman VM, Hoch M, Katz K, Marosevic DV, et al. Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series. Lancet Infect Dis. 2020;20(8):920–8.

27. Cohen-Gihon I, Israeli O, Shifman O, Stein D, Achdout H, Weiss S, et al. Coding-Complete Genome Sequences of Two SARS-CoV-2 Isolates from Early Manifestations of COVID-19 in Israel. Microbiol Resour Announc. 2020 Jul;9(28):e00677-20.

28. Jesus JG de, Sacchi C, Candido D da S, Claro IM, Sales FCS, Manuli ER, et al. Importation and early local transmission of COVID-19 in Brazil, 2020. Rev Inst Med Trop Sao Paulo. 2020;62:e30.

29. do Nascimento VA, Guerra Corado A de L, do Nascimento FO, Araujo da Costa AK, Gomes Duarte DC, Bessa Luz SL, et al. Genomic and phylogenetic characterisation of an imported case of SARS-CoV-2 in Amazonas State, Brazil. Mem Inst Oswaldo Cruz. 2020;115:e200310.

30. Du P, Ding N, Li J, Zhang F, Wang Q, Chen Z, et al. Genomic surveillance of COVID-19 cases in Beijing. Nat Commun. 2020 30;11(1):5503.

31. Garcés-Ayala F, Araiza-Rodríguez A, Mendieta-Condado E, Rodríguez-Maldonado AP, Wong-Arámbula C, Landa-Flores M, et al. Full genome sequence of the first SARS-CoV-2 detected in Mexico. Arch Virol. 2020 Sep;165(9):2095–8.

32. Giandhari J, Pillay S, Wilkinson E, Tegally H, Sinayskiy I, Schuld M, et al. Early transmission of SARS-CoV-2 in South Africa: An epidemiological and phylogenetic report. Int J Infect Dis. 2021 Feb 1;103:234–41.

33. Giovanetti M, Benvenuto D, Angeletti S, Ciccozzi M. The first two cases of 2019-nCoV in Italy: Where they come from? J Med Virol. 2020;92(5):518–21.

34. Gómez-Carballa A, Bello X, Pardo-Seco J, Pérez Del Molino ML, Martinón-Torres F, Salas A. Phylogeography of SARS-CoV-2 pandemic in Spain: a story of multiple introductions, micro-geographic stratification, founder effects, and super-spreaders. Zool Res. 2020 Nov 18;41(6):605–20.

35. Gong Y-N, Tsao K-C, Hsiao M-J, Huang C-G, Huang P-N, Huang P-W, et al. SARS-CoV-2 genomic surveillance in Taiwan revealed novel ORF8-deletion mutant and clade possibly associated with infections in Middle East. Emerg Microbes Infect. 2020 Jan 1;9(1):1457–66.

36. Jia Y, Yang C, Zhang M, Yang X, Li J, Liu J, et al. Characterization of eight novel full-length genomes of SARS-CoV-2 among imported COVID-19 cases from abroad in Yunnan, China. J Infect. 2020;81(2):e96–8.

37. Kouriba B, Dürr A, Rehn A, Sangaré AK, Traoré BY, Bestehorn-Willmann MS, et al. First Phylogenetic Analysis of Malian SARS-CoV-2 Sequences Provides Molecular Insights into the Genomic Diversity of the Sahel Region. Viruses. 2020 Nov;12(11):1251.

38. Kumar P, Pandey R, Sharma P, Dhar MS, A V, Uppili B, et al. Integrated genomic view of SARS-CoV-2 in India. Wellcome Open Res. 2020;5:184.

39. Liu J, Huang J, Xiang D. Large SARS-CoV-2 outbreak caused by asymptomatic traveler, China. Emerg Infect Dis. 2020;26(9):2260–3.

40. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell. 2020 May 28;181(5):997-+.

41. Manning JE, Bohl JA, Lay S, Chea S, Sovann L, Sengdoeurn Y, et al. Rapid metagenomic characterization of a case of imported COVID-19 in Cambodia. BioRxiv Prepr Serv Biol. 2020 Mar 5;

42. Marquez S, Prado-Vivar B, Guadalupe JJ, Gutierrez Granja B, Jibaja M, Tobar M, et al. Genome sequencing of the first SARS-CoV-2 reported from patients with COVID-19 in Ecuador. MedRxiv Prepr Serv Health Sci. 2020 Jun 14;

43. Puenpa J, Suwannakarn K, Chansaenroj J, Nilyanimit P, Yorsaeng R, Auphimai C, et al. Molecular epidemiology of the first wave of severe acute respiratory syndrome coronavirus 2 infection in Thailand in 2020. Sci Rep. 2020 Oct 6;10(1):16602.

44. Rockett RJ, Arnott A, Lam C, Sadsad R, Timms V, Gray K-A, et al. Revealing COVID-19 transmission in Australia by SARS-CoV-2 genome sequencing and agent-based modeling. Nat Med. 2020;26(9):1398–404.

45. Seemann T, Lane CR, Sherry NL, Duchene S, da Silva AG, Caly L, et al. Tracking the COVID-19 pandemic in Australia using genomics. Nat Commun. 2020 Sep 1;11(1):4376.

46. Sekizuka T, Kuramoto S, Nariai E, Taira M, Hachisu Y, Tokaji A, et al. SARS-CoV-2 Genome Analysis of Japanese Travelers in Nile River Cruise. Front Microbiol. 2020 Jun 5;11:1316.

47. Stange M, Mari A, Roloff T, Seth-Smith HM, Schweitzer M, Brunner M, et al. SARS-CoV-2 outbreak in a tri-national urban area is dominated by a B.1 lineage variant linked to mass gathering events. medRxiv. 2020 Nov 4;2020.09.01.20186155.

Tables

Country	Cases		Sequenced samples from cases (passed QC)		Percentage of cases sequenced by country of travel
	N	%	N	%	18.8%
Greece	882	21.0%	166	20.1%	23.6%
Croatia	685	16.3%	162	19.6%	18.0%
Spain	589	14.0%	106	12.8%	20.2%
Unknown	282	6.7%	57	6.9%	23.3%
France	223	5.3%	52	6.3%	11.2%
Turkey	187	4.4%	21	2.5%	18.0%
Portugal	111	2.6%	20	2.4%	15.2%
Malta	99	2.4%	15	1.8%	22.6%
Italy	93	2.2%	21	2.5%	16.5%
Poland	85	2.0%	14	1.7%	16.7%
Romania	78	1.9%	13	1.6%	15.4%
Czech Republic	65	1.5%	10	1.2%	19.7%
Albania	61	1.4%	12	1.5%	27.9%
Hungary	61	1.4%	17	2.1%	24.6%
India	57	1.4%	14	1.7%	21.8%
Pakistan	55	1.3%	12	1.5%	13.2%
Netherlands	38	0.9%	5	0.6%	20.7%
Germany	29	0.7%	6	0.7%	25.0%
Switzerland	28	0.7%	7	0.8%	12.5%
Kosovo	24	0.6%	3	0.4%	18.8%
Total cases	4207		827		19.7%

Table 1: The top 20 countries reported as the travel destination for importations of SARS-CoV-2 into England and the associated number of samples sequenced from travel-related cases

Demographic	Cases	Total contacts of cases	contacts reported per case
Sex			
Male	2193	9835	4.5
Female	1933	8578	4.4
Unknown	82	224	3.1
Age			
0-5	51	183	3.6

6-10	45	124	2.8
11-15	75	321	4.3
16-20	1086	7473	6.9
21-25	843	3536	4.2
26-30	685	2091	3.1
31-35	413	1312	3.2
36-40	278	939	3.4
41-45	185	566	3.1
46-50	169	723	4.3
51-55	130	469	3.6
56-60	121	434	3.6
61-65	57	229	4.0
66-70	30	116	3.9
71-75	20	63	3.2
76-80	7	16	2.3
81-85	7	39	5.6
86-90	3	3	1.0
91-95	0	0	NA
Unknown	2		
Ethnic group			
White			
English/Welsh/Scottish/Northern Irish/British	2509	12745	5.1
Irish	35	121	3.5
Gypsy or Irish Traveller	0	0	NA
Any other White background	583	1755	3.0
Mixed/Multiple ethnic groups			
White and Black Caribbean	42	284	6.8
White and Black African	25	108	4.3
White and Asian	49	216	4.4
Any other Mixed/Multiple ethnic background	44	147	3.3
Asian/Asian British			
Indian	103	481	4.7
Pakistani	79	306	3.9
Bangladeshi	27	82	3.0
Chinese	6	2	0.3
Any other Asian background	62	199	3.2
Black/ African/Caribbean/Black British			
African	58	152	2.6
Caribbean	12	36	3.0

Any other Black/African/Caribbean background	12	28	2.3
Other ethnic group			
Arab	0	0	NA
Any other ethnic group	110	367	3.3
Other			
Prefer not to say	65	135	2.1
Unknown	386	1473	3.8
Region			
London	1205	4275	3.5
South East	622	3211	5.2
North West	584	2323	4.0
East of England	395	2079	5.3
South West	328	1960	6.0
Yorkshire and Humber	327	1411	4.3
West Midlands	299	1259	4.2
East Midlands	251	1351	5.4
North East	161	660	4.1
Not stated	35	108	3.1
Country			
Greece	882	5587	6.3
Croatia	685	3913	5.7
Spain	589	1521	2.6
Unknown	282	988	3.5
France	223	815	3.7
Turkey	187	702	3.8
Portugal	111	439	4.0
Malta	99	492	5.0
Italy	93	390	4.2
Poland	85	417	4.9
Romania	78	189	2.4
Hungary	67	137	2.0
Czech Republic	66	239	3.6
Albania	61	140	2.3
India	57	223	3.9
Pakistan	55	269	4.9
Netherlands	38	166	4.4
Germany	29	101	3.5

