



This is a repository copy of *Genomic epidemiology of SARS-CoV-2 in a UK university identifies dynamics of transmission.*

White Rose Research Online URL for this paper:
<https://eprints.whiterose.ac.uk/183970/>

Version: Published Version

Article:

Aggarwal, D, Warne, B, Jahun, AS et al. (725 more authors) (2022) Genomic epidemiology of SARS-CoV-2 in a UK university identifies dynamics of transmission. *Nature Communications*, 13 (1). 751. ISSN 2041-1723

<https://doi.org/10.1038/s41467-021-27942-w>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Genomic epidemiology of SARS-CoV-2 in a UK university identifies dynamics of transmission

Dinesh Aggarwal ^{1,2,3,4,128}✉, Ben Warne ^{1,3,5,128}, Aminu S. Jahun ⁶, William L. Hamilton ^{1,3,4}, Thomas Fieldman^{1,3}, Louis du Plessis ⁷, Verity Hill ⁸, Beth Blane ¹, Emmeline Watkins⁹, Elizabeth Wright ⁹, Grant Hall ⁶, Catherine Ludden^{1,2}, Richard Myers², Myra Hosmillo^{3,6}, Yasmin Chaudhry ⁶, Malte L. Pinckert ⁶, Iliana Georgana⁶, Rhys Izuagbe ⁶, Danielle Leek¹, Olisaeloka Nsonwu ², Gareth J. Hughes ², Simon Packer ², Andrew J. Page ¹⁰, Marina Metaxaki ¹, Stewart Fuller¹, Gillian Weale ¹¹, Jon Holgate¹², Christopher A. Brown ^{13,14}, The Cambridge Covid-19 testing Centre*, University of Cambridge Asymptomatic COVID-19 Screening Programme Consortium*, The COVID-19 Genomics UK (COG-UK) Consortium*, Rob Howes¹³, Duncan McFarlane¹⁵, Gordon Dougan^{1,5}, Oliver G. Pybus ⁷, Daniela De Angelis^{2,16}, Patrick H. Maxwell^{1,3}, Sharon J. Peacock ^{1,3}, Michael P. Weekes ^{3,17,129}, Chris Illingworth^{16,18,19,129}, Ewan M. Harrison ^{1,2,4,20,129}✉, Nicholas J. Matheson ^{1,3,5,21,129}✉ & Ian G. Goodfellow ^{6,129}✉

Understanding SARS-CoV-2 transmission in higher education settings is important to limit spread between students, and into at-risk populations. In this study, we sequenced 482 SARS-CoV-2 isolates from the University of Cambridge from 5 October to 6 December 2020. We perform a detailed phylogenetic comparison with 972 isolates from the surrounding community, complemented with epidemiological and contact tracing data, to determine transmission dynamics. We observe limited viral introductions into the university; the majority of student cases were linked to a single genetic cluster, likely following social gatherings at a venue outside the university. We identify considerable onward transmission associated with student accommodation and courses; this was effectively contained using local infection control measures and following a national lockdown. Transmission clusters were largely segregated within the university or the community. Our study highlights key determinants of SARS-CoV-2 transmission and effective interventions in a higher education setting that will inform public health policy during pandemics.

The SARS-CoV-2 pandemic has caused substantial morbidity and mortality globally^{1,2}. Universities have been considered conduits for transmission due to extensive social networks of young adults, many of whom live communally, and in-person teaching of large groups³. Outbreaks of SARS-CoV-2 have been observed in a number of higher education institutions, but the drivers for transmission in these settings are poorly understood⁴. It is speculated that infection dynamics are dependent on transmission chains involving student courses, residence, study year and social networks⁵. Understanding these dynamics is essential in order to devise effective infection control measures while minimising disruption to teaching, research and the mental health of students and staff⁶. Furthermore, while university students are less likely to develop severe COVID-19 disease, there is concern that university outbreaks could seed infections in more vulnerable populations, including staff, the local community, and upon returning home to older relatives⁷. Identifying possible sources of cross-transmission is therefore vital.

Although SARS-CoV-2 genome sequencing has clear utility to identify virus emergence and cryptic transmission^{8,9}, no large-scale genomic studies in university settings have been conducted. The United Kingdom has an extensive community genomics surveillance programme through COG-UK¹⁰ which complements traditional contact tracing approaches by providing understanding of circulating viral populations.

We report the results of a genomic epidemiology study of SARS-CoV-2 across a complete term at the University of Cambridge (UoC). Importantly, these findings are from a study period prior to the established circulation of variants of concern and the availability of vaccination, with therefore fewer confounding factors. From 5 October to 6 December 2020, the UoC ran PCR-based symptomatic testing for all staff and students, and offered asymptomatic screening to 15,500 students living in university-managed accommodation. We therefore provide a unique study of SARS-CoV-2 infection that encompasses pre-symptomatic and asymptomatic students¹¹. Positive samples from the UoC were sequenced and compared with systematic surveillance SARS-CoV-2 sequences from the local community. The results were analysed in conjunction with epidemiological data derived from the screening programme and national contact tracing. Overall, we describe introductions of SARS-CoV-2 into a higher education setting, the dynamics of transmission both within the university and between the university and the surrounding community, and the impact of local and national measures to control the spread of SARS-CoV-2 infections.

Results

In total, 972 SARS-CoV-2 cases were identified among university students and staff over the course of term (5 October to 6 December 2020). High-quality genomes were generated from 446/778 (57.3%) positive cases from the university testing programme, from 107/266 (40.2%) cases identified through the Healthcare worker (HCW) screening programme (95 HCWs, 8 students, 4 university staff) and 104 patients identified by hospital testing (71 SARS-CoV-2 positive patients from Cambridge University Hospitals (CUH) and 33 from other medical facilities in Cambridgeshire). A further 797 local cases identified by community testing during the study period were present within the COG-UK dataset, of which 17 were identified as students, 7 as university staff and 26 as HCWs (Fig. 1). Of all identified SARS-CoV-2 cases from Cambridgeshire (university and community) during this period, 8.0% were sequenced (Supplementary Fig. 1).

SARS-CoV-2 lineages and transmission clusters. Over the 9-week term, 62 Pango lineages were identified across the university and community (Fig. 2a, c). In the university, 23 Pango lineages were identified, and 438/482 (90.9%) cases were from just 4 lineages (B.1.60.7, B.1.177, B.1.36, B.1.177.16), all of which were detected by the second week of term. Twelve lineages were only observed after the second week of term and accounted for 6.9% cases. By comparison, 57 lineages were identified in the local community over the same 9-week period. Viral genomes containing mutations in the spike protein that have been linked to decreased sensitivity to antibody-mediated immunity or impact viral transmission were observed in the university population: three sequences from the B.1.258 lineage containing the N439K mutation and Δ H69/ Δ V70; two cases of B.1.1.7/alpha variant and its associated mutations¹²; and 88 cases of B.1.177 with the A222V mutation¹³. Of these, Pango lineage B.1.1.7 is most reliably associated with increased transmission¹⁴; both cases of B.1.1.7 were amongst postgraduate students with no epidemiological links, during national lockdown, and failed to transmit further within the university.

In total, 198 putative transmission clusters were defined by CIVET (<https://github.com/artic-network/civet>). Only 8/36 clusters with university cases contained five or more university members (range 6–337), which together represented 91.3% of all university cases, signifying that the majority of introductions into UoC did not cause ongoing transmission. To further investigate the largest of these, cluster 1 described below, we identified groups of identical samples (0 SNP differences) which produced 19 additional clusters (a total of 34 clusters with >2 university cases) for further analysis.

Determinants of viral spread across the university. To determine transmission dynamics following introduction into the university, we performed a detailed investigation of the largest genomic cluster (Cluster 1), which accounted for 337/484 (69.6%) sequenced university cases (Fig. 3). This was widely dispersed across the university by the middle of term, affecting students from 29/31 colleges, 28 undergraduate courses and 208 households in university accommodation alone (Fig. 4).

Cluster 1 was classified as belonging to Pango lineage B.1.160.7. No mutations previously noted to be associated with increased transmissibility were observed in this lineage compared to other genomes in the study. Interrogation of the entire COG-UK dataset of samples from 2020 showed that this lineage was first identified in the UK on 4 October 2020, in Wales, before becoming predominantly sampled in the UoC (Fig. 3b). The B.1.160.7 lineage was not identified in the local community until term week 3 (19–25 October 2020). This was supported by the median estimate of the time to the most common recent ancestor of cluster 1, in comparison to its most closely related cluster from Cambridgeshire community isolates of 165 days (C.I. 127–207) prior to the start of term (6 October 2020). Together, these results suggest the university cases were introduced from outside Cambridgeshire. Additional analysis with A2B-COVID¹⁵, which uses genomic data alongside timing of infection data to evaluate plausibility of transmission between individuals, we showed that these sequences were consistent with a single introduction into the university (Fig. 3c).

National and university contact tracing data were used to identify the initial source of dispersion of this cluster. Ten students from the first two weeks of term reported visiting the same nightclub (venue A). Nine individuals either had an isolate from cluster 1 or (in the event that their sample did not yield a high-quality sequence) were household contacts of an individual

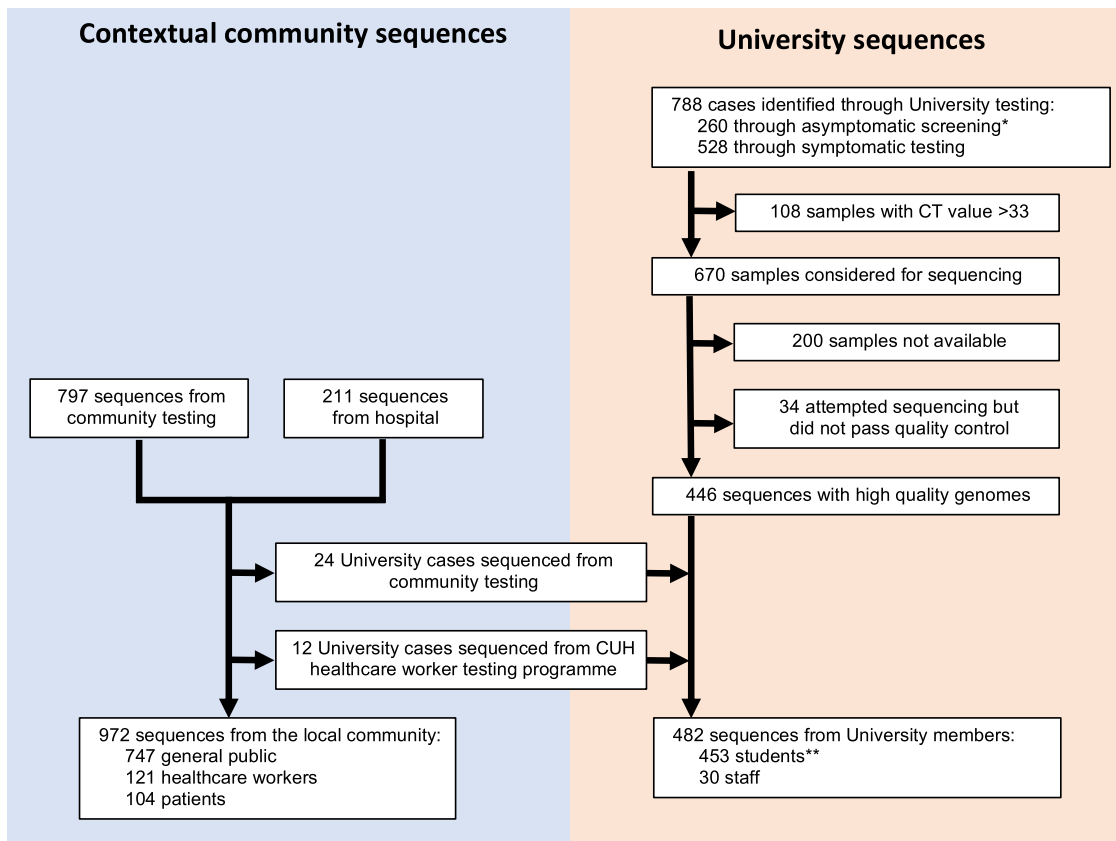


Fig. 1 Study cohort and available genome sequences. *Includes 14 students identified through ad hoc asymptomatic screening conducted as part of an outbreak investigation by the University of Cambridge in conjunction with local public health authorities, responding to increased rates of infection in a block of student accommodation (described in further detail in cluster 2 below). **Includes two students associated with a single sequenced pooled sample (see supplementary methods). CUH Cambridge University Hospitals.

with a sequenced cluster 1 isolate. No information was available for one student (Supplementary Fig. 5).

