

This is an Open Access document downloaded from ORCA, Cardiff University's institutional repository:<https://orca.cardiff.ac.uk/id/eprint/161509/>

This is the author's version of a work that was submitted to / accepted for publication.

Citation for final published version:

Tsui, Joseph L.-H., McCrone, John T., Lambert, Ben, Bajaj, Sumali, Inward, Rhys P. D., Bosetti, Paolo, Pena, Rosario Evans, Tegally, Houriiyah, Hill, Verity, Zarebski, Alexander E., Peacock, Thomas P., Liu, Luyang, Wu, Neo, Davis, Megan, Bogoch, Isaac I., Khan, Kamran, Kall, Meaghan, Abdul Aziz, Nurin Iwani Binti, Colquhoun, Rachel, O'Toole, Áine, Jackson, Ben, Dasgupta, Abhishek, Wilkinson, Eduan, de Oliveira, Tulio, Connor, Thomas R. , Loman, Nicholas J., Colizza, Vittoria, Fraser, Christophe, Volz, Erik, Ji, Xiang, Gutierrez, Bernardo, Chand, Meera, Dellicour, Simon, Cauchemez, Simon, Raghwani, Jayna, Suchard, Marc A., Lemey, Philippe, Rambaut, Andrew, Pybus, Oliver G. and Kraemer, Moritz U. G. 2023. Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1. *Science* 381 (6655) , pp. 336-343. 10.1126/science.adg6605

Publishers page: <http://dx.doi.org/10.1126/science.adg6605>

Please note:

Changes made as a result of publishing processes such as copy-editing, formatting and page numbers may not be reflected in this version. For the definitive version of this publication, please refer to the published source. You are advised to consult the publisher's version if you wish to cite this paper.

This version is being made available in accordance with publisher policies. See <http://orca.cf.ac.uk/policies.html> for usage policies. Copyright and moral rights for publications made available in ORCA are retained by the copyright holders.



Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1

Joseph L.-H. Tsui^{1,†,*}, John T. McCrone^{2,3,†}, Ben Lambert^{4,†}, Sumali Bajaj^{1,†}, Rhys P.D. Inward^{1,†}, Paolo Bosetti^{5,†}, Houriiyah Tegally^{6,7}, Verity Hill^{3,8}, Rosario Evans Pena¹, Alexander E. Zarebski¹, Thomas P. Peacock^{9,10}, Luyang Liu¹¹, Neo Wu¹¹, Megan Davis¹², Isaac I. Bogoch¹³, Kamran Khan^{12,13}, Meaghan Kall¹⁰, Nurin Iwani Binti Abdul Aziz¹⁰, Rachel Colquhoun³, Áine O'Toole³, Ben Jackson³, Abhishek Dasgupta¹, Eduan Wilkinson^{12,13}, Tulio de Oliveira^{6,7}, The COVID-19 Genomics UK (COG-UK) consortium[†], Thomas R. Connor^{14,15,16}, Nicholas J. Loman¹⁷, Vittoria Colizza¹⁸, Christophe Fraser^{19,20}, Erik Volz²¹, Xiang Ji²², Bernardo Gutierrez¹, Meera Chand¹⁰, Simon Dellicour^{23,24}, Simon Cauchemez^{5,†}, Jayna Raghwan^{1,25,†}, Marc A. Suchard^{26,†}, Philippe Lemey^{24,†}, Andrew Rambaut^{3,†}, Oliver G. Pybus^{1,20,25,†,*}, Moritz U.G. Kraemer^{1,20,†,*}

Affiliations

1. Department of Biology, University of Oxford, Oxford, UK.
2. College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter, UK
3. Helix, San Mateo, USA.
4. Institute of Ecology and Evolution, University of Edinburgh, Edinburgh, UK.
5. Institut Pasteur, Université Paris Cité, CNRS, Paris, France.
6. KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa.
7. Centre for Epidemic Response and Innovation (CERI), School for Data Science and Computational Thinking, Stellenbosch University, Stellenbosch, South Africa.
8. Yale University, New Haven, USA.
9. Department of Infectious Disease, Imperial College London, London, UK.
10. UK Health Security Agency, London, UK.
11. Google Research, Mountain View, USA.
12. BlueDot, Toronto, Canada.
13. Department of Medicine, Division of Infectious Diseases, University of Toronto, Toronto, Canada
14. Pathogen Genomics Unit, Public Health Wales NHS Trust, Cardiff, UK.

15. School of Biosciences, The Sir Martin Evans Building, Cardiff University, UK.
16. Quadram Institute, Norwich, UK.
17. Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK.
18. Sorbonne Université, INSERM, Institut Pierre Louis d'Épidémiologie et de Santé Publique (IPLESP), Paris, France.
19. Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, UK.
20. Pandemic Sciences Institute, University of Oxford, UK.
21. MRC Centre of Global Infectious Disease Analysis, Jameel Institute for Disease and Emergency Analytics, Imperial College London, London, UK
22. Department of Mathematics, Tulane University, New Orleans, USA.
23. Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, Bruxelles, Belgium.
24. Department of Microbiology, Immunology and Transplantation, Rega Institute, KU Leuven, Leuven, Belgium.
25. Department of Pathobiology and Population Science, Royal Veterinary College, London, UK.
26. Departments of Biostatistics, Biomathematics and Human Genetics, University of California, Los Angeles, Los Angeles, CA, USA.

†contributed equally as first authors

‡contributed equally as senior authors

[§]Consortium members and affiliations are listed in the Supplementary Material.

*Corresponding authors: lok.tsui@new.ox.ac.uk, oliver.pybus@biology.ox.ac.uk and moritz.kraemer@biology.ox.ac.uk

Abstract

SARS-CoV-2 variants of concern (VOCs) arise against the backdrop of increasingly heterogeneous human connectivity and population immunity. Through a large-scale phylodynamic analysis of 115,622 Omicron genomes, we identified >6,000 independent introductions of the antigenically distinct virus into England and reconstructed the dispersal history of resulting local transmission. We estimate that by the time Omicron BA.1 was reported in southern Africa (November 22nd, 2021) six of the eight largest transmission lineages were already established in England. During that time

internationally well-connected hubs started acting as exporters of the variant which led to continued seeding of the VOC to England where it locally dispersed through the hierarchical travel network. Our results offer a detailed characterisation of processes that drive the invasion of an emerging VOC across multiple spatial scales. Genomic surveillance along the travelling network, coordinated and rapid decision making during the emergence of infectious diseases is necessary to delay their arrival.

One sentence summary: Omicron variant was introduced into England before detection in South Africa and rapidly disseminated from city to city.

Main Text

Since the emergence of SARS-CoV-2 in late 2019, multiple variants of concern (VOCs) have sequentially dominated the pandemic across the world. The Omicron variant (Pango lineage B.1.1.529 later divided into lineages including BA.1 and BA.2) was discovered in late November 2021, through genomic surveillance in Botswana and South Africa and a traveller from South Africa in Hong Kong (1) and designated a VOC by the World Health Organisation on 26 November 2021 (2). An initial surge in Omicron cases in South Africa indicated a higher transmission rate than previous variants (3), which studies later attributed to a shorter serial interval, increased immune evasion and greater intrinsic transmissibility (4–7). The mechanism for this greater transmissibility is hypothesised to be altered tropism and higher replication in the upper respiratory tract (8, 9). Together with waning levels of population immunity from previous infections and vaccination (10), local transmission of Omicron BA.1 was soon reported thereafter in major travel hubs worldwide, including New York City and London by early December 2021, despite travel restrictions on international flights from multiple southern African countries (11, 12).

Following the first confirmed case of Omicron BA.1 in England on 27 November 2021 (13), Omicron prevalence increased rapidly across all regions of England, with Greater London prevalence peaking first in mid-December at ~6%, followed by the South East region (14). Other metropolitan areas in the North West and North East saw similar but delayed increases in prevalence with observed peaks

between early- and mid-January 2022. Incidence of Omicron BA.1 had declined substantially in Greater London and other southern regions by early-January 2022, resulting in a gradient of decreasing prevalence from north to south England (15). This spatiotemporal pattern of early spread was also observed for the Alpha variant in England (16), but is markedly different from that of the Delta variant, which spread initially from the North West and surrounding regions of England (17). Rapid growth in infections during the initial emergence of Omicron in England prompted the UK government to impose interventions including a move to “Plan B” non-pharmaceutical restrictions (a mandatory COVID pass for entry into certain indoor venues, face coverings and work-from-home guidance) on 8 December 2021 (18) and an accelerated program of booster vaccination for all adults by mid-December 2021 (19). The prevalence of SARS-CoV-2 in England decreased later in January 2022, coincident with a falling proportion of BA.1 infections as BA.2 lineage replaced BA.1 as the dominant lineage, which itself was later replaced by the BA.4 and BA.5 lineages (20–22).

Understanding and quantifying the relative contributions of the factors that determined the arrival and spatial dissemination of Omicron BA.1 in England can help inform the design of spatially-targeted interventions against future VOCs (23). Here, we analyse the initial Omicron BA.1 wave in England, using a dataset of 48,748 Omicron BA.1 genomes sampled in England. This dataset represents ~1% of all confirmed Omicron BA.1 cases in England during the study period and is combined with sub-city level aggregated and anonymized human mobility and epidemiological data from 313 lower tier local authorities (LTLAs) in England.

International seeding events and Omicron BA.1 lineage dynamics

To investigate the timing of virus importations into England and the dynamics of their descendent local transmission lineages, we undertook a large-scale phylodynamic analysis of 115,622 SARS-CoV-2 Omicron genomes (BA.1/BA.2 and their descendent lineages), sampled globally between 8 November 2021 and 31 January 2022. About 42% (N=48,748) were sampled from England and sequenced by the COVID-19 Genomics UK (COG-UK) consortium (24). All available genomes (from COG-UK and GISAID (25) on 12 and 9 April 2022 respectively) collected before 28 November

2021 were included; genomes collected after that date were randomly subsampled in proportion to weekly Omicron case incidence while maintaining a roughly 1:1 ratio between English and non-English samples. To reduce potential bias caused by heterogeneous sequencing coverage, we performed a weighted subsampling of the English genomes using a previously developed procedure which accounts for variation in the number of sequences sampled per reported case at the Upper Tier Local Authority (UTLA) level (26) (supplementary material, materials and methods).

We identified at least 6,455 [95% HPD: 6,184 to 6,722] independent importation events. Most imports from outside of England (69.9% [95% HPD: 69.0 to 70.7]) led to singletons (i.e., a single genome sampled in England associated with an importation event, which did not lead to observable local transmission in our dataset). The earliest importation event is estimated between 5 and 18 November (approximated as the midpoint between the inferred times of the most recent common ancestor (MRCA) of the transmission lineage and the parent of the MRCA (PMRCA)). Between the first introduction event and mid-December 2021, we infer an approximately exponential increase in the daily number of imports before a plateau in early January 2022 (Fig. 1C). There is some indication that the daily importation rate was raised between 22 November, when Omicron was first reported and the start of travel restrictions (Fig. 1C). Increased outflow of passengers before (and possibly in anticipation of) travel restrictions has been reported for earlier waves of SARS-CoV-2 (16, 27). This rapid growth in importation continued despite restrictions on incoming international travel from 11 southern African countries and could have originated from Omicron outbreaks in other countries in late November and early December 2021. To explore this hypothesis, we calculate the Estimated Importation Intensity (EII) of Omicron BA.1 from countries with the highest air traffic volumes to the UK capturing 80% of total air travel. We aim to increase the resolution of the global scale analyses in Tegally et al. (2023) (28) by focusing on Omicron imports to England specifically. For each source location, the EII measure combines the weekly average COVID-19 test positivity rate, weekly relative prevalence of Omicron BA.1 genomes, and monthly number of observed air passengers travelling to England (see supplementary material, materials and methods for details and sensitivity analysis using case data and geographic disaggregations; Fig S4-S6). While the earliest imports were mostly inferred

to have come from South Africa, we observe a shift in Omicron BA.1 imports from South Africa to a larger set of countries, by late November/early December 2021 (Fig. 2), during the period of travel restrictions on South Africa. We performed a sensitivity analysis in which EIs are instead calculated using per capita case incidence rather than test positivity and the results are broadly consistent (Fig. S6).

We conclude the exponential growth of BA.1 importations through mid-December is therefore in part due to introductions from countries other than South Africa (Fig. 1B and Fig. 2), which became major contributors due to the Omicron epidemics there and the substantial volume of air travel to England (Fig. S5). At the time when travel restrictions to 11 southern African countries were announced, sequences of Omicron BA.1 from only four countries had been uploaded to GISAID (25). We note that our work is not designed to quantitatively assess the impact of these restrictions on infection numbers in England.

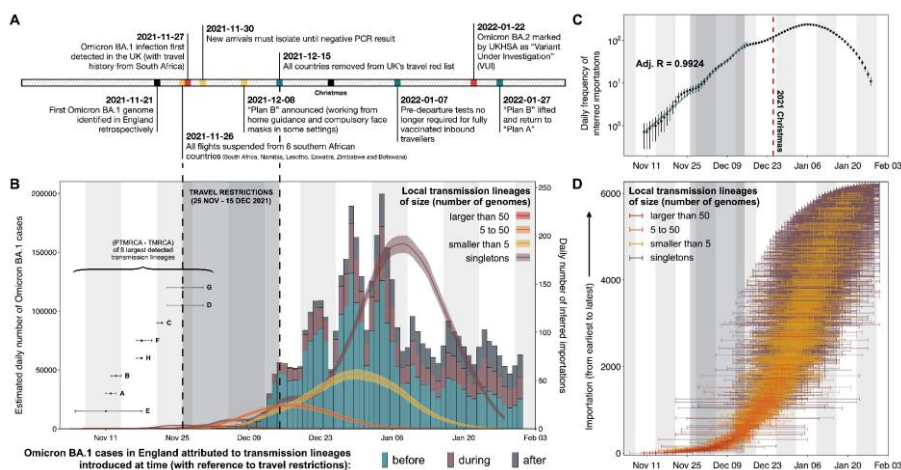


Fig. 1: Dynamics of Omicron BA.1 transmission lineages in England. (A) Timeline of key events during the SARS-CoV-2 Omicron BA.1 wave in England until February 2022. (B) Histogram of the estimated daily number of Omicron BA.1 cases, coloured according to the proportion of cases attributable to importation at different times (shaded region shows period of travel restrictions). Solid

lines represent the daily frequency of inferred importations (7-day rolling average), coloured according to the size of resulting local transmission lineages; shading denotes the 95% HPD across the posterior tree distribution. For each of the eight largest detected transmission lineages (labelled A to H), the estimated time of importation, TPMRCA (inferred time of parent of most recent common ancestor) and TMRCA (inferred time of most recent common ancestor) are shown in the bottom left of the panel. (C) Daily frequency of inferred importations (7-day rolling average), without stratification by size of resulting local transmission lineage (black dots); error bars denote the 95% HPD across the posterior tree distribution. Solid blue line represents the daily number of imports expected from an exponential model fitted to the observed 7-day rolling average importation intensity. (D) Distribution of TPMRCAs and TMRCAs of all 6,455 detected introductions. Each horizontal line represents a single introduction event that led to a transmission lineage or a singleton, with the left limit indicating the TPMRCA and the right limit indicating the TMRCA (or genome collection date for a singleton).

To cross-validate the importation dynamics inferred from viral genomes using an independent data source, we extracted data from the Variant and Mutations (VAM) line list (29) provided by the UK Health Security Agency (UKHSA) and calculated the daily number of incoming travellers who were later tested positive for Omicron BA.1 in community surveillance (Pillar 2) of the UK SARS-CoV-2 testing programme. The temporal profile of importation intensity from these epidemiological data is broadly consistent with that inferred from the phylodynamic analysis, with the latter being temporally expanded and lagged compared to the former (Fig. S2). This observation is consistent with previous studies (30) and the apparent discrepancy represents the time lag between international importation and the first local transmission event that is observable from phylogenetic data.

As with the emergence of previous variants in England (17, 30), we find that transmission lineage sizes are overdispersed (Fig. S3), with most sampled genomes belonging to a small number of large transmission lineages. The eight largest transmission lineages (each with >700 genomes) together comprise >60% of the genomes sampled in England in our dataset (Fig. 1B). Most of these (six of

eight) are inferred to have been imported before restrictions on travel from southern African countries were introduced (26 November 2021), and three could have been introduced before the first epidemiological signal of the new variant (an uptick in S-gene target failure, SGTF, samples identified by a private lab in South Africa on 15 November 2021 (Fig. 1B)). Additionally, we observe a strong association between the size and time of importation of local transmission lineages, with most large transmission lineages attributed to early introductions between 5 and 13 November 2021 (Fig. 1B). This pattern can be recapitulated using a simple mathematical model; if all lineages share the same transmission characteristics, then the date of importation is the main determinant of transmission lineage size when the epidemic in the recipient location is growing exponentially (see materials and methods).

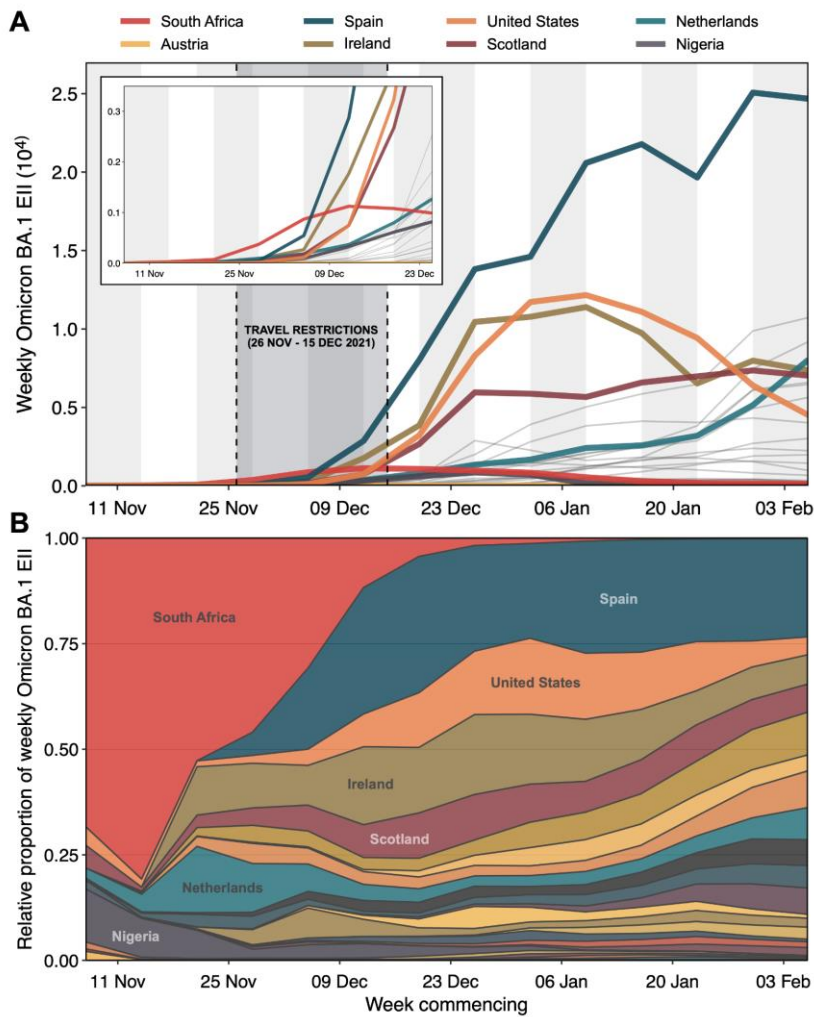


Fig. 2: Estimated Importation Intensity (EII) of Omicron BA.1 from selected potential exporters. Estimated weekly number of Omicron BA.1 cases arriving in England from 27 countries (including Scotland and Northern Ireland independently) with the highest air passenger volumes arriving in England between November 2021 and January 2022 (collectively accounting for ~80% of total air passenger volume in this period). Thick solid lines represent EIIs from eight selected countries with notable contribution to the overall importation intensity at different points during the study period; thin grey lines represent all other countries. Inset shows a magnified view of early trends. (B) Relative

proportion of weekly EII of Omicron BA.1 by country among selected potential exporters. Areas representing countries highlighted in (A) are labelled. See supplementary materials for sensitivity analyses.

We estimate that 399 transmission lineages (including the eight largest) resulted from importation events before the end of restrictions on travel from southern Africa (15 December 2021); 29 of these lineages were introduced before 26 November 2021. Although these early imports account for only a small proportion (~6%) of the estimated total number of introductions, they are responsible collectively for ~80% of Omicron BA.1 infections reported in England to the end of January 2022.

Some transmission lineages from early importations were only detected several weeks after their inferred time of importation. However, we interpret this result cautiously, as we cannot exclude the possibility that these transmission clusters represent the aggregation of multiple independent transmission lineages, as a result of unsampled genetic diversity outside England. Such aggregation would result in earlier estimated dates of importation, potentially explaining the smaller than expected size (compared to predictions from simulations; Fig. S7 and Fig. S8) of these transmission lineages with unusually long importation lag (30). Future analyses incorporating detailed metadata on travel history could help reduce the degree of uncertainty in the number and timing of inferred importation events (31, 32).

Human mobility drives spatial expansion and heterogeneity in Omicron BA.1 growth

The rapid increase in Omicron importations in late 2021 led to the establishment of local transmission chains, initially concentrated in Greater London and neighbouring LTLAs in the South West and East of England. This coincided with early increases in Omicron BA.1 prevalence in the corresponding regions, as observed from SGTF data and other epidemiological studies based on prevalence surveys (15). To further investigate the spatiotemporal dynamics of Omicron transmission lineages in England, we reconstructed the dispersal history of all identified transmission lineages (with >5 genomes) using spatially explicit phylogeographic techniques. Sampling of English genomes was

highly representative of the estimated number of Omicron BA.1 cases at the UTLA level (Fig. S9; comparison with modelled case incidence with adjustment for changes in case reporting is shown in Fig. S10).

We observe multiple distinct stages to the spread of BA.1 across England, with the eight largest identified transmission lineages sharing broadly similar patterns of spatial dispersal. Unlike other variants, we find that the numbers of transmission lineages first detected are fairly evenly distributed among regions, with ~20% in Greater London (followed by 15.4% in the South East and 13.3% in the North West; if only introductions prior to December 2021 are considered, the value for Greater London is 27.3%). However, most of the early cases outside Greater London resulted in limited local spatial diffusion (Fig. 3, Fig. S11 and Fig. S12).

Further, initial long distance viral lineage movements from Greater London repeatedly arrived in multiple urban (according to 2011 Rural-Urban Classification by the UK Office of National Statistics (33)) conurbations in early/mid-December 2021, but local transmission was not established immediately. The fraction of all viral lineage movements that were local (within-city) remained between 25%-50% from December 2021 to January 2022 in all areas except Greater London and Greater Manchester. This fraction grew when local mobility levels recovered after the holiday period (34–37), coinciding with the time when local transmission was established across most LTLAs in England. In contrast, between November and December 2021, local viral movements in Greater London and Greater Manchester comprised ~90% and ~60% of all movements respectively, indicating that epidemics in those locations were driven by multiple locally-established lineages (Fig. S11). Further, cities other than Greater London acted primarily as sinks throughout the BA.1 wave with limited backflow of long-distance viral lineages from North West England to Greater London (e.g. Transmission Lineage-A and Transmission Lineage-B; similar dynamics are also seen for the South West of England; Fig. 3E). We define locations as sinks/sources according to whether there was a net flow of viral lineages into/out of the location over the study period.

Even after the establishment of local transmission in most English LTLAs, Greater London continued to be a source of mid-to-long range viral lineage movements (Fig. 3E). This is expected given that Greater London is a major travel hub in England's mobility network (similar trends were observed during the Alpha wave in 2020 (16)). The importance of Greater London as a source of short range (<50 km) lineage movements declined through time (Fig. 3E, left-top) and we observe a secondary peak in the frequency of mid-to-long range movements (>50 km; Fig. 3E) driven predominantly by virus lineages emanating from the Midlands and southern England (Fig. 3E, middle and right). These observations are consistent with epidemiological data showing that most areas outside of southern England experienced a BA.1 incidence peak only in the last week of December 2021 or the first week of January 2022 (Fig. S13).