Switzerland	28	123	4.4
Kosovo	24	64	2.7

Table 2: Contacts per case related to Sex, Age, Ethnic Group, Region of residence and reported Travel Destination

Figures

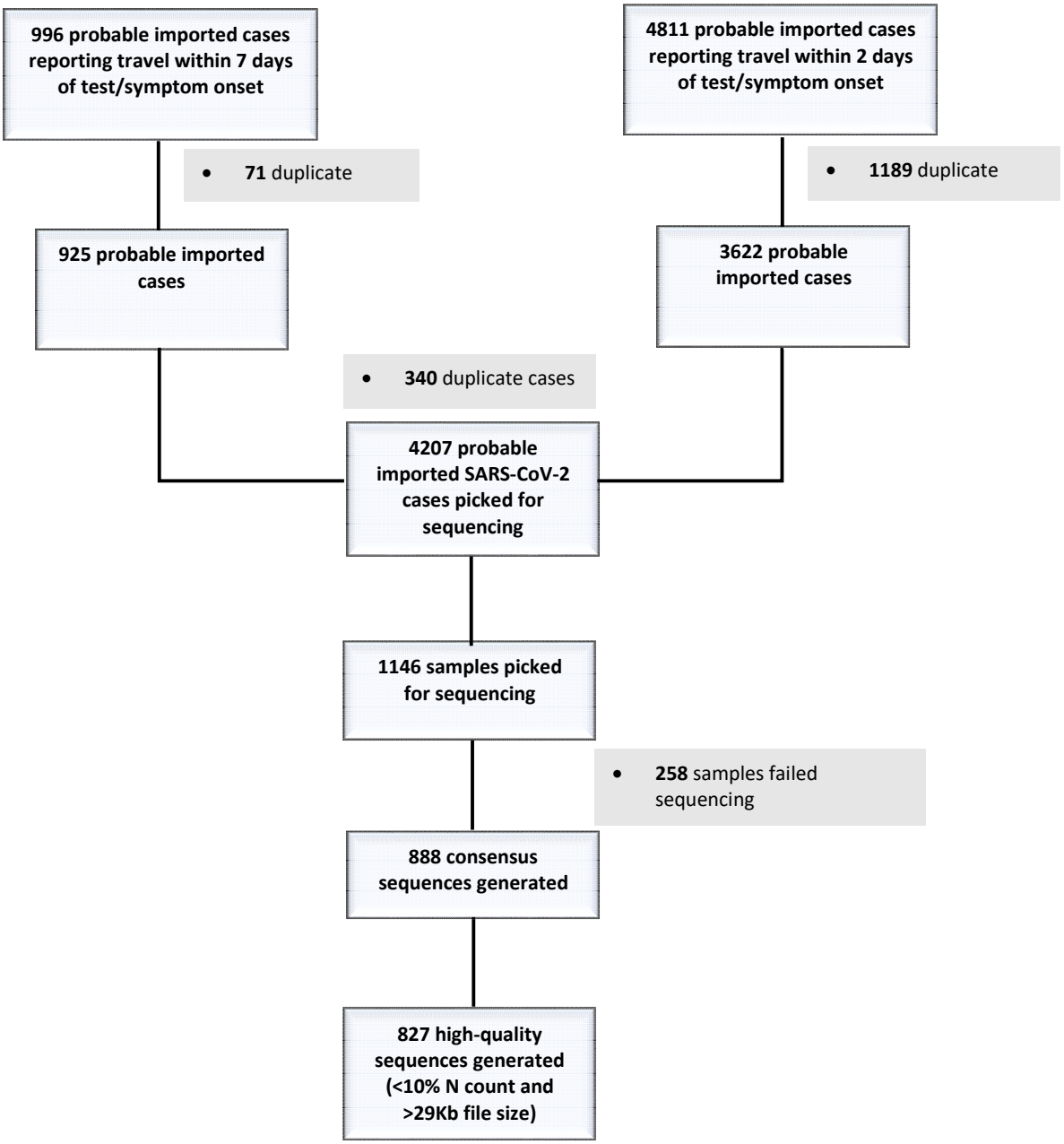


Figure 1a: Flow diagram of travel-related case ascertainment from Test and Trace data and subsequent genome availability. Cases were defined as 'highly probable' and 'probable'. 'Highly probable' travel-related cases were defined as individuals who reported international travel as an activity in the two days before symptom onset/testing. On 12/08/2020 the additional facility to report international travel in the seven days prior to symptom onset/testing became available, and also included in this study and defined as 'probable' travel-related cases.

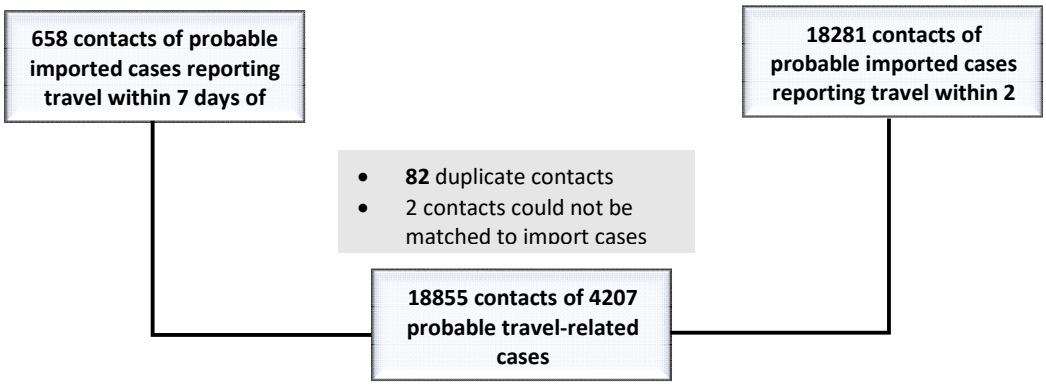


Figure 1b: Flow diagram relaying contacts ascertained of cases from Test and Trace data

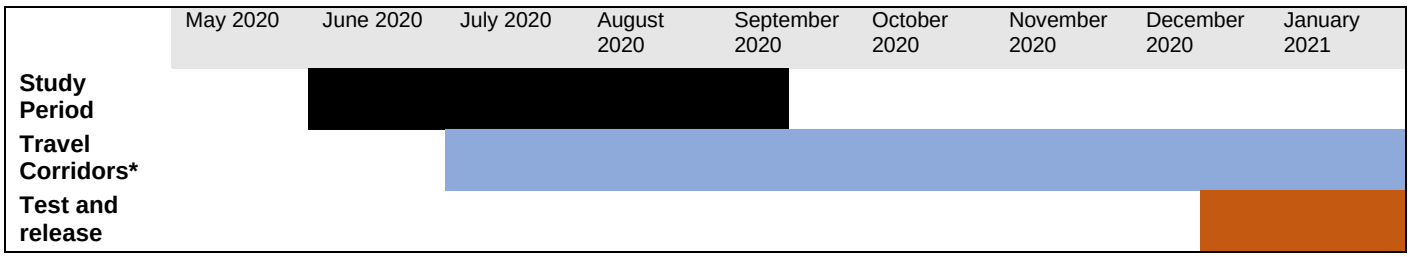


Figure 1c: Timeline of study period (27/05/2020 to 13/09/2020) and associated policy changes on travel introduced in England. Travel restrictions were assigned on a country by country basis from 6 July 2020. Travellers returning from countries that were on the travel restrictions list (4) were required to self-isolate for 14 days (*reduced to 10 days on 15/12/20), or from the 15th December 2020, choose to self-isolate for 5 days and then pay for a SARS-CoV-2 diagnostic test (test and release)

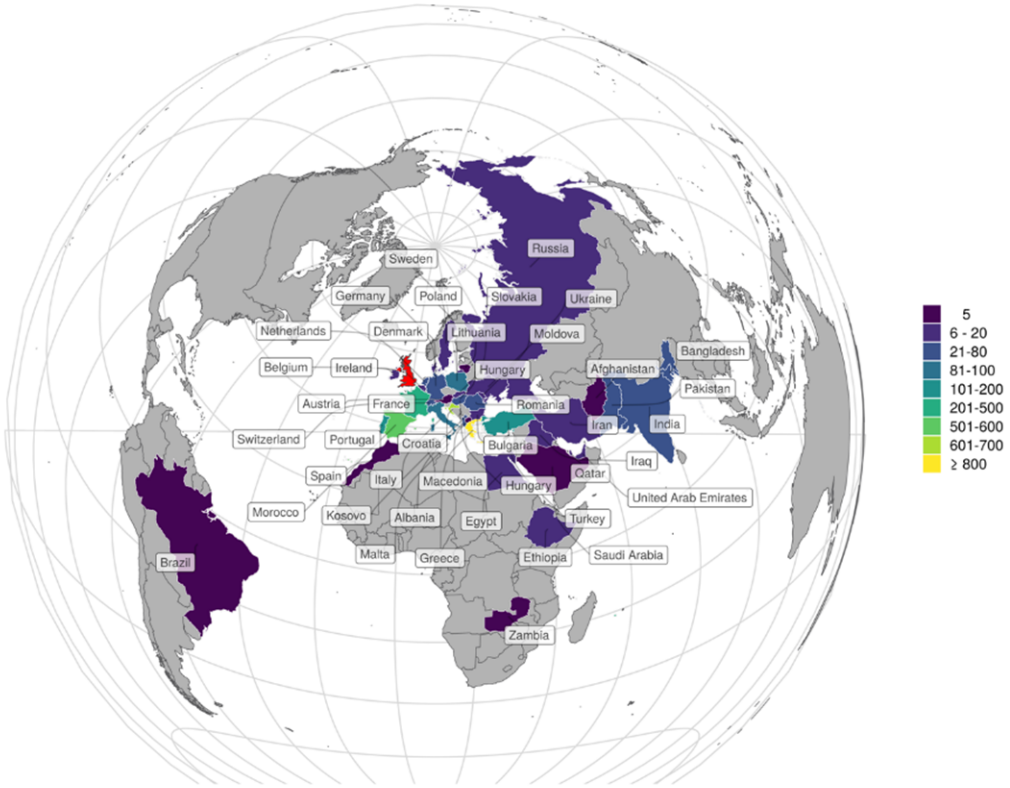


Figure 2a: Countries where importations originated. Countries with less than 5 importations were excluded for confidentiality reasons.