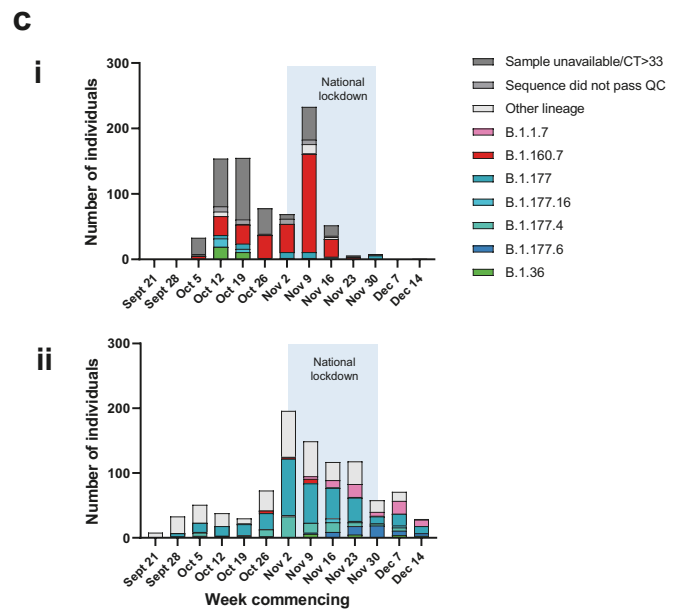
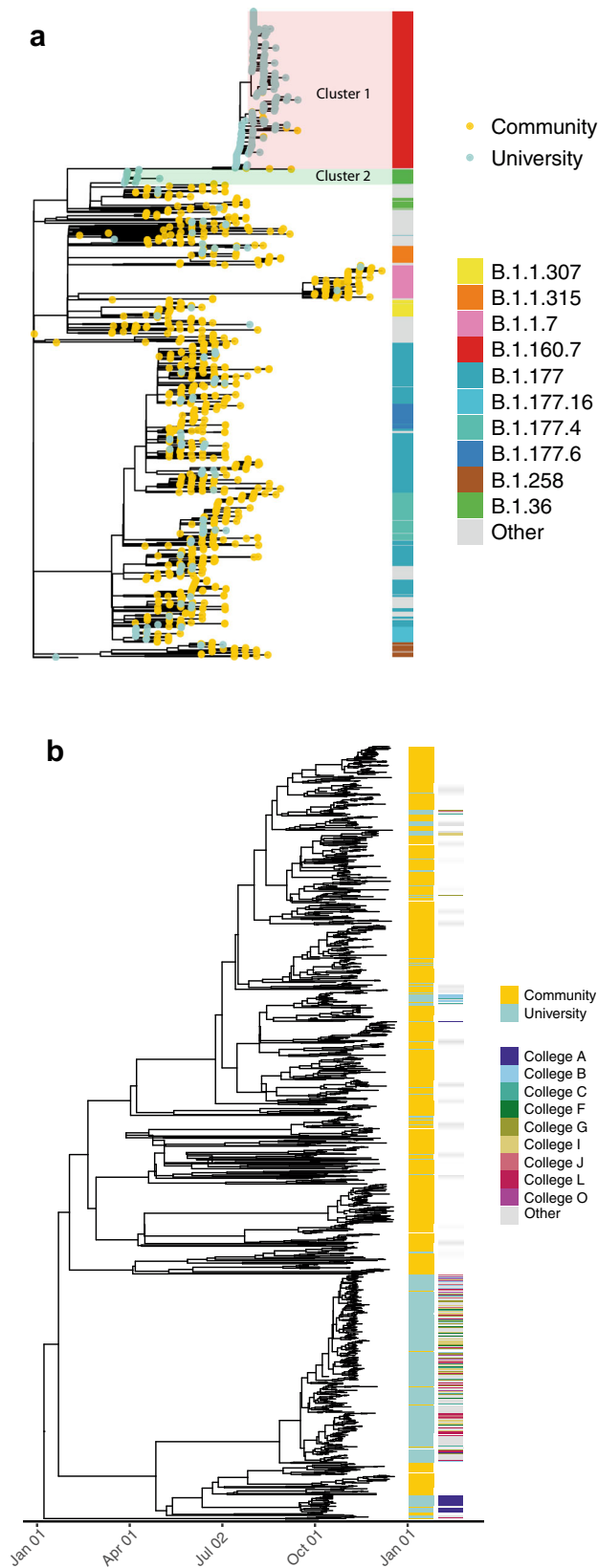
Transmission of cluster 1 was sustained from the first week of term until a national lockdown was enforced on 5th November. Students testing positive in the two weeks around lockdown reported common exposure events predominantly linked to nightclub venues (25/59 (42.4%) of exposures external to the university reported by 48 students). Venue A, identified above as the possible source of dispersion of this cluster at the start of term, was also the most common venue identified in the two weeks around lockdown ($n = 16$). 9/16 cases had sequences in cluster 1, and a further five individuals (where no sequence was available) were household contacts of sequenced cases in cluster 1 (Supplementary Fig. 6).

To determine the impact of lockdown and other control measures within the university, a birth-death skyline model¹⁶ was used to measure changes in the effective reproduction number (R_e) within cluster 1. The model indicated an initial R_e at the start of term that was slightly larger than 1, albeit with wide uncertainty (median 1.14; 95% HPD: 0.27–2.21 on 5 October). Over the next 2 weeks R_e continued to rise (median 1.52; 95% HPD 0.94–2.22 on 15 October) followed by a subsequent gradual decline over the next 2 weeks (Fig. 5a). There was a rise immediately prior to the start of lockdown (median 1.55; 95% HPD 1.25–1.86 on 5 November), followed by a steep decrease thereafter (median 0.23; 95% HPD 0.07–0.41 on 19 November) (Fig. 5a), consistent with declining absolute numbers of SARS-CoV-2 infections seen during this time (Fig. 2c). The model estimated the median effective infectious period for individuals in the cluster at 3.03 days (95% HPD:

2.44–3.59 days) (Fig. 5b). As the model does not explicitly incorporate an incubation period and assumes that individuals cannot transmit after being sampled, the effective infectious period represents the mean time from infection until testing positive and assumes perfect infection control measures thereafter. Estimates of R_e and the effective infectious period are robust to model parameterisations (Supplementary Figs. 8–10). Sampling proportion estimates largely overlap with empirical estimates based on the number of positive cases that were sequenced during each week (Fig. 5c). Although sampling proportion estimates are sensitive to the prior specifications, R_e estimates are unaffected (Supplementary Fig. 11).

Transmission within university households. There was evidence of transmission of SARS-CoV-2 in student accommodation in 18/34 university clusters. In cluster 1, 169/337 (50.1%) students had a virus genome sequence identical to at least one other student living in the same or neighbouring household (sub-clusters within 0 SNPs ranging between 2 and 11 students).

The largest cluster associated with transmission in accommodation was cluster 2 (lineage B.1.36). By term week 3, this cluster involved 30 students, of which 24 (80%) lived in the same accommodation block in College A and 4 students lived in two separate households in the same college (Supplementary Fig. 12). Interventions from the university, supported by local public health authorities, included isolation of all households in the main accommodation block and individual screening offered to all students. Half of all cases in this cluster were diagnosed by asymptomatic screening. No further genomically-related isolates



were identified after term-week 3, indicating a successful intervention, and cessation of transmission.

To quantify the importance of household transmission, a Reed-Frost Chain Binomial Model was employed to estimate the household attack rate. Using A2B-COVID¹⁵, we identified 265 households in which the data were consistent with only 1

introduction of SARS-CoV-2. The per household contact probability that an infected person passed on the virus to an uninfected individual within the same household was estimated at 7.8% (95% C.I. 6.9–8.7%).

Further genomic clusters where transmission between household members was implicated are outlined in Supplementary

Fig. 2 Genomic diversity of SARS-CoV-2 in the university and community. **a** Maximum likelihood tree showing that the majority of lineages from university isolates were distinct from community isolates. The node leaves (branch tips) show case location and global PANGO lineage is illustrated in the vertical bar. **b** Time-scaled coalescent tree including university members and local community isolates from study period with visible segregation between the two groups. College affiliation is shown for university members in the second set of vertical columns, highlighting the 'top nine' colleges by cluster 1 prevalence. **c** Epidemic curves demonstrating a steeper decline in SARS-CoV-2 cases in the University of Cambridge (i) compared to the local community (ii), with associated lineages. Only cases with available genomes are included. University term ran from the week commencing October 5 to the week commencing November 30. The light blue shaded area reflects a 4-week national lockdown in the UK, which was associated with a large fall in COVID-19 cases in University students. Specific lineages highlighted are the four largest lineages within the University (minimum 20 cases over the study period) and the community (minimum 50 cases over the study period). For (i), weekly individual case ascertainment for staff and students testing positive for SARS-CoV-2 through both symptomatic and asymptomatic testing pathways provided at the University of Cambridge is indicated. For (ii), weekly cases with genomes available from the local community are shown. Source data are provided as a Source Data file.

Table 1. They follow similar patterns, with groups of cases confined to a single college not leading to sustained transmission.

Other transmission routes among university members. In addition to household transmission, there was evidence of viral spread between students in the same course and year of study in 14/34 genomic clusters, with the highest proportion being students in their first year of study. In cluster 1, 203/337 (60.2%) students had an identical isolate to at least one other student studying the same course in the same year (cluster size range 2–14 students). Statistical modelling using data from cluster 1 across the term showed a bias towards infections being observed in first year students (p -value = 0.002) (Supplementary Fig. 13, model details in Supplementary Methods). Two further small clusters were comprised of postgraduate students working in the same university department. However, we were not able to determine the probable location of transmission in most cases: there is considerable overlap between course and household clusters, and complex social and study networks exist between students (illustrated in Supplementary Table 1, for example in clusters 3, 4 and 10). Of note, 23/34 clusters with 2 or more genomically linked cases in the dataset contained at least one university member that could not be epidemiologically linked with any other case in their cluster.

The number of SARS-CoV-2 sequences from university staff members were limited in comparison to students ($n = 30$). There was evidence of transmission between staff members working in the same department, college or ancillary role in four genomic clusters. Two clusters contained staff members who shared the same household. There are 8 clusters involving both university staff and students. However, epidemiological associations between these two groups could only be identified in one cluster: a shared household between a student and staff member working in separate university departments.

Transmission between the university and local community. We next sought to address the degree of transmission between the university and the local community. Two distinct phylogenetic approaches, shown in Fig. 2, demonstrate segregation of the majority of community and university cases into separate clusters and therefore a lack of substantial cross-transmission. 29/198 (14.6%) transmission clusters contained both university and community cases. Only six clusters contained five or more university cases and included three or more community cases.

To identify transmission clusters involving university and hospital (patient and healthcare worker) cases, we ran CIVET (<https://github.com/artic-network/civet>) separately with these cases for a focused phylogenetic analysis of this setting. Associations were identified between the university and hospital settings, with 17 clusters involving both university members and either patients or staff. Cluster 1 (69.6% of student cases), contained only 1 patient and 1 healthcare worker with no

