Genomic epidemiology of SARS-CoV-2 in the University of Cambridge identifies dynamics of transmission: an interim report

Report prepared by Dr Dinesh Aggarwal, Dr Tom Fieldman, Dr Ben Warne, COVID-19 Genomics UK (COG-UK) Consortium UK and the University of Cambridge Asymptomatic COVID-19 Screening Programme Consortium, 23rd November 2020.

Introduction

Outbreaks of SARS-CoV-2 have occurred in numerous Universities globally. To our knowledge these have not been characterised with systematic genomic analysis. We present the results of systematic genomic analysis of early SARS-CoV-2 infections amongst students at the University of Cambridge from the first five weeks of Michaelmas term.

Headline Findings

- Limited cross-transmission between University students and the local community is observed during the study period
- Phylogenetic diversity amongst University students is low, indicating few introductions led to established outbreaks of SARS-CoV-2 in the University population
- Outbreaks largely restricted to a single college were observed where cessation of onwards transmission suggests successful intervention to prevent spread of the virus
- Early in the term, continued transmission from outbreaks in 1st year students appears limited and when confined within colleges.
- One cluster, that includes a diverse number of Colleges, courses and years of study is observed. This appears to be the source of the majority of onwards transmission within the University by week 5 of term.

Results

- 219 genomes from 490 positive individuals were sequenced from term weeks 1-5 (Table 1) with samples from week 1 currently unavailable.
- 12 clusters of cases (>2 individuals) can be defined through genomics with 4 clusters >5 cases by term week 5

Community and University Transmission

- 66/73 (90.4%) of University isolates were confined to 3 lineages in week two. This is likely to reflect a small number of seeding events resulting in established clusters at the beginning of the academic year. This also reflects a possibly low incidence of infection amongst international students and/or successful quarantine and compliance with infection prevention control measures by international students on arrival.
- Limited transmission between university students and the 282 community isolates is identified. This remains consistent up until term week 5 (Figure 1).
- It is important to note that inferences drawn from University cases which do appear amongst the community genomes in the Phylogenetic tree (Figure 1), are likely to reflect the limited diversity of SARS-CoV-2 genomes and unlikely to represent community-university transmission.
- By term week 5, a total of 15/282 (5.3%) individuals tested in the community are found to cluster with University students, as defined by a 2 single-nucleotide polymorphism (SNP) difference. 12/282 (4.3%) isolates were within 1 SNP of University isolates. It is unclear if they represent individuals not affiliated with the University, or staff/students accessing Test and Trace services. Contact tracing information for these individuals will allow further characterisation.

Intra-college transmission

- Two large clusters within colleges were identified in term weeks 2 and 3.
- Cluster 2 (College A) involves 19 cases by term week 2 involving 11 households and 10 courses. 15/19 infections are in 1st year students. By term week 3, this cluster involves 32 cases, of which 2 cases are in 2 further Colleges and 2 local samples which have yet to be characterised. Of note, 50% of the individuals in this cluster had an asymptomatic infection. The increase in case ascertainment of asymptomatic infections in term week 3 reflects systematic individual screening of all students in this accommodation, as part of outbreak investigations between the University and local public health teams.
- Cluster 3 (College B) involves 13 cases by term week 2 (n=13). 10/13 cases are represented by College B, with two further cases in College C, and one from College D. 8/10 College B students were 1st year students, whilst all 3 students from College C and D are 1st year and share the same course with a 1st year student from College B. In term week 3, the cluster involves 22 cases, with no further infections within College B, one household infection in College A, and 3 further infections with no clear epidemiological links. 3 local isolates are related to this cluster and are yet to be characterised.
- No further growth of either cluster is seen after week 3 (Figure 2) indicating likely cessation of transmission and successful infection prevention control measures.