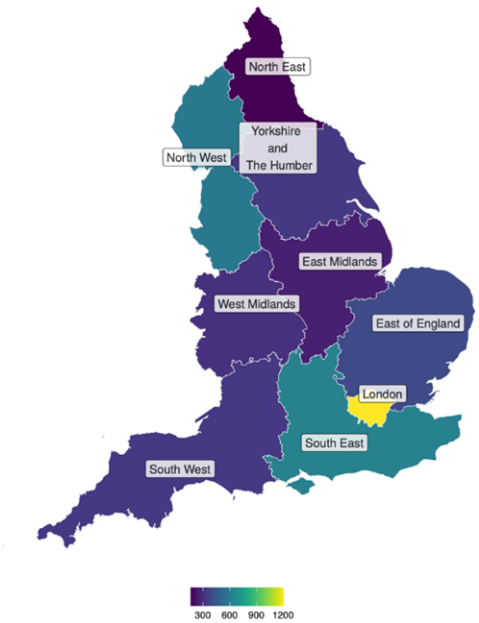


Figure 2b: Destinations of imported cases within England. Areas with less than 3 cases have been excluded

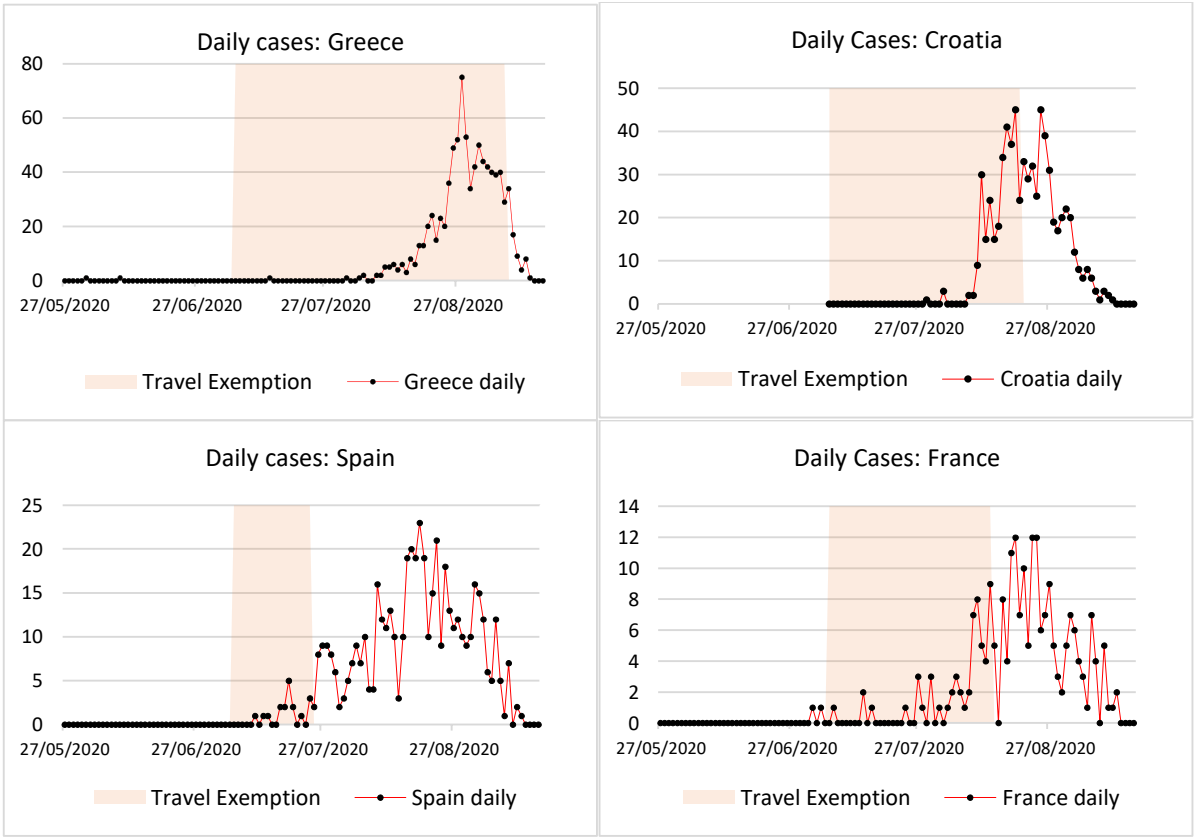


Figure 3: Frequency of importations over time for the top 4 most common countries of travel reported by individuals testing positive for SARS-CoV-2 during the study period. SARS-CoV-2 case numbers in returning travellers by the four most popular countries of travel reported by cases representing 2379/4207 (56.5%) of known travel-related cases. The shaded areas represent the period of time when the countries did not have restrictive travel guidance in place.

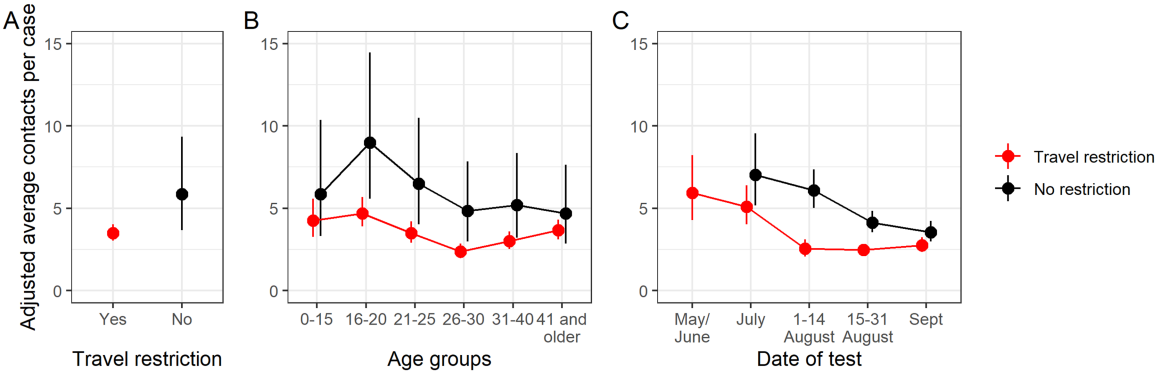


Figure 4: The effect of travel restriction on contacts per imported case of SARS-CoV-2. Estimated marginal mean number of contacts per imported case (a) overall, (b) by age-group and (c) by date of test comparing countries with travel restriction guidance (closed 'travel-corridors') in place and those without (open 'travel-corridors). All estimates are provided with 95% confidence intervals.

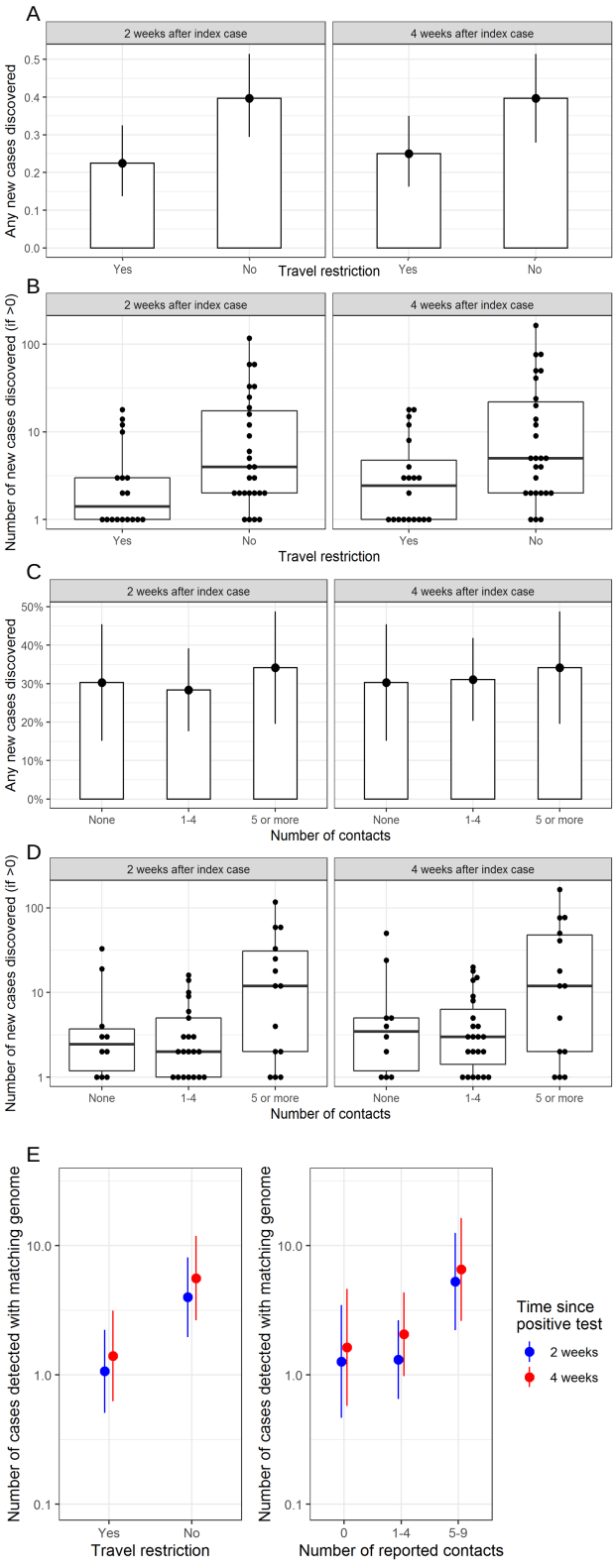


Figure 5: The effect of travel restrictions on the subsequent spread of likely imported cases as determined by genomics. Panel A: The proportion of imported cases with any matching genome detected over the two or four weeks following index test result. Panel B: The number of genomes matching the index case, with zeros excluded. Panel A and B compare countries with travel restriction guidance (closed 'travel-corridors') in place and those without (open 'travel-corridors). Panel C: The proportion of imported cases with any matching genome detected over the two of four weeks following index test result. Panel D:

The number of genomes matching the index case, with zeros excluded. Panel E: Estimated marginal mean number of genomes detected after 2 weeks or 4 weeks matching an index genome, stratified by travel restriction and stratified by number of contacts. In all panels, boxes correspond to median and interquartile range, and error bars correspond to 95% confidence intervals.