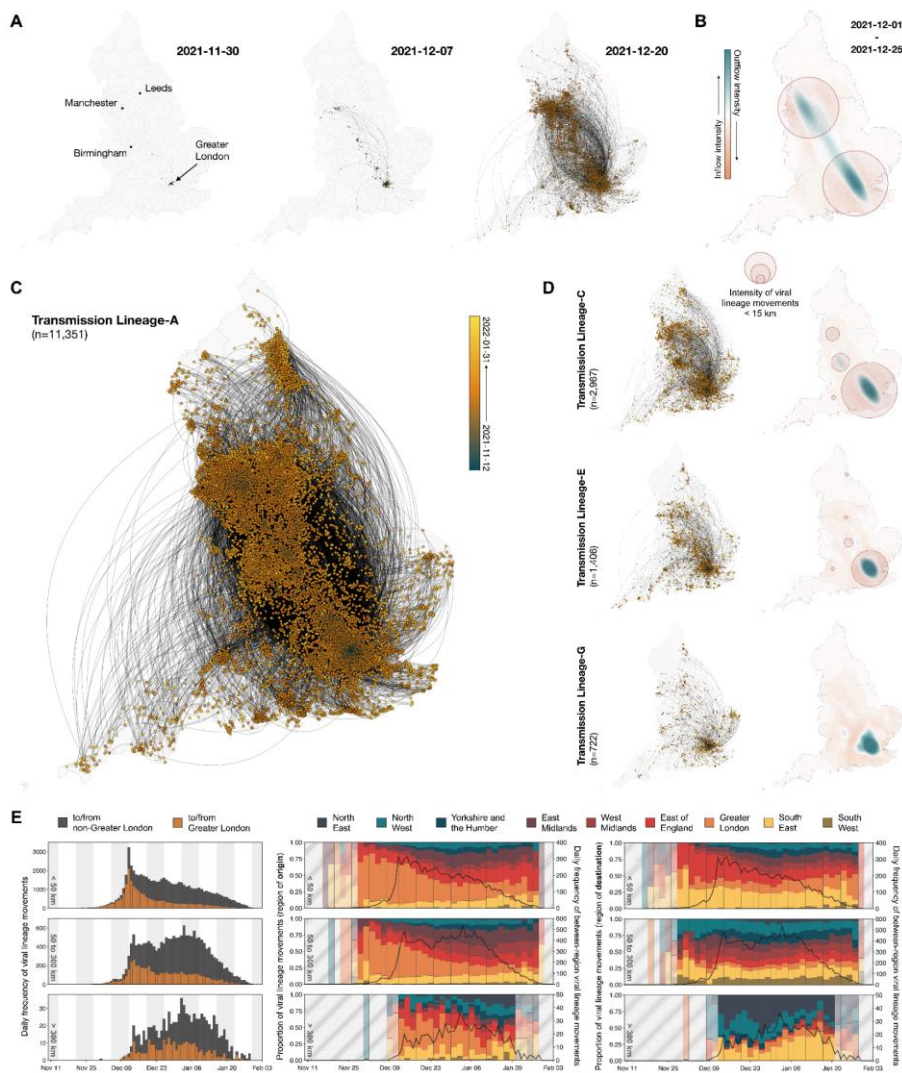


Fig. 3: Spatiotemporal dynamics of Omicron BA.1 transmission lineages in England. (A and C) Continuous phylogeographic reconstruction of the dispersal history of the largest detected BA.1 transmission lineages in England (Transmission Lineage-A). Nodes are coloured according to inferred date of occurrence and the direction of viral lineage movement is indicated by edge curvature (anti-clockwise). Panel A shows the progress of dissemination at three specific times, and panel C shows the complete construction. (B) Geographical distribution of the intensity of inflow and

outflow of viral lineages for Transmission Lineage-A from the beginning of December up to Christmas 2021. Blue colours indicate areas with high intensity of domestic lineage outflow; red colours indicate those with high intensity of inflow. Red circles indicate areas with high densities of local viral movements (distances <15 km); circle radii are proportional to that density. (D) Continuous phylogeographic reconstruction of Transmission Lineages C, E, and G (as per panel C) with corresponding maps of the geographical distribution of the intensity of viral lineage inflow and outflow (as per panel B). Fig. S12 provides an equivalent figure for Transmission Lineages B, D, F and H. (E) Plots in each row correspond to viral lineage movements across different spatial scales (top: <50 km, middle: 50 to 300 km, bottom: >300 km). (Left) Histograms show the daily frequency of viral lineage movements across spatial scales. Colours indicate whether the origin and/or destination of the viral lineage movements are inferred to have occurred in Greater London. (Middle/Right) Solid black lines represent the daily frequency of among-region viral lineage movements across spatial scales. Vertical bars indicate the proportions of viral lineage movements (aggregated at 2-day intervals); coloured according to their origin/destination locations. Shaded grey areas indicate periods when there were <9 inferred viral lineage movements per.

To assess the contribution of demographic, epidemiological and mobility-related factors to the dissemination of Omicron BA.1 in England, we used a discrete phylogeographic generalised linear model (GLM) to test the association of those factors with viral lineage movements among LTLAs, across two periods (before 26 December 2021, and between 26 December 2021 and 31 January 2022; supplementary materials) (35, 36, 38). Using this time-inhomogeneous model we find evidence for a dynamic spatial transmission process, with change through time in the estimated effect sizes of most predictors (Fig. 4B). During the earlier “expansion” period of lineage transmission among cities, we observe consistently strong support for the gravity model components (a spatial interaction model in which travel intensity between pairs of locations increases with origin and destination population sizes but decreases with distance between them). Consistent with results from continuous phylogeography, the early period is characterised by directional viral dissemination; lineage movements tend to

originate from Greater London (Fig 4b) and this is particularly pronounced for smaller transmission lineages (Fig. 3 and Fig. S12).

In three out of four analyses we also find greater dissemination out of LTLAs with earlier times of peak incidence during the expansion period and, conversely, a lower inflow of viral lineages during the post-expansion period in all analyses (Fig. 4 and Fig. S14). These results reflect the dynamic, network-driven nature of Omicron's geographic spread, with variation in the timing of peak incidence reflecting heterogeneity in the underlying human mobility network, i.e. varying degrees of connection to locations with frequent early seeding events (39).

Interestingly, the human mobility matrix predictor is supported consistently only in the post-expansion phase (Fig. 4B), after local transmission had been established in most LTLAs. This reflects a transition from unidirectional long-distance movements to more homogeneous local viral lineage movements. Conversely, support for the gravity model predictors decreased over time (Fig. 4B), consistent with the notion that the gravity model better predicts city-to-city mobility and poorly describes diffusion-like mobility over short distances in urban areas (40). Importantly, the phylogeographic GLM results are consistent among the transmission lineages analysed (Fig 4B), and also when a simpler time-homogenous model is used (Fig. S15). These findings are consistent with our continuous phylogeography analyses (Fig. 3) and with epidemiological studies showing strong local spatial structure of the Omicron BA.1 wave (14, 15). In a supplementary analysis, we included booster vaccine uptake (per capita at the LTLA level) as a predictor under a time-inhomogeneous model, but we did not find it to be significantly supported (supplementary material), possibly due to collinearity with other predictors (peak timing in case incidence and case-sample residuals) or due to limited spatial heterogeneity in vaccine uptake.

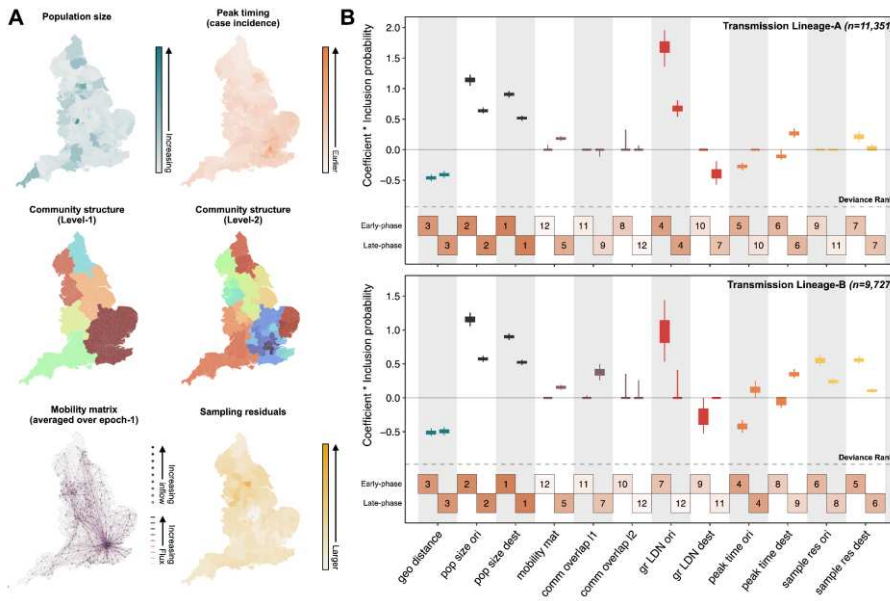


Fig. 4: Predictors of Omicron BA.1 viral lineage movements in England. (A) Map at LTLA level of model predictors included in the discrete phylogeographic GLM analysis, for the largest detected BA.1 transmission lineages (Transmission Lineage-A). (B) For each predictor, the box and whiskers show the posterior distribution of the product of the log predictor coefficient and the predictor inclusion probability; the left hand value represents the expansion period estimate and the right hand value the post-expansion period estimate. Top panel shows estimates for Transmission Lineage-A and bottom panel shows those for Transmission Lineage-B. Posterior distributions are coloured according to predictor type: geographic distances (geo distance, dark blue), population sizes at origin and destination (pop size ori & pop size dest, black), aggregated mobility (mobility mat, purple), mobility-based community membership level 1 and level 2 (comm overlap l1 & l2, purple), Greater London origin and destination (gr LDN ori & gr LDN dest, red), time of peak incidence at origin and destination (peak time ori & peak time dest, orange) and the residual of a regression of sample size against case count regression at either origin and destination (sample res ori & sample res dest, yellow). Boxes at the bottom of each panel are numbered and shaded to represent the rank of

predictors based on their deviance measure, with 1 indicating the largest (most important) and 12 indicating the smallest (least important).

Discussion

We find that a substantial proportion of SARS-CoV-2 infections during the Omicron BA.1 wave in England can be traced back to a small number of introductions inferred to have occurred before or during the early travel restrictions on incoming passengers from southern Africa. Although the rate of importation continued to increase after mid-December, local onward transmission was observed only for a proportion (~25%) of imports that arrived after Christmas 2021. These results augment previous investigations of VOCs in England and other countries (30, 41), highlighting that the impact of international travel restrictions is limited if applied after local exponential growth is established and in the absence of local control measures. Here we conclude that the epidemics of BA.1 in multiple locations outside the country of first detection substantially contributed to the exponential growth of BA.1 importation into England in December 2021 (28). Thus, the practical effect of targeted travel restrictions can be constrained by the existence of multiple pathways between any two countries in the global aviation network, often via highly-connected locations with large travel volumes that can act as early secondary sources (39). UK travel restrictions were intended to delay the rapid expansion of BA.1 locally while offering additional vaccination to at-risk individuals. However, it is likely that Omicron had already spread internationally by the time it was detected in late November 2021, allowing secondary locations of VOC export to become established (28, 42). Therefore any proposed global systems that intend to rapidly detect and respond to new VOCs (and emerging infectious diseases in general) need to be designed around the connection structure of human mobility networks. Despite this, there are likely to be scenarios under which travel restriction can help control, contain, or delay the spread of emerging infections (43, 44) and much further theoretical and empirical work is needed to improve and inform rapid decision making concerning travel during public health emergencies.

Our continuous and discrete phylogeographic analyses (Fig. 3 & 4) jointly show how Omicron BA.1 disseminated rapidly across England, with Greater London playing a central role in its initial dissemination. Early viral movements outside of Greater London were dominated by medium-to long-distance travel from there; local transmission chains in recipient locations were observed later, coinciding with an increase in human mobility after the winter holidays (Fig. S18). The epidemic is revealed to be a network-driven phenomenon with an initial expansion phase that is well described by a gravity model, followed by a period of sustained local transmission propagated by local human mobility (39).

With this study, we can now compare the transmission histories of three successive VOC waves in England (Alpha (16), Delta (17), and Omicron) and contrast the factors that influenced their dispersals. First, Omicron and Delta were both introduced through international importation, whereas Alpha appeared to have originated in England (45). For both Omicron and Delta, early introductions from their presumed location of origin were followed by an increase in importation intensity from secondary locations. However, early transmission clusters for Delta were observed mainly in North West England, whereas most early Omicron infections were found in Greater London (15, 20). Second, different local NPIs and restrictions on within-country travel were implemented during the VOC waves. Although the introduction of Delta occurred during a period of relaxed NPIs, initial spreading was delayed due to a lower level of mobility following a national lockdown (17), whereas Omicron was introduced when human mobility had mostly recovered to a pre-pandemic level (Fig. S18). For the Alpha wave, rapid expansion from the South East was observed as a result of holiday travels (16) and was subsequently brought under control when NPIs were introduced, leading to reduced levels of local mobility (16). Third, spatial variations in population immunity from prior infections are likely to have impacted the dissemination of each VOC differently. For example, we expect the spread of Delta to be relatively unaffected by population immunity due to widespread infections and vaccination, and similarly for Omicron due to the antigenic novelty of the variant (9, 46, 47); whereas the initial growth rates of Alpha were affected by local variations in previous attack rates (16). These findings highlight two key questions for future work: how do the spatiotemporal interactions between importation and local transmission shape the spread of an invading VOC, and

how can we efficiently evaluate the interplay of factors that drive the dissemination of an emerging VOC within a country.

Findings from our phylodynamic analysis should be interpreted in the context of several limitations. First, as discussed previously (30), the number of importation events identified is likely to underestimate the true number of independent introductions due to incomplete sampling and uneven sequencing coverage worldwide (48). Nevertheless, we were able to cross-validate our phylogenetic results using independent epidemiological data (Fig. S7 and Fig. S8). Second, to maintain computational tractability and remove potential sampling bias in the phylogeographic reconstruction of local transmission lineages, we included only a subset (about 7%) of the available English SARS-CoV-2 genomes from COG-UK, while accounting for geographical variations in sequencing coverage and COVID-19 prevalence. Despite this subsampling procedure, we note that spatial and temporal sampling was not perfectly representational (Fig. 4A and Fig. S9). This could be caused by geographical variation in case reporting rate or because the maximum sequencing capacity was exceeded in locations with exceptionally high case incidence. Third, our phylogenetic GLM analysis that explores the association of factors with virus lineage movement should be interpreted in light of potential biases in the mobility data. For example, movements in sparsely populated locations may be poorly captured due to censoring to protect user anonymity, and the degree to which smartphone mobility data is representative of the whole population could be affected by variation in smartphone use among locations. Work is ongoing to assess the benefit of human mobility data in the prediction and description of infectious diseases invasion dynamics.

Omicron BA.1 was replaced by lineage BA.2 in February 2022 and later by lineage BA.5 in June 2022 (20, 21). While the public health burden of COVID-19 has lessened due to reduced average disease severity and increased population immunity, the continued antigenic evolution of SARS-CoV-2 means that future variants with increased virulence remain possible. One priority in preparing for the next SARS-CoV-2 variant or novel pathogen emergence is to develop and implement robust pipelines for large-scale genomic and epidemiological analyses supported by unified data

infrastructures (49, 50) a challenging task that will be realised only through the coordination of public health efforts worldwide.

Acknowledgements: We acknowledge the UK Health Security Agency (UKHSA), members of the COVID-19 Genomics UK (COG-UK) consortium, NHS labs, GISAID contributors (acknowledgment table of genomes used is provided on our GitHub repository) for sharing of genomic data of Omicron BA.1. **Funding:** COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) [grant code: MC_PC_19027], and Genome Research Limited, operating as the Wellcome Sanger Institute. The authors acknowledge use of data generated through the COVID-19 Genomics Programme funded by the Department of Health and Social Care. M.U.G.K. acknowledges funding from The Rockefeller Foundation, Google.org, the Oxford Martin School Pandemic Genomics programme (also O.G.P. and A.E.Z.), European Union Horizon 2020 project MOOD (#874850) (also supports R.I., S.D. and P.L.), The John Fell Fund, and a Branco Weiss Fellowship. This work was also supported by the Foreign, Commonwealth & Development Office and Wellcome [225288/Z/22/Z] and [226052/Z/22/Z] (to M.U.G.K.). V.H. was supported by the Biotechnology and Biological Sciences Research Council (BBSRC) [grant number BB/M010996/1]. J.T.M, R.C. and A.R. acknowledge support from the Wellcome Trust [Collaborators Award 206298/Z/17/Z - ARTIC network]. S.D. also acknowledges support from the *Fonds National de la Recherche Scientifique* (F.R.S.-FNRS, Belgium; grant n°F.4515.22) and from the Research Foundation – Flanders (*Fonds voor Wetenschappelijk Onderzoek – Vlaanderen*, FWO, Belgium; grant n°G098321N). A.R. is also supported by the European Research Council [grant agreement number 725422 - ReservoirDOCS] and Bill & Melinda Gates Foundation [OPP1175094 – HIV-PANGAEA II]. J.L.H.T is supported by the Yeotown Scholarship from New College, University of Oxford. S.B. is supported by the Clarendon Scholarship, University of Oxford and NERC DTP [grant number NE/S007474/1]. M.A.S. acknowledges support from US National Institutes of Health grants R01 AI153044 and R01 AI162611. M.A.S. and X.J. gratefully acknowledge support from NVIDIA Corporation and Advanced Micro Devices, Inc. with the donation of parallel computing resources used for this research. S.C. acknowledges Labex IBEID

(grant ANR-10-LABX-62-IBEID), European Union Horizon 2020 projects VEO (874735) and RECOVER (101003589), AXARF, Groupama, EMERGEN (ANRS0151) and INCEPTION (PIA/ANR-16-CONV-0005). E.W., H.T. and T.dO are supported in part by grants from the Rockefeller Foundation (HTH 017), the Abbott Pandemic Defense Coalition (APDC), the African Society for Laboratory Medicine, the National Institute of Health USA (U01 AI151698) for the United World Antivirus Research Network (UWARN) and the INFORM Africa project through IHVN (U54 TW012041). The views expressed are those of the author and not necessarily those of the Department of Health and Social Care, UKHSA, or European commission or any of the other funders.

Author contributions: J.L.H.T., O.G.P., M.U.G.K. conceived and planned the research. J.L.H.T., S.B., B.L., V.H., J.T.M., P.B., R.E.P., J.R., P.L., S.D. analysed the data. A.E.Z., T.P.P., B.J., R.C., A.O'T, A.D., J.T.M., B.L., S.C., S.D., J.R., X.J., M.A.S., M.U.G.K., O.G.P., A.R. advised on methodologies. I.I.B., K.K., M.D. contributed international flight passenger data. J.L.H.T., O.G.P., M.U.G.K. wrote the initial manuscript draft. All authors edited, read and approved the manuscript.

Competing interests: K.K. is the founder of BlueDot, a social enterprise that develops digital technologies for public health. M.D. is employed at BlueDot. All other authors declare no competing interests. L.L. and N.W. are employed by Google and own equity in Alphabet. J.T.M is employed by Helix.

Data and materials availability: UK genome sequences used were generated by the COVID-19 Genomics UK consortium (COG-UK, <https://www.cogconsortium.uk/>). Data linking COG-IDs to location have been removed to protect privacy, however if you require this data please visit

<https://www.cogconsortium.uk/contact/> for information on accessing consortium-only data. The

Google COVID-19 Aggregated Mobility Research Dataset used for this study is available with

permission from Google LLC. Code to reproduce the analyses will be made available on our [GitHub repository](#). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY

4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit

<https://creativecommons.org/licenses/by/4.0/>. This license does not apply to figures/photos/artwork or

other content included in the article that is credited to a third party; obtain authorization from the

rights holder before using such material.

Commented [1]: TODO

Supplementary Materials

Materials and Methods

Figs. S1 - S18

Tables S1 - S4

References (44 - 65)



Supplementary Materials for

Genomic assessment of invasion dynamics of SARS-CoV-2 Omicron BA.1

Joseph L.-H. Tsui^{1,†*}, John T. McCrone^{2,3,†}, Ben Lambert^{4,†}, Sumali Bajaj^{1,†}, Rhys P.D. Inward^{1,†}, Paolo Bosetti^{5,†}, Houriiyah Tegally^{6,7}, Verity Hill^{3,8}, Rosario Evans Pena¹, Alexander E. Zarebski¹, Thomas P. Peacock^{9,10}, Luyang Liu¹¹, Neo Wu¹¹, Megan Davis¹², Isaac I. Bogoch¹³, Kamran Khan^{12,13}, Meaghan Kall¹⁰, Nurin Iwani Binti Abdul Aziz¹⁰, Rachel Colquhoun³, Áine O'Toole³, Ben Jackson³, Abhishek Dasgupta¹, Eduan Wilkinson^{12,13}, Tulio de Oliveira^{6,7}, The COVID-19 Genomics UK (COG-UK) consortium[†], Thomas R. Connor^{14,15,16}, Nicholas J. Loman¹⁷, Vittoria Colizza¹⁸, Christophe Fraser^{19,20}, Erik Volz²¹, Xiang Ji²², Bernardo Gutierrez¹, Meera Chand¹⁰, Simon Dellicour^{23,24}, Simon Cauchemez^{5,†}, Jayna Raghwanjani^{1,25,†}, Marc A. Suchard^{26,†}, Philippe Lemey^{24,†}, Andrew Rambaut^{3,†}, Oliver G. Pybus^{1,20,25,†*}, Moritz U.G. Kraemer^{1,20,†*}

†contributed equally as first authors

‡contributed equally as senior authors

[†]Consortium members and affiliations are listed in the Supplementary Material.

*Corresponding authors: lok.tsui@new.ox.ac.uk, oliver.pybus@biology.ox.ac.uk and moritz.kraemer@biology.ox.ac.uk

The PDF file includes:

Materials and Methods
Figs. S1 - S18
Tables S1 - S4
References

Materials and Methods

Genomic data

All SARS-CoV-2 sequences used in this study were downloaded on 12 April 2022. All available international (non-England, including Wales, Scotland and Northern Ireland independently) sequences were downloaded from GISAID (25) while English samples marked as community surveillance (pillar 2) were acquired from COG-UK. Historically, pillar 2 testing sites were instructed to select a number of 96 well plates for sequencing proportional to the fraction of total tests that week. Pillar 2 surveillance is intended to represent a random sample of community cases in the UK, with only 8% of being associated with testing for special reasons, i.e. 'attended-event', 'attended-outbreak-venue', 'confirmatory-test-borders', 'contact-testing-study', 'test-for-contact-self-referral', 'test-for-contact-tracing', 'test-for-contact-tracing-app', 'venue-outbreak'. However, given the changes in testing behaviour and regulations that occurred during the study period we can not rule out the possibility that there are some biases in the data set. These were partially addressed in the subsampling mentioned below.

Sequences were aligned and filtered as part of the COG-UK datapipe analysis hosted by CLIMB. This analysis removed duplicate and environmental sequences, and flagged samples with improbable collection dates (see <https://github.com/COG-UK/datapipe> for details). All sequences with impossible or improbable collection dates were removed. To further minimise dating errors caused by retrospective sequencing, only samples published to COG-UK or GISAID (25) within four weeks of sample collection were included. Sequences were aligned to the reference Wuhan-Hu-1 (genbank accession MN908947.3) with minimap2 and samples with less than 93% coverage were discarded. Sequence coverage weights were calculated for English sequences (https://github.com/robj411/sequencing_coverage) to ensure they could be subsampled proportional to the number of reported cases in each Upper Tier Local Authority using a two week sliding window. Scorpio was run as part of Pangolin and sequences identified as BA.1 or BA.2 were selected for further analysis.

Variant and Mutations (VAM) line list (from UK Health Security Agency)

The variant and mutations (VAM) line list compiled and provided by the UK Health Security Agency (UKHSA) contains epidemiological metadata of specimens sequenced from the pillar 2 mass testing programme by the COVID-19 Genomics UK Consortium (COG-UK). From the variant line list we extracted the traveller status (Traveller, Contact of Traveller, Not Travel-associated, Refused or Uncontactable, Awaiting Information) and specimen date of all sampled individuals who were tested positive for Omicron B.1.1.529 (BA.1) between 1 November 2021 and 31 January 2022.