Continued Inter-College Transmission

- Cluster 1 is the largest identified cluster and the source of ongoing transmission within the University.
- From term week 2 to term week 5, this cluster grew from 30 cases to 139 cases (135 university students, 1 known staff member and 3 local isolates).
- The earliest available genomes are from term week 2 (n=29); 15 individuals are 2nd year students, 5 students are from College E and 5 from College F, and 6 students share a course subject (Course 1). A total of 13 courses, 14 colleges, 23 households and 4 academic years are represented.
- No clear epidemiological link identifies a common source event and therefore sourcing term week 1 samples and enhanced contact tracing is required to accurately ascertain this seeding event.
- As of term week 5, this cluster had grown to 139 cases, with 3 represented by local samples and 1 confirmed staff member. Most infections are in College G (n=20, the largest contribution appearing to be an outbreak amongst students sharing a course (Course 2) in term weeks 4 and 5, and linked to 1 large household), with a third course, Course 3, being the most represented course (n=16). A total of 41 courses 29 Colleges and 101 households are involved.
- The cluster is represented by a unique phylotype (COG-UK metadata accessed 17/11/2020). After placing this cluster in the national and international context, we have identified the likely source region presenting a route to assist contact tracing efforts.
- *Please see addendum for an additional update to this cluster.**

Conclusions

Our findings highlight the paucity of introduction events that led to established transmission, and emphasise the importance of outbreak containment. We identify little evidence of substantial transmission between University of Cambridge students and the local community from the isolates included in this analysis and during the study-period. Together, this may reflect the efficacy of a structured University wide screening programme. There are a small number of lineages accounting for the vast majority of University cases, likely to reflect successful control of transmission by compliance with infection prevention control measures. Early outbreaks appear to have halted when confined within Colleges and predominantly affect 1st year students. Speculatively, this may be due to minimal intercollege contact of 1st year students in their first term. One cluster, that is diverse by college, course and year of matriculation is observed. This has become the dominant cluster, accounting for the of onwards transmission within the University. The widespread distribution of

members of this cluster likely hampers efforts to control transmission; the widespread distribution of this cluster facilitates greater opportunity for onward transmission than if it were mostly restricted to single college or course. Critically, rapid sequencing and analysis of cases has provided information on genomically linked clusters that should prompt contact tracing to provide insight into transmission pathways.

Recommendations

- Use of genomic data can help characterise community-university transmission.
- Investigation of transmission between courses and colleges should be a priority; once established, this results in continued cluster growth.
- Early use of genomics can identify propagation of diverse clusters by college, course and academic year and help direct prevention strategies.
- The national COG-UK genome dataset can be utilised to potentially identify the source of a cluster that should inform contact-tracing and better characterise transmission.
- Finer resolution contact tracing is required to further understand dominant transmission sources within the University that do not conform to Colleges and households.

Methods

Isolates for this study were derived from the symptomatic and asymptomatic COVID-19 testing programmes within the University of Cambridge. Testing for all symptomatic students and staff within the University has been available on all weekdays from 5th October. Asymptomatic screening is offered to all students resident in College accommodation. To optimise efficiency of testing, swabs are pooled at the time of sampling into the same tube of viral transport medium. Pools vary in size from 1 to 10 students, based on student households. In term weeks 1 and 2, 2 students from each testing pool were invited to submit swabs. In term weeks 3-5, half of the students in each household were invited to submit swabs, alternating weekly such that all students participating in the programme are screened once every 2 weeks. Consent rates are persistently >75% of eligible students. Further details, including protocols, participation and positivity rates, can be found here: https://www.cam.ac.uk/coronavirus/stay-safe-cambridge-uni/asymptomatic-covid-19-screening-programme

All SARS-CoV-2 tests were performed by PCR in one of the UK's lighthouse laboratories located on the Cambridge Biomedical Campus, using the same procedures as those used in national testing. All plates containing extracted RNA from University samples were shipped to the University of Cambridge Department of Medicine, so that positive samples with a Ct value ≤33 could be picked and sequenced using the GridION platform (Oxford Nanopore).