Supplementary Tables

Study	Month (2020)	Country	Imported from	No. imports	Genomes
(25)	March	China	US (2), Germany (1)	3	7
(26)	January	Germany	China	1	
(27)		Israel	Japan (1), Italy (1)	2	
(28)		Brazil	Italy (4)		6
(29)	March	Brazil	Spain	1	
(30)	Before May	China	South America (2), North America (15), Europe (101), Other Asian countries (3)		102
(31)	February	Mexico	Italy	1	
(32)	February - March	South Africa	Europe	13	27
(33)	January	Italy	China	2	
(34)	Before June	Spain		> 34	
(35)	January - March	Taiwan	China, Germany, UK, Turkey, Iran, Middle East, Europe		20
(36)	March	China	Spain (2), France (1), Cambodia (1), Sri Lanka (1), US (3)	8	
(37)	April	Mali			21
(38)		India	China, South Asia, Middle East, Italy, Spain, UK, France and USA		104
(39)	April	China	US	1	
(40)	January - March	China	19 different countries	102	53 new + 177 publicly available sequences
(41)	January	Cambodia		1	
(42)	March	Ecuador	Netherlands	1	4
(43)	<May	Thailand	China	1	40
(44)	January - March	Australia	Asia, Western Europe and North America		
(45)		Australia	Asia, Europe, North America	193	76 clusters, 34 only international travel, 34 mixed local/international
(46)	March	Japan	Egypt	10	26
(47)	February - March	Switzerland	Italy (2), France (1), Austria (refers to a previous study)		486
(Tapfumani et al 2020)	February - March	Zimbabwe	UK, US, South Africa, Dubai		97

Supplementary Table 1: Studies using genomics to as part of epidemiological investigations of importations of SARS-CoV-2.

Ethnic group	Cases		Census
White	**No.**	**%**	%
English/Welsh/Scottish/Northern Irish/British	2509	66.8%	80.5%
Irish	35	0.9%	0.9%
Gypsy or Irish Traveller	0	0.0%	0.1%
Any other White background	583	15.5%	4.4%
Mixed/Multiple ethnic groups			
White and Black Caribbean	42	1.1%	0.8%
White and Black African	25	0.7%	0.3%
White and Asian	49	1.3%	0.6%
Any other Mixed/Multiple ethnic background	44	1.2%	0.5%
Asian/Asian British			
Indian	103	2.7%	2.5%
Pakistani	79	2.1%	2.0%
Bangladeshi	27	0.7%	0.8%
Chinese	6	0.2%	0.7%
Any other Asian background	62	1.7%	1.5%
Black/ African/Caribbean/Black British			
African	58	1.5%	1.8%
Caribbean	12	0.3%	1.1%
Any other Black/African/Caribbean background	12	0.3%	0.5%
Other ethnic group			
Arab	0	0.0%	0.4%
Any other ethnic group	110	2.9%	0.6%
Other			
Prefer not to say	65		
Unknown	386		

Supplementary Table 2: Self-identified ethnicity of cases (UK Government Statistical Service ethnic groups). The 2011 census data for England and Wales was used.

Demographic	Cases		Contacts		Cases with Genomes that passed QC	
Sex						
Male	2193	56.2%	6088	49.7%	394	47.8%
Female	1933	46.8%	6160	50.3%	414	50.2%
Unknown	81		6607		19	
Age						
0-5	51	1.2%	303	2.5%	9	1.1%
6-10	45	1.1%	361	2.9%	12	1.5%

11-15	75	1.8%	467	3.8%	14	1.7%
16-20	1086	25.8%	2274	18.5%	228	27.6%
21-25	843	20.0%	1866	15.2%	165	20.0%
26-30	685	16.3%	1350	11.0%	135	16.3%
31-35	413	9.8%	849	6.9%	88	10.6%
36-40	278	6.6%	721	5.9%	46	5.6%
41-45	185	4.4%	691	5.6%	35	4.2%
46-50	169	4.0%	984	8.0%	34	4.1%
51-55	130	3.1%	1168	9.5%	22	2.7%
56-60	121	2.9%	692	5.6%	23	2.8%
61-65	57	1.4%	264	2.1%	7	0.8%
66-70	30	0.7%	141	1.1%	4	0.5%
71-75	20	0.5%	101	0.8%	4	0.5%
76-80	7	0.2%	35	0.3%	0	0.0%
81-85	7	0.2%	30	0.2%	1	0.1%
86-90	3	0.1%	7	0.1%	0	0.0%
91-95	0	0.0%	4	0.0%	0	0.0%
Unknown	2		6547		0	
Ethnic group						
White						
English/Welsh/Scottish/Northern Irish/British	2509	66.8%	8370	73.8%	499	68.4%
Irish	35	0.9%	134	1.2%	4	0.5%
Gypsy or Irish Traveller	0	0.0%	0	0.0%	0	0.0%
Any other White background	583	15.5%	1261	11.1%	120	16.4%
Mixed/Multiple ethnic groups						
White and Black Caribbean	42	1.1%	88	0.8%	4	0.5%
White and Black African	25	0.7%	60	0.5%	8	1.1%
White and Asian	49	1.3%	112	1.0%	9	1.2%
Any other Mixed/Multiple ethnic background	44	1.2%	124	1.1%	9	1.2%
Asian/Asian British						
Indian	103	2.7%	318	2.8%	13	1.8%
Pakistani	79	2.1%	183	1.6%	4	0.5%
Bangladeshi	27	0.7%	50	0.4%	3	0.4%
Chinese	6	0.2%	33	0.3%	1	0.1%
Any other Asian background	62	1.7%	197	1.7%	13	1.8%
Black/ African/Caribbean/Black British						
African	58	1.5%	148	1.3%	12	1.6%
Caribbean	12	0.3%	35	0.3%	3	0.4%
Any other Black/African/Caribbean background	12	0.3%	25	0.2%	2	0.3%
Other ethnic group						

Arab	0	0.0	0	0.0%	0	0.0%
Any other ethnic group	110	2.9%	195	1.7%	26	3.6%
Other						
Prefer not to say	65		139		85	
Unknown	386		7382		12	
Region						
London	1205	28.6%	4681	24.9%	298	36.3%
South East	623	14.8%	3201	17.0%	175	21.3%
North West	584	13.9%	2503	13.3%	70	8.5%
East of England	395	9.4%	1958	10.4%	94	11.4%
South West	328	7.8%	1755	9.3%	74	9.0%
Yorkshire and Humber	327	7.8%	1488	7.9%	21	2.6%
West Midlands	299	7.1%	1410	7.5%	20	2.4%
East Midlands	251	6.0%	1157	6.2%	29	3.5%
North East	161	3.8%	646	3.4%	40	4.9%
Not stated	34	0.8%	56		6	

Supplementary Table 3: Demographics of cases, contacts, and cases with genomes that pass quality control available

	Effect of travel restriction (ratio of mean contacts)	**Adjusted mean contacts**	
		With travel restriction	**Without travel restriction**
Overall	0.60 (0.37-0.95)	3.50 (3.04-4.02)	5.85 (3.67-9.34)
By age-group			
0-15	0.73 (0.39-1.34)	4.3 (3.3-5.6)	5.9 (3.3-10.3)
16-20	0.52 (0.32-0.85)	4.7 (3.9-5.7)	9.0 (5.6-14.5)
21-25	0.54 (0.33-0.88)	3.5 (2.9-4.2)	6.5 (4.0-10.5)
26-30	0.49 (0.30-0.80)	2.4 (2.0-2.9)	4.8 (3.0-7.8)
31-40	0.58 (0.36-0.94)	3.0 (2.5-3.6)	5.2 (3.2-8.3)
41 and older	0.78 (0.48-1.29)	3.7 (3.1-4.3)	4.7 (2.9-7.6)
By calendar date			
May/June		5.9 (4.3-8.2)	Insufficient data
July	0.72 (0.51-1.03)	5.1 (4.0-6.4)	7.0 (5.2-9.5)
August 1-14	0.42 (0.33-0.53)	2.5 (2.1-3.1)	6.1 (5.0-7.4)
August 15-31	0.60 (0.52-0.69)	2.5 (2.1-2.8)	4.1 (3.5-4.8)
September	0.78 (0.65-0.93)	2.8 (2.4-3.2)	3.6 (3.0-4.2)

Supplementary Table 4: The effect of travel restriction on reported contacts per imported case, and the estimated marginal mean number of reported cases when imported from a country with or without a travel restriction in place. Figures are reported for the overall dataset, and then stratified by age-group and calendar date of positive test. All estimates are provided with 95% confidence intervals

	Effect of travel restriction on subsequent cases with matching genome (rate ratio with 95% confidence interval)	
	2 weeks after index case	**4 weeks after index case**
Model without contacts		
Quarantine	0.18 (0.06-0.55)	0.17 (0.05-0.52)
Model including contacts		
Quarantine	0.27 (0.09-0.77)	0.25 (0.08-0.81)
1-4 contacts (vs none)	1.03 (0.30-3.58)	1.26 (0.34-4.75)
5 or more contacts (vs none)	4.14 (1.14-15.02)	4.01 (1.06-15.1)

Supplementary Table 5: Effect of travel restriction on number of genomes detected. Effect of travel restriction on number of genomes detected (measured by rate ratio with 95% confidence interval). All estimates are adjusted for calendar date.