Estimated Omicron BA.1 case incidence (from COVID-19 case count and S-gene target failure data)

Daily number of new COVID-19 cases by specimen date in each LTLA were downloaded from <https://coronavirus.data.gov.uk/details/download> (last assessed on 26 June 2022). S-gene target failure data were provided by UKHSA via a data sharing agreement. The presence of a genetic deletion on the spike protein of the Omicron BA.1 sub-variant produces SGTG in most PCR tests which can be used as a proxy for BA.1 infections. We used daily SGTF PCR-positive tests as a proxy (because these were time- and cost-effective as a test compared to genetic sequencing to ascertain variants) for Omicron BA.1 infection in conjunction with reported case data to estimate daily number of new BA.1 cases. However, small sample sizes in the SGTF dataset could lead to extreme scaling, i.e. zero or 100% of cases could be attributed to BA.1 infections if for example none or all samples were SGTF positive. Hence we calculated BA.1 cases in a Bayesian framework using uninformative Beta(1, 1) priors and the observed proportion of BA.1 infections (from the SGTF dataset) to estimate the

posterior proportion of BA.1 cases which was then scaled up by the number of reported cases from the coronavirus data download. We can also use the estimated uncertainty from the posterior distribution to get lower and upper bounds in the scaled up BA.1 case numbers.

International passenger flight data arriving in England

We evaluated travel data generated from the International Air Transport Association (IATA) to quantify passenger volumes originating from international airports and arriving in England. IATA data accounts for approximately 90% of passenger travel itineraries on commercial flights, excluding transportation via unscheduled charter flights (the remainder is modelled using market intelligence).

Estimated importation intensity of Omicron BA.1 from potential exporters

We estimated and compared the weekly importation intensity of SARS-CoV-2 Omicron BA.1 from 27 countries (including Scotland and Northern Ireland independently) with the highest air passenger volumes arriving in England between 7 Nov 2021 and 26 March 2022 (collectively accounting for ~80% of the total air passenger volume during this period). The weekly importation intensity is an estimate of the number of Omicron BA.1 cases imported during a given week from a specified source location, calculated by multiplying together the estimated weekly relative prevalence of Omicron BA.1 at the source location and the number of air passengers arriving from the source location.

We estimated the weekly number of air passengers arriving in England using monthly air traffic data, assuming a uniform daily distribution of passengers throughout the month and aggregating to a weekly level. To account for potential biases that might result from differences in case reporting rate, we used test positivity as a proxy for the underlying weekly prevalence at the source locations. Daily test positivity rates at the country level were downloaded from OWID (<https://ourworldindata.org/>; last accessed on 3 April 2023) and their weekly averages were computed. For Scotland and Northern Ireland, daily test positivity rates were downloaded from GOV.UK COVID-19 Dashboard (<https://coronavirus.data.gov.uk/>; last accessed on 23 April 2023). To account for local (within-country) heterogeneities in Omicron BA.1 prevalence and air traffic volume, we calculated EIIs from Spain and the United States at the autonomous community- and state-level. Weekly test positivity rates for autonomous communities in Spain were downloaded from the European Centre for Disease Prevention and Control Data Dashboard (<https://www.ecdc.europa.eu/en/publications-data/archive-historical-data-testing-volume-covid-19>; last assessed on 22 April 2023). For the US, weekly test positivity rates at the state level were calculated using data from https://github.com/govex/COVID-19/blob/master/data_tables/testing_data/time_series_covid19_US.csv (last assessed on 4 April 2023). We note that three different approaches were used to calculate test positivity for the US states depending on the availability of different test statistics. The three different approaches are: (A) positive specimens / total specimens, i.e. the number of positive PCR tests divided by the total number of PCR tests given, (B) positive people / total encounters, i.e. number of people who tested positive (PCR) divided by the total number of PCR tests given, and (C) positive people / total specimens, i.e. the number of people who tested positive (PCR) divided by the total number of PCR tests given. For states where multiple measures of the positivity rates are possible, the optimal approach was applied according to the order (A), (B) and (C), with approach (A) being the optimal approach. For Washington state (WA) in particular, no appropriate measure of the denominator in the calculation of positivity rate is available using any of the approaches, and as a result the national average positivity rates were used instead. We also note that no reliable testing data for Egypt can be found; Egypt is therefore omitted in this EII analysis and only included in subsequent sensitivity analysis as detailed below.

In a sensitivity analysis, we separately calculated EIs using weekly COVID-19 case incidence per capita as a proxy for prevalence at the source locations (Fig. S6). Data sources are the same as in the analysis using test positivity rates (with the inclusion of Egypt using data from OWID (<https://ourworldindata.org/>; last accessed on 3 April 2023)). Similarly, EIs for Spain and the US were calculated at the autonomous community- and state-level. To highlight the potential bias that might have resulted from variations in case reporting rate between countries, the weekly number of tests performed per capita is calculated for each country (Fig. S4)

UK population estimates

Mid-year population estimates for England in 2020 at the LTLA level were downloaded from <https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland>. Population sizes were used as the denominator in calculating numbers of COVID-19 cases per capita and normalised local mobility in each LTLA.

Vaccination data with age breakdown

Daily vaccination data with age breakdown at the lower tier local authority (LTLA) level were downloaded from <https://coronavirus.data.gov.uk/metrics/doc/vaccinationsAgeDemographics>. The dataset consists of daily cumulative number and percentages of people who have received either a 1st dose, 2nd dose, or booster dose (of any type) since the start of the pandemic in each LTLA, with age breakdown by roughly 5-year intervals (5-11, 12-15, 16-17, 18-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85-89, 90+).

Aggregated and anonymised human mobility data

We used the Google COVID-19 Aggregated Mobility Research Dataset described in detail in (51, 52), which contains anonymized relative mobility flows aggregated over users who have turned on the *Location History* setting, which is turned off by default. This is similar to the data used to show how busy certain types of places are in Google Maps — helping identify when a local business tends to be the most crowded. The mobility flux is aggregated per week, between pairs of approximately 5km² cells worldwide, and for the purpose of this study further aggregated for LTLAs in the United Kingdom (<https://geoportal.statistics.gov.uk/datasets/lower-tier-local-authority-to-upper-tier-local-authority-december-2016-lookup-in-england-and-wales/explore>) for the time period of November 2019 to January 31st, 2022.

To produce this dataset, machine learning is applied to log data to automatically segment it into semantic trips. To provide strong privacy guarantees (53), all trips were anonymized and aggregated using a differentially private mechanism to aggregate flows over time (see <https://policies.google.com/technologies/anonymization>). This research is done on the resulting heavily aggregated and differentially private data. No individual user data was ever manually inspected, only heavily aggregated flows of large populations were handled. All anonymized trips are processed in aggregate to extract their origin and destination location and time. For example, if n users travelled from location a to location b within time interval t , the corresponding cell (a,b,t) in the tensor would be $n \mp \text{err}$, where err is Laplacian noise. The automated Laplace mechanism adds random noise drawn from a zero mean Laplacian distribution and yields (ϵ, δ) -differential privacy guarantee of $\epsilon = 0.66$ and $\delta = 2.1 \times 10^{-29}$ per metric. Specifically, for each week W and each location pair (A,B) , we compute the number of unique users who took a trip from location A to location B during week W . To each of these metrics, we add Laplace noise from a zero-mean distribution of scale $1/0.66$. We then remove all metrics for which the noisy number of users is lower than 100, following

the process described in (53), and publish the rest. This yields that each metric we publish satisfies (ϵ, δ) -differential privacy with values defined above. The parameter ϵ controls the noise intensity in terms of its variance, while δ represents the deviation from pure ϵ -privacy. The closer they are to zero, the stronger the privacy guarantees.

These results should be interpreted in light of several important limitations. First, the Google mobility data is limited to smartphone users who have opted in to Google's *Location History* feature, which is off by default. These data may not be representative of the population as whole, and furthermore their representativeness may vary by location. Importantly, these limited data are only viewed through the lens of differential privacy algorithms, specifically designed to protect user anonymity and obscure fine detail. Moreover, comparisons across rather than within locations are only descriptive since these regions can differ in substantial ways.

Changes in case reporting rate in the United Kingdom

To assess the degree of changes in case reporting rate in the United Kingdom, we compared the weekly national case incidence downloaded from the GOV.UK COVID-19 Dashboard with that estimated by the UK Office of National Statistics (ONS). Specifically, since we were interested in the relative changes over time rather than the absolute values, a linear regression of the ONS case incidence estimates against case incidence from the GOV.UK COVID-19 Dashboard was performed and the residuals from the model were examined (Fig. S10).

As described in further details below, Omicron sequences from England were subsampled with sample weights calculated from the ratio between the cumulative number of reported cases and the cumulative number of sequences collected in the preceding two weeks for any given date. Therefore, to assess the potential bias that might have resulted from changes in case reporting rate in the context of the subsampling of English genomes, a similar linear regression analysis as above was performed, with additional (two-preceding-weeks) smoothing applied to both the ONS and GOV.UK case incidences (Fig. S10).

Phylogenetic analysis and importation analysis

We developed a large-scale phylogenetic analysis pipeline following a similar approach as in du Plessis et al. (2021) (54) with additional extensions and modifications to ensure the computational tractability of analyses of up to hundreds of thousands of SARS-CoV-2 sequences (55) (Fig. S1).

First, the study period was divided into two phases: (1) from 21 November 2021 (sample date of the earliest known genome of the Omicron variant in England, sequenced retrospectively) to 28 November 2021, and (2) from 29 November 2021 to 31 January 2022. The time of division between the two phases was chosen on the basis of an expected change in importation intensity as a result of the implementation of travel restrictions targeted at multiple southern African countries starting on 28 November 2021. With the relatively few genomes available from the first phase and to account for an increased risk of importations prior to the travel restrictions, all 874 available sequences (from both England and non-England locations) were included. Owing to the large number of genome samples collected during the second phase, a downsampling strategy was applied to ensure that the analysis was computationally tractable. To generate a manageable dataset of global sequences, first we computed a crude estimate of the number of new Omicron cases in each country in each epi-week by multiplying the number of reported COVID-19 cases (downloaded from <https://github.com/owid>; last accessed on 4 May 2022) by the proportion of sampled genomes that were of the Omicron variant PANGO lineages BA.1 and BA.2, using metadata available from GISAID (25) (<https://gisaid.org/>;

last accessed on 12 April 2022). The number of global sequences to be sampled in each epi-week was then allocated in proportion to the estimated total number of Omicron lineages BA.1 and BA.2 cases in the week whilst maintaining a dataset size of ~50,000. In a given epi-week, countries with an estimated number of Omicron cases that accounted for at least 0.5% of the estimated global total were considered as potential exporters. Genome samples were then allocated in proportion to the estimated number of cases among these potential exporters, with the remaining allocation randomly distributed among the non-exporter countries. There was a slight enrichment for samples collected in the early phase of the Omicron wave (early December 2021), where we ensured that a minimum of 4,000 genomes were sampled for each epi-week where available. A similar approach was used to curate a dataset of 21,039 Omicron genomes sampled from Wales, Scotland and Northern Ireland, again using relevant metadata from GISAID (25) and epidemiological data available on (<https://api.coronavirus.data.gov.uk/v1/data>; last accessed on 4 May 2022). This downsampling procedure resulted in a dataset of 59,647 global (non-English) sequences. To generate a dataset of English genomes of roughly the same size, 60,000 sequences were randomly sampled from the COG-UK master alignment whilst accounting for variations in sequencing coverage and prevalence amongst UTLAs over time, using the same method as in Volz et al. (56). This resulted in a combined dataset of 140,686 genomes of which 42.6% were sampled in England with the remaining from non-England locations.

Despite substantial downsampling, estimating a phylogenetic tree for hundreds of thousands of SARS-CoV-2 sequences remains a challenge, with most standard programs only able to handle up to thousands of sequences. To tackle this, we first estimated a maximum likelihood (ML) tree for the 874 sequences collected during the first phase of the study period using IQTREE (57) with the GTR+G substitution model, rooted with reference genome Wuhan-Hu-1 (GenBank accession MN908947.3) as an outgroup. Five molecular clock outliers were identified and subsequently removed, after examining the root-to-tip regression plot from TreeTime (58). The resulting tree was then used as a starting tree from which a parsimony tree was estimated by inserting individual sequences sequentially and in chronological order according to sample dates, using the recently developed USHER placement tool (59). During each step in the iterative process, all sequences sampled on a given date were considered for placement whilst excluding sequences with 5 or more equally parsimonious placements. Sequences excluded in a previous step were appended to the next batch for reconsideration. The resulting tree was then optimised through 6 iterations of matOptimize (60) with SPR radius of 40 and 100 for the first 5 and final iteration respectively. This iterative tree building process resulted in a phylogeny of 115,634 sequences (with 25,921 (18.3%) sequences excluded due to uncertainty in sample placement). Next we used Chronumetal (61) (a recently developed time-tree estimation tool for handling large phylogenies) to estimate a randomly resolved time-calibrated tree, with inferred tip dates that maximise the evidence lower bound under a probabilistic model. By comparing the inferred tip dates with sample dates and examining a root-to-tip plot, 12 molecular clock outliers were further removed, resulting in a final phylogeny of 115,622 sequences.

To further reduce the computational resources and time required, we divided the phylogeny estimated above into smaller tree partitions according to sub-lineage (of Omicron) assignment as defined by the Pango nomenclature (62). Using a custom Python script, subtrees with a high degree of clustering of sequences of the same descendant lineage of Omicron were identified, whilst accounting for some level of ambiguity in lineage assignment (e.g. a tree partition may contain up to 25% of sequences that are of a minority sub-lineage before it is subdivided into multiple partitions), as would be expected given the high sampling density and variations in sequencing quality. Further merging of these

identified subtrees resulted in five final tree partitions, labelled BA.1 (n=38,522), BA.1.1 (n=37,028), BA.1.15 (n=12,229), BA.1.17 (n=21,549), and BA.2 (6,294) according to the sub-lineage represented by the majority of sequences in each partition. Given that the primary focus of this study is the invasion dynamics of Omicron BA.1 in England, the BA.2 partition was omitted in all further downstream analyses.

Having divided the phylogeny into smaller tree partitions of computationally manageable size, we then performed time-calibration of the subtrees using a recently implemented model in BEAST v1.10 (63) which replaces the traditional tree-likelihood with a more efficient likelihood based on a simple Poisson model, thus allowing Bayesian phylogenetic analyses of up to tens of thousands of sequences. In this approach, the tree operators are constrained such that only node heights and polytomy resolutions are sampled, whilst the tree topology is fixed to that of a data tree which we generated using Treetime (58) with a fixed clock rate of 7.5×10^{-4} substitutions/site/yr. Using a Skygrid coalescent tree prior (64) with grid points every two weeks, we ran between 2 and 6 MCMC chains of 3×10^8 to 2.4×10^9 iterations for each tree partition independently. The first 33% to 40% of each chain was discarded as burn-in and resampled every 1×10^6 to 2.4×10^8 states before merging using LogCombiner, resulting in 1,200 posterior tree samples for each tree partition. Model convergence and mixing was assessed using Tracer (65).

To reconstruct the importation dynamics of Omicron BA.1, we then used a two-state asymmetric discrete trait analysis (DTA) model implemented in BEAST v1.10 (63), using the posterior tree samples estimated above as the empirical tree distributions. For each tree partition, we ran two MCMC chains of 5 million iterations each, resampled every 9,000 states and with the first 10% discarded as burn-in. TreeAnnotator 1.10 (63) was used to generate a maximum clade credibility (MCC) tree for each subtree, in which each internal node is assigned a posterior probability of representing a transmission event in England. Nodes with a posterior probability of >0.5 were identified as introductions; a small number of nodes with ambiguous location assignment (posterior probability = 0.5) were ignored in downstream analyses. To identify the local transmission lineage resulting from each of the introductions, a depth-first search was performed following the same procedure as in du Plessis et al. (2021) (54), where a path starting from each internal node that corresponds to an introduction is traversed forwards in time until a non-England node is encountered or there are no more nodes to be explored. By convention, introductions that led to only a single sampled English sequence were labelled as singletons; only introductions that led to more than one observed local transmission event were labelled as transmission lineages. The time of importation of each transmission lineage was estimated by taking the mid-point between the internal node corresponding to the introduction and its parent.

Our methodology estimating the time of importation of transmission lineages is likely to result in an apparent “expansion” of the temporal profile of inferred importation intensity (daily number of infected travellers arriving in England) relative to its true underlying distribution. This could be explained by an increase in importation lag (time elapsed between when a lineage is inferred to have been imported and the first detected local transmission event) over time as shown previously by du Plessis et al. (30), due to transmission lineages from later importations having fewer genomes as they have less time to grow, and are therefore less likely to be captured by genomic surveillance assuming a constant sampling intensity. To verify this effect, we compared the inferred importation intensity from the above phylodynamic analysis with empirical observations from testing data recording international travellers, extracted from the Variant Mutation line list provided by UKHSA (Fig. S2).

However, we note that the robustness of this comparison is potentially limited by variations in sampling intensity in the epidemiological data as a result of rapidly-changing testing policies for arriving travellers in the United Kingdom during the latter part of January 2022 (66).

Exponential growth of daily frequency of importations

In the absence of any travel restrictions and changes in human mobility as a result of the emergence of a new VOC, the importation intensity during the initial phase of the invasion would be expected to follow a pattern of exponential growth that mirrors the increase in number of infections in the exporting countries. To verify and examine any potential deviation from this pattern, we fitted a simple exponential model to the 7-day rolling average daily number of importations inferred from the phylodynamic analysis. Specifically, we fitted the model using least-squares regression to the inferred daily numbers of importations during the period between the beginning of November 2021 and a range of cut-off dates. The cut-off date that resulted in the highest adjusted R^2 value can be interpreted as an estimate of the time when the growth of importation intensity began to deviate from an exponential trajectory.

Continuous phylogeographic analysis

To reconstruct the spatiotemporal patterns of the Omicron BA.1 wave in England, all local transmission lineages (as identified from the MCC trees generated from the 2-state discrete trait analysis above) with five or more sequences were extracted for continuous phylogeographic analyses. Each sequence was assigned a latitude and a longitude randomly from within the postal district (metadata provided by COG-UK) where the sample was collected. For each transmission lineage, we performed the continuous phylogeographic reconstruction on a fixed (pruned from the MCC tree) using a relaxed random walk model (67) implemented in BEAST 1.10.4 (63), with a Cauchy distribution to model the among-branch heterogeneity in dispersal velocity. Following a similar approach as in McCrone et al. (17), the eight largest transmission lineages (labelled A to H, from largest to smallest) containing >700 sequences were inferred independently, with the remaining smaller transmission lineages (n=524) inferred in a single joint analysis with a shared diffusion model (i.e. same parameter estimates for likelihood, precision matrix, correlation, etc, but independent estimates for diffusion rate and trait likelihood). Owing to variations in the extent of spatial dispersal among these smaller transmission lineages (with larger lineages being more spatially dispersed in general), 30 were subsequently removed from the joint analysis and inferred independently. Model convergence and mixing was assessed using Tracer v1.7 (65). For the independent analyses of the eight largest transmission lineages, we ran between 2 and 5 MCMC chains of 200 to 300 million iterations, sampling every 10,000 to 80,000 states and removing the first 10% to 33% of each chain as burn-in, resulting in 10,000 to 13,5000 trees sampled from the posterior distribution. For the independent analyses of the 30 smaller transmission lineages with fewer than 700 sequences, we ran 2 MCMC chains each of 200 million iterations which we then merged after resampling every 30,000 states and removing the first 10% as burn-in, giving 12,000 posterior trees per transmission lineage. Finally, in the joint analysis, 8 independent chains of 200 million were run with sampling every 120,000 states. They were combined after removing the first 10% as burn-in, again resulting in 12,000 posterior tree samples for each transmission lineage. These posterior tree samples were then used to generate an annotated MCC tree for each transmission lineage using TreeAnnotator (63).

To facilitate subsequent analyses of viral lineage movements at the LTLA level, we mapped the inferred location of each internal node in the transmission lineages to its corresponding LTLA by checking whether the inferred coordinates are contained within the associated polygon. In the case where an enclosing polygon could not be found (e.g. a small proportion of internal nodes were

inferred to lie in the small spaces between neighbouring polygons), the polygon that is geographically closest to the inferred location was assigned.

Discrete phylogeography with generalised linear model (GLM) parameterisation

We used the approach of discrete phylogeography with generalised linear model (GLM) to parameterise transition rates between locations and test the association of viral lineage dispersal with a number of geographical, demographic, epidemiological and human mobility-related factors (see Table S2 for full list of predictors). Specifically, to test the gravity model as a predictor of viral lineage movements, we considered in the GLM analysis the population size at the origin and destination location of each movement and the geographical distance between them. To further capture any heterogeneities in aggregated human mobility at the city-level (which are unlikely to be adequately described by the gravity model), we also included the aggregated mobility matrix and community memberships as predictors. We allowed these mobility-related predictors to vary across different phases in the time-inhomogeneous model to test for temporal variations in aggregated human mobility patterns and also potentially time-varying effect of mobility on viral dispersal. We observed from both epidemiological data and continuous phylogeography that many LTLAs in Greater London experienced an earlier uptick in Omicron BA.1 cases compared to most LTLAs with other regions of England. To capture this asynchronicity in local epidemic dynamics and investigate its impact on viral dispersal, we considered in the GLM analysis whether each viral movement started or ended in the Greater London region and additionally the time of first peak in Omicron BA.1 case incidence at the origin and destination location. Furthermore, we also tested for the impact of sampling bias by including a predictor based on the residuals from a simple regression of sample size against Omicron BA.1 cases for both the origin and destination location. Due to the small number of sequences collected in some LTLAs especially during later phases of the epidemic, the regression residuals were computed using sample sizes and case counts aggregated over the whole study period in both the time-homogeneous and time-inhomogeneous models.

Unlike continuous phylogeography where each sequence is assigned a unique set of coordinates in continuous space, discrete phylogeography requires that sequences are grouped into discrete geographical units. The level of granularity of these geographical units depends on a number of factors including (i) the desired level of resolution at which the dispersal history is to be reconstructed, (ii) the amount of heterogeneities present within each geographical unit, and (iii) the maximum number of geographical units beyond which the analysis becomes computationally intractable. To capture heterogeneities in viral movements at the city-level and to allow comparisons with results from continuous phylogeography, we allocated sequences to their corresponding LTLAs using a lookup table which provides unique mapping between postal districts and LTLAs.

The current computational architecture and implementation of the discrete phylogeographic GLM model limits the number of discrete units possible to 256, which is smaller than the number of LTLAs across which sequences were sampled for some of the larger transmission lineages. To tackle this, we aggregated LTLAs where appropriate to reduce the number of geographic units. In order to minimise the resulting information loss, we first considered LTLAs with the fewest sequences and performed a merging operation if an adjacent LTLA with at least one sampled genome could be found. In the case where multiple adjacent LTLAs were available, the LTLA with the most number of sampled genomes was chosen for the merger. After each merging operation, the list of LTLAs (or geographical units after merging) ranked by the number of sampled genomes was recalculated for the next iterative step (it is therefore possible for an LTLA to be involved with multiple merging operations). This process continued until there were only 253 geographic units in each transmission lineage. For the

geographical units consisting of multiple LTLAs, each statistic of interest was averaged over the relevant LTLAs, weighted by population size where appropriate. For transmission lineages with sequences sampled in fewer than 256 LTLAs, no merging was performed.

The discrete phylogeographic GLM model parameterizes the log of between-location transition rates as a log linear function of the predictors. Continuous predictors (geographical distances, population sizes, aggregated mobility matrices, peak timing in case incidence, sampling residuals) were therefore log-transformed and standardised after adding a pseudo-count to each entry where appropriate. Binary variables (community memberships, Greater London/non-Greater London) were encoded as 0 and 1. In the mobility-related predictors, there was missing data for one or two geographical units in some transmission lineages (due to mobility data being unavailable for South Tyneside and City of London), which we labelled as NA and later integrated out in our Bayesian inference. For the aggregated mobility matrix predictor with continuous values in the large-scale transmission analyses, we confronted this using a new Hamiltonian Monte Carlo (HMC) kernel to jointly sample all missing covariates from their posterior distributions building on similar efforts in the BEAST framework (68, 69). The HMC kernel produces distant proposals with relatively high acceptance rate for the Metropolis algorithm by exploiting numerical solutions to the Hamiltonian dynamics. We performed the analyses using the code available in the `hmc-clock` branch of the BEAST codebase (available at <https://github.com/beast-dev/beast-mcmc/tree/hmc-clock>) in conjunction with the BEAGLE code available in the `hmc-clock` branch of the codebase (available at <https://github.com/beagle-dev/beagle-lib/tree/hmc-clock>). We ran the analyses on a set of 100 empirical trees for each transmission lineage extracted from the BEAST importation analysis and ran sufficiently long chains sampling every 500 generations, or combined multiple chains (excluding adequate burn-ins), to ensure effective sample sizes (ESSs) > 100 for the continuous parameters as diagnosed using Tracer (65). A custom R script was used to summarise and visualise the posterior coefficient estimates and inclusion probabilities of each predictor.