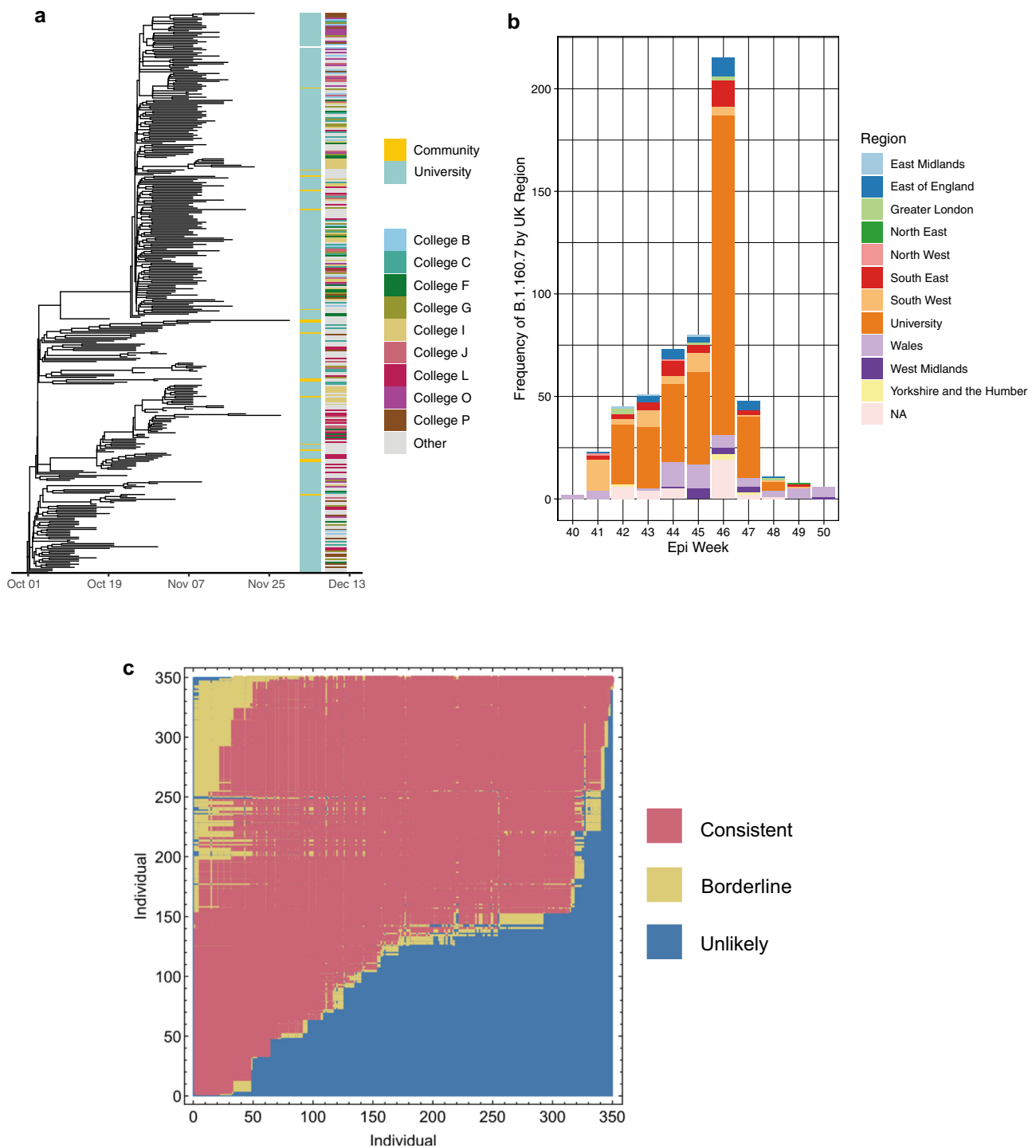
identifiable epidemiological link to students. The remaining 16 clusters comprised 133 individuals, including 26 patients, 55 hospital staff or their family members and 52 university members (including 18 staff and 15 clinical medical students). The second-largest cluster of university members ($n = 21$ university and hospital cases) included nine medical students, five healthcare workers and two patients. Phylogenetically, the medical students and one of the healthcare workers were closely linked (Supplementary Fig. 14) and analysis of these cases with A2B-COVID¹⁵ confirmed the plausibility of transmission. All 9 medical students were on clinical rotations at the time of diagnosis of the index case; 7/9 lived in neighbouring households in the same college and the remaining two were named contacts of the index student. Plausible transmission events between this group and the other cluster members were refuted using A2B-COVID (Supplementary Fig. 14).

To further investigate epidemiological associations in clusters involving university members and the local community, 1243/1455 of the cases sequenced over the sampling period were linked to national contact tracing data (excluding hospital cases). 219 (17.6%) cases reported 127 common exposure events. Cluster 1, representing 69.6% of cases within the university, included only 17/976 (1.7%) community cases; only one community case had a common exposure with a university student, dining at the same restaurant. No other epidemiological links were identified in all other genomic clusters. Transmission suspected in 19 epidemiologically linked clusters defined by common exposures was refuted by phylogenetic variation (i.e. identified in separate transmission clusters as defined by CIVET).

Discussion

We report the first comprehensive and integrated epidemiological and genomic analysis of SARS-CoV-2 transmission in a higher education setting. Following a limited number of introductions, the majority of cases were linked to a single genetic cluster, that was likely to have dispersed across the university following multiple social gatherings at a nightclub. There was considerable transmission associated with student accommodation and student courses, but minimal evidence of transmission within departments, or between students and staff. We observe the great majority of transmissions occur either within the university or within the local community. Finally, we present evidence demonstrating the efficacy of university measures and national lockdown in reducing COVID-19 cases.

Nearly 70% of all university cases belonged to one genetic cluster (cluster 1), introduced into the UoC by the arrival of students and likely forming a single transmission chain. A nightclub was implicated as an important transmission event at the start of term and again prior to lockdown. This corroborates previous studies identifying such venues as a risk factor for substantial SARS-CoV-2 transmission^{17,18}. We urge a cautious approach to the access of such venues during a SARS-CoV-2



pandemic, particularly in the context of a young susceptible student population.

Our data suggest a substantial change in case numbers and the effective reproduction number over the course of the term. This likely reflects a combination of changes in student behaviour and effective interventions to reduce transmission. Overall, we note that incidence and the effective reproductive number within the university are lower than in other higher education settings and the general UK young adult population during the study period¹⁹. We highlight a limited number of introductions and low lineage diversity in the university compared to the surrounding community. While the natural extinction of lineages is relatively

common²⁰, multiple genetically diverse clusters may be expected given the congregation of students from across the globe (international students make up 35% of students in college accommodation)¹¹. The lack of diversity may reflect the impact of robust and widely implemented university infection control measures maintained throughout the term, full details of which are provided in the Supplementary Materials, but include social distancing, mask wearing and quarantine of international students at the beginning of term.

There was an initial rise in cases over the first two weeks, coinciding with the first week of term and university Freshers week. This is known to be a period of more intense social mixing

Fig. 3 Emergence and transmission of SARS-CoV-2 in a large university cluster. **a** Time-scaled phylogenetic tree of largest university cluster (cluster 1) derived from the BDSKY model implemented in BEAST 2.6 (Fig. 5). The left-sided heatmap is coloured by case location, and the right-sided heatmap is coloured by student college affiliation, highlighting the top nine colleges by cluster 1 prevalence. Cluster 1 was widely dispersed across the university with limited transmission into the community. **b** Frequency of Lineage B.1.160.7 (to which cluster 1 belongs) in each region of the UK and the University of Cambridge. Regions are defined as 'Nomenclature of territorial units for statistics' (NUTS) regions, where the UK has 9 regions. It is visible that the lineage B.1.160.7 was first sequenced in Wales, and then in the neighbouring South West of England, before becoming prevalent within the University of Cambridge. The lineage remained infrequently detected in the community populating the wider surrounding region (Cambridgeshire, East Anglia, Bedfordshire and Hertfordshire, and Essex, making up East of England) throughout the university term. **c** A continuous transmission chain of SARS-CoV-2 infections in cluster 1 commenced with a single introduction. Relationships between individuals in cluster 1 were calculated within A2B-COVID. Colours denote potential transmission events from the donor (vertical axis) to the recipient (horizontal axis) that are consistent with transmission¹² or which are borderline possibilities (yellow). The plot shows that the data are consistent with a continuous transmission chain of SARS-CoV-2 infections in cluster 1 occurring via a single introduction; there are multiple potential networks of transmission events between these individuals for which each event would be consistent with a statistical model of direct transmission. We note that individuals in this plot are ordered by the date of the first positive COVID test. Source data are provided as a Source Data file.

between students in venues both inside and outside university premises. Between term weeks three and five there was a fall in the effective reproductive number, which coincides with both a reduction in social mixing and the identification of, and subsequent university measures to control, transmission events identified in college residences. In multiple clusters, transmission in student households was successfully interrupted through a combination of measures provided by the university, including rapid case identification through asymptomatic screening, readily available symptomatic testing, contact tracing and comprehensive support provided by colleges for cases and their contacts while in isolation. Further details, including the elaboration of the specific measures to control cluster 2, an outbreak associated with a large accommodation block described above, are provided in the Supplementary Materials. Although we have demonstrated that transmission between students in the same accommodation block is an important factor in the spread of SARS-CoV-2, we report a lower secondary household attack rate (7.8%) than that identified in domestic households (16.6–21.1%) and a lower than expected effective infectious period (3.0 days)²¹.

University measures may have been less successful in controlling transmission in settings outside colleges. There was a rise in the effective reproduction number coinciding with the announcement of a national lockdown on 31 October, to begin on 5 November 2020. This announcement prior to implementation of a major socially restrictive public health measure, alongside existing Halloween festivities, may have led to increased levels of behaviour associated with a higher risk of transmission. This supports either reducing the time from announcement to implementation of socially restrictive measures, or the need for a targeted public health campaign to limit high-risk activities where this is not possible. In addition, having identified considerable transmission between students on the same course, we suggest that further mitigation of viral spread may be obtained by implementing shared student accommodation based on university courses.

The national lockdown dramatically reduced case numbers within the university, at a faster rate than the local community, demonstrating high levels of compliance from our study population with an effective control strategy. Contemporary studies conducted elsewhere in the UK have demonstrated that adherence to COVID-19 prevention measures, such as national lockdown, are mixed²². Although young age is a risk factor for poor adherence, other associations are less common within the university population, such as having a dependent child in the household, financial hardship and working in a key sector. Although no direct incentives were provided to students, the expectation of individuals to adhere to rules was communicated

widely in both national and university media. We also believe that the key to the successful implementation of lockdown was the additional support provided by the collegiate university, ranging from the practical provision of food and drink through to the pastoral and community support provided by established networks of staff, tutors and student representatives.

Finally, we observed limited transmission between the university and the local community. The largest university cluster, accounting for the majority of student infections, was largely phylogenetically distinct from community cases. Further, epidemiological evidence describing common exposures for community and university cases was sparse. However, clinical medical students were disproportionately represented within community clusters. This is an important epidemiological link between secondary care and the university; we highlight this group as being at-risk for both acquisition and transmission of SARS-CoV-2 and medical students should therefore be prioritised for interventions such as vaccination.

A combination of contact tracing and genomics was instrumental to understanding transmission within the university and with its surrounding population; notably in refuting transmission within epidemiologically linked clusters. We advocate for a combined genomic epidemiological approach to inform outbreak investigations as used in other settings^{8,23}.

This study has a number of limitations. Incomplete sampling and subsequent sequence filtering in both the university and community should be considered when interpreting transmission; the asymptomatic and active case ascertainment in this study should mitigate this discrepancy. The lower community case ascertainment may result in unobserved transmission chains (such as those when assessing the introduction of Pango lineage B.1.160.7 into the university). Further, epidemiological links are dependent on self-reporting and therefore some data will be missing; whilst a lack of epidemiological association between groups in clusters is important and reassuring (such as between staff and students), it does not confirm a lack of transmission. We highlight shared student courses as a risk factor for transmission; this does not take into account the setting of transmission, i.e., during educational or social activities. Finally, the UoC is distinct in its collegiate structure with limited integration with the community; any generalisation of conclusions should be tempered by the study setting.

We present the first comprehensive integrated epidemiological and genomic evaluation of transmission of SARS-CoV-2 within a university. The insights gained will inform public policy regarding infection control measures in higher education settings. We find containment of transmission in student accommodation necessary to mitigate onward propagation. We highlight the

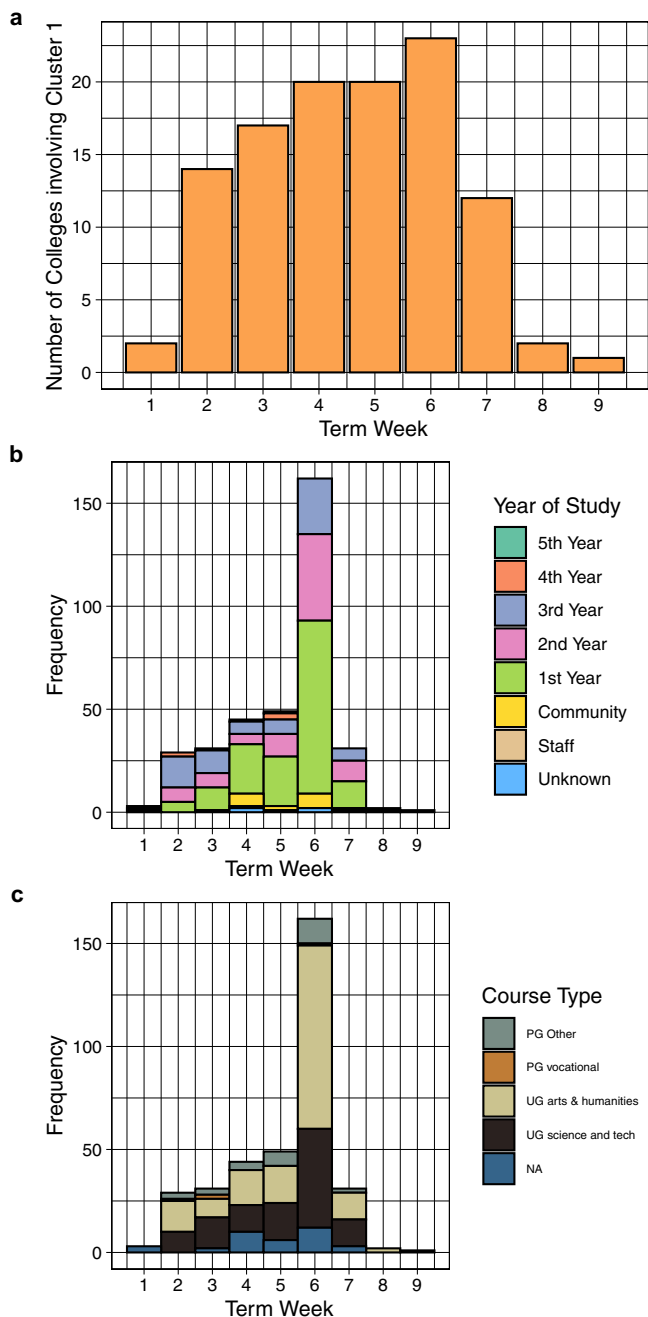


Fig. 4 Demographics of Cluster 1 across the first university term.

a Cumulative number of colleges involved in the cluster. Cases included in this cluster were between a number of colleges early during the university term. **b** Frequency of cases involved in the cluster by year of study. **c** Frequency of cases involved in the cluster by course type. Source data are provided as a Source Data file.

importance of targeted public health measures towards nightclub venues to limit transmission. Critically, these findings are likely to be informative for future pandemic preparedness.