Lineage	No. Samples	Percentage
UK5	152	18.4%
UK1897	73	8.8%
UK461	66	8.0%
UK2229	28	3.4%
UK1249	22	2.7%
UK649	22	2.7%
UK1506	22	2.7%
UK1031	12	1.5%
UK1205	10	1.2%
UK2347	10	1.2%
UK761	10	1.2%
UK1780	8	1.0%
UK1791	8	1.0%
UK1569	7	0.8%
UK831	7	0.8%
UK669	7	0.8%
UK1018	6	0.7%
UK1219	6	0.7%
UK2683	6	0.7%
UK2726	6	0.7%
UK778	5	0.6%
UK1535	5	0.6%
UK1581	5	0.6%
UK2268	5	0.6%
214 lineages with <5 cases	319	38.6%

Supplementary Table 6: The number of samples with each UK lineage

Lineage	No. Samples	Percentage
B.1.1	159	19.2%
B.1.177	128	15.5%
D.1	87	10.5%
B.1.160	75	9.1%
B.1.5	72	8.7%
B.1	72	8.7%
B.1.1.1	55	6.7%
B.1.1.37	55	6.7%
B.1.78	36	4.4%
B.1.1.70	17	2.1%
B.1.36	14	1.7%
B.1.5.12	6	0.7%
B.1.36.1	6	0.7%
B.1.1.34	5	0.6%
25 lineages with <5 cases	40	4.8%

Supplementary Table 7: The number of travel-related samples with each Global lineage

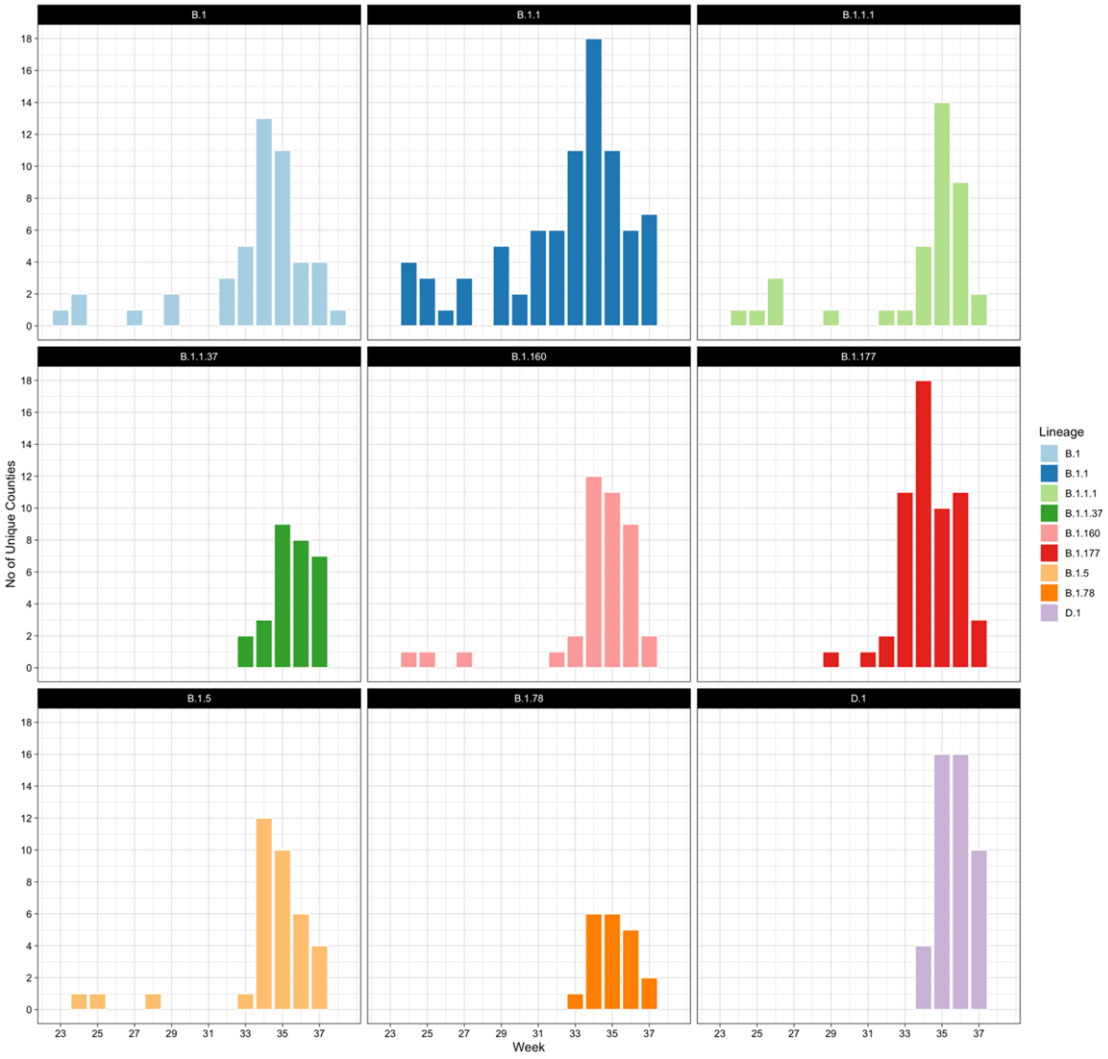
Lineage	Number	Percentage
B.1.1	7673	37.2
B.1.177	1862	9.0
B.1	1547	7.5
B.1.5	1299	6.3
B.1.1.37	1173	5.7
B.1.1.35	996	4.8
B.1.1.1	805	3.9
D.1	647	3.1
B.1.160	405	2.0
B.1.36.1	362	1.8
B.1.1.4	335	1.6
B	321	1.6
B.1.1.51	319	1.5
B.1.1.30	233	1.1
B.1.36	232	1.1
B.1.1.15	189	0.9
B.1.78	181	0.9
B.1.1.55	145	0.7
C.3	125	0.6
B.1.1.70	112	0.5

Supplementary Table 8: The number of samples with each Global lineage from the COG-UK dataset during the study period. This table includes the 'top 20' lineages sequenced during the study period.

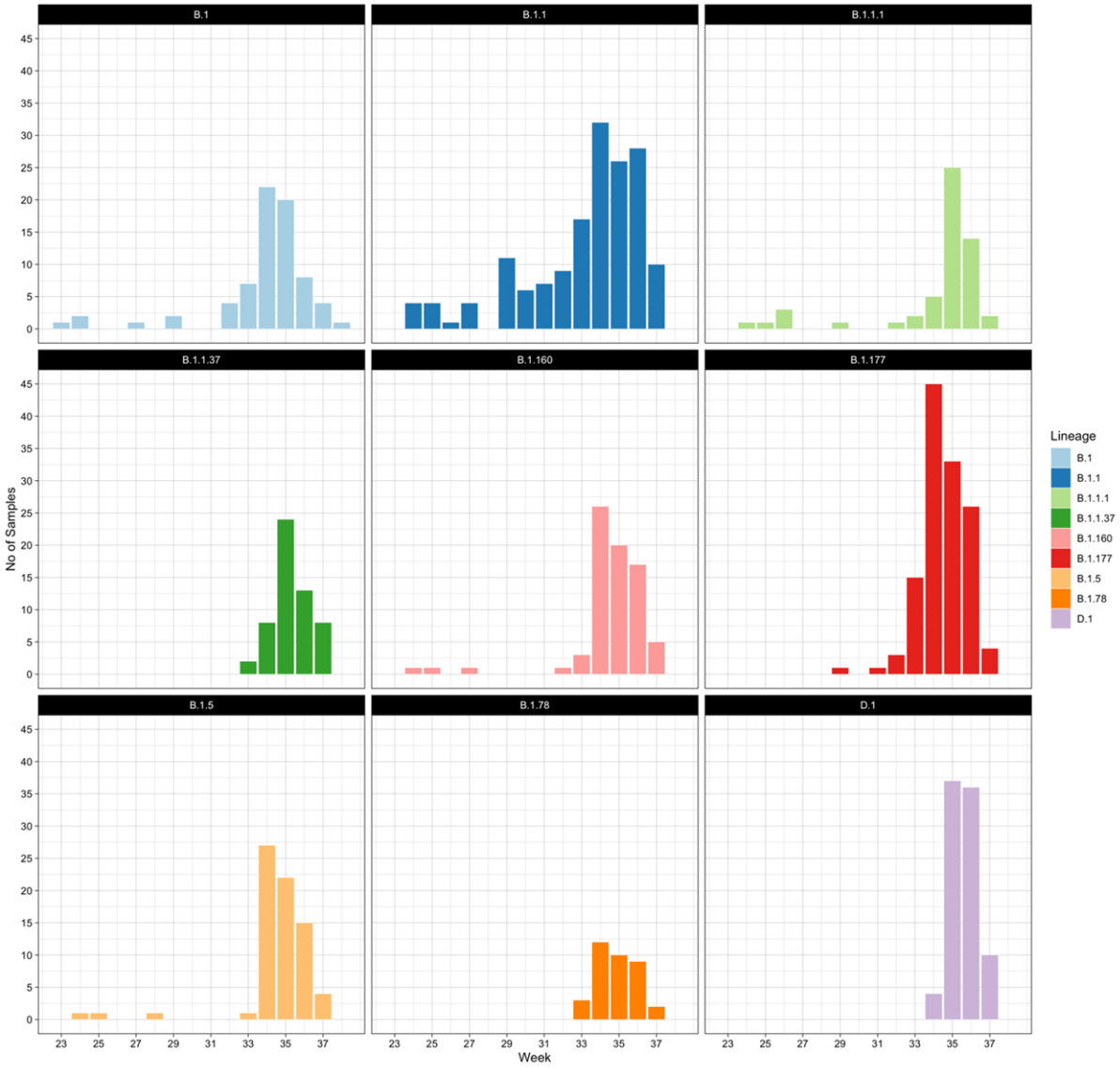
Mutation	Cases with Mutant Variant		Wild Type Variant		Inconclusive	
D614G	824	99.64%	3	0.36%	0	0.00%
P323L*	4	0.48%	815	98.55%	7	0.85%
N439K	65	7.86%	758	91.66%	4	0.48%
A222V	131	15.84%	694	83.92%	2	0.24%
Y453F	0	0.00%	826	99.88%	1	0.12%
Total Cases						1114