Genomic data was filtered to exclude sequences with <90% 'N's and those of spuriously low file sizes (<29KB). Sequence alignment was performed using MAAFT including a reference genome from Wuhan, China collected December 2019 and used to root the tree (GISAID ID: EPI_ISL_402123). Phylogenetic trees were generated using IQ-TREE (version 2.1.2 COVID-edition, and visualised using Microreact online tool (Argimon et al, 2016). Viral lineages were assigned using Pangolin COVID-19 Lineage Assigner web utility (COG-UK, 2020). Samples were also analysed using the CIVET tool (version 2.0) on 17/11/2020. Collapsed nodes from trees generated from the CIVET tool were inspected to visualise data in the context of the COG-UK national database (https://www.cogconsortium.uk/) and some background information from the international sequence repository, GISAID (www.gisaid.org).

282 samples were used to represent the Cambridgeshire local population. All local isolates between 27/09/2020 and 26/10/2020 from the COG-UK database were included in the analysis (n=158), accessed through MRC-CLIMB system (Connor et al, 2016) and sequences derived from SARS-CoV-2 isolates sampled in Cambridge University Hospitals included those sampled between 24/10/2020

and 09/11/2020 (n=124). All local samples were subject to the same quality control measures as the University isolates.

Table 1. Number of positive University of Cambridge isolates and subsequent sequencing by term week

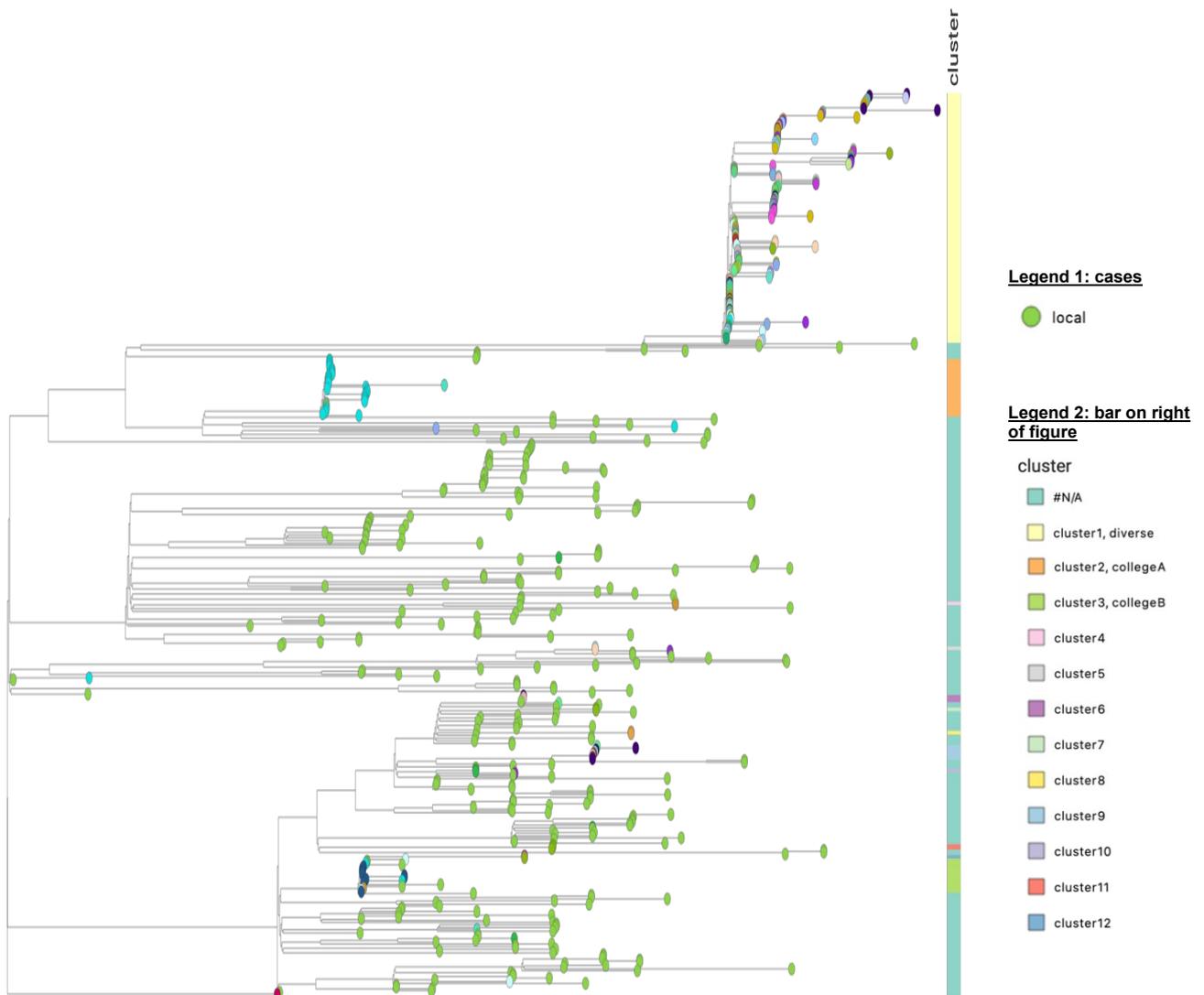| Term week	Number of student tests performed	Positive samples	Positive samples (Ct ≤32)	Samples Sequenced (% of positive samples)	Genomes pass QC
1 (5/10-11/10)	3693	34	27 (79.4%)	0 (0%)	0
2 (12/10-18/10)	4086	154	145 (94.2%)	81 (52.6%)	73
3 (19/10-25/10)	5069	155	133 (85.8%)	61 (39.3%)	55*
4 (26/10-1/11)	5784	78	57 (73.1%)	39 (50.0%)	37
5 (2/11-8/11)	4825	69	63 (91.3%)	62 (89.9%)	54
Total	23457	490	425 (86.7%)	243 (49.6%)	219