Methods

Ethics. The COG-UK study protocol was approved by the Public Health England Research Ethics Governance Group (reference: R&D NR0195). Public Health England affiliated authors had access to identifiable Cambridgeshire community case data. This data was processed under Regulation 3 of The Health Service (Control of Patient Information) Regulations 2002- permitting the processing of confidential patient information for communicable disease and other risks to public health and as such, individual patient consent is not required. Other authors only had access to anonymised or summarised data. Ethical approval for the UoC asymptomatic COVID-19

screening programme was granted by the UoC Human Biology Research Ethics Committee (HBREC.2020.35) with informed consent gained from participants.

Study setting. The UoC has ~23,000 students and 12,600 staff. The university is divided into 31 colleges and 150 departments, faculties and other institutions. Students belong to a college community, as well as being members of the university and an academic faculty/department. Colleges provide residential accommodation for approximately two thirds of students, either on campuses or in off-site housing, and offer social and sports activities, pastoral and academic support for each individual²⁴. All colleges have membership from students across multiple courses. The university is based in the City of Cambridge (which has an estimated population of 123,900²⁵), in the county of Cambridgeshire (estimated population 855,796 people in 2019²⁶) in the East of England.

Participants and samples. Samples were derived from university symptomatic testing and asymptomatic COVID-19 screening programmes between 5 October 2020 and 6 December 2020, covering the full term. Testing for all symptomatic students and staff was available on weekdays. The asymptomatic screening programme has been described in detail elsewhere¹¹. In brief, screening was offered on a voluntary basis to all students residing in accommodation owned or managed by a college or the Cambridge Theological Federation. In total, 15,561 students were eligible to participate. To optimise testing efficiency, multiple swabs were pooled into the same tube of viral transport medium at the time of sample collection. Testing pools varied in size from 1 to 10 students, with each devised to include one or more student households as far as possible¹¹. In this study, households are defined as individuals who share a kitchen, bathroom and/or lounge facilities. The members of any pool testing positive were re-tested using individual confirmatory PCR tests to confirm the result and identify the positive subject(s) (see Supplementary Methods for further details including infection prevention control measures). Only samples from individuals that were confirmed positive upon the re-testing were used for sequencing.

SARS-CoV-2 strains circulating in the local community were identified from the COG-UK dataset for Cambridgeshire. These data were derived from local community samples from non-hospitalised, symptomatic individuals, who requested a free diagnostic test via national community testing. Other samples were derived from patients treated at three Cambridgeshire hospital trusts: Cambridge University Hospitals NHS Foundation Trust (a teaching hospital providing secondary care services for Cambridge and the surrounding area as well as tertiary referral services for the East of England and surge capacity for COVID-19); Royal Papworth Hospital NHS Foundation Trust (specialist heart and lung hospital, also providing surge capacity for COVID-19); Cambridgeshire and Peterborough NHS Foundation Trust (provider of community, mental health and learning disability services in Cambridgeshire). Hospital samples were obtained from both asymptomatic screening and those exhibiting COVID-19 symptoms. Finally, samples were derived from the asymptomatic HCW programme at Cambridge University Hospitals²⁷.

Sequencing. Positive samples from UoC testing with a PCR cycle threshold value ≤ 33 were selected and sequenced using the GridION platform (Oxford Nanopore). All Cambridgeshire samples sequenced between 24th September and 21st December 2020 were included to overlap with the university term. Samples from the local Cambridgeshire community and hospital cases (described above) were collected as part of national SARS-CoV-2 testing, and sequenced at one of seventeen COG-UK sequencing sites (further details in Supplementary Methods). The samples were prepared using either the ARTIC²⁸ or veSeq²⁹ protocols, and were sequenced using Illumina or Oxford Nanopore platforms. Genomic data were filtered to exclude sequences with $>5\%$ Ns and those of spuriously low file sizes (<29 KB). Genomes were aligned with minimap²³⁰ to the Wuhan Hu-1 reference genome (MN908947.3), collected December 2019. All samples were processed through COVID-CLIMB pipelines^{31,32}. Protocols are available at <https://github.com/COG-UK>.

Phylogenetic analysis. Maximum likelihood phylogenetic trees were estimated using IQ-TREE (version 2.1.2 COVID-edition)³³ and rooted using Wuhan Hu-1 (MN908947.3) as an outgroup. Trees were constructed using the GTR + Γ substitution model³⁴, as determined by ModelFinder³⁵. Branch support statistics were generated using the ultrafast bootstrap method³⁶. TempEst³⁷ was used to explore the temporal signal in the data. Trees were visualised, explored, and labelled with associated metadata using Microreact³⁸ to identify epidemiological links supported by the genomic data. Specified mutations were identified using type_variants (https://github.com/cov-ert/type_variants). Possible transmission clusters were defined by extracting phylogenetic neighbourhoods identified using the CIVET tool (version 2.1.0) on 11 January 2021 (<https://github.com/artic-network/civet>). In selected clusters, further evaluation was conducted using A2B-COVID¹⁵. A2B-COVID evaluates data from individuals in a pairwise manner. Using viral genome sequences from two individuals, alongside data describing the timing of infection, it evaluates whether or not these data are consistent with a hypothesis that SARS-CoV-2 was transmitted directly from one individual to the other; data from each pair is described as being either consistent, borderline, or unlikely to have been observed given this hypothesis. Where indicated, collapsed nodes from trees generated from CIVET were inspected to visualise data in the context of the COG-UK national database (<https://www.cogconsortium.uk/>). For further evaluation of

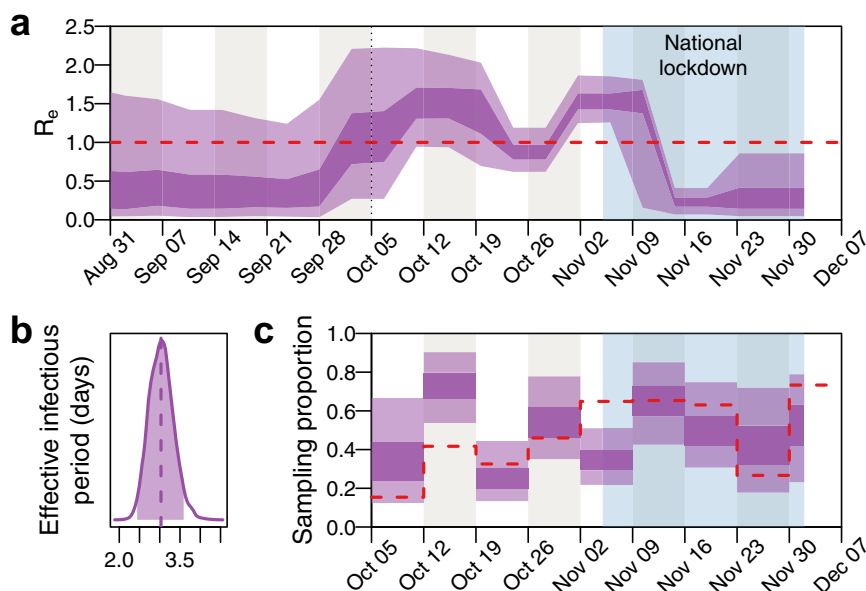


Fig. 5 Effective reproduction number and infectious period of SARS-CoV-2 from a dominant university cluster. A 20-epoch birth-death skyline model shows the effect of local infection control measures and the national lockdown on the effective reproduction number (R_e), and estimates of the mean effective infectious period as 3.03 (95% HPD = 2.44–3.59) days. **a** R_e posterior estimates (dark shading = 50% HPD; light shading = 95% HPD). The dotted line indicates the start of term and the light blue shaded area the 4-week national lockdown in the UK, which was associated with a large fall in COVID-19 cases in University students. The red dashed line indicates $R_e = 1$. **b** Effective infectious period posterior estimates (shaded region = 95% HPD; dashed line = median). **c** Weekly sampling proportion posterior estimates (dark shading = 50% HPD; light shading = 95% HPD). The red dashed line indicates the empirical sampling proportion estimates for each week in term (number of sequenced genomes from all University clusters divided by the number of positive tests among University staff and students). Source data are provided as a Source Data file.

transmission in the largest cluster identified by CIVET, pairwise SNP differences between sequences were determined using SNP-dist (<https://github.com/tseemann/snp-dists/releases/tag/v0.7.0>).

Lineages. Global Pango Lineages³⁹ were assigned to each genome using Pangolin (<https://github.com/cov-lineages/pangolin/releases/tag/v2.1.6>) with analyses performed on COVID-CLIMB³² (further details in Supplementary Methods).

Molecular clock and phylodynamic analyses. BEAST v1.10.4⁴⁰ was used to perform a time-scaled phylogenetic analysis using an exponential growth coalescent tree prior and a GTR + Γ substitution model including all university and community high-quality genomes from the study period. As there was a lack of clear temporal signal in our dataset due to the relatively short time period analysed, the substitution rate was fixed to 8×10^{-4} substitutions per site per year (s/s/y) under a strict clock model in line with previous SARS-CoV-2 analyses^{13,41–44}. Two chains of 100 million iterations were run independently to ensure convergence to the correct posterior distribution. Convergence was assessed using Tracer⁴⁵, and 10% of states were removed to account for burn-in. Finally, a maximum clade credibility (MCC) tree was generated using TreeAnnotator.

To estimate the effective reproduction number (R_e) and infectious period of SARS-CoV-2 over the term, a dominant clade (representing 69.6% of all university genomes) was selected and all community genome sequences that cluster with it incorporated, resulting in a total of 354 genomes. A Bayesian birth-death skyline (BDSKY) model¹⁶ was employed using BEAST v2.6⁴⁶. A GTR + Γ substitution model was used along with a strict clock model, placing a lognormal prior with mean 8×10^{-4} s/s/y (in real space) and standard deviation 0.1 on the clock rate. A lognormal prior with mean 0 and standard deviation 1 was placed on R_e and a Beta prior with $a = 5$ and $\beta = 5$ was placed on the sampling proportion. R_e was parameterised into 20 epochs, equidistantly spaced between the origin time and the most recent sequence collection date. The sampling proportion was fixed to 0 before the first week of term and estimated for each week thereafter. The rate at which infected patients become non-infectious was assumed to be constant and a lognormal prior with mean 48.7 years^{-1} (in real space) and standard deviation 0.25 was placed on it, resulting in a prior mean effective infectious period between ~5 and ~15 days. To test the robustness of the posterior estimates different parameterisations were used for R_e and the sampling proportion, and the sampling proportion prior was varied. Further details are provided in the supplementary methods. To test the robustness of posterior estimates to the clock rate prior all analyses were repeated using a lognormal prior with mean 1×10^{-3} s/s/y (in real space) and standard deviation 0.1 on the clock rate. Finally, to test the assumption of a strict clock model, analyses were repeated using an uncorrelated lognormally distributed relaxed clock model⁴⁷. In these analyses the 95% HPD interval of the

coefficient of variation of the clock rate did not exclude 0, indicating poor support for a relaxed clock model in this dataset. Furthermore, estimates of the BDSKY model parameters did not differ significantly from estimates under a strict clock model. Therefore, we only show results under a strict clock model. For all models three chains of 200 million iterations were run independently. Convergence was assessed using the R-package coda⁴⁸, and 10% of states were removed to account for burn-in. MCC trees were generated using TreeAnnotator.

Household attack rates. A2B-COVID¹⁵ was used to exclude households for which the sequence and epidemiological data were inconsistent with a single viral introduction to the household. A chain binomial model was then used to estimate the probability that an infected person transmitted the virus to an uninfected person within the same household (further details in supplementary methods).