Supplementary Table 9: Mutant variants identified in the travel-related cases during the study period. *F mutation found in 1 case

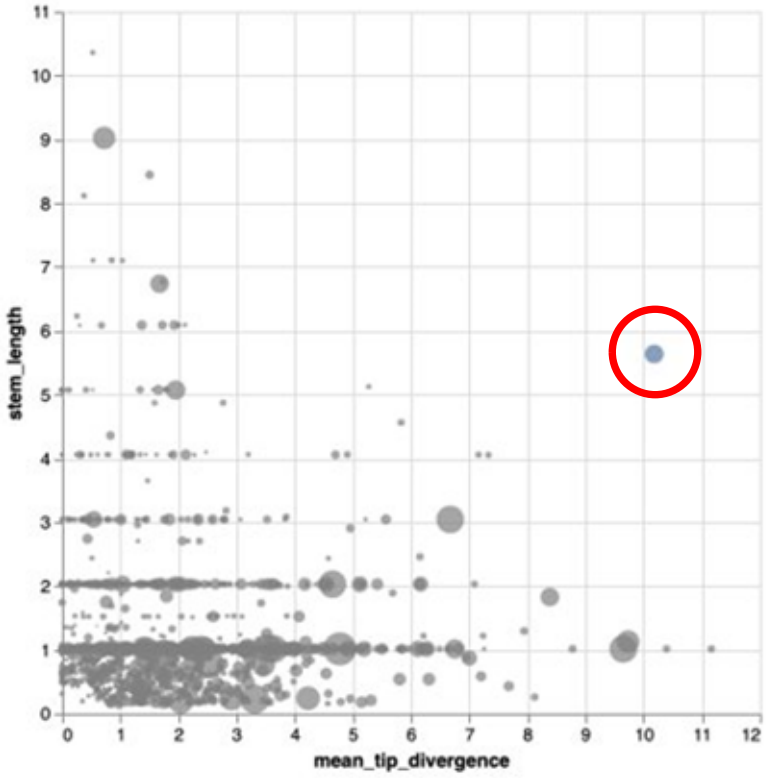
Supplementary Figures



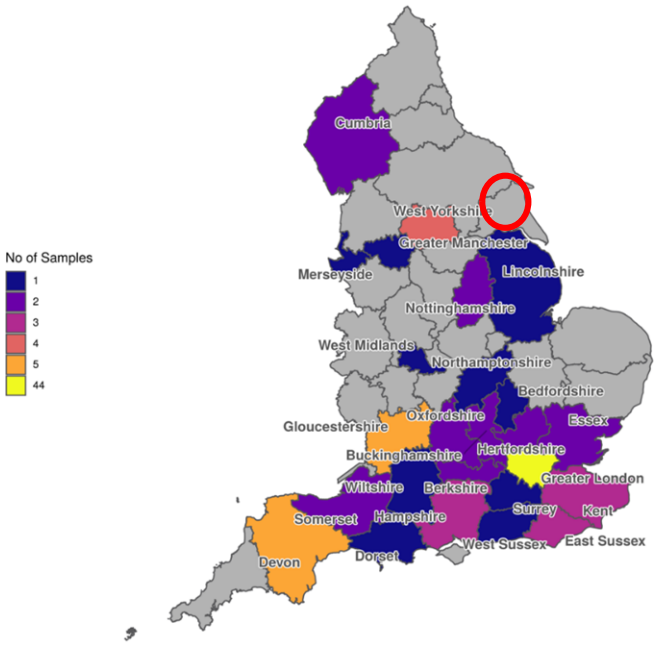
Supplementary Figure 1: The dispersion of importations of different lineages throughout England per week. This represents the top 9 global lineages versus the number of unique counties the lineage is found in, using the county provided by the case. The counties are the lieutenancies or ceremonial counties of which there are 48.



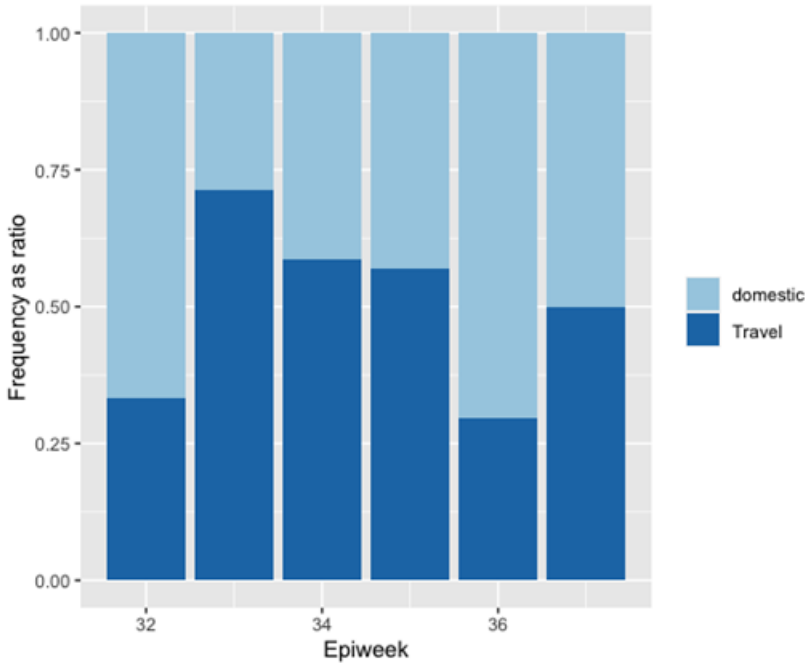
Supplementary Figure 2: The number of importations of each global lineage per week of 2020. This Figure represents the Top 9 global lineages.



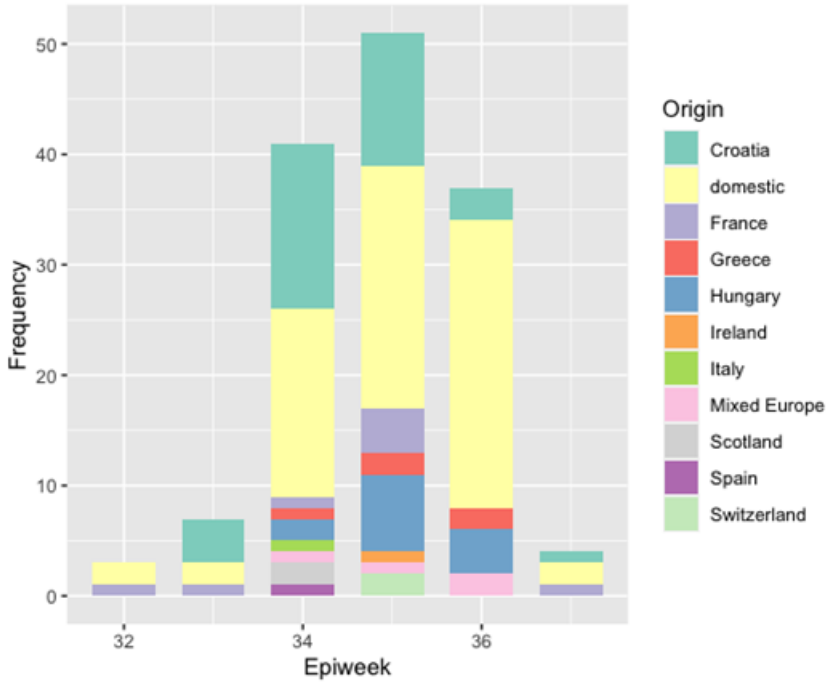
Supplementary Figure 3: Polecat cluster analysis with the likely travel-related cluster highlighted.



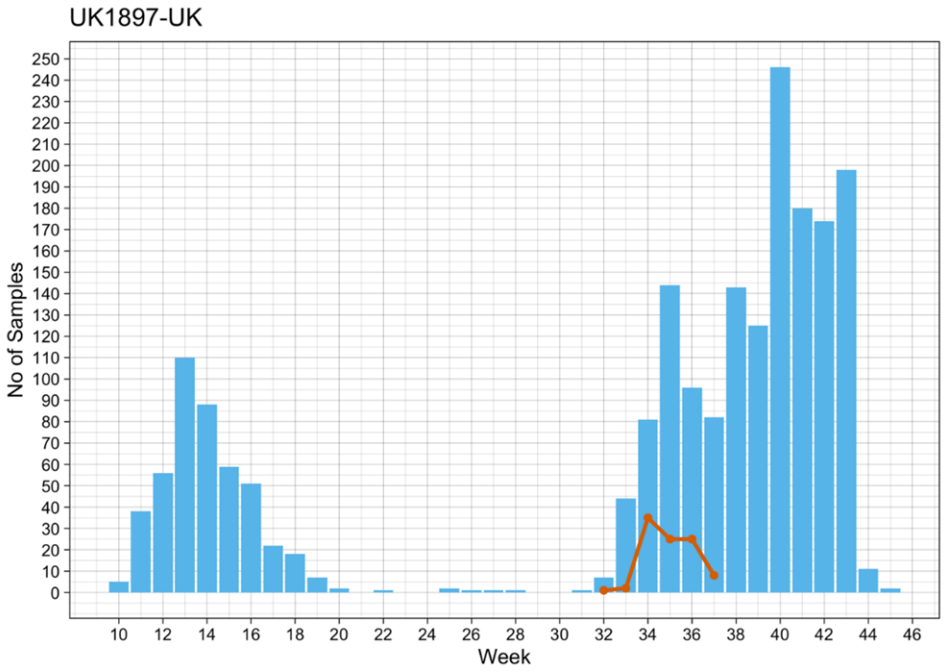
Supplementary Figure 4: No. of genomes of importations or their contacts of lineage UK1897 per county in England.