Discrete phylogeography with GLM: effect of booster uptake

The rollout of booster vaccination in the United Kingdom began in September 2021 (70) and was initially prioritised for those aged 50 and above as they are at a higher risk of severe symptoms and hospitalisation from infection. Eligibility for boosters was extended to those aged 40 and above on 22 November 2021 (71), and subsequently to all adults on 30 November 2021 (72). This resulted in spatial variations in booster uptake that are strongly correlated with the underlying age structure of the population (Fig. S16, A and C), which is in turn correlated with Omicron BA.1 prevalences due age-dependent transmission patterns as shown by Elliott et al. (2022) (15) (Fig. S16, B and D).

To adjust for age structure as a confounder, here we devise an effective measure of the booster uptake that is independent of the underlying age structure of the population. Using vaccination data (downloaded from the GOV.UK COVID-19 Dashboard) consisting of the percentage of people in different age groups who have received a booster dose, we calculate the overall booster uptake in each Lower Tier Local Authority (LTLA) assuming an age distribution that is the same as the national population-weighted average age distribution (computed from mid-2020 population estimates published by the UK Office of National Statistics). This is equivalent to the overall proportion of the population who would have received a booster dose in an LTLA given its observed age-specific booster uptake (with roughly 5-year grouping), assuming that it has the same age structure as the national average. Similar to other covariates included in the GLM analysis, for the geographical units consisting of multiple LTLAs, the effective booster uptake is averaged over the relevant LTLAs weighted by population size.

Discrete phylogeography with GLM: likelihood-deviance measure

To evaluate the relative importance of the different predictors in the time-inhomogeneous GLM analysis, we have developed and implemented a new phylogeographic model-fit measure which builds upon standard, permutation-based machine learning approaches to assessing variable importance (73).

Starting from the posterior $p(\theta|Y, x_{11}, \dots, x_{ke}, \dots, x_{K2})$ with phylogeographic likelihood $p(Y|\theta, x_{11}, \dots, x_{ke}, \dots, x_{K2})$ where θ represents all model parameters and x_{ke} represents the vector of covariate values for covariate $k \in \{1, \dots, K\}$ (K being the total number of predictors in the model) in epoch $e \in \{1, 2\}$, we define the deviance for this covariate as $d_{ke} = \log p(Y|\theta, x_{11}^\pi, \dots, x_{ke}^\pi, \dots, x_{K2}^\pi) - \log p(Y|\theta, x_{11}, \dots, x_{ke}, \dots, x_{K2})$ where x_{ke}^π is a random permutation of the observed covariate vector x_{ke} . To estimate the posterior distribution of d_{ke} , we computed values of d_{ke} for each MCMC sample, s , with realised model parameters $\theta^{(s)}$ by setting $\theta = \theta^{(s)}$, and randomly permuting x_{ke} with equal probability for all possible permutations. For each covariate and in each epoch, we estimated the posterior distribution of the likelihood-deviance resulting from a random permutation of the covariate values. We then compared the resulting posterior distributions and ranked the importance of each covariate in predicting the geographic locations Y . The covariate marginal posterior modal (most probable) ranking was then reported, as shown in Fig. 4 and Fig. S14 (with 1 being most important). While this approach averages over all possible marginal permutations and therefore has improved performance over earlier permutation-based measures in machine learning (74), it may nevertheless return limited discrimination among highly correlated covariates. Permute-and-relearn importance methods (75) are able to overcome this limitation but remain computationally impractical given the numbers of tips and the size of the state-spaces considered in this study.

Using the above approach, we find that the predictor rankings do not always reflect differences in absolute effect size and that they help to identify similarities and differences between transmission lineages, as well as between epochs for a given transmission lineage. In the two largest transmission lineages, the gravity model covariates are consistently the most important covariates, in both the early- and late-epoch. For Transmission Lineage-A, the Greater London origin predictor is the next important predictor throughout the study period. The Greater London origin predictor is also more important in Transmission Lineage-A than in Transmission Lineage-B, for which a change in importance of this predictor between the early- and late-epoch is observed. In Transmission Lineage-B, the origin peak time predictor is the next important predictor after the gravity model predictors. For both transmission lineages, we observe a large and consistent increase in the importance of the mobility matrix predictor between the early- and late-epoch.

We also note that the magnitude of the likelihood-deviance estimates scales with the size of the dataset (and therefore the number of tips in the transmission lineages). As such, the deviance estimates do not provide a relative measure of fit across transmission lineages.

Branching process model and comparison of transmission lineage size distributions

To verify that the time of importation is the key determinant of transmission lineage size, we compared the size distribution of empirically observed transmission lineages with that from a model that simulates the branching process of transmission lineages following importation. Simulated importation dates are set to the dates estimated from the phylodynamic analysis and simulated

transmissions occur at the spatially homogeneous growth rates estimated from the daily number of reported COVID-19 cases (from the GOV.UK COVID-19 Dashboard) and SGTF data in England. Due to the low number of Omicron BA.1 cases at the beginning of the epidemic, which can lead to unreliable estimates of the initial growth rate, we performed a series of simulations with a range of different starting growth rates (taken from estimates during early parts of the invasion). We computed the Kullback-Leibler (KL) distance between the size distribution of simulated lineages and that of lineages inferred from phylodynamic analysis (Table S1). The growth rate that minimised the KL distance was then used to impute the initial growth rate in the best-fit model. We note that, given the simple nature of the model, we did not take into account any uncertainties associated with the case growth rates but relied only on the central estimates. As a sensitivity analysis for the potential bias that might have resulted from this and also any changes in case reporting rate during the study period, we repeated the simulations using case incidence estimates from the UK Office of National Statistics (ONS) (see Fig. S10 for comparisons between case incidence data from UK.GOV COVID-19 Dashboard and ONS estimates). We observed consistent results as those obtained using the case incidence data from the GOV.UK COVID19 Dashboard (Fig. S7 and Fig. S8).

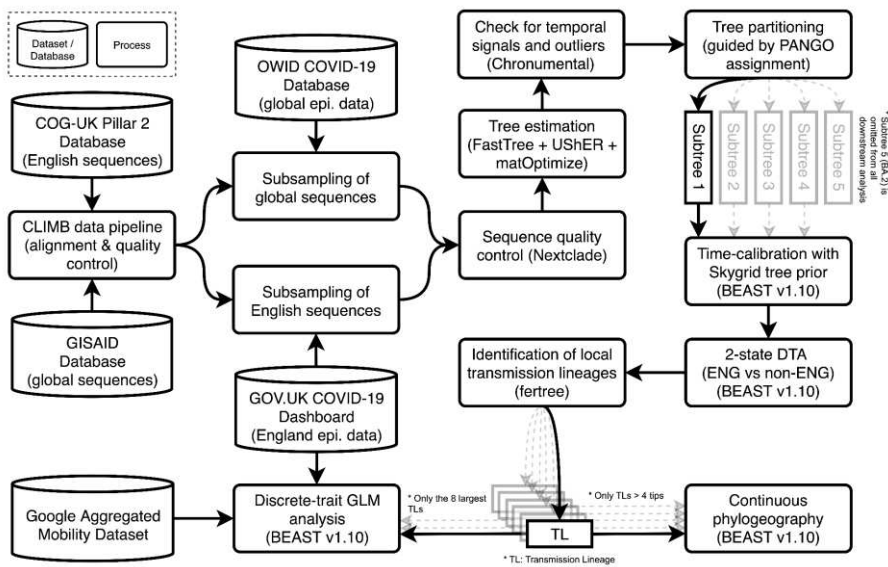


Fig. S1: Outline of phylogenetic analysis pipeline. A high-level overview of the various processes and phylogenetic analyses performed, as well as any relevant programs and packages for each step. Note that each subtree (except for the subtree containing only Omicron BA.2 sequences which we have omitted from further downstream analysis) from the tree-partitioning procedure is passed onto further analysis independently. Please refer to materials and methods for a more detailed description of each analysis.

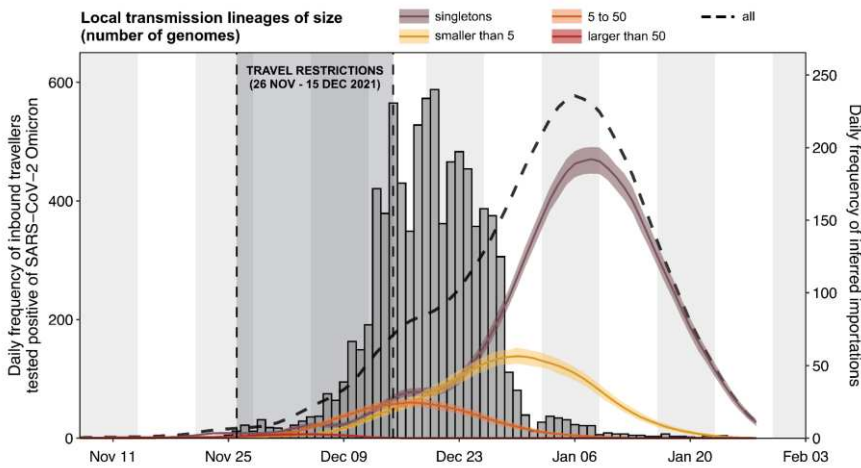


Fig. S2: Comparison of Omicron BA.1 importation intensity as observed empirically from Variant Mutation (VAM) line list versus estimates from phylodynamic analysis. Histogram (grey bars) shows the daily number of incoming travellers who were later tested positive for Omicron BA.1 under the UKHSA mass testing programme (by specimen date). Solid lines represent the daily frequency of importations (7-day rolling average) as inferred from the phylodynamic analysis, coloured according to the size of resulting local transmission lineages (with the black dashed line representing the total numbers irrespective of size); shading denotes the 95% HPD across the posterior tree distribution.

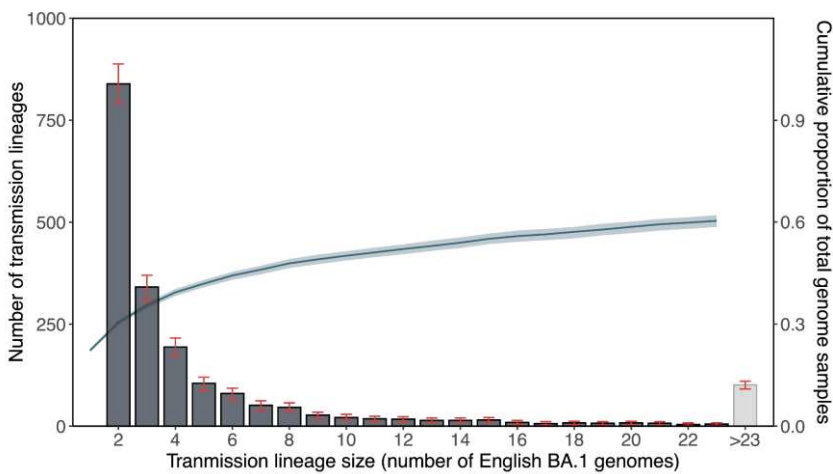


Fig. S3. Distribution of local transmission lineage sizes from phylodynamic analysis. Grey bars show the number of transmission lineages of different sizes; red error bars denote the 95% HPDs across the posterior tree distribution. Blue solid line represents the cumulative proportion of English Omicron BA.1 genomes in our dataset accounted for by transmission lineages up to a certain size; shading denotes the 95% HPD across the posterior tree distribution.

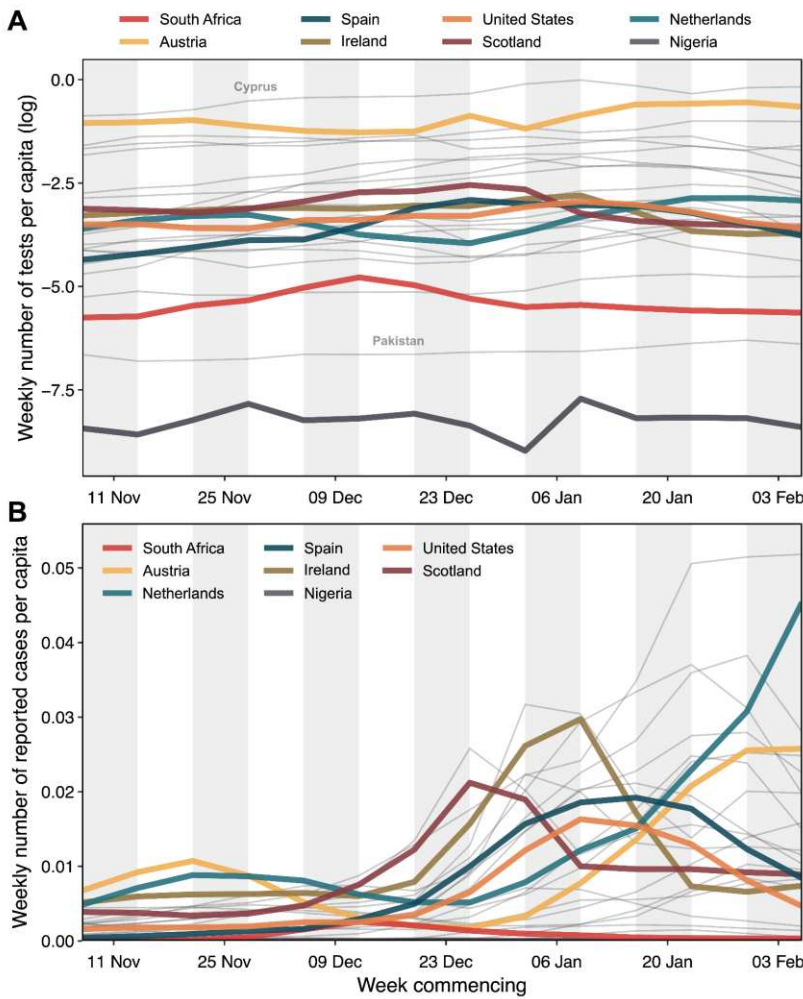


Fig. S4: Variations in case reporting rates between countries. (A) Weekly number of tests performed per capita (log-transformed) and (B) weekly number of reported cases per capita for 27 countries (including Scotland and Northern Ireland independently) with the highest air passenger volumes arriving in England between November 2021 and January 2022 (collectively accounting for ~80% of total air passenger volume in this period). Thick solid lines represent a subset of eight selected countries with notable contribution to the overall intensity of Omicron BA.1 importation into England at different points during the study period; thin grey lines represent all other countries.

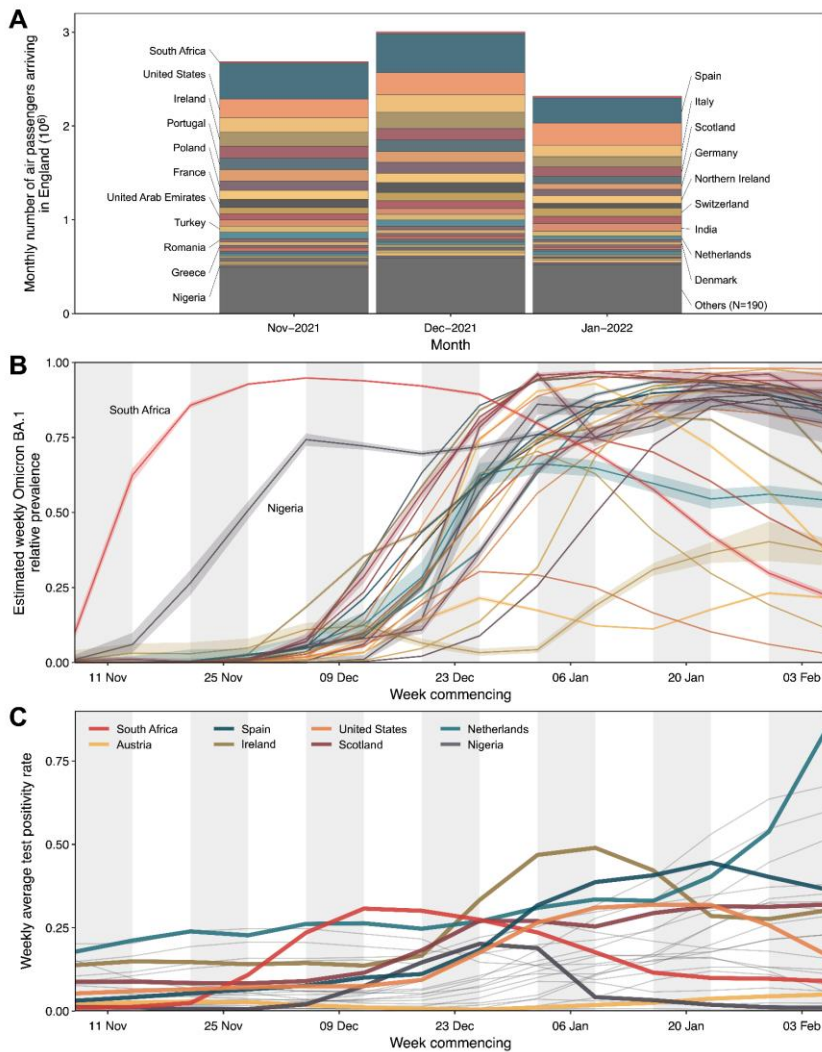


Fig. S5: Components of Estimated Importation Intensity (EII). (A) Monthly number of air passengers arriving in England from all countries (N=217) between November 2021 and January 2022. Area of each coloured block indicates the number of air passengers arriving from a given country (out of the 27 countries for which EIIs are calculated) during a given month; grey blocks at the bottom represent air traffic volume from all other countries (N=190). (B) Estimated weekly relative prevalence of Omicron BA.1 in the 27 selected countries during the study period; shaded region represents the 95% CI. (C) Weekly average test positivity rate in the 27 selected countries during the study period. Thick solid lines represent a subset of eight selected countries with notable contribution to the overall intensity of Omicron BA.1 importation into England at different points during the study period; thin grey lines represent all other countries.

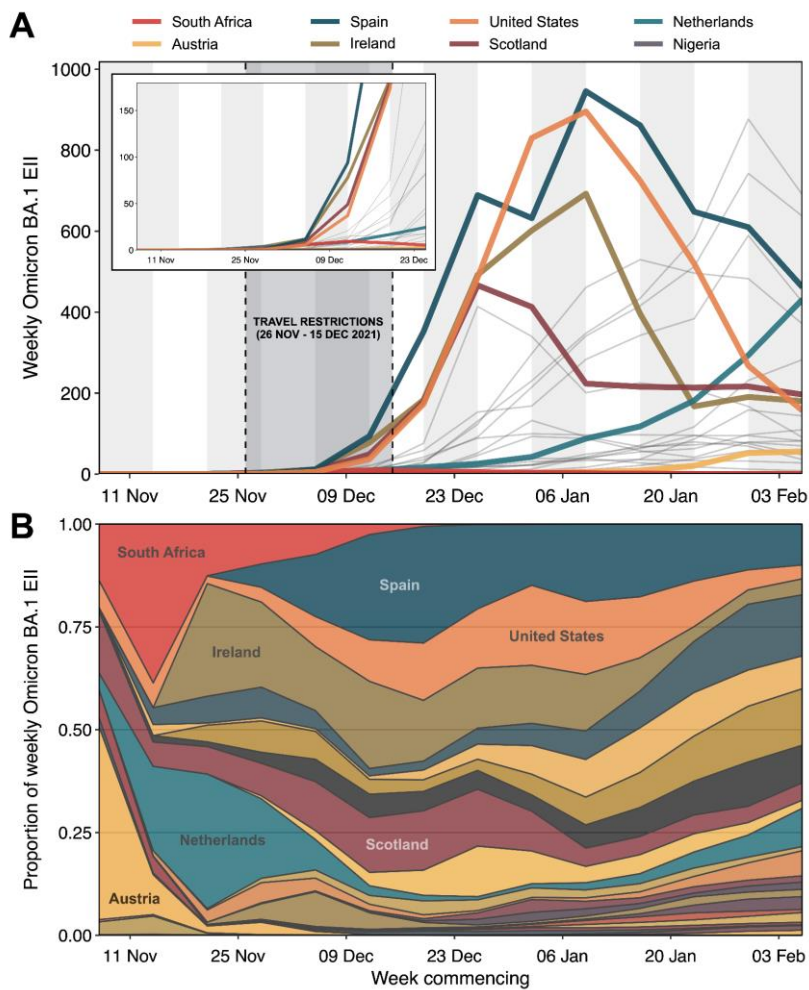


Fig. S6: Estimated Importation Intensity (EII) of Omicron BA.1 from selected potential exporters, using case incidence per capita as proxy for underlying prevalence. Estimated weekly number of Omicron BA.1 cases arriving in England from 27 countries (including Scotland and Northern Ireland independently) with the highest air passenger volumes arriving in England between November 2021 and January 2022 (collectively accounting for ~80% of total air passenger volume in this period), using weekly number of reported cases as a proxy for trends in the underlying prevalence. Thick solid lines represent EIIs from eight selected countries with notable contribution to the overall intensity of Omicron BA.1 importation into England at different points during the study period; thin grey lines represent all other countries. Inset shows a magnified view of early trends. Grey shaded region represents the period (26 November to 15 December 2021) when travel restrictions on international arrivals from multiple southern African countries were implemented. (B) Relative proportion of weekly EII of Omicron BA.1 by location/country among selected potential exporters. Areas representing countries highlighted in (A) are labelled.

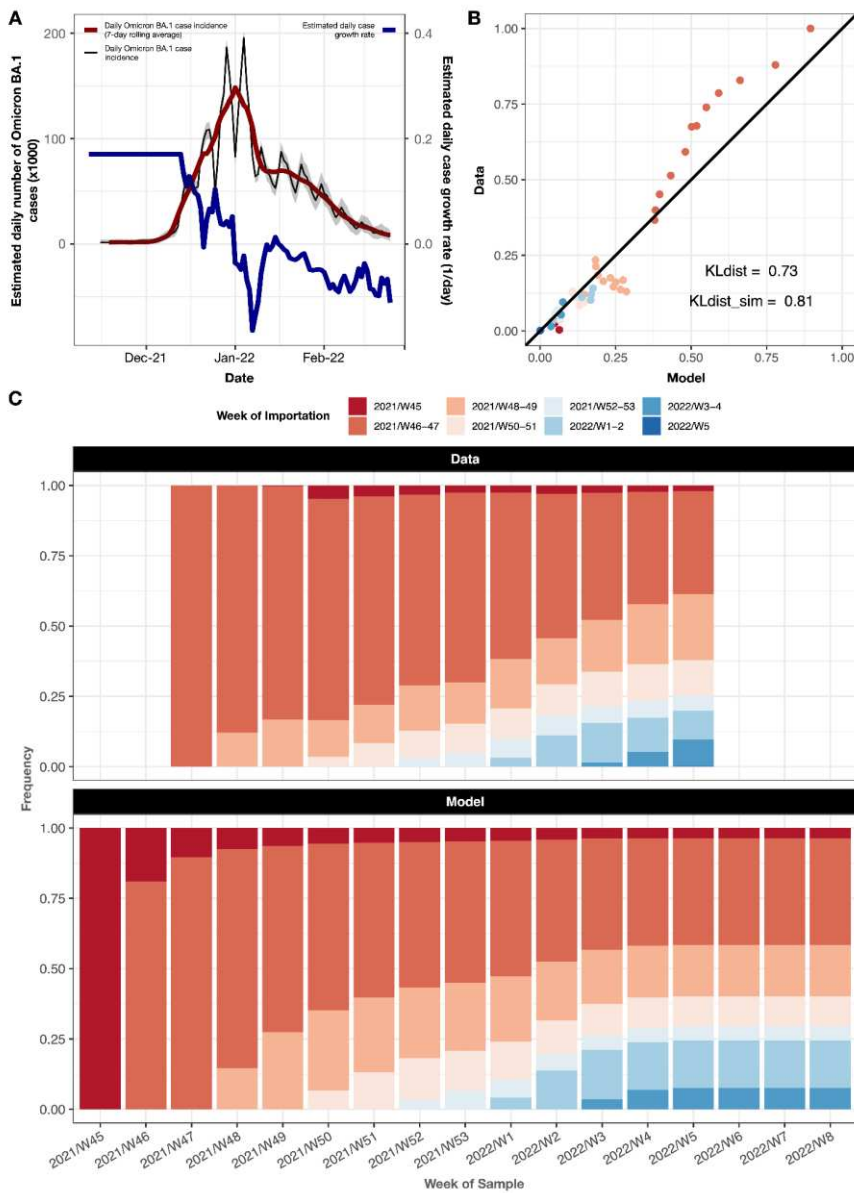


Fig. S7. Comparison of transmission lineage size distribution from phylodynamic analysis versus simulated results from a branching process model. (A) Black and red solid lines represent the estimated daily and 7-day rolling average daily number of Omicron BA.1 cases in England. Grey

shaded region represents the 95% CI associated with the estimated daily number of Omicron BA.1 cases. Blue solid line represents the estimated daily growth rate, with the initial values imputed using an estimate of the growth rate on 13 December 2021. (B and C) Weekly proportion of local Omicron BA.1 infections resulting from importations at different times throughout the epidemic, with comparison between empirical observations from the phylodynamic analysis (C, top) and predictions from the branching process model (C, bottom).