*These include University of Cambridge students and one known staff member

Table 2. Local Coverage of SARS-CoV-2 Sequencing in Cambridgeshire

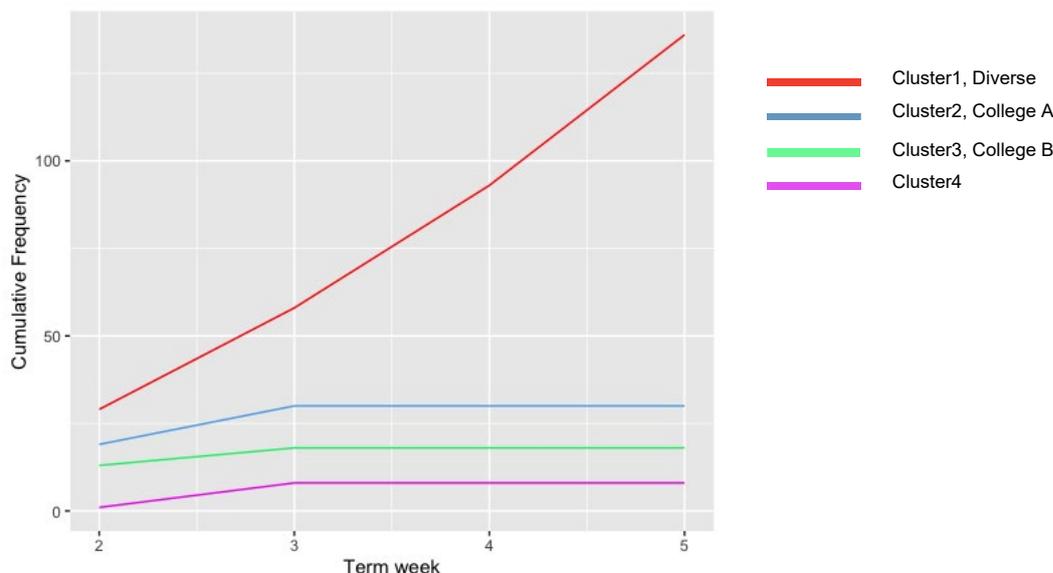| Week Commencing	Numbers sequenced/Positive Samples
28/09/2020	31/308 (10.1%)
05/10/2020	45/445 (10.1%)
12/10/2020	61/562 (10.9%)
19/10/2020	144/814 (17.7%)
02/10/2020	62/816 (7.6%)