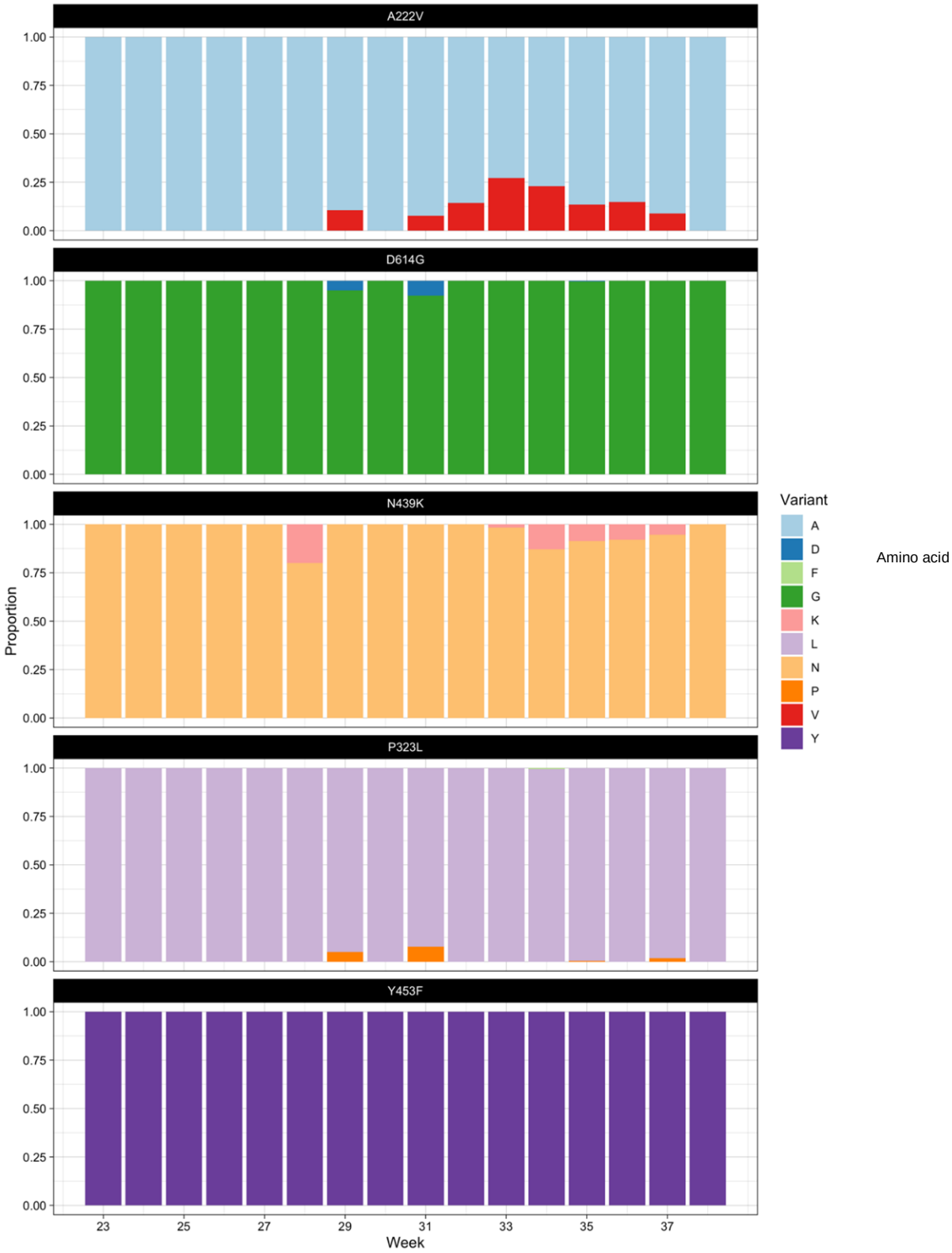
Supplementary Figure 5: Frequency of individuals identifying a travel or domestic source of SARS-CoV-2 acquisition within the suspected travel-related cluster of genomes highlighted by the Polecat tool, represented by epiweek



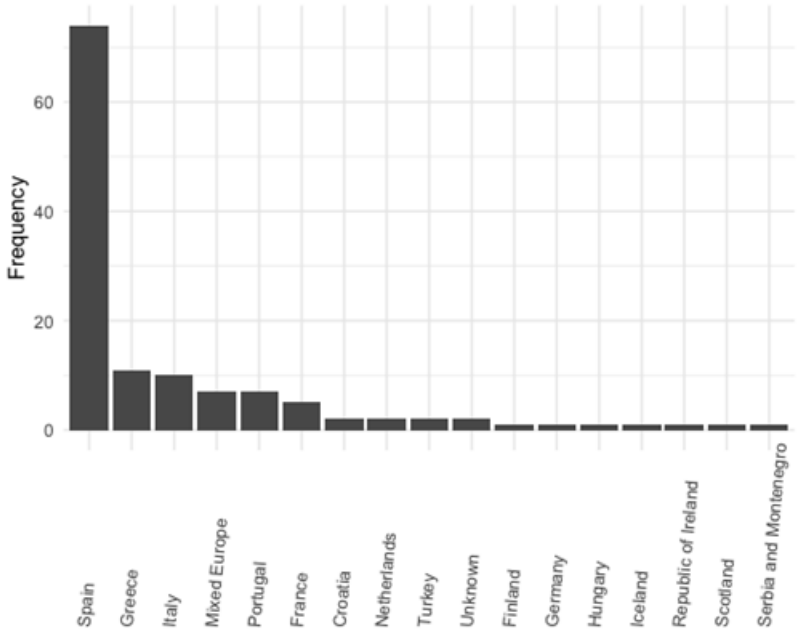
Supplementary Figure 6: Source country of SARS-CoV-2 acquisition for individuals identified within a suspected travel-related cluster highlighted by the Polecat tool, represented by epiweek



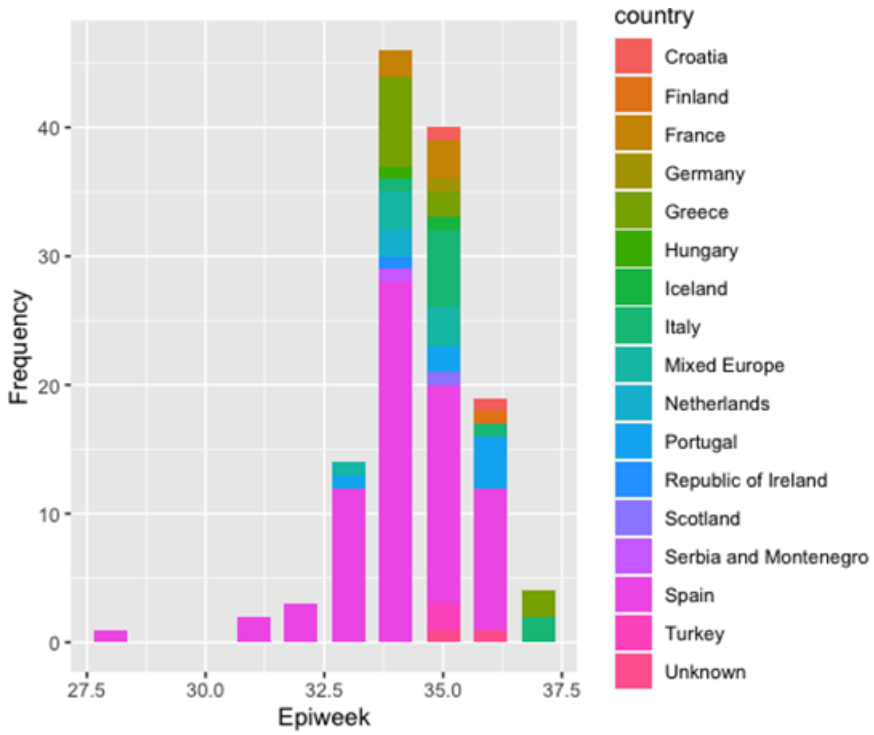
Supplementary Figure 7: UK1897 SARS-CoV-2 lineage in the United Kingdom by epiweek. The line (orange) is the number of genomes which are confirmed importations from the lineage UK1897 per week of 2020. The blue bars indicate the number of genomes of this lineage seen per week anywhere in the UK (including the importations).