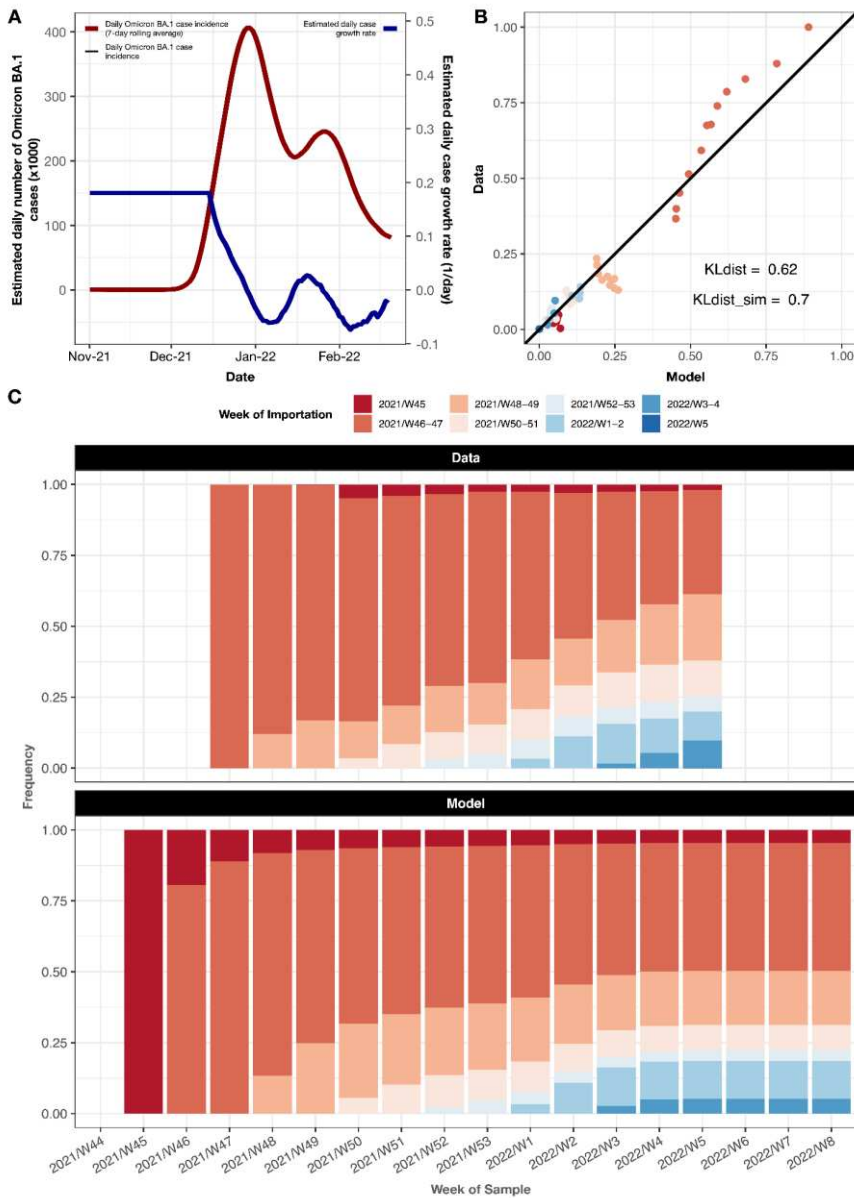


Fig. S8. Comparison of transmission lineage size distribution from phylodynamic analysis versus simulated results from a branching process model (sensitivity analysis using ONS case incidence estimates). (A) Black and red solid lines represent the estimated daily and 7-day rolling average daily number of Omicron BA.1 cases in England, from the UK Office of National Statistics (ONS). Blue solid line represents the estimated daily growth rate, with the initial values imputed using an estimate of the growth rate on 13 December 2021. (B and C) Weekly proportion of local Omicron BA.1

infections resulting from importations at different times throughout the epidemic, with comparison between empirical observations from the phylodynamic analysis (C, top) and predictions from the branching process model (C, bottom).

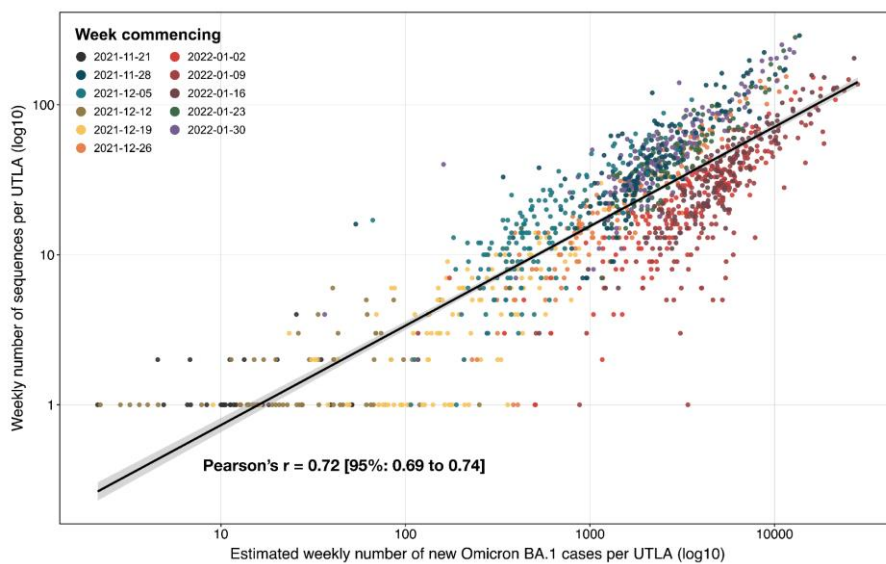


Fig. S9. Correlation between estimated number of Omicron BA.1 cases and number of Omicron BA.1 genomes sampled across UTLAs in England. Circles are coloured by week commencing date. Solid black line represents the line of best-fit; shaded region represents the 95% CI. We note in particular the clustering of circles corresponding to the same week along the line of best-fit, indicating small changes in sequencing coverage across time but not across UTLAs.

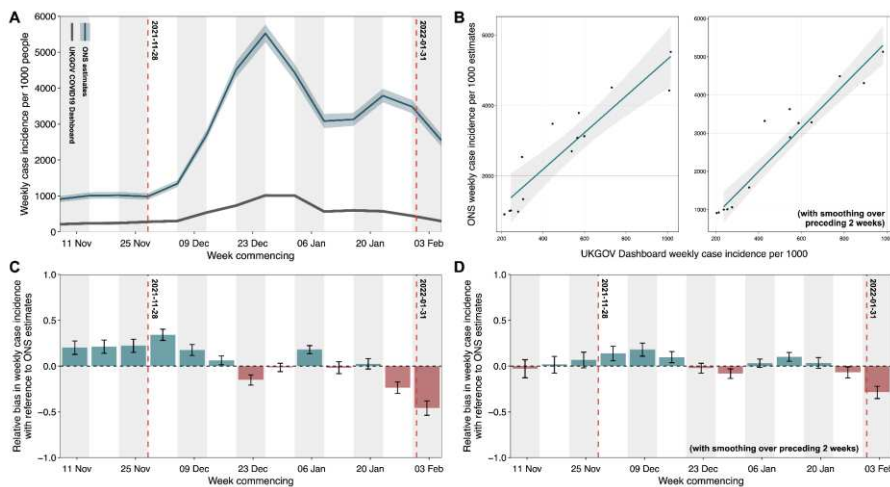


Fig. S10: Comparison of case incidence from the GOV.UK COVID19 Dashboard against estimates from UK Office of National Statistics. (A) Weekly number of positive COVID-19 cases in England as reported by the GOV.UK COVID19 Dashboard (<https://coronavirus.data.gov.uk/>) (solid black line); weekly number of COVID-19 cases in England as estimated from positivity rates by the UK Office of National Statistics (ONS) (solid blue line; shading denotes the associated 95% CI). Vertical red dashed lines indicate the start date and end date of the period during which English genomes were sampled in proportional to weekly number of reported cases (see supplementary materials). (B) Weekly number of COVID-19 cases per 1000 people as estimated by ONS versus that from the GOV.UK COVID19 Dashboard, with (right) and without (left) smoothing over the preceding two weeks for each given date. Blue lines show the least-squares fit and the shading denotes the 95% CI. (C, D) Residuals from a linear regression between the weekly number of COVID-19 cases per capita as estimated by ONS versus that from the GOV.UK COVID19 Dashboard (during the period of interest from 28 November 2021 to 31 January 2022), with (D) and without (C) smoothing over the preceding two weeks for each given date. Error bars denote the 95% confidence interval associated with uncertainty in the ONS estimates; boxes are coloured red (negative) or blue (positive) according to the sign of the residuals.

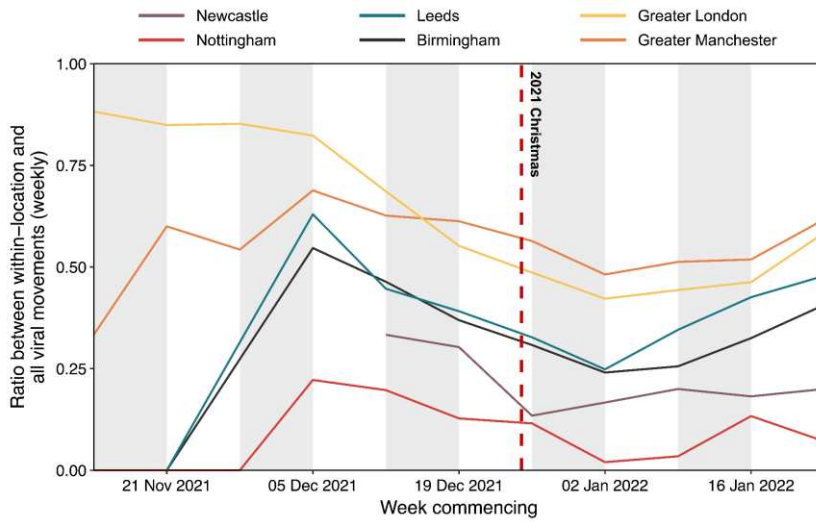


Fig. S11. Within-location versus all viral lineage movements for major cities in England. Each solid line represents the ratio between the frequency of within-location and all viral lineage movements per week, as inferred from continuous phylogeography for 6 major cities in England. For Greater Manchester and Greater London, viral lineages associated with multiple lower tier local authorities were aggregated in the calculation of these ratios. The timing of each viral lineage movement was assumed to be half-way between the inferred time of the nodes corresponding to the origin and destination.

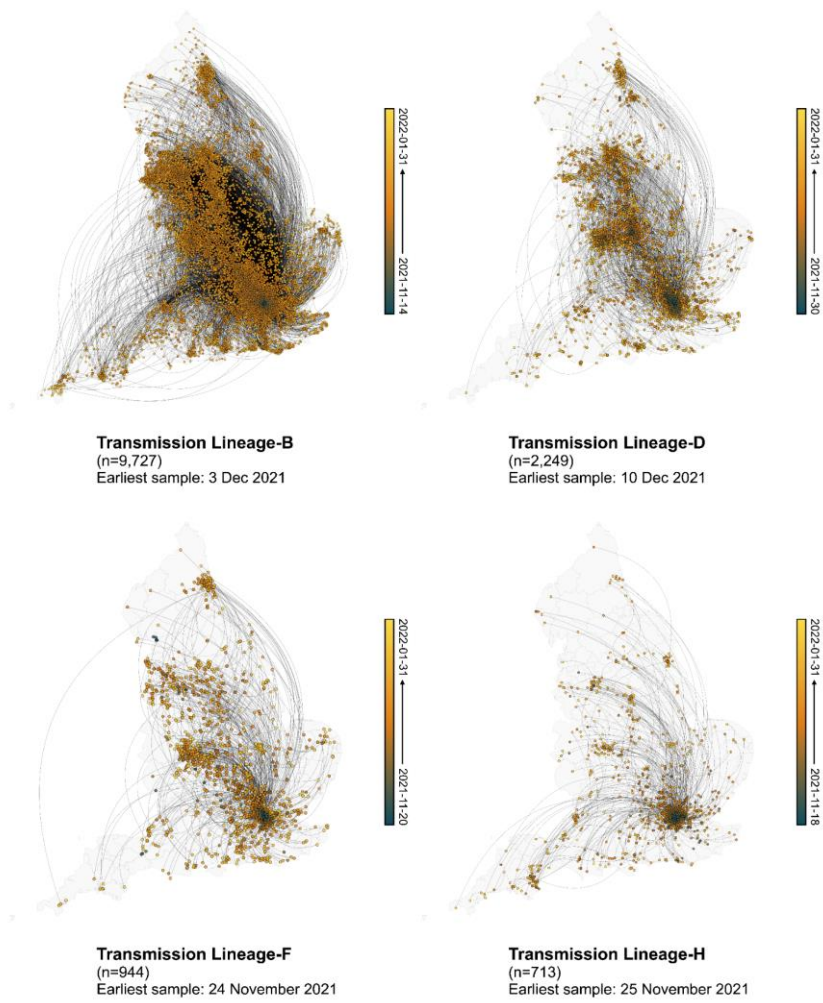


Fig. S12. Spatiotemporal dynamics of Omicron BA.1 transmission lineages in England (Transmission Lineage-B, D, F and H). Maps showing viral lineage movements inferred from continuous phylogeography for Transmission Lineage-B, D, F and H. Nodes are coloured according to inferred date of occurrence and the direction of viral lineage movement is indicated by edge curvature (anti-clockwise).

Commented [2]: @joetsui1994@gmail.com missing labels a-d
Assigned to Joseph Tsui_

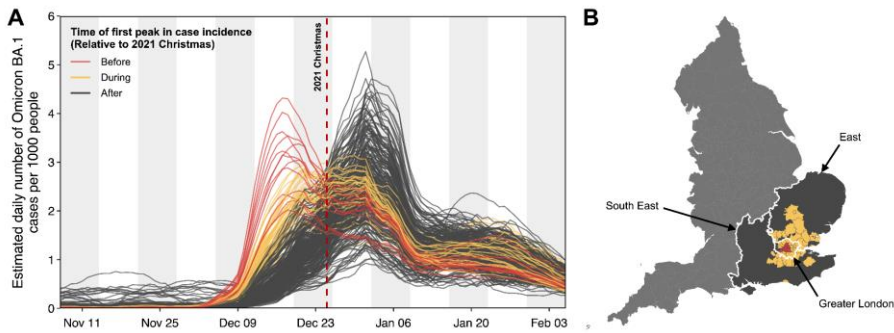


Fig. S13. Spatial variations in timing of first peak of Omicron BA.1 case incidence across England. Estimated daily number of Omicron BA.1 cases per 1000 people at the Lower Tier Local Authority (LTLA) level (7-day rolling average), coloured according to the timing of their first peak relative to Christmas 2021 (specifically, whether the interval during which the daily number of Omicron BA.1 cases exceed 85% of the peak incidence lies entirely before (red), after (dark grey), or encloses (yellow) 25 December 2021 (Christmas). (B) Map showing the spatial distribution of the timing of the first peak in Omicron BA.1 case incidence at the LTLA level, following the same colour scheme as in (A).

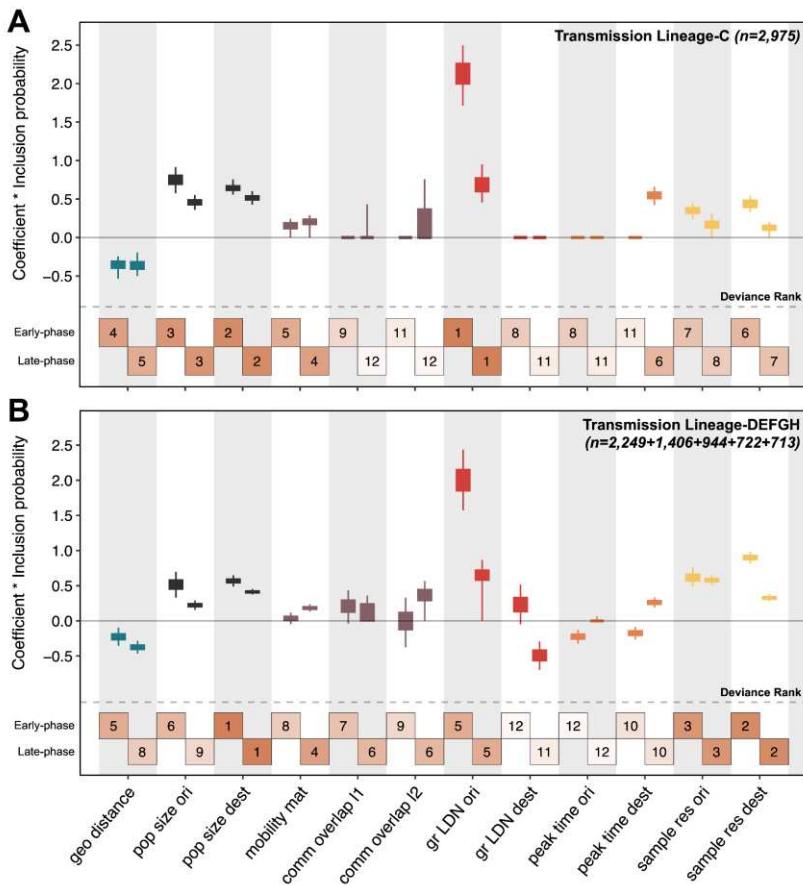


Fig. S14. Predictors of Omicron BA.1 viral lineage movements in England in the time-inhomogeneous discrete-phylogeography with GLM model. For each predictor, the box and whiskers show the posterior distribution of the product of the log predictor coefficient and the predictor inclusion probability; the left hand value represents the expansion period estimate and the right hand value the post-expansion period estimate. Top panel (A) shows estimates for Transmission Lineage-C and bottom panel (B) shows those for Transmission Lineages D, E, F, G, and H analysed in a joint model. Posterior distributions are coloured according to predictor type: geographic distances (geo distance, dark blue), population sizes at origin and destination (pop size ori & pop size dest, black), aggregated mobility (mobility mat, purple), mobility-based community membership level 1 and level 2 (comm overlap l1 & l2, purple), Greater London origin and destination (gr LDN ori & gr LDN dest, red), time of peak incidence at the origin and destination (peak time ori & peak time dest, orange) and the residual of a regression of sample size against case count regression at either the origin and destination (sample res ori & sample res dest, yellow). Boxes at the bottom of each panel are numbered and shaded to represent the rank of predictors based on their deviance measure, with 1 indicating the largest (most important) and 12 indicating the smallest (least important).

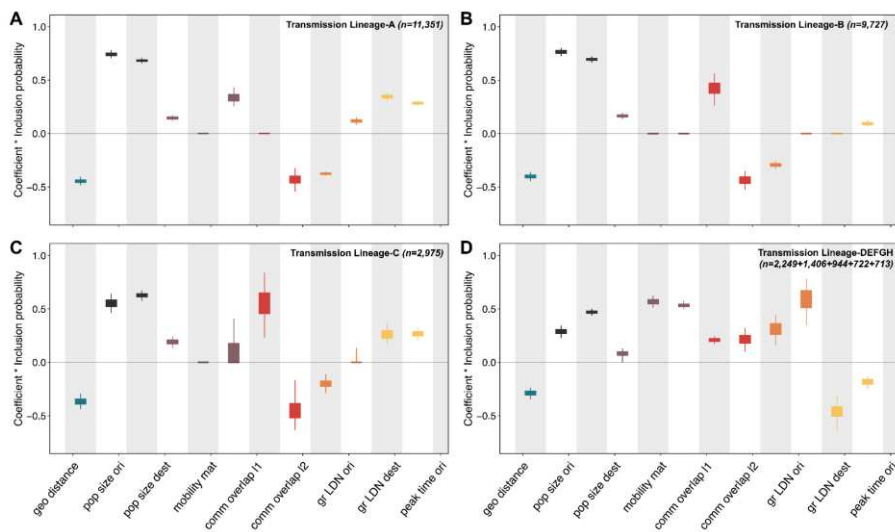


Fig. S15. Predictors of Omicron BA.1 viral lineage movements in England in the time-homogeneous discrete-phylogeography with GLM model. Each panel corresponds to an independent analysis for Transmission Lineage-A (A), Transmission Lineage-B (B), Transmission Lineage-C (C), and Transmission Lineages D, E, F, G and H together in a joint model (D). For each predictor within a panel, the box and whiskers show the posterior distributions of the product of the log predictor coefficient and the predictor inclusion probability. Posterior distributions are coloured according to predictor type: geographic distances (geo distance, dark blue), population sizes at origin and destination (pop size ori & pop size dest, black), aggregated mobility (mobility mat, purple), mobility-based community membership level 1 and level 2 (comm overlap l1 & l2, purple), Greater London origin and destination (gr LDN ori & gr LDN dest, red), time of peak incidence at origin and destination (peak time ori & peak time dest, orange) and the residual of a regression of sample size against case count regression at either the origin and destination (sample res ori & sample res dest, yellow).

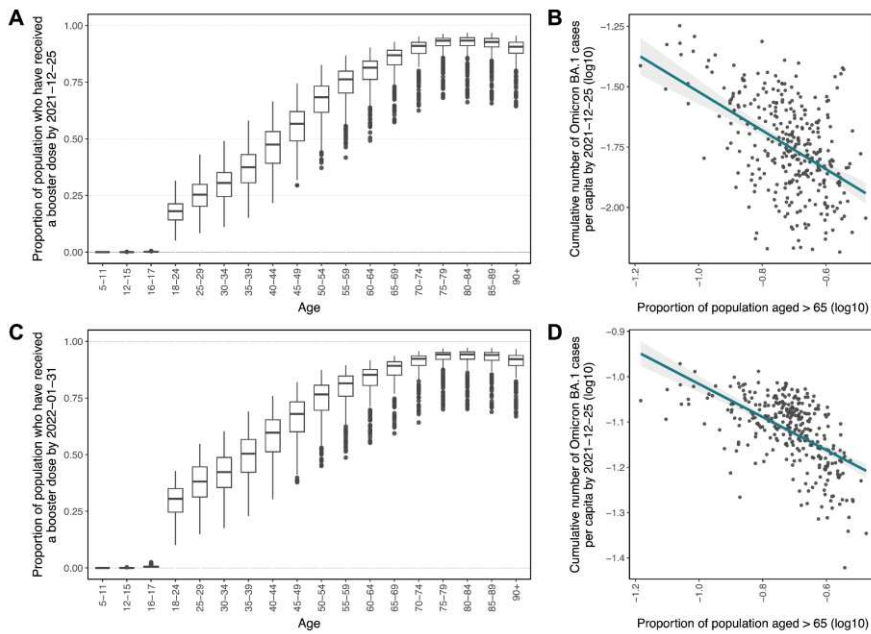


Fig. S16: Dependency of booster uptakes and cumulative Omicron BA.1 case counts on population age structure. (A, C) Distribution of the proportion of age-specific population in each Lower Tier Local Authority (LTLA) who have received a booster dose by 25 December 2021 and 31 January 2022, respectively. Each box extends from the 25th to 75th percentile of the distribution for the corresponding age group; the midline within each box represents the median; the vertical lines represent the lower and upper limits and the dots denote the outliers. (B, D) Cumulative number of Omicron BA.1 cases per capita (log₁₀-transformed) versus proportion of the population aged above 65 (log₁₀-transformed). Each dot represents an LTLA. Blue lines show the least-squares fit and the shading denotes the associated 95% CI.