Limitations

Due to incomplete sampling and quality control screening of sequences, not all positive University isolates were included in this analysis. This should be factored into interpretation of cluster size and subsequent epicurves. The proportion of University, Pillar 2 and Pillar 1 samples sequenced will differ and any comparison of absolute numbers should take this into account (Table 2). Furthermore, individuals affiliated with the University may have accessed the Test and Trace service and therefore provide a false signal for community transmission. Use of Pillar 1 samples may introduce a bias into any conclusions drawn regarding transmission between University and the community. Local sample selection in this preliminary analysis was limited by data availability. This will be addressed with the availability of further Cambridgeshire Pillar 2 samples in the COG-UK database. Attempts to place local isolates in the context of the national COG-UK database should factor in heterogenous sampling across the country; inferences drawn may be over-interpreted due to a sampling bias. This report is a descriptive account clusters based on phylogenetic analysis; with the current number of samples included, we believe it is not possible to draw strong conclusions of causal relationships between specific colleges or courses and cluster propagation. Finally, this analysis describes the genomic epidemiology of SARS-CoV-2 for the first five weeks of the academic term at the University of Cambridge; any generalisation of conclusions drawn should be tempered by the incomplete nature of the data and study setting.

Figure 1. Phylogenetic tree including University students and Local Community isolates from term weeks 2-5. Each individual case is represented by a "Node Leaf" (coloured dot), which are coloured by college affiliation (legend not provided) and community (lime green, "local", Legend 1). Bar on right of figure coloured by cluster (Legend 2). Positive inferences of transmission between the local community and University students through examination of University cases interspersed with community isolates should be tempered due to the limited diversity of SARS-CoV-2 across the UK. The phylogenetic tree includes University samples up to term week 5; clusters of SARS-CoV-2 infections identified are spread across time and inferences of super-spreading events should not be drawn.
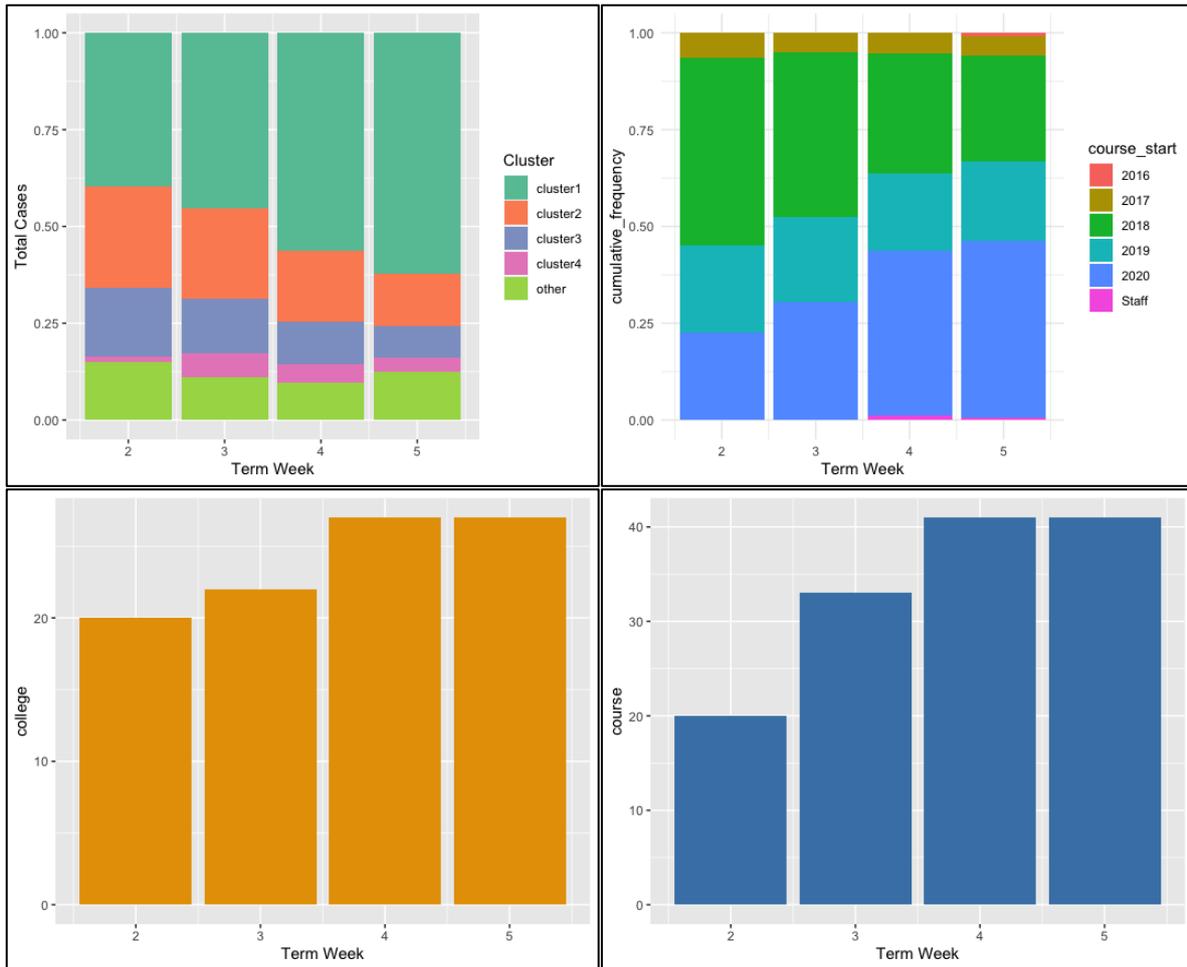
Figure 2. Epicurve demonstrating growth of 4 large clusters from term week 0 to 5