Epidemiological data. University student demographic data were derived from the UoC student electronic record system CamSIS, and household structure and membership data from the UoC asymptomatic screening programme. To identify university affiliated cases (students and staff) and hospital staff accessing the national SARS-CoV-2 testing service, Second Generation Surveillance System (SGSS) and contact-tracing data provided by NHS Test and Trace (T&T) data were interrogated. Epidemiologically linked common exposures for students, university staff and the local community were identified through T&T data. Common exposures were defined by T&T as locations or events that two or more people testing positive for COVID-19 visited in the same two to seven day period before symptom onset or positive test. Additional contact tracing information was also provided by the UoC COVID helpdesk. These data were compared with observed phylogenetic clusters to determine potential sources of transmission and determine the extent of transmission between the university and community.

Epidemiological data from UoC were initially compiled in Microsoft Azure SQL and Excel 2013 (Microsoft) and analysed in STATA 14.2 (College Station, TX, USA). Further data manipulation, statistical analysis and figure generation was undertaken with RStudio (version 1.3.1093) using R (version 4.0.2). Network diagrams were produced with R package iGraph (v1.2.6).

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

The Assembled/consensus genomes generated in this study have been deposited in the GISAID⁴⁹ database and raw reads are available from European Nucleotide Archive

(ENA)⁵⁰ under accession PRJEB37886. Pooled sample sequence raw reads and assembled sequences are deposited in the NCBI Sequence Read Archive Database (SRA; <https://www.ncbi.nlm.nih.gov/sra>) under the BioProject accession number PRJNA779279.

ENA and Genbank accession codes for individual sequences used in this study are available in supplementary materials (Supplementary Data 1 and 2). All genomes, phylogenetic trees and basic metadata are available from the COG-UK consortium website (<https://www.cogconsortium.uk/data>). Limited public metadata, analysis files, and processed genomic data for this work are available from GitHub at <https://github.com/COG-UK/camb-uni-phylo/> (<https://doi.org/10.5281/zenodo.5643354>⁵¹), which also contains a list of ENA and Genbank study sequence accession numbers for this study. For confidentiality reasons, extended metadata⁵² are under restricted access for confidentiality reasons and in line with study ethics; requests for access should be directed to corresponding authors and specifically for Public Health England data, to the Public Health England office of data release (<https://www.gov.uk/government/publications/accessing-public-health-england-data/about-the-phe-odr-and-accessing-data>) with an estimated 60 working days turnaround time. Processed metadata generated for figures in this study are provided in the Source Data file. Source data are provided with this paper.

Code availability

Custom code used in this analysis is available at <https://github.com/COG-UK/camb-uni-phylo/>. Please direct further queries to the corresponding authors.

Received: 19 May 2021; Accepted: 17 December 2021;

Published online: 08 February 2022

References

- Zhou, F. et al. Clinical course and risk factors for mortality of adult inpatients with COVID-19 in Wuhan, China: a retrospective cohort study. *Lancet* **395**, 1054–1062 (2020).
- Yang, J. et al. Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: a systematic review and meta-analysis. *Int J. Infect. Dis.* **94**, 91–95 (2020).
- Amoolya Vusirikala, H. W. et al. Gayatri Amirthalingam. Seroprevalence of SARS-CoV-2 Antibodies in University Students: cross-sectional study, December 2020, England. (2021).
- Yamey, G. & Walensky, R. P. Covid-19: re-opening universities is high risk. *BMJ* **370**, m3365 (2020).
- Group, C.s.T.F. Risks associated with the reopening of education settings in September. 2021 (2020).
- Sahu, P. Closure of universities due to coronavirus disease 2019 (COVID-19): impact on education and mental health of students and academic staff. *Cureus* **12**, e7541 (2020).
- Education, T.a.F.G.o.H.E.F. Principles for managing SARS-CoV-2 transmission associated with higher education - 3 September 2020. Vol. 2021 (2020).
- Meredith, L. W. et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. *Lancet Infect. Dis.* **20**, 1263–1271 (2020).
- Moreno, G. K. et al. Severe acute respiratory syndrome coronavirus 2 transmission in intercollegiate athletics not fully mitigated with daily antigen testing. *Clin Infect Dis.* **73**, S45–S53 (2021).
- consortiumcontact@cogconsortium.uk, C.-G.U. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**, e99–e100 (2020).
- Ben Warne, J. E. et al. Feasibility and efficacy of mass testing for SARS-CoV-2 in a UK university using swab pooling and PCR. Preprint at <https://www.researchsquare.com/article/rs-520626/v1> (2021).
- Andrew Rambaut, N. L. et al., On behalf of COVID-19 Genomics Consortium UK (CoG-UK). Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Vol. 2021 (2020).
- Hodcroft, E. B. et al. Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. Preprint at <https://www.medrxiv.org/content/10.1101/2020.10.25.20219063v3> (2020).
- Volz, E. et al. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* **593**, 266–269 (2021).
- Illingworth, C. J. R. et al. A2B-COVID: a method for evaluating potential SARS-CoV-2 transmission events. Preprint at <https://www.medrxiv.org/content/10.1101/2020.10.26.20219642v2> (2020).
- Stadler, T., Kuhnert, D., Bonhoeffer, S. & Drummond, A. J. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc. Natl Acad. Sci. USA* **110**, 228–233 (2013).
- Muller, N. et al. Severe acute respiratory syndrome coronavirus 2 outbreak related to a Nightclub, Germany, 2020. *Emerg. Infect. Dis.* **27**, 645–648 (2020).
- Choi, H., Cho, W., Kim, M. H. & Hur, J. Y. Public health emergency and crisis management: case study of SARS-CoV-2 outbreak. *Int. J. Environ. Res. Public Health* **17**, 3984 (2020).
- Group, T.C.s.T.a.f. Children’s Task and Finish Group: Paper on higher education settings. Vol. 2021 (2021).
- du Plessis, L. et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).
- Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). *Science* **368**, 489–493 (2020).
- Smith, L. E. et al. Adherence to the test, trace, and isolate system in the UK: results from 37 nationally representative surveys. *BMJ* **372**, n608 (2021).
- Seemann, T. et al. Tracking the COVID-19 pandemic in Australia using genomics. *Nat. Commun.* **11**, 4376 (2020).
- Cambridge, U.o. How the University and Colleges work. <https://www.cam.ac.uk/about-the-university/how-the-university-and-colleges-work> (2013).
- Statistics, O.f.N. 2011 ONS Census. <https://www.cambridge.gov.uk/media/1170/census-2011-cambridge-data.pdf> (2011).
- Statistics, O.f.N. Cambridgeshire and Peterborough Population Overview Report. <https://cambridgeshireinsight.org.uk/population/> (2019).
- Rivett, L. et al. Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission. *Elife* **9**, e58728 (2020).
- Quick, J. nCoV-2019 sequencing protocol v3 (LoCost) V.3. Vol. 2021 (2020).
- Bonsall, D. et al. A comprehensive genomics solution for HIV surveillance and clinical monitoring in low-income settings. *J. Clin. Microbiol.* **58**, e00382 (2020).
- Li, H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* **34**, 3094–3100 (2018).
- Nicholls, S. M. et al. CLIMB-COVID: continuous integration supporting decentralised sequencing for SARS-CoV-2 genomic surveillance. *Genome Biol.* **22**, 196 (2021).
- Connor, T. R. et al. CLIMB (the Cloud Infrastructure for Microbial Bioinformatics): an online resource for the medical microbiology community. *Micro. Genom.* **2**, e000086 (2016).
- Minh, B. Q. et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol. Biol. Evol.* **37**, 1530–1534 (2020).
- Tavare, S. Some probabilistic and statistical problems in the analysis of DNA sequences. *Lect. math. life sci.* **17**, 57–86 (1986).
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
- Minh, B. Q., Nguyen, M. A. & von Haeseler, A. Ultrafast approximation for phylogenetic bootstrap. *Mol. Biol. Evol.* **30**, 1188–1195 (2013).
- Rambaut, A., Lam, T. T., Max Carvalho, L. & Pybus, O. G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evol.* **2**, vew007 (2016).
- Argimon, S. et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Micro. Genom.* **2**, e000093 (2016).
- Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
- Suchard, M. A. et al. Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* **4**, vey016 (2018).
- Ghafari, M. et al. Purifying selection determines the short-term time dependency of evolutionary rates in SARS-CoV-2 and pH1N1 influenza. *Mol. Biol. Evol.* msac009. <https://doi.org/10.1093/molbev/msac009> (2022).
- Vaughan, T. G., Scire, J., Nadeau, S. A. & Stadler, T. Estimates of outbreak-specific SARS-CoV-2 epidemiological parameters from genomic data. Preprint at <https://www.medrxiv.org/content/10.1101/2020.09.12.20193284v1> (2020).
- Nadeau, S. A., Vaughan, T. G., Scire, J., Huisman, J. S. & Stadler, T. The origin and early spread of SARS-CoV-2 in Europe. *Proc. Natl. Acad. Sci. USA* **118**, e2012008118 (2021).
- Geoghegan, J. L. et al. Genomic epidemiology reveals transmission patterns and dynamics of SARS-CoV-2 in Aotearoa New Zealand. *Nat. Commun.* **11**, 6351 (2020).
- Rambaut, A., Drummond, A. J., Xie, D., Baele, G. & Suchard, M. A. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst. Biol.* **67**, 901–904 (2018).
- Bouckaert, R. et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
- Drummond, A. J., Ho, S. Y., Phillips, M. J. & Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol.* **4**, e88 (2006).
- Plummer, M., Best, N., Cowles, K. & Vines, K. CODA: convergence diagnosis and output analysis for MCMC. *R. N.* **6**, 7–11 (2006).

49. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, 30494 (2017).
50. Toribio, A. L. et al. European Nucleotide Archive in 2016. *Nucleic Acids Res.* **45**, D32–D36 (2017).
51. Aggarwal, D. et al. Genomic epidemiology of SARS-CoV-2 in a UK university identifies dynamics of transmission. Github <https://doi.org/10.5281/zenodo.5643354> (2021).
52. Griffiths, E. J. T. et al. The PHA4GE SARS-CoV-2 contextual data specification for open genomic epidemiology. Preprint at <https://www.preprints.org/manuscript/202008.0220/v1> (2020).

Acknowledgements

Authors A.S.J., W.H. and T.F., and authors L.d.P. and V.H. contributed equally. We thank members of the COVID-19 Genomics Consortium UK and NHS Test and Trace contact tracers for their contributions to generating data used in this study. We thank the Sanger Covid Team for assisting with Samples and Logistics. We are grateful to all students and staff at the University of Cambridge who have contributed to the COVID-19 response during Michaelmas Term. We are grateful to all staff members of the Cambridge COVID-19 Testing Centre for generating qPCR data. D.A. is a Wellcome Clinical PhD Fellow and gratefully supported by the Wellcome Trust (Grant number: 222903/Z/21/Z). B.W. receives funding from the University of Cambridge and the National Institute for Health Research (NIHR) Cambridge Biomedical Research Centre (BRC) at the Cambridge University Hospitals NHS Foundation Trust. I.G. is a Wellcome Senior Fellow and is supported by the Wellcome Trust (Grant number: 207498/Z/17/Z and 206298/B/17/Z). E.M.H. is supported by a UK Research and Innovation (UKRI) Fellowship: MR/S00291X/1. C.J.R.I. acknowledges Medical Research Council (MRC) funding (ref. MC_UU_00002/11). NJM is supported by the MRC (CSF MR/P008801/1) and NHSBT (WPA15-02). A.J.P. gratefully acknowledge the support of the Biotechnology and Biological Sciences Research Council (BBSRC); their research was funded by the BBSRC Institute Strategic Programme Microbes in the Food Chain BB/R012504/1 and its constituent project BBS/E/F/000PR10352, also Quadram Institute Bioscience BBSRC funded Core Capability Grant (project number BB/CCG1860/1). L.d.P. and O.G.P. were supported by the Oxford Martin School. This research was supported by the NIHR Cambridge BRC. The views expressed are those of the authors and not necessarily those of the NHS, the NIHR, or the Department of Health and Social Care. The COVID-19 Genomics UK Consortium is supported by funding from the MRC part of UK Research & Innovation (UKRI), the National Institute of Health Research and Genome Research Limited, operating as the Wellcome Sanger Institute. The Cambridge Covid-19 testing Centre is funded by the Department of Health and Social Care, UK Government. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. For the purpose of Open Access, the author has applied a CC-BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Author contributions

All authors read the manuscript and consented to its publication. Where a = conceptualization; b = methodology; c = software; d = validation; e = formal analysis; f = investigation; g = resources; h = data curation; i = writing—original draft preparation; j = writing—review and editing; k = visualization; l = supervision; m = project administration; n = funding acquisition, D.A. contributed a, b, c, d, e, f, h, i, j, k, m; B.B. contributed f, g, h, j, m; CB f, g, h; D.D.A. contributed j, l; G.D. contributed l; L.d.P. contributed b, c, e, g, h, j, k; T.F. contributed a, b, d, e, f, h, j; S.F. contributed f, g, h; I.G. contributed e, f, h; Y.C.

contributed e, f, h; I.G.G. contributed a, b, f, g, j, l, m, n; Gr.H. contributed e, f, h; W.L.H. contributed a, b, c, e, j, k; E.M.H. contributed a, b, g, j, l, m, n; V.H. contributed b, c, e, j; J.H. contributed f, g, h; M.H. contributed e, f, h; R.H. contributed b, d, g, h; Ga.H. contributed f, h; R.I. contributed e, f, h; C.I. contributed b, c, e, h, j, k, l; A.J. contributed b, e, f, h, j; D.L. contributed f; C.L. contributed a, g, j, n; D.M. contributed g, l, m; N.J.M. contributed a, b, j, l, m, n; P.H.M. contributed j, l, m, n; M.M. contributed f, g, h; R.M. contributed j, l, m; O.N. contributed f, h; S.P. contributed b, d, g, h; A.J.P. contributed l; S.J.P. contributed a, b, g, j, l, m, n; M.L.P. contributed e, f, h; O.G.P. contributed j, l; B.W. contributed a, b, d, e, f, h, j, k, m; Em.W. contributed a, j; G.W. contributed f, g, h; M.W. contributed j, m, n; EL.W. contributed f, h, j; Cambridge Covid-19 Testing Centre contributed b, d, g, h; University of Cambridge Asymptomatic COVID-19 Screening Programme Consortium contributed b, d, g, h; The COVID-19 Genomics UK (COG-UK) Consortium contributed b, d, g, h.