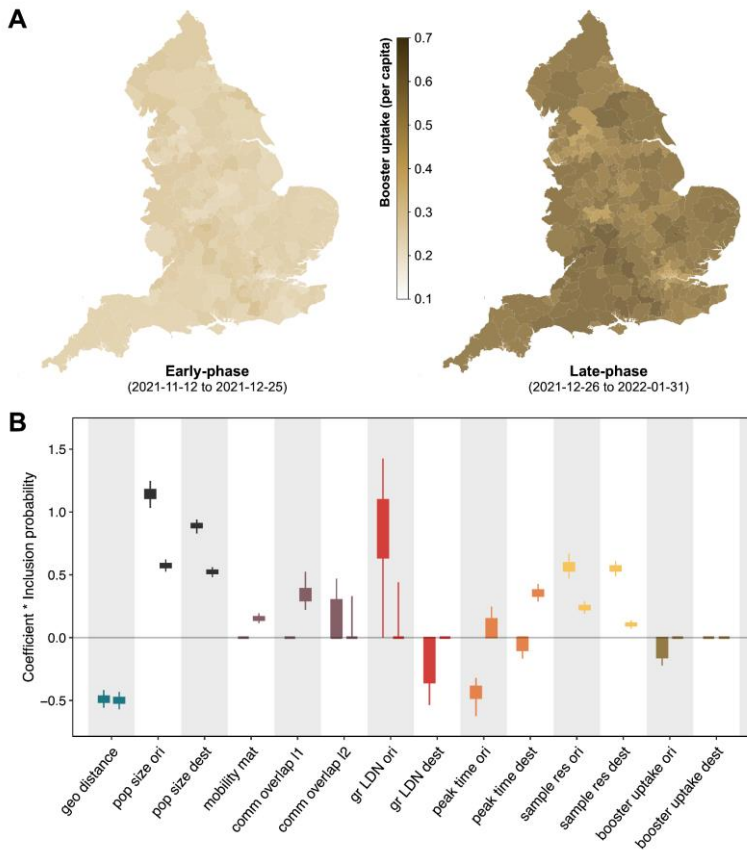


Fig. S17: Booster uptake as a predictor of Omicron BA.1 viral lineage movements in England. (A) Map of age-corrected effective booster uptake at the Lower Tier Local Authority (LTLA) level, averaged over the early-phase (12 November 2021 to 25 December 2021) (left) and the late-phase of the epidemic (26 December 2021 to 31 January 2022) (right). The effective booster uptake is defined as the proportion of the population who would have received a booster dose having accounted for age-specific booster uptakes, assuming the national average population age structure. (B) For each predictor, the box and whiskers show the posterior distribution of the product of the log predictor coefficient and the predictor inclusion probability; the left hand value represents the expansion period estimate and the right hand value the post-expansion period estimate. Posterior distributions are coloured according to predictor type: geographic distances (geo distance, dark blue), population sizes at origin and destination (pop size ori & pop size dest, black), aggregated mobility (mobility mat, purple), mobility-based community membership level 1 and level 2 (comm overlap l1 & l2, purple), Greater London origin and destination (gr LDN ori & gr LDN dest, red), time of peak incidence at origin and destination (peak time ori & peak time dest, orange), the residual of a regression of sample size against case count regression at either origin and destination (sample res ori & sample res dest, yellow), and effective booster uptake at origin and destination (booster uptake ori & booster uptake dest, brown).

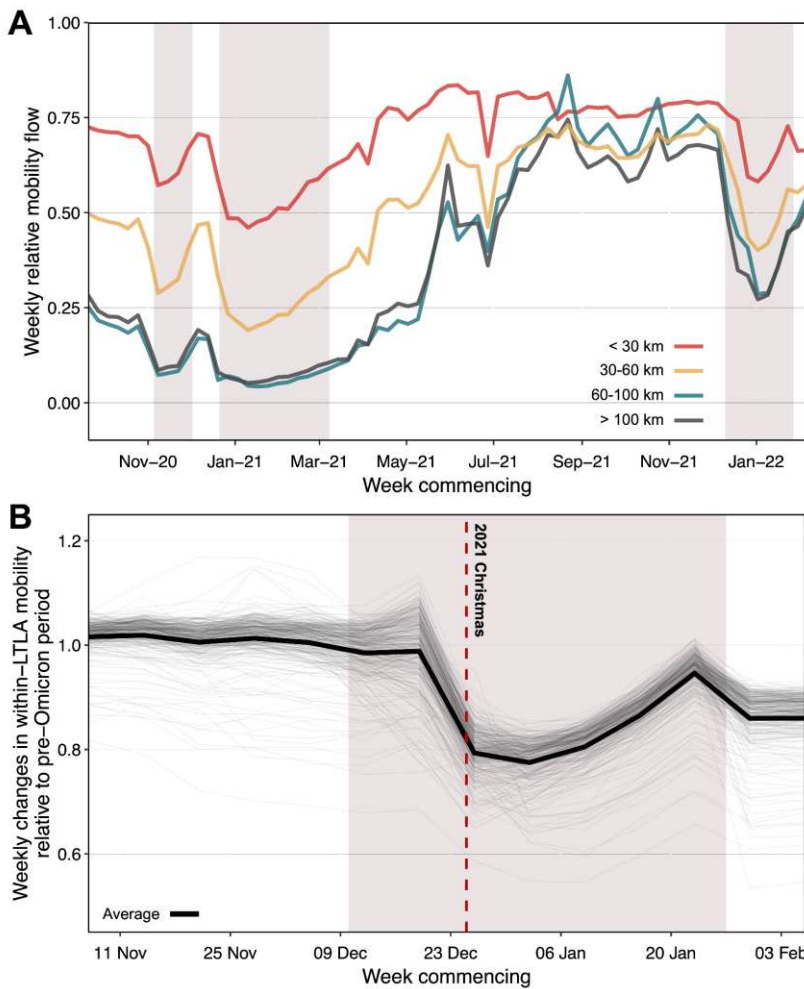


Fig. S18. Trends in human mobility across England. (A) Weekly human mobility flows relative to pre-pandemic levels (averaged over period from 3 November 2019 to 28 December 2019) across different spatial scales (red: <30 km, yellow: 30-60 km, blue: 60-100 km, dark grey: >100 km). (B) Weekly changes in within-LTLA mobility relative to pre-Omicron levels (averaged over period from 12 September 2021 to 6 November 2021 for each LTLA individually). Thick black line represents the weekly mobility changes averaged over all LTLAs; each thin grey line represents the weekly mobility changes for a single LTLA. Vertical line shows 24th of December 2021.

Table S1. Imputation of initial growth rate in branching process model. Kullback-Leibler (KL) divergence values computed from comparisons of the weekly proportion of local Omicron BA.1 infections resulting from importations at different times as inferred from the phylodynamic analysis, versus predictions from branching process models with a range of initial growth rates. The initial growth rates in the branching models were imputed using estimates taken from the early growth phase of the epidemic (between 2021-12-04 and 2021-12-15), and the growth rate that minimised the KL divergence (shown in bold) was used in the best-fit model.

Initial growth rate taken from	Kullback–Leibler divergence
2021-12-04	2.84
2021-12-05	2.04
2021-12-06	1.33
2021-12-07	1.42
2021-12-08	2.95
2021-12-09	12.54
2021-12-10	13.13
2021-12-11	8.31
2021-12-12	2.02
2021-12-13	0.81
2021-12-14	2.76
2021-12-15	5.17

Table S2. Imputation of initial growth rates in branching process model (sensitivity analysis using incidence estimates from the UK Office of National Statistics). Kullback-Leibler (KL) divergence values computed from comparisons of the weekly proportion of local Omicron BA.1 infections resulting from importations at different times as inferred from the phylodynamic analysis, versus predictions from branching process models with a range of initial growth rates. The initial growth rates in the branching models were imputed using estimates taken from the early growth phase of the epidemic (between 2021-11-20 and 2021-12-20), and the growth rate that minimised the KL divergence (shown in bold) was used in the best-fit model. Note that the UK Office of National Statistics (ONS) case incidence (central) estimates are used here for the estimation of daily case growth rates, instead of case incidence data from the GOV.UK COVID-19 Dashboard.

Initial growth rate taken from	Kullback–Leibler divergence
2021-11-20	5.03
2021-11-21	5.16
2021-11-22	5.29
2021-11-23	4.70
2021-11-24	4.92
2021-11-25	5.17
2021-11-26	3.77
2021-11-27	3.78
2021-11-28	3.18
2021-11-29	3.33
2021-11-30	3.9
2021-12-01	6.63
2021-12-02	17.5
2021-12-03	20.29
2021-12-04	19.00
2021-12-05	14.91
2021-12-06	12.32
2021-12-07	9.89
2021-12-08	13.77
2021-12-09	13.33
2021-12-10	13.16
2021-12-11	8.96
2021-12-12	5.60
2021-12-13	2.77
2021-12-14	1.51
2021-12-15	0.7
2021-12-16	1.22
2021-12-17	2.59
2021-12-18	4.18
2021-12-19	5.96

2021-12-20	7.24
------------	------

Table S3. Predictors of viral lineage movements in England. Summary table and descriptions of predictors considered in the discrete phylogeographic GLM analysis. For location-specific predictors (as indicated by an asterisk), we included both an origin and a destination covariate in the GLM model.

Predictor	Domain	Time-varying	Description
<i>Geographical distance between origin and destination</i>	Geography	No	Geographical distance between the origin and destination calculated using the Haversine formula in km
<i>Population size*</i>	Demography	No	Population size at the origin/destination, from population estimates obtained by the Office of National Statistics in mid-year 2020
<i>Aggregated mobility matrix</i>	Human mobility	Yes	Average weekly number of trips taken between origin and destination estimated from Google COVID-19 Aggregated Mobility Research Dataset; given that the mobility flux between two locations does not in general differ from symmetry in a statistically significant manner (i.e. the magnitude of mobility flux in either direction is generally very similar for a given connection), we considered a mobility matrix that was symmetrised
<i>Community memberships from mobility network (level-1/2)*</i>	Human mobility	Yes	A binary variable [0,1] indicating whether the origin and destination belong to the same community at level-1/2; community structures were identified from the human mobility network as described by the aggregated mobility matrix, using the community detection algorithm Infomap (76, 77), with level-1/2 corresponding to the tree-depth at which the communities were extracted; level-1 has a higher level of aggregation (fewer communities) compared to level-2 (more communities)
<i>Greater London / non-Greater London indicator*</i>	Geography/ Epidemiology	No	A binary variable [0,1] indicating whether the origin/destination is in the Greater London region
<i>Timing of peak in Omicron BA.1 case incidence*</i>	Epidemiology	No	Timing of first peak in Omicron BA.1 case incidence at the origin/destination, measured as the number of days from 1 December 2021
<i>Sampling residuals*</i>	Sampling	No	Residuals from a regression of sample size against Omicron BA.1 case count at origin/destination; time-invariant residuals were

			used due to the small number of samples in some locations during the post-expansion phase
--	--	--	---

Table S4. Definitions of Lower Tier Local Authorities (LTLAs). Table of LTLAs that have been deprecated and aggregated into newly defined LTLAs, according to recent definitions used in the report of population estimates for the UK in mid-2020, compiled by the Office of National Statistics, UK.

Most recent LTLA definition	Deprecated LTLA definition(s)
E06000058 (Bournemouth, Christchurch and Poole)	E06000028, E06000029, E07000048
E06000060 (Buckinghamshire)	E07000004, E07000005, E07000006, E07000007
E06000059 (Dorset)	E07000049, E07000050, E07000051, E07000052, E07000053
E07000244 (East Suffolk)	E07000205, E07000206
E07000244 (West Suffolk)	E07000201, E07000204
E07000246 (Somerset West and Taunton)	E07000190, E07000191

Table S6. The COVID-19 Genomics UK (COG-UK) consortium, June 2021 V4

**The COVID-19 Genomics UK (COG-UK) consortium
FINAL (06-2021 V4)**

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

Samuel C Robson^{13,84}

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

Thomas R Connor^{11,74} and Nicholas J Loman⁴³

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, Software and analysis tools, and Visualisation:

Tanya Golubchik⁵

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:

Rocio T Martinez Nunez⁴⁶

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

David Bonsall⁵

Funding acquisition, Leadership and supervision, Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:

Andrew Rambaut¹⁰⁴

Funding acquisition, Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

Luke B Snell¹²

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, Software and analysis tools, and Visualisation:

Rich Livett¹¹⁶

Funding acquisition, Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:

Catherine Ludden^{20,70}

Funding acquisition, Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:

Sally Corden⁷⁴ and Eleni Nastouli^{96,95,30}

Funding acquisition, Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:

Gaia Nebbia ¹²

Funding acquisition, Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:

Ian Johnston ¹¹⁶

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:

Katrina Lythgoe ⁵, M. Estee Torok ^{19, 20} and Ian G Goodfellow ²⁴

Leadership and supervision, Metadata curation, Project administration, Samples and logistics, and Visualisation:

Jacqui A Prieto ^{97, 82} and Kordo Saeed ^{97, 83}

Leadership and supervision, Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:

David K Jackson ¹¹⁶

Leadership and supervision, Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:

Catherine Houlihan ^{96, 94}

Leadership and supervision, Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:

Dan Frampton ^{94, 95}

Metadata curation, Project administration, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

William L Hamilton ¹⁹ and Adam A Witney ⁴¹

Funding acquisition, Samples and logistics, Sequencing and analysis, and Visualisation:

Giselda Bucca ¹⁰¹

Funding acquisition, Leadership and supervision, Metadata curation, and Project administration:

Cassie F Pope ^{40, 41}

Funding acquisition, Leadership and supervision, Metadata curation, and Samples and logistics:

Catherine Moore ⁷⁴

Funding acquisition, Leadership and supervision, Metadata curation, and Sequencing and analysis:

Emma C Thomson ⁵³

Funding acquisition, Leadership and supervision, Project administration, and Samples and logistics:

Teresa Cutino-Moguel ², Ewan M Harrison ^{116, 102}

Funding acquisition, Leadership and supervision, Sequencing and analysis, and Visualisation:

Colin P Smith ¹⁰¹

Leadership and supervision, Metadata curation, Project administration, and Sequencing and analysis:

Fiona Rogan ⁷⁷

Leadership and supervision, Metadata curation, Project administration, and Samples and logistics:

Shaun M Beckwith ⁶, Abigail Murray ⁶, Dawn Singleton ⁶, Kirstine Eastick ³⁷, Liz A Sheridan ⁹⁸, Paul Randell ⁹⁹, Leigh M Jackson ¹⁰⁵, Cristina V Ariani ¹¹⁶ and Sónia Gonçalves ¹¹⁶

Leadership and supervision, Metadata curation, Samples and logistics, and Sequencing and analysis:

Derek J Fairley ^{3,77}, Matthew W Loose ¹⁸ and Joanne Watkins ⁷⁴

Leadership and supervision, Metadata curation, Samples and logistics, and Visualisation:

Samuel Moses ^{25,106}

Leadership and supervision, Metadata curation, Sequencing and analysis, and Software and analysis tools:

Sam Nicholls ⁴³, Matthew Bull ⁷⁴ and Roberto Amato ¹¹⁶

Leadership and supervision, Project administration, Samples and logistics, and Sequencing and analysis:

Darren L Smith ^{36,65,66}

Leadership and supervision, Sequencing and analysis, Software and analysis tools, and Visualisation:

David M Aanensen ^{14,116} and Jeffrey C Barrett ¹¹⁶

Metadata curation, Project administration, Samples and logistics, and Sequencing and analysis:

Beatrix Kele ², Dinesh Aggarwal ^{20,116,70}, James G Shepherd ⁵³, Martin D Curran ⁷¹ and Surendra Parmar ⁷¹

Metadata curation, Project administration, Sequencing and analysis, and Software and analysis tools:

Matthew D Parker ¹⁰⁹

Metadata curation, Samples and logistics, Sequencing and analysis, and Software and analysis tools:

Catryn Williams ⁷⁴

Metadata curation, Samples and logistics, Sequencing and analysis, and Visualisation:

Sharon Glaysher ⁶⁸

Metadata curation, Sequencing and analysis, Software and analysis tools, and Visualisation:

Anthony P Underwood ^{14,116}, Matthew Bashton ^{36,65}, Nicole Pacchiarini ⁷⁴, Katie F Loveson ⁸⁴ and Matthew Byott ^{95,96}

Project administration, Sequencing and analysis, Software and analysis tools, and Visualisation:

Alessandro M Carabelli ²⁰

Funding acquisition, Leadership and supervision, and Metadata curation:

Kate E Templeton ^{56,104}

Funding acquisition, Leadership and supervision, and Project administration:

Sharon J Peacock ^{20,70}, Thushan I de Silva ¹⁰⁹, Dennis Wang ¹⁰⁹, Cordelia F Langford ¹¹⁶ and John Sillitoe ¹¹⁶

Funding acquisition, Leadership and supervision, and Samples and logistics:

Rory N Gunson ⁵⁵

Funding acquisition, Leadership and supervision, and Sequencing and analysis:

Simon Cottrell ⁷⁴, Justin O'Grady ^{75,103} and Dominic Kwiatkowski ^{116,108}

Leadership and supervision, Metadata curation, and Project administration:

Patrick J Lillie ³⁷

Leadership and supervision, Metadata curation, and Samples and logistics:

Nicholas Cortes ³³, Nathan Moore ³³, Claire Thomas ³³, Phillipa J Burns ³⁷, Tabitha W Mahungu ⁸⁰ and Steven Liggett ⁸⁶

Leadership and supervision, Metadata curation, and Sequencing and analysis:

Angela H Beckett ^{13,81} and Matthew TG Holden ⁷³

Leadership and supervision, Project administration, and Samples and logistics:

Lisa J Levett ³⁴, Husam Osman ^{70,35} and Mohammed O Hassan-Ibrahim ⁹⁹

Leadership and supervision, Project administration, and Sequencing and analysis:

David A Simpson ⁷⁷

Leadership and supervision, Samples and logistics, and Sequencing and analysis:

Meera Chand ⁷², Ravi K Gupta ¹⁰², Alistair C Darby ¹⁰⁷ and Steve Paterson ¹⁰⁷

Leadership and supervision, Sequencing and analysis, and Software and analysis tools:

Oliver G Pybus ²³, Erik M Volz ³⁹, Daniela de Angelis ⁵², David L Robertson ⁵³, Andrew J Page ⁷⁵ and Inigo Martincorena ¹¹⁶

Leadership and supervision, Sequencing and analysis, and Visualisation:

Louise Aigrain ¹¹⁶ and Andrew R Bassett ¹¹⁶

Metadata curation, Project administration, and Samples and logistics:

Nick Wong ⁵⁰, Yusri Taha ⁸⁹, Michelle J Erkiert ⁹⁹ and Michael H Spencer Chapman ^{116,102}

Metadata curation, Project administration, and Sequencing and analysis:

Rebecca Dewar ⁵⁶ and Martin P McHugh ^{56,111}

Metadata curation, Project administration, and Software and analysis tools:

Siddharth Mookerjee ^{38,57}

Metadata curation, Project administration, and Visualisation:

Stephen Aplin ⁹⁷, Matthew Harvey ⁹⁷, Thea Sass ⁹⁷, Helen Umpleby ⁹⁷ and Helen Wheeler ⁹⁷

Metadata curation, Samples and logistics, and Sequencing and analysis:

James P McKenna ³, Ben Warne ⁹, Joshua F Taylor ²², Yasmin Chaudhry ²⁴, Rhys Izuagbe ²⁴, Aminu S Jahun ²⁴, Gregory R Young ^{36,65}, Claire McMurray ⁴³, Clare M McCann ^{65,66}, Andrew Nelson ^{65,66} and Scott Elliott ⁶⁸

Metadata curation, Samples and logistics, and Visualisation:

Hannah Lowe ²⁵

Metadata curation, Sequencing and analysis, and Software and analysis tools:

Anna Price ¹¹, Matthew R Crown ⁶⁵, Sara Rey ⁷⁴, Sunando Roy ⁹⁶ and Ben Temperton ¹⁰⁵

Metadata curation, Sequencing and analysis, and Visualisation:

Sharif Shaaban ⁷³ and Andrew R Hesketh ¹⁰¹

Project administration, Samples and logistics, and Sequencing and analysis:

Kenneth G Laing⁴¹, Irene M Monahan⁴¹ and Judith Heaney^{95,96,34}

Project administration, Samples and logistics, and Visualisation:

Emanuela Pelosi⁹⁷, Siona Silveira⁹⁷ and Eleri Wilson-Davies⁹⁷

Samples and logistics, Software and analysis tools, and Visualisation:

Helen Fryer⁵

Sequencing and analysis, Software and analysis tools, and Visualization:

Helen Adams⁴, Louis du Plessis²³, Rob Johnson³⁹, William T Harvey^{53,42}, Joseph Hughes⁵³, Richard J Orton⁵³, Lewis G Spurgin⁵⁹, Yann Bourgeois⁸¹, Chris Ruis¹⁰², Áine O'Toole¹⁰⁴, Marina Gourtovaia¹¹⁶ and Theo Sanderson¹¹⁶

Funding acquisition, and Leadership and supervision:

Christophe Fraser⁵, Jonathan Edgeworth¹², Judith Breuer^{96,29}, Stephen L Michell¹⁰⁵ and John A Todd¹¹⁵

Funding acquisition, and Project administration:

Michaela John¹⁰ and David Buck¹¹⁵

Leadership and supervision, and Metadata curation:

Kavitha Gajee³⁷ and Gemma L Kay⁷⁵

Leadership and supervision, and Project administration:

David Heyburn⁷⁴

Leadership and supervision, and Samples and logistics:

Themoula Charalampous^{12,46}, Adela Alcolea-Medina^{32,112}, Katie Kitchman³⁷, Alan McNally^{43,93}, David T Pritchard⁵⁰, Samir Dervisevic⁵⁸, Peter Muir⁷⁰, Esther Robinson^{70,35}, Barry B Vipond⁷⁰, Newara A Ramadan⁷⁸, Christopher Jeanes⁹⁰, Danni Weldon¹¹⁶, Jana Catalan¹¹⁸ and Neil Jones¹¹⁸

Leadership and supervision, and Sequencing and analysis:

Ana da Silva Filipe⁵³, Chris Williams⁷⁴, Marc Fuchs⁷⁷, Julia Miskelly⁷⁷, Aaron R Jeffries¹⁰⁵, Karen Oliver¹¹⁶ and Naomi R Park¹¹⁶

Metadata curation, and Samples and logistics:

Amy Ash¹, Cherian Koshy¹, Magdalena Barrow⁷, Sarah L Buchan⁷, Anna Mantzouratou⁷, Gemma Clark¹⁵, Christopher W Holmes¹⁶, Sharon Campbell¹⁷, Thomas Davis²¹, Ngee Keong Tan²², Julianne R Brown²⁹, Kathryn A Harris^{29,2}, Stephen P Kidd³³, Paul R Grant³⁴, Li Xu-McCrae³⁵, Alison Cox^{38,63}, Pinglawathee Madona^{38,63}, Marcus Pond^{38,63}, Paul A Randell^{38,63}, Karen T Withell⁴⁸, Cheryl Williams⁵¹, Clive Graham⁶⁰, Rebecca Denton-Smith⁶², Emma Swindells⁶², Robyn Turnbull⁶², Tim J Sloan⁶⁷, Andrew Bosworth^{70,35}, Stephanie Hutchings⁷⁰, Hannah M Pymont⁷⁰, Anna Casey⁷⁶, Liz Ratcliffe⁷⁶, Christopher R Jones^{79,105}, Bridget A Knight^{79,105}, Tanzina Haque⁸⁰, Jennifer Hart⁸⁰, Dianne Irish-Tavares⁸⁰, Eric Witele⁸⁰, Craig Mower⁸⁶, Louisa K Watson⁸⁶, Jennifer Collins⁸⁹, Gary Eltringham⁸⁹, Dorian Crudgington⁹⁸, Ben Macklin⁹⁸, Miren Iturriza-Gomara¹⁰⁷, Anita O Lucaci¹⁰⁷ and Patrick C McClure¹¹³

Metadata curation, and Sequencing and analysis:

Matthew Carlile¹⁸, Nadine Holmes¹⁸, Christopher Moore¹⁸, Nathaniel Storey²⁹, Stefan Rooke⁷³, Gonzalo Yebra⁷³, Noel Craine⁷⁴, Malorie Perry⁷⁴, Nabil-Fareed Alikhan⁷⁵, Stephen Bridgett⁷⁷, Kate F Cook⁸⁴, Christopher Fearn⁸⁴, Salman Goudarzi⁸⁴, Ronan A Lyons⁸⁸, Thomas Williams¹⁰⁴, Sam T Haldenby¹⁰⁷, Jillian Durham¹¹⁶ and Steven Leonard¹¹⁶

Metadata curation, and Software and analysis tools:

Robert M Davies ¹¹⁶

Project administration, and Samples and logistics:

Rahul Batra ¹², Beth Blane ²⁰, Moira J Spyer ^{30,95,96}, Perminder Smith ^{32,112}, Mehmet Yavus ^{85,109}, Rachel J Williams ⁹⁶, Adhyana IK Mahanama ⁹⁷, Buddhini Samaraweera ⁹⁷, Sophia T Girgis ¹⁰², Samantha E Hansford ¹⁰⁹, Angie Green ¹¹⁵, Charlotte Beaver ¹¹⁶, Katherine L Bellis ^{116,102}, Matthew J Dorman ¹¹⁶, Sally Kay ¹¹⁶, Liam Prestwood ¹¹⁶ and Shavanthi Rajatileka ¹¹⁶

Project administration, and Sequencing and analysis:

Joshua Quick ⁴³

Project administration, and Software and analysis tools:

Radoslaw Poplawski ⁴³

Samples and logistics, and Sequencing and analysis:

Nicola Reynolds ⁸, Andrew Mack ¹¹, Arthur Morriss ¹¹, Thomas Whalley ¹¹, Bindi Patel ¹², Iliana Georgana ²⁴, Myra Hosmillo ²⁴, Malte L Pinckert ²⁴, Joanne Stockton ⁴³, John H Henderson ⁶⁵, Amy Hollis ⁶⁵, William Stanley ⁶⁵, Wen C Yew ⁶⁵, Richard Myers ⁷², Alicia Thornton ⁷², Alexander Adams ⁷⁴, Tara Annett ⁷⁴, Hibo Asad ⁷⁴, Alec Birchley ⁷⁴, Jason Coombes ⁷⁴, Johnathan M Evans ⁷⁴, Laia Fina ⁷⁴, Bree Gatica-Wilcox ⁷⁴, Lauren Gilbert ⁷⁴, Lee Graham ⁷⁴, Jessica Hey ⁷⁴, Ember Hilvers ⁷⁴, Sophie Jones ⁷⁴, Hannah Jones ⁷⁴, Sara Kumziene-Summerhayes ⁷⁴, Caoimhe McKerr ⁷⁴, Jessica Powell ⁷⁴, Georgia Pugh ⁷⁴, Sarah Taylor ⁷⁴, Alexander J Trotter ⁷⁵, Charlotte A Williams ⁹⁶, Leanne M Kermack ¹⁰², Benjamin H Foulkes ¹⁰⁹, Marta Gallis ¹⁰⁹, Hailey R Hornsby ¹⁰⁹, Stavroula F Louka ¹⁰⁹, Manoj Pohare ¹⁰⁹, Paige Wolverson ¹⁰⁹, Peijun Zhang ¹⁰⁹, George MacIntyre-Cockett ¹¹⁵, Amy Trebes ¹¹⁵, Robin J Moll ¹¹⁶, Lynne Ferguson ¹¹⁷, Emily J Goldstein ¹¹⁷, Alasdair Maclean ¹¹⁷ and Rachael Tomb ¹¹⁷

Samples and logistics, and Software and analysis tools:

Igor Starinskij ⁵³

Sequencing and analysis, and Software and analysis tools:

Laura Thomson ⁵, Joel Southgate ^{11,74}, Moritz UG Kraemer ²³, Jayna Raghvani ²³, Alex E Zarebski ²³, Olivia Boyd ³⁹, Lily Geidelberg ³⁹, Chris J Illingworth ⁵², Chris Jackson ⁵², David Pascall ⁵², Sreenu Vattipally ⁵³, Timothy M Freeman ¹⁰⁹, Sharon N Hsu ¹⁰⁹, Benjamin B Lindsey ¹⁰⁹, Keith James ¹¹⁶, Kevin Lewis ¹¹⁶, Gerry Tonkin-Hill ¹¹⁶ and Jaime M Tovar-Corona ¹¹⁶

Sequencing and analysis, and Visualisation:

MacGregor Cox ²⁰

Software and analysis tools, and Visualisation:

Khalil Abudahab ^{14,116}, Mirko Menegazzo ¹⁴, Ben EW Taylor MEng ^{14,116}, Corin A Yeats ¹⁴, Afrida Mukaddas ⁵³, Derek W Wright ⁵³, Leonardo de Oliveira Martins ⁷⁵, Rachel Colquhoun ¹⁰⁴, Verity Hill ¹⁰⁴, Ben Jackson ¹⁰⁴, JT McCrone ¹⁰⁴, Nathan Medd ¹⁰⁴, Emily Scher ¹⁰⁴ and Jon-Paul Keatley ¹¹⁶

Leadership and supervision:

Tanya Curran ³, Sian Morgan ¹⁰, Patrick Maxwell ²⁰, Ken Smith ²⁰, Sahar Eldirdiri ²¹, Anita Kenyon ²¹, Alison H Holmes ^{38,57}, James R Price ^{38,57}, Tim Wyatt ⁶⁹, Alison E Mather ⁷⁵, Timofey Skvortsov ⁷⁷ and John A Hartley ⁹⁶

Metadata curation:

Martyn Guest ¹¹, Christine Kitchen ¹¹, Ian Merrick ¹¹, Robert Munn ¹¹, Beatrice Bertolusso ³³, Jessica Lynch ³³, Gabrielle Vernet ³³, Stuart Kirk ³⁴, Elizabeth Wastnedge ⁵⁶, Rachael Stanley ⁵⁸, Giles Idle ⁶⁴, Declan T Bradley ^{69,77}, Nicholas F Killough ⁶⁹, Jennifer Poyner ⁷⁹ and Matilde Mori ¹¹⁰

Project administration:

Owen Jones¹¹, Victoria Wright¹⁸, Ellena Brooks²⁰, Carol M Churcher²⁰, Laia Delgado Callico²⁰, Mireille Fragakis²⁰, Katerina Galai^{20,70}, Andrew Jermy²⁰, Sarah Judges²⁰, Anna Markov²⁰, Georgina M McManus²⁰, Kim S Smith²⁰, Peter M D Thomas-McEwen²⁰, Elaine Westwick²⁰, Stephen W Attwood²³, Frances Bolt^{38,57}, Alisha Davies⁷⁴, Elen De Lacy⁷⁴, Fatima Downing⁷⁴, Sue Edwards⁷⁴, Lizzie Meadows⁷⁵, Sarah Jeremiah⁹⁷, Nikki Smith¹⁰⁹ and Luke Foulser¹¹⁶

Samples and logistics:

Amita Patel¹², Louise Berry¹⁵, Tim Boswell¹⁵, Vicki M Fleming¹⁵, Hannah C Howson-Wells¹⁵, Amelia Joseph¹⁵, Manjinder Khakh¹⁵, Michelle M Lister¹⁵, Paul W Bird¹⁶, Karlie Fallon¹⁶, Thomas Helmer¹⁶, Claire L McMurray¹⁶, Mina Odedra¹⁶, Jessica Shaw¹⁶, Julian W Tang¹⁶, Nicholas J Willford¹⁶, Victoria Blakey¹⁷, Veena Raviprakash¹⁷, Nicola Sheriff¹⁷, Lesley-Anne Williams¹⁷, Theresa Feltwell²⁰, Luke Bedford²⁶, James S Cargill²⁷, Warwick Hughes²⁷, Jonathan Moore²⁸, Susanne Stonehouse²⁸, Laura Atkinson²⁹, Jack CD Lee²⁹, Dr Divya Shah²⁹, Natasha Ohemeng-Kumi^{32,112}, John Ramble^{32,112}, Jasveen Sehmi^{32,112}, Rebecca Williams³³, Wendy Chatterton³⁴, Monika Pusok³⁴, William Everson³⁷, Anibolina Castigador⁴⁴, Emily Macnaughton⁴⁴, Kate El Bouzidi⁴⁵, Temi Lampejo⁴⁵, Malur Sudhanva⁴⁵, Cassie Breen⁴⁷, Graciela Sluga⁴⁸, Shazaad SY Ahmad^{49,70}, Ryan P George⁴⁹, Nicholas W Machin^{49,70}, Debbie Binns⁵⁰, Victoria James⁵⁰, Rachel Blacow⁵⁵, Lindsay Coupland⁵⁸, Louise Smith⁵⁹, Edward Barton⁶⁰, Debra Padgett⁶⁰, Garren Scott⁶⁰, Aidan Cross⁶¹, Mariyam Mirfenderesky⁶¹, Jane Greenaway⁶², Kevin Cole⁶⁴, Phillip Clarke⁶⁷, Nichola Duckworth⁶⁷, Sarah Walsh⁶⁷, Kelly Bicknell⁶⁸, Robert Impey⁶⁸, Sarah Wyllie⁶⁸, Richard Hopes⁷⁰, Chloe Bishop⁷², Vicki Chalker⁷², Ian Harrison⁷², Laura Gifford⁷⁴, Zoltan Molnar⁷⁷, Cressida Auckland⁷⁹, Cariad Evans^{85,109}, Kate Johnson^{85,109}, David G Partridge^{85,109}, Mohammad Raza^{85,109}, Paul Baker⁸⁶, Stephen Bonner⁸⁶, Sarah Essex⁸⁶, Leanne J Murray⁸⁶, Andrew I Lawton⁸⁷, Shirelle Burton-Fanning⁸⁹, Brendan AI Payne⁸⁹, Sheila Waugh⁸⁹, Andrea N Gomes⁹¹, Maimuna Kimuli⁹¹, Darren R Murray⁹¹, Paula Ashfield⁹², Donald Dobie⁹², Fiona Ashford⁹³, Angus Best⁹³, Liam Crawford⁹³, Nicola Cumley⁹³, Megan Mayhew⁹³, Oliver Megram⁹³, Jeremy Mirza⁹³, Emma Moles-Garcia⁹³, Benita Percival⁹³, Megan Driscoll⁹⁶, Leah Ensell⁹⁶, Helen L Lowe⁹⁶, Laurentiu Maftai⁹⁶, Matteo Mondani⁹⁶, Nicola J Chaloner⁹⁹, Benjamin J Cogger⁹⁹, Lisa J Easton⁹⁹, Hannah Huckson⁹⁹, Jonathan Lewis⁹⁹, Sarah Lowdon⁹⁹, Cassandra S Malone⁹⁹, Florence Munemo⁹⁹, Manasa Mutingwende⁹⁹, Roberto Nicodemi⁹⁹, Olga Podplomyk⁹⁹, Thomas Somassa⁹⁹, Andrew Beggs¹⁰⁰, Alex Richter¹⁰⁰, Claire Cormie¹⁰², Joana Dias¹⁰², Sally Forrest¹⁰², Ellen E Higginson¹⁰², Mailis Maes¹⁰², Jamie Young¹⁰², Rose K Davidson¹⁰³, Kathryn A Jackson¹⁰⁷, Alexander J Keeley¹⁰⁹, Jonathan Ball¹¹³, Timothy Byaruhanga¹¹³, Joseph G Chappell¹¹³, Jayasree Dey¹¹³, Jack D Hill¹¹³, Emily J Park¹¹³, Arezou Fanaie¹¹⁴, Rachel A Hilson¹¹⁴, Geraldine Yaze¹¹⁴ and Stephanie Lo¹¹⁶

Sequencing and analysis:

Safiah Afifi¹⁰, Robert Beer¹⁰, Joshua Maksimovic¹⁰, Kathryn McCluggage¹⁰, Karla Spellman¹⁰, Catherine Bresner¹¹, William Fuller¹¹, Angela Marchbank¹¹, Trudy Workman¹¹, Ekaterina Shelest^{13,81}, Johnny Debebe¹⁸, Fei Sang¹⁸, Sarah Francois²³, Bernardo Gutierrez²³, Tetyana I Vasylyeva²³, Flavia Flaviani³¹, Manon Ragonnet-Cronin³⁹, Katherine L Smollett⁴², Alice Broos⁵³, Daniel Mair⁵³, Jenna Nichols⁵³, Kyriaki Nomikou⁵³, Lily Tong⁵³, Ioulia Tsatsani⁵³, Sarah O'Brien⁵⁴, Steven Rushton⁵⁴, Roy Sanderson⁵⁴, Jon Perkins⁵⁵, Seb Cotton⁵⁶, Abbie Gallagher⁵⁶, Elias Allara^{70,102}, Clare Pearson^{70,102}, David Bibby⁷², Gavin Dabrera⁷², Nicholas Ellaby⁷², Eileen Gallagher⁷², Jonathan Hubb⁷², Angie Lackenby⁷², David Lee⁷², Nikos Manesis⁷², Tamyo Mbisa⁷², Steven Platt⁷², Katherine A Twohig⁷², Mari Morgan⁷⁴, Alp Aydin⁷⁵, David J Baker⁷⁵, Ebenezer Foster-Nyarko⁷⁵, Sophie J Prosolek⁷⁵, Steven Rudder⁷⁵, Chris Baxter⁷⁷, Sílvia F Carvalho⁷⁷, Deborah Lavin⁷⁷, Arun Mariappan⁷⁷, Clara Radulescu⁷⁷, Aditi Singh⁷⁷, Miao Tang⁷⁷, Helen Morcrette⁷⁹, Nadua Bayzid⁹⁶, Marius Cotic⁹⁶, Carlos E Balcazar¹⁰⁴, Michael D Gallagher¹⁰⁴, Daniel Maloney¹⁰⁴, Thomas D Stanton¹⁰⁴, Kathleen A Williamson¹⁰⁴, Robin Manley¹⁰⁵, Michelle L Michelsen¹⁰⁵, Christine M Sambles¹⁰⁵, David J Studholme¹⁰⁵, Joanna Warwick-Dugdale¹⁰⁵, Richard Eccles¹⁰⁷, Matthew Gemmell¹⁰⁷, Richard Gregory¹⁰⁷, Margaret Hughes¹⁰⁷, Charlotte Nelson¹⁰⁷, Lucille Rainbow¹⁰⁷, Edith E Vamos¹⁰⁷, Hermione J Webster¹⁰⁷, Mark Whitehead¹⁰⁷, Claudia Wierzbicki¹⁰⁷, Adrienn Angyal¹⁰⁹, Luke R Green¹⁰⁹, Max Whiteley¹⁰⁹, Emma Betteridge¹¹⁶, Iraad F Bronner¹¹⁶, Ben W Farr¹¹⁶, Scott Goodwin¹¹⁶, Stefanie V Lensing¹¹⁶, Shane

A McCarthy ^{116,102}, Michael A Quail ¹¹⁶, Diana Rajan ¹¹⁶, Nicholas M Redshaw ¹¹⁶, Carol Scott ¹¹⁶, Lesley Shirley ¹¹⁶ and Scott AJ Thurston ¹¹⁶

Software and analysis tools:

Will Rowe ⁴³, Amy Gaskin ⁷⁴, Thanh Le-Viet ⁷⁵, James Bonfield ¹¹⁶, Jennifer Liddle ¹¹⁶ and Andrew Whitwham ¹¹⁶

1 Barking, Havering and Redbridge University Hospitals NHS Trust, **2** Barts Health NHS Trust, **3** Belfast Health & Social Care Trust, **4** Betsi Cadwaladr University Health Board, **5** Big Data Institute, Nuffield Department of Medicine, University of Oxford, **6** Blackpool Teaching Hospitals NHS Foundation Trust, **7** Bournemouth University, **8** Cambridge Stem Cell Institute, University of Cambridge, **9** Cambridge University Hospitals NHS Foundation Trust, **10** Cardiff and Vale University Health Board, **11** Cardiff University, **12** Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, **13** Centre for Enzyme Innovation, University of Portsmouth, **14** Centre for Genomic Pathogen Surveillance, University of Oxford, **15** Clinical Microbiology Department, Queens Medical Centre, Nottingham University Hospitals NHS Trust, **16** Clinical Microbiology, University Hospitals of Leicester NHS Trust, **17** County Durham and Darlington NHS Foundation Trust, **18** Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, **19** Department of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust, **20** Department of Medicine, University of Cambridge, **21** Department of Microbiology, Kettering General Hospital, **22** Department of Microbiology, South West London Pathology, **23** Department of Zoology, University of Oxford, **24** Division of Virology, Department of Pathology, University of Cambridge, **25** East Kent Hospitals University NHS Foundation Trust, **26** East Suffolk and North Essex NHS Foundation Trust, **27** East Sussex Healthcare NHS Trust, **28** Gateshead Health NHS Foundation Trust, **29** Great Ormond Street Hospital for Children NHS Foundation Trust, **30** Great Ormond Street Institute of Child Health (GOS ICH), University College London (UCL), **31** Guy's and St. Thomas' Biomedical Research Centre, **32** Guy's and St. Thomas' NHS Foundation Trust, **33** Hampshire Hospitals NHS Foundation Trust, **34** Health Services Laboratories, **35** Heartlands Hospital, Birmingham, **36** Hub for Biotechnology in the Built Environment, Northumbria University, **37** Hull University Teaching Hospitals NHS Trust, **38** Imperial College Healthcare NHS Trust, **39** Imperial College London, **40** Infection Care Group, St George's University Hospitals NHS Foundation Trust, **41** Institute for Infection and Immunity, St George's University of London, **42** Institute of Biodiversity, Animal Health & Comparative Medicine, **43** Institute of Microbiology and Infection, University of Birmingham, **44** Isle of Wight NHS Trust, **45** King's College Hospital NHS Foundation Trust, **46** King's College London, **47** Liverpool Clinical Laboratories, **48** Maidstone and Tunbridge Wells NHS Trust, **49** Manchester University NHS Foundation Trust, **50** Microbiology Department, Buckinghamshire Healthcare NHS Trust, **51** Microbiology, Royal Oldham Hospital, **52** MRC Biostatistics Unit, University of Cambridge, **53** MRC-University of Glasgow Centre for Virus Research, **54** Newcastle University, **55** NHS Greater Glasgow and Clyde, **56** NHS Lothian, **57** NIHR Health Protection Research Unit in HCAI and AMR, Imperial College London, **58** Norfolk and Norwich University Hospitals NHS Foundation Trust, **59** Norfolk County Council, **60** North Cumbria Integrated Care NHS Foundation Trust, **61** North Middlesex University Hospital NHS Trust, **62** North Tees and Hartlepool NHS Foundation Trust, **63** North West London Pathology, **64** Northumbria Healthcare NHS Foundation Trust, **65** Northumbria University, **66** NU-OMICS, Northumbria University, **67** Path Links, Northern Lincolnshire and Goole NHS Foundation Trust, **68** Portsmouth Hospitals University NHS Trust, **69** Public Health Agency, Northern Ireland, **70** Public Health England, **71** Public Health England, Cambridge, **72** Public Health England, Colindale, **73** Public Health Scotland, **74** Public Health Wales, **75** Quadram Institute Bioscience, **76** Queen Elizabeth Hospital, Birmingham, **77** Queen's University Belfast, **78** Royal Brompton and Harefield Hospitals, **79** Royal Devon and Exeter NHS Foundation Trust, **80** Royal Free London NHS Foundation Trust, **81** School of Biological Sciences, University of Portsmouth, **82** School of Health Sciences, University of Southampton, **83** School of Medicine, University of Southampton, **84** School of Pharmacy & Biomedical Sciences, University of Portsmouth, **85** Sheffield Teaching Hospitals NHS Foundation Trust, **86** South Tees Hospitals NHS Foundation Trust, **87** Southwest Pathology Services, **88** Swansea University, **89** The Newcastle upon Tyne Hospitals NHS Foundation Trust, **90** The Queen Elizabeth Hospital King's Lynn NHS Foundation Trust, **91** The Royal Marsden NHS Foundation Trust, **92** The

Royal Wolverhampton NHS Trust, **93** Turnkey Laboratory, University of Birmingham, **94** University College London Division of Infection and Immunity, **95** University College London Hospital Advanced Pathogen Diagnostics Unit, **96** University College London Hospitals NHS Foundation Trust, **97** University Hospital Southampton NHS Foundation Trust, **98** University Hospitals Dorset NHS Foundation Trust, **99** University Hospitals Sussex NHS Foundation Trust, **100** University of Birmingham, **101** University of Brighton, **102** University of Cambridge, **103** University of East Anglia, **104** University of Edinburgh, **105** University of Exeter, **106** University of Kent, **107** University of Liverpool, **108** University of Oxford, **109** University of Sheffield, **110** University of Southampton, **111** University of St Andrews, **112** Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, **113** Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, **114** Watford General Hospital, **115** Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, **116** Wellcome Sanger Institute, **117** West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, **118** Whittington Health NHS Trust

References

1. R. Viana, S. Moyo, D. G. Amoako, H. Tegally, C. Scheepers, C. L. Althaus, U. J. Anyaneji, P. A. Bester, M. F. Boni, M. Chand, W. T. Choga, R. Colquhoun, M. Davids, K. Deforche, D. Doolabh, L. du Plessis, S. Engelbrecht, J. Everatt, J. Giandhari, M. Giovanetti, D. Hardie, V. Hill, N.-Y. Hsiao, A. Iranzadeh, A. Ismail, C. Joseph, R. Joseph, L. Koopile, S. L. Kosakovsky Pond, M. U. G. Kraemer, L. Kuate-Lere, O. Laguda-Akingba, O. Lesetedi-Mafoko, R. J. Lessells, S. Lockman, A. G. Lucaci, A. Maharaj, B. Mahlangu, T. Maponga, K. Mahlakwane, Z. Makatini, G. Marais, D. Maruapula, K. Masupu, M. Matshaba, S. Mayaphi, N. Mbhele, M. B. Mbulawa, A. Mendes, K. Mlisana, A. Mnguni, T. Mohale, M. Moir, K. Moruisi, M. Mosepele, G. Motsatsi, M. S. Motsaledi, T. Mphoyakgosi, N. Msomi, P. N. Mwangi, Y. Naidoo, N. Ntuli, M. Nyaga, L. Olubayo, S. Pillay, B. Radibe, Y. Ramphal, U. Ramphal, J. E. San, L. Scott, R. Shapiro, L. Singh, P. Smith-Lawrence, W. Stevens, A. Strydom, K. Subramoney, N. Tebeila, D. Tshiabuila, J. Tsui, S. van Wyk, S. Weaver, C. K. Wibmer, E. Wilkinson, N. Wolter, A. E. Zarebski, B. Zuze, D. Goedhals, W. Preiser, F. Treurnicht, M. Venter, C. Williamson, O. G. Pybus, J. Bhiman, A. Glass, D. P. Martin, A. Rambaut, S. Gaseitsiwe, A. von Gottberg, T. de Oliveira, Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*. **603**, 679–686 (2022).
2. Classification of Omicron (B.1.1.529): SARS-CoV-2 variant of Concern, (available at [https://www.who.int/news-room/statements/26-11-2021-classification-of-omicron-\(b.1.1.529\)-sars-cov-2-variant-of-concern](https://www.who.int/news-room/statements/26-11-2021-classification-of-omicron-(b.1.1.529)-sars-cov-2-variant-of-concern)).
3. J. R. C. Pulliam, C. van Schalkwyk, N. Govender, A. von Gottberg, C. Cohen, M. J. Groome, J. Dushoff, K. Mlisana, H. Moultrie, Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science*. **376**, eabn4947 (2022).
4. A. Rössler, L. Riepler, D. Bante, D. von Laer, J. Kimpel, SARS-CoV-2 Omicron Variant Neutralization in Serum from Vaccinated and Convalescent Persons. *N. Engl. J. Med.* **386**, 698–700 (2022).
5. F. P. Lyngse, L. H. Mortensen, M. J. Denwood, L. E. Christiansen, C. H. Møller, R. L. Skov, K. Spiess, A. Fomsgaard, R. Lassaunière, M. Rasmussen, M. Stegger, C. Nielsen, R. N. Sieber, A. S. Cohen, F. T. Møller, M. Overvad, K. Mølbak, T. G. Krause, C. T. Kirkeby, Household transmission of the SARS-CoV-2 Omicron variant in Denmark. *Nat. Commun.* **13**, 5573 (2022).
6. J. A. Backer, D. Eggink, S. P. Andeweg, I. K. Veldhuijzen, N. van Maarseveen, K. Vermaas, B.