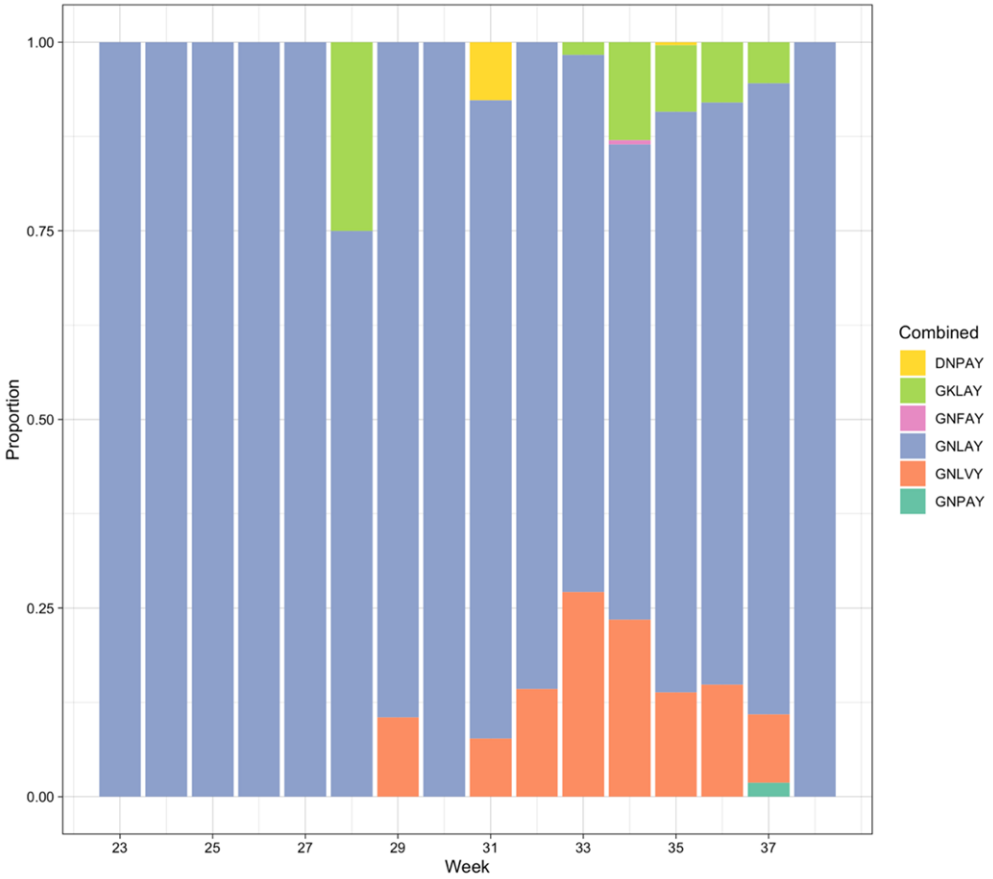
Supplementary Figure 8: The percentage of each major mutation observed per week in the imported genomes. Letters in the amino acid substitution nomenclature correspond to: A, alanine; D, aspartic acid; F, phenylalanine; G, glycine; K, lysine; L, leucine; N, asparagine; P, proline; V, valine; Y, tyrosine. The mutations are named as following: the letter preceding number (the amino acid site of substitution) represents the wild-type amino acid, the letter following the number is the observed amino acid in the sample ('a mutation', if different from the wild-type). The figure legend represents the observed amino acid at the site of interest, e.g. 'A' in the panel representing the A222V mutation shows cases observing alanine at site 222.



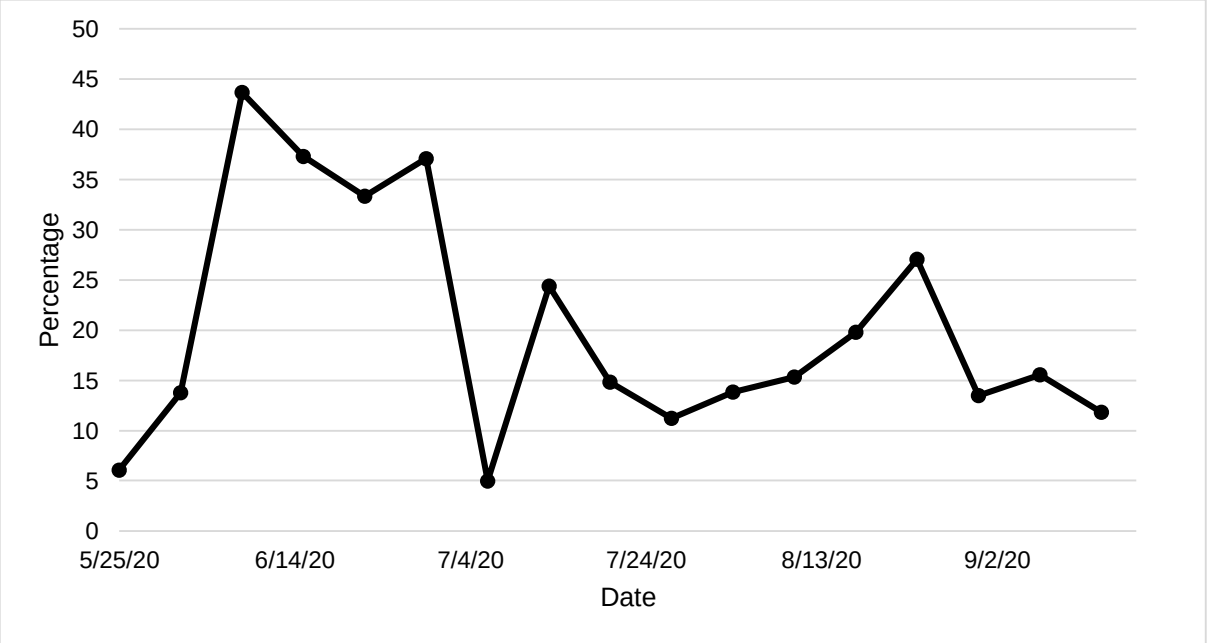
Supplementary Figure 9: Destination country of travel-related SARS-CoV-2 with the A222V variant identified, during the study period



Supplementary Figure 10: Reported country of travel for cases with the A222V variant of SARS-CoV-2, imported over time



Supplementary Figure 11: Combination of SARS-CoV-2 mutations seen in imported SARS-CoV-2 genomes by epiweek. The combinations of co-occurring variants, where the variants are in the order: 1) D614G, 2) N439K,3) P323L, 4) A222V and 5) Y453F.



Supplementary Figure 12: Percentage of known SARS-CoV-2 cases sequenced in England from 25/05/2020 to 14/09/2020

Supplementary Methods and Definitions

Travel guidance
During the time period of the study all non-essential travel outside of the UK was advised against. Varying restrictions were applied to travellers returning from different countries or regions of countries, changing over the course of the study period. Travellers were required to quarantine for 2 weeks if they had visited a restricted region in the previous 2 weeks. There were exceptions for particular classes of individuals such as freight drivers and flight crews. Designated regions were exempt from the quarantine requirements, and commonly referred to as open 'travel-corridors'. These restrictions changed over time for different regions (https://www.gov.uk/guidance/travel-advice-novel-coronavirus).

Contact tracing and case identification
Contact-tracing data was obtained from T&T. Case data gathered from testing laboratories is enriched with data provided by NHS Spine, prior to arrival at the contact tracing advise service system. All cases and contacts had a field for demographic data, but this was not always reported (Table 2 and Supplementary Table 3). 'Highly probable' travel-related cases were defined as individuals who reported international travel as an activity in the two days before symptom onset/testing. On 12/08/2020 the additional facility to report international travel in the seven days prior to symptom onset/testing became available, and also included in this study and defined as 'probable' travel-related cases.

Cases are asked to provide details of all contacts for activities in the 2 days prior to onset/testing up to completing the system which were gathered. If any contacts become cases they would then also be included in T&T as a case separately but if they did not report direct travel themselves, then they would not meet the definition for a travel-associated case.

Cases identified reporting travel in 7-day period prior to symptom onset or positive test:
Test and Trace data included destination city, and a free-text search was run with a custom python script to convert city to associated country of destination. All fields were manually cross-checked and any errors corrected (142 corrections). A further 103 countries were manually inputted due to spelling errors in the free-text Test and Trace data provided and 1 country by searching flight numbers provided by the case when country or city not available. 22 cases reporting travel-related activities did not have an associated destination clearly identified.

Cases identified reporting travel in 2-day period prior to symptom onset or positive test:
A free-text country and city search with a custom python script on travel-related T&T data was used to identify destination country. This yielded 1898 destination countries, and a further 1182 by city search. All fields were manually cross-checked and errors corrected (210 corrected). 542 case-country associations were manually entered where spelling mistakes were present in the free-text entries, including 98 entered by flight number searches where this was the only available data.

Lineages
Global and UK Lineages (14) were assigned to each genome using Pangolin (https://github.com/cov-lineages/pangolin) with analysis performed on COVID-CLIMB (13). Global lineages, reflecting genomically distinct identifiable importations into a new region, are denoted with a letter followed by a hierarchy of up to 4 numbers such as B.1.2.3, providing for a stable and consistent naming of clusters. These lineages are manually curated and assigned. UK lineages represent the subsequent regional and local spread within the UK, taking the form UK1234, providing an identifier for a cluster for a given phylogeny. These identifiers are assigned programmatically are unstable. Labelled phylogenetic trees were created using CIVET tool (version 2.0) (https://github.com/cog-uk/civet).

Identification of extinct and unique genomes
The 827 high-quality travel-related genomes were compared to the COG-UK dataset on 16/10/2020. Genomes were only compared to other genomes with the same UK lineage assigned by COG-UK, since we assume that no relatedness relevant to transmission exists between genomes of different UK lineages. A unique genome in the community was deemed to be one that was known to be from a travel-related case and either: (1) A UK lineage that had not been sampled in the previous 4 weeks in the UK, (2) >3 SNPs distance to the closest relative in the COG-UK dataset.
Within the same UK lineage we identified those genomes sampled within 4 weeks prior to the genome of interest. We determined the minimum SNP distance between the sequence of interest and these genomes. This identified 207/827 genomes with a minimum SNP distance of >3 SNPs to its closest relative in the COG-UK dataset. These constitute genomes for which no close relative was sampled in the UK at the time of importation. The analysis was then repeated on 05/12/2020 on these 207 ‘unique genomes’ to account for delays in genomes uploaded to MRC CLIMB. 195/207 were included in this analysis, with 12/207 genomes excluded due to the large UK phylotypes they belonged to and the subsequent computational requirements. At this time a further 8 genomes were determined to have a close relative sampled in the UK in the 4 weeks before importation. This wasn't detected earlier, because their close relative was uploaded with a significant delay.
The remaining 186 genomes were ‘Unique’ genomes were compared to sequences that were generated in the COG-UK dataset within 2 and 4 weeks after their sampling date, to identify samples with the same UK lineage and within 2 SNPs. These would represent onward transmission or further introductions of similar genomes. The analysis was run with an in-house custom Python script developed by US and RM.