Competing interests

R.H. is an employee of AstraZeneca AB. The remaining authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41467-021-27942-w>.

Correspondence and requests for materials should be addressed to Dinesh Aggarwal, Ewan M. Harrison, Nicholas J. Matheson or Ian G. Goodfellow.

Peer review information *Nature Communications* thanks Sébastien Calvignac-Spencer, Joep de Ligt and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022

¹University of Cambridge, Department of Medicine, Cambridge, UK. ²Public Health England, 61 Colindale Ave, London NW9 5EQ, UK. ³Cambridge University Hospital NHS Foundation Trust, Cambridge, UK. ⁴Wellcome Sanger Institute, Hinxton, Cambridge, UK. ⁵Cambridge Institute for Therapeutic Immunology and Infectious Disease, University of Cambridge, Cambridge, UK. ⁶University of Cambridge, Department of Pathology, Division of Virology, Cambridge, UK. ⁷Department of Zoology, University of Oxford, Oxford, UK. ⁸Institute of Evolutionary Virology, University of Edinburgh, Edinburgh, UK. ⁹Public Health Directorate, Cambridgeshire County Council and Peterborough City Council, Peterborough, UK. ¹⁰Quadram Institute Bioscience, Norwich Research Park, Norwich NR4 7UQ, UK. ¹¹Health, Safety & Regulated Facilities Division, University of Cambridge, Cambridge, UK. ¹²University Information Services, University of Cambridge, Cambridge, UK. ¹³Cambridge Covid-19 Testing Centre, Discovery Sciences, R&D, AstraZeneca, Cambridge, UK. ¹⁴Charles River Laboratories, Chesterford Research Park, Saffron Walden CB10 1XL, UK. ¹⁵Institute for Manufacturing, University of Cambridge, Cambridge, UK. ¹⁶MRC Biostatistics Unit, University of Cambridge, East Forvie Building, Forvie Site, Robinson Way, Cambridge CB2 0SR, UK. ¹⁷Cambridge Institute for Medical Research, University of Cambridge, Cambridge, UK. ¹⁸Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK. ¹⁹MRC-University of Glasgow Centre for Virus Research, Glasgow, UK. ²⁰Department of Public Health and Primary Care, University of Cambridge, Cambridge, UK. ²¹NHS Blood and Transplant, Cambridge, UK. ²²These authors contributed equally: Dinesh Aggarwal, Ben Warne. ²³These authors jointly supervised this work: Michael P. Weekes, Chris Illingworth, Ewan M. Harrison, Nicholas J. Matheson, Ian G. Goodfellow. *Lists of authors and their affiliations appear at the end of the paper. ✉email: dinesh.aggarwal@nhs.net; eh6@sanger.ac.uk; njm25@cam.ac.uk; ig299@cam.ac.uk

The Cambridge Covid-19 testing Centre

Rob Howes¹³, Alexandra Orton¹³, Julie Douthwaite¹³, Steve Rees¹³, Christopher Brown^{13,14}, Roger Clark¹⁴, Daniel R. Jones¹⁴, Fred Kuenzi¹⁴, Jennifer Rankin¹⁴ & Ian Waddell¹⁴

University of Cambridge Asymptomatic COVID-19 Screening Programme Consortium

Patrick Maxwell^{1,3}, Nicholas Matheson^{1,3,5,21}, Chris Abell²², Vickie Braithwaite^{9,23}, Craig Brierley²⁴, Jon Crowcroft²², Aastha Dahal²⁵, Kathryn Faulkner⁹, Michael Glover²², Ian Goodfellow⁶, Jane Grotzinger²⁶, Jon Holgate¹², Rob Howes¹³, Laura James²², Paul Lehner^{1,3,5}, Ian Leslie¹², Kathleen Liddell²⁷, Ben Margolis²⁵, Duncan McFarlane¹⁵, Sally Morgan²⁸, Linda Sheridan⁹, Sally Valletta²², Anna Vignoles²², Martin Vinnell²⁹, Ben Warne^{1,3,5,128}, Michael P. Weekes^{3,17,129}, Mark Wills¹, Stewart Fuller¹, Marina Metaxaki¹, Sarah Hilborne¹, Sarah Berry¹, Mahin Bagheri Kahkeshi¹, Dawn Hancock¹, Jennifer Winstler¹, Jessica Enright³⁰, Richard Samworth³¹, Vijay Samtani¹², Gabriela Ahmadi-Assalemi¹², Tom Feather¹², Robin Goodall¹², Steve Hoensch¹², Dean Johnson¹², Martin Hunt¹², Nick Mathieson¹², Katya Nikitina¹², Zara Sheldrake¹², Martin Keen³², Aris Sato³², David Connor³², Jonathan Tolhurst³², Jack Williman³², Victoria Hollamby¹, Gillian Weale¹¹, Sinead Jordan³³, Tania Fatseas³³, Peter Taylor³³, Christine Georgiou³³, Michelle Caspersz³³, Claire McNulty³³, Richard Davies³³, Rebecca Clarke¹⁵, Darius Danaei¹⁵, Rory Dyer¹⁵, Rob Glew¹⁵, Oliver Lambson¹⁵, Karen Gibbs²², Barbara Mozdzen²², Gabor Raub²², Asako Radecki²², Phil White²², Robert Hughes²², Lucie Gransden²², Matt Ceaser²², Robert Sing²², Karl Wilson²², Ajith Parlikad¹⁵, Maharshi Dhada¹⁵, Tom Ridgman¹⁵, Diane Mungovan³⁴, Steve Matthews³⁴, Paul Searle³⁴, John Mills³⁴, Andy Neely³⁵, Robert Henderson²⁸, Edna Murphy³⁶, Matthew Russell³⁶, Anthony Freeling²⁸, Steve Poppitt³⁷, Jo Tynan³⁷, James Knapton³⁸, Filippo Marchetti³⁸, Sharon J. Peacock^{1,3}, Ian G. Goodfellow^{6,129}, Ewan M. Harrison^{1,2,4,20,129}, Dinesh Aggarwal^{1,2,3,4,128}, Thomas Fieldman^{1,3}, Beth Blane¹, Yasmin Chaudhry⁶, Daniela De Angelis¹⁶, Theresa Feltwell², Iliana Georgana⁶, Nazreen F. Hadjirin¹, Grant Hall⁶, William L. Hamilton^{1,3,4}, Myra Hosmillo^{3,6}, Chris Illingworth^{16,18,19,129}, Aminu Jahun⁶, Danielle Leek¹, Catherine Ludden^{1,2}, Malte Pinckert⁶, Ashley Shaw³, Afzal Chaudhry³, Nicholas M. Brown³, Lenette Mactavous³, Sophie Hannan³, Aleksandra Hosaja³, Clare Leong³, Jo Wright³⁹, Natalie Quinnell³⁹, Chris Workman³⁹, Mark Ferris³⁹, Giles Wright³⁹, Emmeline Watkins⁹ & Elizabeth Wright⁹

²²University of Cambridge, Cambridge, UK. ²³MRC Epidemiology Unit, University of Cambridge, Cambridge, UK. ²⁴Office of External Affairs and Communications, University of Cambridge, Cambridge, UK. ²⁵Cambridge Students' Union, Cambridge, UK. ²⁶Lucy Cavendish College, University of Cambridge, Cambridge, UK. ²⁷Centre for Law, Medicine and Life Sciences, Faculty of Law, University of Cambridge, Cambridge, UK. ²⁸Fitzwilliam College, University of Cambridge, Cambridge, UK. ²⁹Occupational Health and Safety Services, University of Cambridge, Cambridge, UK. ³⁰School of Computing Science, University of Glasgow, Glasgow, UK. ³¹Statistical Laboratory, Centre for Mathematical Sciences, University of Cambridge, Cambridge, UK. ³²Clinical School Computing Service, School of Clinical Medicine, University of Cambridge, Cambridge, UK. ³³COVID-19 Operations Centre, University of Cambridge, Cambridge, UK. ³⁴University Messenger Service, University of Cambridge, Cambridge, UK. ³⁵Vice Chancellor's Office, University of Cambridge, Cambridge, UK. ³⁶Office of Intercollegiate Services Ltd., Cambridge, UK. ³⁷St John's College, University of Cambridge, Cambridge, UK. ³⁸Governance and Compliance Division, University of Cambridge, Cambridge, UK. ³⁹Occupational Health and Wellbeing, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK.

The COVID-19 Genomics UK (COG-UK) Consortium

Dinesh Aggarwal¹, Beth Blane¹, Ellena Brooks¹, Alessandro M. Carabelli¹, Carol M. Churcher¹, Katerina Galai¹, Sophia T. Girgis¹, Ravi K. Gupta¹, Nazreen F. Hadjirin¹, Danielle Leek¹, Catherine Ludden^{1,2}, Georgina M. McManus¹, Sophie Palmer¹, Sharon J. Peacock^{1,2,3,4}, Kim S. Smith¹, Elias Allara², David Bibby², Chloe Bishop², Andrew Bosworth², Daniel Bradshaw², Vicki Chalker², Meera Chand², Gavin Dabrera², Nicholas Ellaby², Eileen Gallagher², Natalie Groves², Ian Harrison², Hassan Hartman², Richard Hopes², Jonathan Hubb², Stephanie Hutchings², Angie Lackenby², Juan Ledesma², David Lee², Nikos Manesis², Carmen Manso², Tamyo Mbisa², Shahjahan Miah², Peter Muir², Richard Myers², Husam Osman², Vineet Patel²,