Addendum (dated 02/12/2020)*

This heterogeneous cluster (Cluster 1) is unlikely to be the result of multiple introductions of a similar phylotype for many reasons:

1. University students arrive from multiple regions in the UK and we have not observed multiple lineages propagating in our population of students. The likelihood of only one sub-lineage propagating is unlikely.

2. After acquisition of a limited number of term week 1 samples, we have identified the earliest University case corresponding to this cluster, which was sampled on the 8th of October. The earliest case in the COG-UK dataset of this phylotype (as of 17/11/2020) is dated 03/11/2020, and predominantly from a defined region within the UK rather than dispersed across the UK.

3. The cases of this phylotype in the COG-UK metadata are now dominated by Cambridge University samples, though there does now appear to be limited dissemination across the UK.

4. The earliest isolate, local to Cambridgeshire, that was sampled and sequenced through Pillar 2 testing is dated 19/11/2020. This provides some level of certainty this sub-lineage was not circulating in the local community prior to the outbreak observed within the University, though these conclusions must be tempered by bias introduced through discrepant sequencing coverage of University student isolates and local community isolates (Table 2).

5. Furthermore, we have now identified a possible seeding event that links cases in term week 1 samples with individuals subsequently testing positive in term week 2, distributed across Colleges and all linked to Cluster 1. This is through an ongoing epidemiological investigation and initiated after the identification of Cluster 1.

6. The demographic make-up of this cluster has become increasingly diverse over a period of time that has included restrictions of movement within the UK (Figure 3). We would suggest it is unlikely for students to be travelling outside of Cambridgeshire during term-time and expect limited non-essential visits; the increase in Colleges, courses and academic years involved in the cluster is therefore likely as a result of student-student transmission.

Figure 3. Progression of Cluster 1 from term week 2 to term week 5. *Top Left:* Growth of Cluster 1 when compared to all cases. *Top Right:* Distribution of cases in Cluster 1 by course start year. *Bottom Left:* Cumulative frequency of Colleges involved in Cluster 1. *Bottom Right:* Cumulative frequency of courses involved in Cluster 1