- Vlaemynck, R. Scheepers, S. van den Hof, C. B. Reusken, J. Wallinga, Shorter serial intervals in SARS-CoV-2 cases with Omicron BA.1 variant compared with Delta variant, the Netherlands, 13 to 26 December 2021. *Eurosurveillance*. **27** (2022), . doi:10.2807/1560-7917.es.2022.27.6.2200042.
7. T. Briefing, SARS-CoV-2 variants of concern and variants under investigation in England (2022), (available at https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1056487/Technical-Briefing-36-22.02.22.pdf).
 8. K. P. Y. Hui, J. C. W. Ho, M.-C. Cheung, K.-C. Ng, R. H. H. Ching, K.-L. Lai, T. T. Kam, H. Gu, K.-Y. Sit, M. K. Y. Hsin, T. W. K. Au, L. L. M. Poon, M. Peiris, J. M. Nicholls, M. C. W. Chan, SARS-CoV-2 Omicron variant replication in human bronchus and lung ex vivo. *Nature*. **603**, 715–720 (2022).
 9. T. P. Peacock, J. C. Brown, J. Zhou, N. Thakur, K. Sukhova, J. Newman, R. Kugathasan, A. W. C. Yan, W. Furnon, G. De Lorenzo, Others, The altered entry pathway and antigenic distance of the SARS-CoV-2 Omicron variant map to separate domains of spike protein. *bioRxiv*. **15**, e0241955 (2022).
 10. S. Cele, L. Jackson, D. S. Khoury, K. Khan, T. Moyo-Gwete, H. Tegally, J. E. San, D. Cromer, C. Scheepers, D. G. Amoako, F. Karim, M. Bernstein, G. Lustig, D. Archary, M. Smith, Y. Ganga, Z. Jule, K. Reedoy, S.-H. Hwa, J. Giandhari, J. M. Blackburn, B. I. Gosnell, S. S. Abdool Karim, W. Hanekom, NGS-SA, COMMIT-KZN Team, A. von Gottberg, J. N. Bhiman, R. J. Lessells, M.-Y. S. Moosa, M. P. Davenport, T. de Oliveira, P. L. Moore, A. Sigal, Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature*. **602**, 654–656 (2022).
 11. S. Mallapaty, Omicron-variant border bans ignore the evidence, say scientists. *Nature*. **600**, 199 (2021).
 12. Prime Minister’s Office, Prime Minister sets out new measures as Omicron variant identified in UK: 27 November 2021. *GOV.UK* (2021), (available at <https://www.gov.uk/government/news/prime-minister-sets-out-new-measures-as-omicron-variant-identified-in-uk-27-november-2021>).
 13. UK Health Security Agency, COVID-19 variants identified in the UK – latest updates. *GOV.UK* (2021), (available at <https://www.gov.uk/government/news/covid-19-variants-identified-in-the-uk-latest-updates>).
 14. P. Elliott, B. Bodinier, O. Eales, H. Wang, D. Haw, J. Elliott, M. Whitaker, J. Jonnerby, D. Tang, C. E. Walters, C. Atchison, P. J. Diggle, A. J. Page, A. J. Trotter, D. Ashby, W. Barclay, G. Taylor, H. Ward, A. Darzi, G. S. Cooke, M. Chadeau-Hyam, C. A. Donnelly, Rapid increase in Omicron infections in England during December 2021: REACT-1 study. *Science*. **375**, 1406–1411 (2022).
 15. P. Elliott, O. Eales, N. Steyn, D. Tang, B. Bodinier, H. Wang, J. Elliott, M. Whitaker, C. Atchison, P. J. Diggle, A. J. Page, A. J. Trotter, D. Ashby, W. Barclay, G. Taylor, H. Ward, A. Darzi, G. S. Cooke, C. A. Donnelly, M. Chadeau-Hyam, Twin peaks: The Omicron SARS-CoV-2 BA.1 and BA.2 epidemics in England. *Science*. **376**, eabq4411 (2022).
 16. M. U. G. Kraemer, V. Hill, C. Ruis, S. Dellicour, S. Bajaj, J. T. McCrone, G. Baele, K. V. Parag, A. L. Battle, B. Gutierrez, B. Jackson, R. Colquhoun, A. O’Toole, B. Klein, A. Vespignani, COVID-19 Genomics UK (COG-UK) Consortium, E. Volz, N. R. Faria, D. M. Aanensen, N. J. Loman, L. du Plessis, S. Cauchemez, A. Rambaut, S. V. Scarpino, O. G. Pybus, Spatiotemporal invasion dynamics of SARS-CoV-2 lineage B.1.1.7 emergence. *Science*. **373**, 889–895 (2021).

17. J. T. McCrone, V. Hill, S. Bajaj, R. E. Pena, B. C. Lambert, R. Inward, S. Bhatt, E. Volz, C. Ruis, S. Dellicour, G. Baele, A. E. Zarebski, A. Sadilek, N. Wu, A. Schneider, X. Ji, J. Raghwani, B. Jackson, R. Colquhoun, Á. O'Toole, T. P. Peacock, K. Twohig, S. Thelwall, G. Dabrera, R. Myers, COVID-19 Genomics UK (COG-UK) Consortium, N. R. Faria, C. Huber, I. I. Bogoch, K. Khan, L. du Plessis, J. C. Barrett, D. M. Aanensen, W. S. Barclay, M. Chand, T. Connor, N. J. Loman, M. A. Suchard, O. G. Pybus, A. Rambaut, M. U. G. Kraemer, Context-specific emergence and growth of the SARS-CoV-2 Delta variant. *Nature*. **610**, 154–160 (2022).
18. Prime Minister's Office, Prime Minister confirms move to Plan B in England. *GOV.UK* (2021), (available at <https://www.gov.uk/government/news/prime-minister-confirms-move-to-plan-b-in-england>).
19. N. H. S. England, NHS England » NHS sets out next steps to accelerate COVID-19 booster rollout, (available at <https://www.england.nhs.uk/2021/12/nhs-sets-out-next-steps-to-accelerate-covid-19-booster-rollout/>).
20. P. Elliott, O. Eales, B. Bodinier, D. Tang, H. Wang, J. Jonnerby, D. Haw, J. Elliott, M. Whitaker, C. E. Walters, C. Atchison, P. J. Diggle, A. J. Page, A. J. Trotter, D. Ashby, W. Barclay, G. Taylor, H. Ward, A. Darzi, G. S. Cooke, M. Chadeau-Hyam, C. A. Donnelly, Post-peak dynamics of a national Omicron SARS-CoV-2 epidemic during January 2022. *bioRxiv* (2022), , doi:10.1101/2022.02.03.22270365.
21. M. Chadeau-Hyam, D. Tang, O. Eales, B. Bodinier, H. Wang, J. Jonnerby, M. Whitaker, J. Elliott, D. Haw, C. E. Walters, C. Atchinson, P. J. Diggle, A. J. Page, D. Ashby, W. Barclay, G. Taylor, G. Cooke, H. Ward, A. Darzi, C. A. Donnelly, P. Elliott, The Omicron SARS-CoV-2 epidemic in England during February 2022, , doi:10.1101/2022.03.10.22272177.
22. SARS-CoV-2 evolution, post-Omicron. *Virological* (2022), (available at <https://virological.org/t/sars-cov-2-evolution-post-omicron/911>).
23. S. Chang, D. Vrabac, J. Leskovec, J. Ugander, Estimating Geographic Spillover Effects of COVID-19 Policies From Large-Scale Mobility Networks. *arXiv [cs.CY]* (2022), (available at <http://arxiv.org/abs/2212.06224>).
24. COVID-19 Genomics UK (COG-UK) consortiumcontact@cogconsortium.uk, An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe*. **1**, e99–e100 (2020).
25. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data – from vision to reality. *Euro Surveill*. **22** (2017), doi:10.2807/1560-7917.es.2017.22.13.30494.
26. robj, *robj411/sequencing_coverage: for B.1.1.7 phylodynamic analysis* (2021); <https://zenodo.org/record/4599180>).
27. M. U. G. Kraemer, C.-H. Yang, B. Gutierrez, C.-H. Wu, B. Klein, D. M. Pigott, Open COVID-19 Data Working Group, L. du Plessis, N. R. Faria, R. Li, W. P. Hanage, J. S. Brownstein, M. Layan, A. Vespignani, H. Tian, C. Dye, O. G. Pybus, S. V. Scarpino, The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science*. **368**, 493–497 (2020).
28. H. Tegally, E. Wilkinson, D. Martin, M. Moir, A. Brito, M. Giovanetti, K. Khan, C. Huber, I. I. Bogoch, J. E. San, J. L.-H. Tsui, J. Poongavanan, J. S. Xavier, D. da S. Candido, F. Romero, C. Baxter, O. G. Pybus, R. Lessells, N. R. Faria, M. U. G. Kraemer, T. de Oliveira, Global Expansion of SARS-CoV-2 Variants of Concern: Dispersal Patterns and Influence of Air Travel. *medRxiv* (2022), p. 2022.11.22.22282629.
29. Variants: distribution of case data, 18 February 2022. *GOV.UK*, (available at

<https://www.gov.uk/government/publications/covid-19-variants-genomically-confirmed-case-numbers/variants-distribution-of-case-data-18-february-2022>).

30. L. du Plessis, J. T. McCrone, A. E. Zarebski, V. Hill, C. Ruis, B. Gutierrez, J. Raghvani, J. Ashworth, R. Colquhoun, T. R. Connor, N. R. Faria, B. Jackson, N. J. Loman, Á. O'Toole, S. M. Nicholls, K. V. Parag, E. Scher, T. I. Vasylyeva, E. M. Volz, A. Watts, I. I. Bogoch, K. Khan, COVID-19 Genomics UK (COG-UK) Consortium, D. M. Aanensen, M. U. G. Kraemer, A. Rambaut, O. G. Pybus, Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*. **371**, 708–712 (2021).
31. P. Lemey, S. L. Hong, V. Hill, G. Baele, C. Poletto, V. Colizza, Á. O'Toole, J. T. McCrone, K. G. Andersen, M. Worobey, M. I. Nelson, A. Rambaut, M. A. Suchard, Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat. Commun.* **11**, 5110 (2020).
32. D. Aggarwal, A. J. Page, U. Schaefer, G. M. Savva, R. Myers, E. Volz, N. Ellaby, S. Platt, N. Groves, E. Gallagher, N. M. Tumelty, T. Le Viet, G. J. Hughes, C. Chen, C. Turner, S. Logan, A. Harrison, COVID-19 Genomics UK (COG-UK) Consortium, S. J. Peacock, M. Chand, E. M. Harrison, Genomic assessment of quarantine measures to prevent SARS-CoV-2 importation and transmission. *Nat. Commun.* **13**, 1012 (2022).
33. 2011 rural/urban classification, (available at <https://www.ons.gov.uk/methodology/geography/geographicalproducts/ruralurbanclassifications/2011ruralurbanclassification>).
34. V. Charu, S. Zeger, J. Gog, O. N. Bjørnstad, S. Kissler, L. Simonsen, B. T. Grenfell, C. Viboud, Human mobility and the spatial transmission of influenza in the United States. *PLoS Comput. Biol.* **13**, e1005382 (2017).
35. M. U. G. Kraemer, N. R. Faria, R. C. Reiner Jr, Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015–16: a modelling study. *Lancet Infect. Dis.* (2017) (available at <https://www.sciencedirect.com/science/article/pii/S1473309916305138>).
36. M. U. G. Kraemer, N. Golding, D. Bisanzio, S. Bhatt, D. M. Pigott, S. E. Ray, O. J. Brady, J. S. Brownstein, N. R. Faria, D. A. T. Cummings, O. G. Pybus, D. L. Smith, A. J. Tatem, S. I. Hay, R. C. Reiner Jr, Utilizing general human movement models to predict the spread of emerging infectious diseases in resource poor settings. *Sci. Rep.* **9**, 5151 (2019).
37. B. Finkenstädt, B. Grenfell, Empirical determinants of measles metapopulation dynamics in England and Wales. *Proc. Biol. Sci.* **265**, 211–220 (1998).
38. P. Nouvellet, S. Bhatia, A. Cori, K. E. C. Ainslie, M. Baguelin, S. Bhatt, A. Boonyasiri, N. F. Brazeau, L. Cattarino, L. V. Cooper, H. Coupland, Z. M. Cucunuba, G. Cuomo-Dannenburg, A. Dighe, B. A. Djaafara, I. Dorigatti, O. D. Eales, S. L. van Elsland, F. F. Nascimento, R. G. FitzJohn, K. A. M. Gaythorpe, L. Geidelberg, W. D. Green, A. Hamlet, K. Hauck, W. Hinsley, N. Imai, B. Jeffrey, E. Knock, D. J. Laydon, J. A. Lees, T. Mangal, T. A. Mellan, G. Nedjati-Gilani, K. V. Parag, M. Pons-Salort, M. Ragonnet-Cronin, S. Riley, H. J. T. Unwin, R. Verity, M. A. C. Vollmer, E. Volz, P. G. T. Walker, C. E. Walters, H. Wang, O. J. Watson, C. Whittaker, L. K. Whittles, X. Xi, N. M. Ferguson, C. A. Donnelly, Reduction in mobility and COVID-19 transmission. *Nat. Commun.* **12**, 1090 (2021).
39. D. Brockmann, D. Helbing, The Hidden Geometry of Complex, Network-Driven Contagion Phenomena. *Science*. **342** (2013), pp. 1337–1342.
40. M. U. G. Kraemer, N. R. Faria, R. C. Reiner Jr, N. Golding, B. Nikolay, S. Stasse, M. A. Johansson, H. Salje, O. Faye, G. R. W. Wint, M. Niedrig, F. M. Shearer, S. C. Hill, R. N.

- Thompson, D. Bisanzio, N. Taveira, H. H. Nax, B. S. R. Pradelski, E. O. Nsoesie, N. R. Murphy, I. I. Bogoch, K. Khan, J. S. Brownstein, A. J. Tatem, T. de Oliveira, D. L. Smith, A. A. Sall, O. G. Pybus, S. I. Hay, S. Cauchemez, Spread of yellow fever virus outbreak in Angola and the Democratic Republic of the Congo 2015-16: a modelling study. *Lancet Infect. Dis.* **17**, 330–338 (2017).
41. C. L. Murall, E. Fournier, J. H. Galvez, A. N’Guessan, S. J. Reiling, P.-O. Quirion, S. Naderi, A.-M. Roy, S.-H. Chen, P. Stretenowich, M. Bourgey, D. Bujold, R. Gregoire, P. Lepage, J. St-Cyr, P. Willet, R. Dion, H. Charest, M. Lathrop, M. Roger, G. Bourque, J. Ragoussis, B. J. Shapiro, S. Moreira, A small number of early introductions seeded widespread transmission of SARS-CoV-2 in Québec, Canada. *Genome Med.* **13**, 169 (2021).
 42. T. S. Brett, P. Rohani, Containing novel SARS-CoV-2 variants at source is possible with high-intensity sequencing. *PNAS Nexus.* **1**, gac159 (2022).
 43. N. M. Ferguson, D. A. T. Cummings, S. Cauchemez, C. Fraser, S. Riley, A. Meeyai, S. Iamsirithaworn, D. S. Burke, Strategies for containing an emerging influenza pandemic in Southeast Asia. *Nature.* **437**, 209–214 (2005).
 44. C. Fraser, S. Riley, R. M. Anderson, N. M. Ferguson, Factors that make an infectious disease outbreak controllable. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 6146–6151 (2004).
 45. V. Hill, L. Du Plessis, T. P. Peacock, D. Aggarwal, R. Colquhoun, A. M. Carabelli, N. Ellaby, E. Gallagher, N. Groves, B. Jackson, J. T. McCrone, Á. O’Toole, A. Price, T. Sanderson, E. Scher, J. Southgate, E. Volz, W. S. Barclay, J. C. Barrett, M. Chand, T. Connor, I. Goodfellow, R. K. Gupta, E. M. Harrison, N. Loman, R. Myers, D. L. Robertson, O. G. Pybus, A. Rambaut, The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evol.* **8**, veac080 (2022).
 46. B. J. Willett, J. Grove, O. A. MacLean, C. Wilkie, G. De Lorenzo, W. Furnon, D. Cantoni, S. Scott, N. Logan, S. Ashraf, M. Manali, A. Szemiel, V. Cowton, E. Vink, W. T. Harvey, C. Davis, P. Asamaphan, K. Smollett, L. Tong, R. Orton, J. Hughes, P. Holland, V. Silva, D. J. Pascall, K. Puxty, A. da Silva Filipe, G. Yebra, S. Shaaban, M. T. G. Holden, R. M. Pinto, R. Gunson, K. Templeton, P. R. Murcia, A. H. Patel, P. Klenerman, S. Dunachie, PITCH Consortium, COVID-19 Genomics UK (COG-UK) Consortium, J. Haughney, D. L. Robertson, M. Palmirini, S. Ray, E. C. Thomson, SARS-CoV-2 Omicron is an immune escape variant with an altered cell entry pathway. *Nat Microbiol.* **7**, 1161–1179 (2022).
 47. C. Stein, H. Nassereldine, R. J. D. Sorensen, J. O. Amlag, C. Bisignano, S. Byrne, E. Castro, K. Coberly, J. K. Collins, J. Dalos, F. Daoud, A. Deen, E. Gakidou, J. R. Giles, E. N. Hulland, B. M. Huntley, K. E. Kinzel, R. Lozano, A. H. Mokdad, T. Pham, D. M. Pigott, R. C. Reiner Jr, T. Vos, S. I. Hay, C. J. L. Murray, S. S. Lim, Past SARS-CoV-2 infection protection against reinfection: a systematic review and meta-analysis. *Lancet.* **401**, 833–842 (2023).
 48. A. F. Brito, E. Semenova, G. Dudas, G. W. Hassler, C. C. Kalinich, M. U. G. Kraemer, J. Ho, H. Tegally, G. Githinji, C. N. Agoti, L. E. Matkin, C. Whittaker, Bulgarian SARS-CoV-2 sequencing group, Communicable Diseases Genomics Network (Australia and New Zealand), COVID-19 Impact Project, Danish Covid-19 Genome Consortium, Fiocruz COVID-19 Genomic Surveillance Network, GISAID core curation team, Network for Genomic Surveillance in South Africa (NGS-SA), Swiss SARS-CoV-2 Sequencing Consortium, B. P. Howden, V. Sintchenko, N. S. Zuckerman, O. Mor, H. M. Blankenship, T. de Oliveira, R. T. P. Lin, M. M. Siqueira, P. C. Resende, A. T. R. Vasconcelos, F. R. Spilki, R. S. Aguiar, I. Alexiev, I. N. Ivanov, I. Philipova, C. V. F. Carrington, N. S. D. Sahadeo, B. Branda, C. Gurry, S. Maurer-Stroh, D. Naidoo, K. J. von Eije, M. D. Perkins, M. van Kerkhove, S. C. Hill, E. C. Sabino, O. G. Pybus, C. Dye, S. Bhatt, S. Flaxman, M. A. Suchard, N. D. Grubaugh, G. Baele, N. R. Faria, Global disparities in SARS-CoV-2 genomic surveillance. *Nat. Commun.* **13**, 7003 (2022).

49. V. Moorthy, O. Morgan, C. Ihekweazu, S. Swaminathan, WHO principles speed up ethical sharing of pathogen genomic data. *Nature*. **611**, 449 (2022).
50. V. Hill, C. Ruis, S. Bajaj, O. G. Pybus, M. U. G. Kraemer, Progress and challenges in virus genomic epidemiology. *Trends Parasitol.* **37**, 1038–1049 (2021).
51. M. U. G. Kraemer, A. Sadilek, Q. Zhang, N. A. Marchal, G. Tuli, E. L. Cohn, Y. Hswen, T. A. Perkins, D. L. Smith, R. C. Reiner Jr, J. S. Brownstein, Mapping global variation in human mobility. *Nat Hum Behav.* **4**, 800–810 (2020).
52. P. Lemey, N. Ruktanonchai, S. L. Hong, V. Colizza, C. Poletto, F. Van den Broeck, M. S. Gill, X. Ji, A. Levasseur, B. B. Oude Munnink, M. Koopmans, A. Sadilek, S. Lai, A. J. Tatem, G. Baele, M. A. Suchard, S. Dellicour, Untangling introductions and persistence in COVID-19 resurgence in Europe. *Nature* (2021), doi:10.1038/s41586-021-03754-2.
53. R. J. Wilson, C. Y. Zhang, W. Lam, D. Desfontaines, D. Simmons-Marengo, B. Gipson, Differentially Private SQL with Bounded User Contribution. *arXiv [cs.CR]* (2019), (available at <http://arxiv.org/abs/1909.01917>).
54. L. Du Plessis, J. T. McCrone, A. E. Zarebski, V. Hill, C. Ruis, Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* (2021) (available at <https://www.science.org/doi/abs/10.1126/science.abf2946>).
55. *beast-mcmc* (Github).
56. E. Volz, S. Mishra, M. Chand, J. C. Barrett, R. Johnson, L. Geidelberg, W. R. Hinsley, D. J. Laydon, G. Dabrera, Á. O’Toole, R. Amato, M. Ragonnet-Cronin, I. Harrison, B. Jackson, C. V. Ariani, O. Boyd, N. J. Loman, J. T. McCrone, S. Gonçalves, D. Jorgensen, R. Myers, V. Hill, D. K. Jackson, K. Gaythorpe, N. Groves, J. Sillitoe, D. P. Kwiatkowski, S. Flaxman, O. Ratmann, S. Bhatt, S. Hopkins, A. Gandy, A. Rambaut, N. M. Ferguson, Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature*, 1–17 (2021).
57. B. Q. Minh, H. A. Schmidt, O. Chernomor, D. Schrempf, M. D. Woodhams, A. von Haeseler, R. Lanfear, IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
58. P. Sagulenko, V. Puller, R. A. Neher, TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
59. Y. Turakhia, B. Thornlow, A. S. Hinrichs, N. De Maio, L. Gozashti, R. Lanfear, D. Haussler, R. Corbett-Detig, Ultrafast Sample placement on Existing tRees (USHER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. *Nat. Genet.* **53**, 809–816 (2021).
60. C. Ye, B. Thornlow, A. Hinrichs, A. Kramer, C. Mirchandani, D. Torvi, R. Lanfear, R. Corbett-Detig, Y. Turakhia, matOptimize: a parallel tree optimization method enables online phylogenetics for SARS-CoV-2. *Bioinformatics.* **38**, 3734–3740 (2022).
61. T. Sanderson, Chronumental: time tree estimation from very large phylogenies. *bioRxiv* (2022), p. 2021.10.27.465994.
62. 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).
63. 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).

64. 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).
65. 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).
66. Travel to England from another country during coronavirus (COVID-19). *GOV.UK*, (available at <https://www.gov.uk/guidance/travel-to-england-from-another-country-during-coronavirus-covid-19>).
67. P. Lemey, A. Rambaut, J. J. Welch, M. A. Suchard, Phylogeography takes a relaxed random walk in continuous space and time. *Mol. Biol. Evol.* **27**, 1877–1885 (2010).
68. X. Ji, Z. Zhang, A. Holbrook, A. Nishimura, G. Baele, A. Rambaut, P. Lemey, M. A. Suchard, Gradients Do Grow on Trees: A Linear-Time $O(N)$ -Dimensional Gradient for Statistical Phylogenetics. *Mol. Biol. Evol.* **37**, 3047–3060 (2020).
69. G. Baele, M. S. Gill, P. Lemey, M. A. Suchard, Hamiltonian Monte Carlo sampling to estimate past population dynamics using the skygrid coalescent model in a Bayesian phylogenetics framework. *Wellcome Open Res.* **5**, 53 (2020).
70. I. TJCoVa, JCVI statement regarding a COVID-19 booster vaccine programme for winter 2021 to 2022. *The Joint Committee on Vaccination and Immunisation*.
71. Department of Health, Social Care, People urged to get booster jabs to keep your family protected this Christmas. *GOV.UK* (2021), (available at <https://www.gov.uk/government/news/people-urged-to-get-booster-jabs-to-keep-your-family-protected-this-christmas>).
72. Department of Health, Social Care, All adults to be offered COVID-19 boosters by end of January. *GOV.UK* (2021), (available at <https://www.gov.uk/government/news/all-adults-to-be-offered-covid-19-boosters-by-end-of-january>).
73. L. Breiman, Random Forests. *Mach. Learn.* **45**, 5–32 (2001).
74. A. Fisher, C. Rudin, F. Dominici, All Models are Wrong, but Many are Useful: Learning a Variable’s Importance by Studying an Entire Class of Prediction Models Simultaneously. *J. Mach. Learn. Res.* **20** (2019), doi:10.1080/01621459.1963.10500830.
75. L. Mentch, G. Hooker, Quantifying uncertainty in random forests via confidence intervals and hypothesis tests. *arXiv [stat.ML]* (2014), (available at <https://jmlr.org/papers/volume17/14-168/14-168.pdf>).
76. M. Rosvall, D. Axelsson, C. T. Bergstrom, The map equation. *Eur. Phys. J. Spec. Top.* **178**, 13–23 (2009).
77. M. Rosvall, C. T. Bergstrom, Multilevel compression of random walks on networks reveals hierarchical organization in large integrated systems. *PLoS One.* **6**, e18209 (2011).