Clare Pearson², Steven Platt², Hannah M. Pymont², Mary Ramsay², Esther Robinson², Ulf Schaefer², Alicia Thornton², Katherine A. Twohig², Ian B. Vipond², David Williams², William L. Hamilton^{1,3,4}, Ben Warne^{1,3,5,128}, Louise Aigrain⁴, Alex Alderton⁴, Roberto Amato⁴, Cristina V. Ariani⁴, Jeff Barrett⁴, Andrew R. Bassett⁴, Mathew A. Beale⁴, Charlotte Beaver⁴, Katherine L. Bellis⁴, Emma Betteridge⁴, James Bonfield⁴, Iraad F. Bronner⁴, Michael H. S. Chapman⁴, John Danesh⁴, Robert Davies⁴, Matthew J. Dorman⁴, Eleanor Drury⁴, Jillian Durham⁴, Ben W. Farr⁴, Luke Foulser⁴, Sonia Goncalves⁴, Scott Goodwin⁴, Marina Gourtovaia⁴, Ewan M. Harrison^{1,2,4,20,129}, David K. Jackson⁴, Keith James⁴, Dorota Jamroz⁴, Ian Johnston⁴, Leanne Kane⁴, Sally Kay⁴, Jon-Paul Keatley⁴, Dominic Kwiatkowski⁴, Cordelia F. Langford⁴, Mara Lawniczak⁴, Stefanie V. Lensing⁴, Steven Leonard⁴, Laura Letchford⁴, Kevin Lewis⁴, Jennifer Liddle⁴, Rich Livett⁴, Stephanie Lo⁴, Alex Makunin⁴, Inigo Martincorena⁴, Shane McCarthy⁴, Samantha McGuigan⁴, Robin J. Moll⁴, Rachel Nelson⁴, Karen Oliver⁴, Steve Palmer⁴, Naomi R. Park⁴, Minal Patel⁴, Liam Prestwood⁴, Christoph Puethe⁴, Michael A. Quail⁴, Diana Rajan⁴, Shavanthi Rajatileka⁴, Nicholas M. Redshaw⁴, Carol Scott⁴, Lesley Shirley⁴, John Sillitoe⁴, Scott A. J. Thurston⁴, Gerry Tonkin-Hill⁴, Jaime M. Tovar-Corona⁴, Danni Weldon⁴, Andrew Whitwham⁴, Yasmin Chaudhry⁶, Iliana Georgana⁶, Ian G. Goodfellow^{6,129}, Grant Hall⁶, Myra Hosmillo^{3,6}, Aminu S. Jahun⁶, Malte L. Pinckert⁶, Stephen W. Attwood⁷, Louis du Plessis⁷, Marina Escalera Zamudio⁷, Sarah Francois⁷, Bernardo Gutierrez⁷, Moritz U. G. Kraemer⁷, Oliver G. Pybus⁷, Jayna Raghwan⁷, Tetyana I. Vasylyeva⁷, Alex E. Zarebski⁷, Verity Hill⁸, Nabil-Fareed Alikhan¹⁰, Alp Aydin¹⁰, David J. Baker¹⁰, Leonardo de Oliveira Martins¹⁰, Gemma L. Kay¹⁰, Thanh Le-Viet¹⁰, Alison E. Mather¹⁰, Lizzie Meadows¹⁰, Justin O'Grady¹⁰, Andrew J. Page¹⁰, Steven Rudder¹⁰, Alexander J. Trotter¹⁰, Chris J. Illingworth¹⁶, Chris Jackson¹⁶, Elihu Aranday-Cortes¹⁹, Patawee Asamaphan¹⁹, Alice Broos¹⁹, Stephen N. Carmichael¹⁹, Ana da Silva Filipe¹⁹, Joseph Hughes¹⁹, Natasha G. Jesudason¹⁹, Natasha Johnson¹⁹, Kathy K. Li¹⁹, Daniel Mair¹⁹, Jenna Nichols¹⁹, Seema Nickbakhsh¹⁹, Marc O. Niebel¹⁹, Kyriaki Nomikou¹⁹, Richard J. Orton¹⁹, David L. Robertson¹⁹, Rajiv N. Shah¹⁹, James G. Shepherd¹⁹, Joshua B. Singer¹⁹, Igor Starinskij¹⁹, Emma C. Thomson¹⁹, Lily Tong¹⁹, Sreenu Vattipally¹⁹, Amy Ash⁴⁰, Cherian Koshy⁴⁰, Nick Cortes⁴¹, Stephen Kidd⁴¹, Jessica Lynch⁴¹, Nathan Moore⁴¹, Matilde Mori⁴¹, Emma Wise⁴¹, Tanya Curran⁴², Derek J. Fairley⁴², James P. McKenna⁴², Helen Adams⁴³, David Bonsall⁴⁴, Christophe Fraser⁴⁴, Tanya Golubchik⁴⁴, Benjamin J. Cogger⁴⁵, Mohammed O. Hassan-Ibrahim⁴⁵, Cassandra S. Malone⁴⁵, Nicola Reynolds⁴⁶, Michelle Wantoch⁴⁶, Safiah Afifi⁴⁷, Robert Beer⁴⁷, Michaela John⁴⁷, Joshua Maksimovic⁴⁷, Kathryn McCluggage⁴⁷, Sian Morgan⁴⁷, Karla Spellman⁴⁷, Catherine Bresner⁴⁸, Thomas R. Connor⁴⁸, William Fuller⁴⁸, Martyn Guest⁴⁸, Huw Gulliver⁴⁸, Christine Kitchen⁴⁸, Angela Marchbank⁴⁸, Ian Merrick⁴⁸, Robert Munn⁴⁸, Anna Price⁴⁸, Joel Southgate⁴⁸, Trudy Workman⁴⁸, Amita Patel⁴⁹, Luke B. Snell⁴⁹, Rahul Batra⁵⁰, Themoula Charalampous⁵⁰, Jonathan Edgeworth⁵⁰, Gaia Nebbia⁵⁰, Angela H. Beckett⁵¹, Samuel C. Robson⁵¹, David M. Aanensen⁵², Khalil Abudahab⁵², Mirko Menegazzo⁵², Ben E. W. Taylor⁵², Anthony P. Underwood⁵², Corin A. Yeats⁵², Louise Berry⁵³, Tim Boswell⁵³, Gemma Clark⁵³, Vicki M. Fleming⁵³, Hannah C. Howson-Wells⁵³, Carl Jones⁵³, Amelia Joseph⁵³, Manjinder Khakh⁵³, Michelle M. Lister⁵³, Wendy Smith⁵³, Iona Willingham⁵³, Paul Bird⁵⁴, Karlie Fallon⁵⁴, Thomas Helmer⁵⁴, Christopher Holmes⁵⁴, Julian Tang⁵⁴, Victoria Blakey⁵⁵, Sharon Campbell⁵⁵, Veena Raviprakash⁵⁵, Nicola Sheriff⁵⁵, Lesley-Anne Williams⁵⁵, Matthew Carlile⁵⁶, Johnny Debebe⁵⁶, Nadine Holmes⁵⁶, Matthew W. Loose⁵⁶, Christopher Moore⁵⁶, Fei Sang⁵⁶, Victoria Wright⁵⁶, Francesc Coll⁵⁷, Gilberto Betancor⁵⁸, Adrian W. Signell⁵⁸, Harry D. Wilson⁵⁸, Thomas Davis⁵⁹, Sahar Eldirdiri⁵⁹, Anita Kenyon⁵⁹, M. Estee Torok⁶⁰, Hannah Lowe⁶¹, Samuel Moses⁶¹, Luke Bedford⁶², Jonathan Moore⁶³, Susanne Stonehouse⁶³, Ali R. Awan⁶⁴, Chloe L. Fisher⁶⁴,

John BoYes⁶⁵, Laura Atkinson⁶⁶, Judith Breuer⁶⁶, Julianne R. Brown⁶⁶, Kathryn A. Harris⁶⁶, Jack C. D. Lee⁶⁶, Divya Shah⁶⁶, Nathaniel Storey⁶⁶, Flavia Flaviani⁶⁷, Adela Alcolea-Medina⁶⁸, Gabrielle Vernet⁶⁹, Rebecca Williams⁶⁹, Michael R. Chapman⁷⁰, Wendy Chatterton⁷¹, Judith Heaney⁷¹, Lisa J. Levett⁷¹, Monika Pusok⁷¹, Li Xu-McCrae⁷², Matthew Bashton⁷³, Darren L. Smith⁷³, Gregory R. Young⁷³, Frances Bolt⁷⁴, Alison Cox⁷⁴, Alison Holmes⁷⁴, Pinglawathee Madona⁷⁴, Siddharth Mookerjee⁷⁴, James Price⁷⁴, Paul A. Randell⁷⁴, Olivia Boyd⁷⁵, Fabricia F. Nascimento⁷⁵, Lily Geidelberg⁷⁵, Rob Johnson⁷⁵, David Jorgensen⁷⁵, Manon Ragonnet-Cronin⁷⁵, Aileen Rowan⁷⁵, Igor Siveroni⁷⁵, Graham P. Taylor⁷⁵, Erik M. Volz⁷⁵, Katherine L. Smollett⁷⁶, Nicholas J. Loman⁷⁷, Claire McMurray⁷⁷, Alan McNally⁷⁷, Sam Nicholls⁷⁷, Radoslaw Poplawski⁷⁷, Joshua Quick⁷⁷, Will Rowe⁷⁷, Joanne Stockton⁷⁷, Rocio T. Martinez Nunez⁷⁸, Cassie Breen⁷⁹, Angela Cowell⁷⁹, Jenifer Mason⁷⁹, Elaine O'Toole⁷⁹, Trevor I. Robinson⁷⁹, Joanne Watts⁷⁹, Graciela Sluga⁸⁰, Shazaad S. Y. Ahmad⁸¹, Ryan P. George⁸¹, Nicholas W. Machin⁸¹, Fenella Halstead⁸², Wendy Hogsden⁸², Venkat Sivaprakasam⁸², Holli Carden⁸³, Antony D. Hale⁸³, Katherine L. Harper⁸³, Louissa R. Macfarlane-Smith⁸³, Shirelle Burton-Fanning⁸⁴, Jennifer Collins⁸⁴, Gary Eltringham⁸⁴, Brendan Al. Payne⁸⁴, Yusri Taha⁸⁴, Sheila Waugh⁸⁴, Sarah O'Brien⁸⁵, Steven Rushton⁸⁵, Rachel Blacow⁸⁶, Amanda Bradley⁸⁶, Alasdair Maclean⁸⁶, Guy Mollett⁸⁶, Rebecca Dewar⁸⁷, Martin P. McHugh⁸⁷, Kate E. Templeton⁸⁷, Elizabeth Wastenge⁸⁷, Lindsay Coupland⁸⁸, Samir Dervisevic⁸⁸, Emma J. Meader⁸⁸, Rachael Stanley⁸⁸, Louise Smith⁸⁹, Edward Barton⁹⁰, Clive Graham⁹⁰, Debra Padgett⁹⁰, Garren Scott⁹⁰, Jane Greenaway⁹¹, Emma Swindells⁹¹, Clare M. McCann⁹², Andrew Nelson⁹², Wen C. Yew⁹², Monique Andersson⁹³, Derrick Crook⁹³, David Eyre⁹³, Anita Justice⁹³, Timothy Peto⁹³, Nichola Duckworth⁹⁴, Tim J. Sloan⁹⁴, Sarah Walsh⁹⁴, Kelly Bicknell⁹⁵, Anoop J. Chauhan⁹⁵, Scott Elliott⁹⁵, Sharon Glaysher⁹⁵, Robert Impey⁹⁵, Allyson Lloyd⁹⁵, Sarah Wyllie⁹⁵, Nick Levene⁹⁶, Lynn Monaghan⁹⁶, Declan T. Bradley⁹⁷, Tim Wyatt⁹⁷, Martin D. Curran⁹⁸, Surendra Parmar⁹⁸, Matthew T. G. Holden⁹⁹, Sharif Shaaban⁹⁹, Alexander Adams¹⁰⁰, Hibo Asad¹⁰⁰, Alec Birchley¹⁰⁰, Matthew Bull¹⁰⁰, Jason Coombes¹⁰⁰, Sally Corden¹⁰⁰, Simon Cottrell¹⁰⁰, Noel Craine¹⁰⁰, Michelle Cronin¹⁰⁰, Alisha Davies¹⁰⁰, Elen De Lacy¹⁰⁰, Fatima Downing¹⁰⁰, Sue Edwards¹⁰⁰, Johnathan M. Evans¹⁰⁰, Laia Fina¹⁰⁰, Amy Gaskin¹⁰⁰, Bree Gatica-Wilcox¹⁰⁰, Laura Gifford¹⁰⁰, Lauren Gilbert¹⁰⁰, Lee Graham¹⁰⁰, David Heyburn¹⁰⁰, Ember Hilvers¹⁰⁰, Robin Howe¹⁰⁰, Hannah Jones¹⁰⁰, Rachel Jones¹⁰⁰, Sophie Jones¹⁰⁰, Sara Kumziene-SummerhaYes¹⁰⁰, Caoimhe McKerr¹⁰⁰, Catherine Moore¹⁰⁰, Mari Morgan¹⁰⁰, Nicole Pacchiarini¹⁰⁰, Malorie Perry¹⁰⁰, Amy Plimmer¹⁰⁰, Sara Rey¹⁰⁰, Giri Shankar¹⁰⁰, Sarah Taylor¹⁰⁰, Joanne Watkins¹⁰⁰, Chris Williams¹⁰⁰, Anna Casey¹⁰¹, Liz Ratcliffe¹⁰¹, Erwan Acheson¹⁰², Zoltan Molnar¹⁰², David A. Simpson¹⁰², Thomas Thompson¹⁰², Cressida Auckland¹⁰³, Sian Ellard¹⁰³, Christopher R. Jones¹⁰³, Bridget A. Knight¹⁰³, Jane A. H. Masoli¹⁰³, Tanzina Haque¹⁰⁴, Jennifer Hart¹⁰⁴, Dianne Irish-Tavares¹⁰⁴, Tabitha W. Mahungu¹⁰⁴, Eric Witele¹⁰⁴, Ashok Dadrah¹⁰⁵, Melisa L. Fenton¹⁰⁵, Tranprit Saluja¹⁰⁵, Amanda Symmonds¹⁰⁵, Yann Bourgeois¹⁰⁶, Garry P. Scarlett¹⁰⁶, Kate Cook¹⁰⁷, Hannah Dent¹⁰⁷, Christopher Fearn¹⁰⁷, Salman Goudarzi¹⁰⁷, Katie F. Loveson¹⁰⁷, Hannah Paul¹⁰⁷, Cariad Evans¹⁰⁸, Kate Johnson¹⁰⁸, David G. Partridge¹⁰⁸, Mohammad Raza¹⁰⁸, Paul Baker¹⁰⁹, Stephen Bonner¹⁰⁹, Sarah Essex¹⁰⁹, Steven Liggett¹⁰⁹, Ronan A. Lyons¹¹⁰, Adhyana I. K. Mahanama¹¹¹, Kordo Saeed¹¹¹, Buddhini Samaraweera¹¹¹, Siona Silveira¹¹¹, Eleri Wilson-Davies¹¹¹, P. Emanuela¹¹¹, Nadua Bayzid¹¹², Marius Cotic¹¹², Leah Ensell¹¹², John A. Hartley¹¹², Riaz Jannoo¹¹², Angeliki Karamani¹¹², Mark Kristiansen¹¹², Helen L. Lowe¹¹², Sunando Roy¹¹², Adam P. Westhorpe¹¹², Rachel J. Williams¹¹², Charlotte A. Williams¹¹², Sarah Jeremiah¹¹³, Jacqui A. Prieto¹¹³, Lisa Berry¹¹⁴, Dimitris Grammatopoulos¹¹⁴, Katie Jones¹¹⁴, Sarojini Pandey¹¹⁴, Andrew Beggs¹¹⁵, Alex Richter¹¹⁵, Fiona Ashcroft¹¹⁶, Angus Best¹¹⁶, Liam Crawford¹¹⁶, Nicola Cumley¹¹⁶, Megan Mayhew¹¹⁶, Oliver Megram¹¹⁶, Jeremy Mirza¹¹⁶, Emma Moles-Garcia¹¹⁶,

Benita Percival¹¹⁶, Giselda Bucca¹¹⁷, Andrew R. Hesketh¹¹⁷, Colin P. Smith¹¹⁷, Rose K. Davidson¹¹⁸, Carlos E. Balcazar¹¹⁹, Michael D. Gallagher¹¹⁹, Áine O'Toole¹¹⁹, Andrew Rambaut¹¹⁹, Stefan Rooke¹¹⁹, Thomas D. Stanton¹¹⁹, Thomas Williams¹¹⁹, Kathleen A. Williamson¹¹⁹, Claire M. Bewshea¹²⁰, Audrey Farbos¹²⁰, James W. Harrison¹²⁰, Aaron R. Jeffries¹²⁰, Robin Manley¹²⁰, Stephen L. Michell¹²⁰, Michelle L. Michelsen¹²⁰, Christine M. Sambles¹²⁰, David J. Studholme¹²⁰, Ben Temperton¹²⁰, Joanna Warwick-Dugdale¹²⁰, Alistair C. Darby¹²¹, Richard Eccles¹²¹, Matthew Gemmell¹²¹, Richard Gregory¹²¹, Sam T. Haldenby¹²¹, Julian A. Hiscox¹²¹, Margaret Hughes¹²¹, Miren Iturriza-Gomara¹²¹, Kathryn A. Jackson¹²¹, Anita O. Lucaci¹²¹, Charlotte Nelson¹²¹, Steve Paterson¹²¹, Lucille Rainbow¹²¹, Lance Turtle¹²¹, Edith E. Vamos¹²¹, Hermione J. Webster¹²¹, Mark Whitehead¹²¹, Claudia Wierzbicki¹²¹, Adrienn Angyal¹²², Rebecca Brown¹²², Thushan I. de Silva¹²², Timothy M. Freeman¹²², Marta Gallis¹²², Luke R. Green¹²², Danielle C. Groves¹²², Alexander J. Keeley¹²², Benjamin B. Lindsey¹²², Stavroula F. Louka¹²², Matthew D. Parker¹²², Paul J. Parsons¹²², Nikki Smith¹²², Rachel M. Tucker¹²², Dennis Wang¹²², Max Whiteley¹²², Matthew Wyles¹²², Peijun Zhang¹²², Mohammad T. Alam¹²³, Laura Baxter¹²³, Hannah E. Bridgewater¹²³, Paul E. Brown¹²³, Jeffrey K. J. Cheng¹²³, Chrystala Constantinidou¹²³, Lucy R. Frost¹²³, Sascha Ott¹²³, Richard Stark¹²³, Grace Taylor-Joyce¹²³, Meera Unnikrishnan¹²³, Alberto C. Cerda¹²⁴, Tammy V. Merrill¹²⁴, Rebekah E. Wilson¹²⁴, Jonathan Ball¹²⁵, Joseph G. Chappell¹²⁵, Patrick C. McClure¹²⁵, Theocharis Tsoleridis¹²⁵, David Buck¹²⁶, Mariateresa de Cesare¹²⁶, Angie Green¹²⁶, George MacIntyre-Cockett¹²⁶, John A. Todd¹²⁶, Amy Trebes¹²⁶, Rory N. Gunson¹²⁷, Claire Cormie²², Joana Dias²², Sally Forrest²², Harmeet K. Gill²², Ellen E. Higginson²², Leanne M. Kermack²², Mailis Maes²², Chris Ruis²², Sushmita Sridhar²² & Jamie Young²²

⁴⁰Barking, Havering and Redbridge University Hospitals NHS Trust, Romford, UK. ⁴¹Basingstoke Hospital, Basingstoke, UK. ⁴²Belfast Health & Social Care Trust, Belfast, UK. ⁴³Betsi Cadwaladr University Health Board, Bangor, UK. ⁴⁴Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ⁴⁵Brighton and Sussex University Hospitals NHS Trust, Brighton, UK. ⁴⁶Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. ⁴⁷Cardiff and Vale University Health Board, Cardiff, UK. ⁴⁸Cardiff University, Cardiff, UK. ⁴⁹Centre for Clinical Infection & Diagnostics Research, St. Thomas' Hospital and Kings College London, London, UK. ⁵⁰Centre for Clinical Infection and Diagnostics Research, Department of Infectious Diseases, Guy's and St Thomas' NHS Foundation Trust, London, UK. ⁵¹Centre for Enzyme Innovation, University of Portsmouth (PORT), Portsmouth, UK. ⁵²Centre for Genomic Pathogen Surveillance, University of Oxford, Oxford, UK. ⁵³Clinical Microbiology Department, Queens Medical Centre, Nottingham, UK. ⁵⁴Clinical Microbiology, University Hospitals of Leicester NHS Trust, Leicester, UK. ⁵⁵County Durham and Darlington NHS Foundation Trust, Darlington, UK. ⁵⁶Deep Seq, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK. ⁵⁷Department of Infection Biology, Faculty of Infectious & Tropical Diseases, London School of Hygiene & Tropical Medicine, London, UK. ⁵⁸Department of Infectious Diseases, King's College London, London, UK. ⁵⁹Department of Microbiology, Kettering General Hospital, Kettering, UK. ⁶⁰Departments of Infectious Diseases and Microbiology, Cambridge University Hospitals NHS Foundation Trust, Cambridge, UK. ⁶¹East Kent Hospitals University NHS Foundation Trust, Canterbury, UK. ⁶²East Suffolk and North Essex NHS Foundation Trust, Colchester, UK. ⁶³Gateshead Health NHS Foundation Trust, Gateshead, UK. ⁶⁴Genomics Innovation Unit, Guy's and St. Thomas' NHS Foundation Trust, London, UK. ⁶⁵Gloucestershire Hospitals NHS Foundation Trust, Cheltenham, UK. ⁶⁶Great Ormond Street Hospital for Children NHS Foundation Trust, London, UK. ⁶⁷Guy's and St. Thomas' BRC, London, UK. ⁶⁸Guy's and St. Thomas' Hospitals, London, UK. ⁶⁹Hampshire Hospitals NHS Foundation Trust, Basingstoke, UK. ⁷⁰Health Data Research UK Cambridge, Cambridge, UK. ⁷¹Health Services Laboratories, London, UK. ⁷²Heartlands Hospital, Birmingham, UK. ⁷³Hub for Biotechnology in the Built Environment, Northumbria University, Newcastle, UK. ⁷⁴Imperial College Hospitals NHS Trust, London, UK. ⁷⁵Imperial College London, London, UK. ⁷⁶Institute of Biodiversity, Animal Health & Comparative Medicine, Glasgow, UK. ⁷⁷Institute of Microbiology and Infection, University of Birmingham, Birmingham, UK. ⁷⁸King's College London, London, UK. ⁷⁹Liverpool Clinical Laboratories, Liverpool, UK. ⁸⁰Maidstone and Tunbridge Wells NHS Trust, Tunbridge Wells, UK. ⁸¹Manchester University NHS Foundation Trust, Manchester, UK. ⁸²Microbiology Department, Wye Valley NHS Trust, Hereford, UK. ⁸³National Infection Service, PHE and Leeds Teaching Hospitals Trust, Leeds, UK. ⁸⁴Newcastle Hospitals NHS Foundation Trust, Newcastle, UK. ⁸⁵Newcastle University, Newcastle, UK. ⁸⁶NHS Greater Glasgow and Clyde, Glasgow, UK. ⁸⁷NHS Lothian, Edinburgh, UK. ⁸⁸Norfolk and Norwich University Hospital, Norwich, UK. ⁸⁹Norfolk County Council, Norwich, UK. ⁹⁰North Cumbria Integrated Care NHS Foundation Trust, Carlisle, UK. ⁹¹North Tees and Hartlepool NHS Foundation Trust, Stockton on Tees, UK. ⁹²Northumbria University, Newcastle, UK. ⁹³Oxford University Hospitals NHS Foundation Trust, Oxford, UK. ⁹⁴PathLinks, Northern Lincolnshire & Goole NHS Foundation Trust, Grimsby, UK. ⁹⁵Portsmouth Hospitals University NHS Trust, Portsmouth, UK. ⁹⁶Princess Alexandra Hospital Microbiology Dept, Harlow, UK. ⁹⁷Public Health Agency, Belfast, UK. ⁹⁸Public Health England, Clinical Microbiology and Public Health Laboratory, Cambridge, UK. ⁹⁹Public Health Scotland, Edinburgh, UK. ¹⁰⁰Public Health Wales NHS Trust, Cardiff, UK. ¹⁰¹Queen Elizabeth Hospital, London, UK. ¹⁰²Queen's University Belfast, Belfast, UK. ¹⁰³Royal Devon and Exeter NHS Foundation Trust, Exeter, UK. ¹⁰⁴Royal Free NHS Trust, London, UK. ¹⁰⁵Sandwell and West Birmingham NHS Trust, Birmingham, UK. ¹⁰⁶School of Biological Sciences, University of Portsmouth (PORT), Portsmouth, UK. ¹⁰⁷School of Pharmacy and Biomedical Sciences, University of Portsmouth (PORT), Portsmouth, UK. ¹⁰⁸Sheffield Teaching Hospitals, Sheffield, UK. ¹⁰⁹South Tees Hospitals NHS Foundation Trust, Middlesbrough, UK. ¹¹⁰Swansea University, Swansea, UK. ¹¹¹University Hospitals Southampton NHS Foundation Trust, Southampton, UK. ¹¹²University College London, London, UK. ¹¹³University Hospital Southampton NHS Foundation Trust, Southampton, UK. ¹¹⁴University Hospitals Coventry and Warwickshire, Coventry, UK. ¹¹⁵University of Birmingham, Birmingham, UK. ¹¹⁶University of Birmingham Turnkey Laboratory, Birmingham, UK. ¹¹⁷University of Brighton, Brighton, UK. ¹¹⁸University of East Anglia, Norwich, UK. ¹¹⁹University of Edinburgh, Edinburgh, UK. ¹²⁰University of Exeter, Exeter, UK. ¹²¹University of

Liverpool, Liverpool, UK. ¹²²University of Sheffield, Sheffield, UK. ¹²³University of Warwick, Warwick, UK. ¹²⁴Viapath, Guy's and St Thomas' NHS Foundation Trust, and King's College Hospital NHS Foundation Trust, London, UK. ¹²⁵Virology, School of Life Sciences, Queens Medical Centre, University of Nottingham, Nottingham, UK. ¹²⁶Wellcome Centre for Human Genetics, Nuffield Department of Medicine, University of Oxford, Oxford, UK. ¹²⁷West of Scotland Specialist Virology Centre, NHS Greater Glasgow and Clyde, Glasgow, UK.