

USE OF SARS-CoV-2 GENOMICS TO UNDERSTAND TRANSMISSION

A report by the COG-UK Consortium. 28 June 2020

This paper supplements 9 reports submitted to SAGE in the last 3 months, which are available on the COG-UK website <https://www.cogconsortium.uk>

TRANSMISSION AND INTERNATIONAL TRAVEL

The COVID-19 epidemic experienced in the UK is an integral part of a Europe-wide epidemic. The pace and early distribution of the UK epidemic was predicated on the number of independent introductions from other European countries. Two viral genomic studies reported in the last month have described the introduction of SARS-CoV-2 into the UK and Scotland and the establishment of transmission lineages.

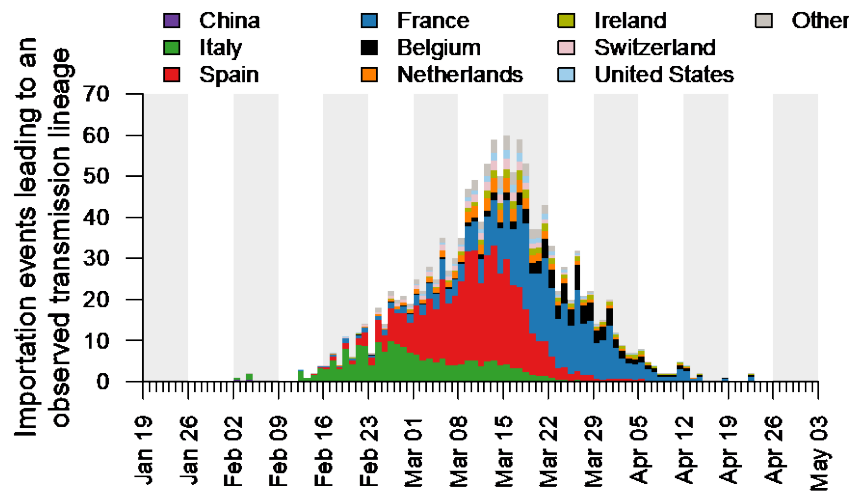
Ana da Silva Filipe, Emma Thomson and others undertook a study of SARS-CoV-2 spread in Scotland after the first case on the 1st March 2020. They sequenced 452 SARS-CoV-2 isolates, representing 20% of all confirmed diagnoses in Scotland up to 1st April 2020 (n=2310). Introductions were estimated based on a combined analysis of travel history, sampling date and genetic phylogeny.

From this they estimated that at least 113 separate introductions of SARS-CoV-2 occurred into Scotland within the first four weeks of the outbreak, although this will represent a significant underestimate of the actual number. A total of 86 independent clusters were identified, of which 48 (56%) contained multiple sequences indicative of onward transmission. Half of clusters did not contain a case with a known international travel history, indicating that the index case of the cluster was not detected. The first local transmission event was detected three days after the first apparent introduction, and a shift from travel-associated to sustained community transmission was evident after 11 days. The majority of viral sequences were most closely related to those circulating in other European countries, including Italy, Austria and Spain. The dataset included 20 viral sequences from lineage A, one of the primary lineages in China at the beginning of the outbreak, but now globally distributed. However, the introductions to Scotland of lineage A appear to have occurred via Spain, and cases with direct links to China and other countries in South-East Asia were rare.

Oliver Pybus & Andrew Rambaut and others undertook a preliminary analysis to estimate trends through time in the number and sources of SARS-CoV-2 introductions into the UK. They obtain these estimates by combining three data sources: (i) daily numbers of inbound travellers to the UK, (ii) estimated numbers of infections worldwide, and (iii) large-scale virus genome sequencing undertaken by the COG-UK consortium. The key conclusions of the analysis are that:

1. The UK epidemic comprises a very large number of importations due to inbound international travel. They detect 1356 independently introduced transmission lineages, but they expect this number to be an underestimate.
2. The rate and source of introduction of SARS-CoV-2 lineages into the UK changed substantially and rapidly through time.
3. The rate peaked in mid-March (median = 25th March) and most introductions occurred during March 2020 (interquartile range = 17th March-1st April).
4. They estimate that ≈34% of detected UK transmission lineages arrived via inbound travel from Spain, ≈29% from France, ≈14% from Italy, and ≈23% from other countries.
5. The relative contributions of the source locations rapidly fluctuated through time.
6. Many UK transmission lineages now appear to be very rare or extinct, as they have not been detected by genome sequencing for >4 weeks.
7. The increasing rates and shifting source locations of SARS-CoV-2 importation were not fully captured by early contact tracing.

The figure below shows the estimated number of importation events attributed to inbound travellers from different countries of embarkation.



In early March there was a high volume of arrivals into the UK, but countries from which most of these arrivals originated had comparatively small numbers of active infections. Towards the end of March, the situation was reversed with large epidemics in many countries but a low volume of international arrivals. The mid-March peak in importation occurred because moderate-to-high levels of inbound travel coincided with growing or peak case numbers in several European countries. Most early importations (before mid-Feb) likely originated from China and east/southeast Asia, but these constitute a very small fraction of all importation events that resulted in detectable UK transmission lineages.

At present, it is not possible to reliably determine the source location of SARS-CoV-2 transmission lineage using virus genomes alone, because (i) the virus exhibits insufficient genetic variation, and (ii) rates of genome sampling vary greatly among countries, which will bias reconstructions of lineage introduction. It is therefore necessary to combine multiple data sources to estimate the dynamics of this process. Further analyses are ongoing, but this preliminary analysis suggests that by combining multiple data sources it should be possible to rapidly evaluate future trends in virus introduction.

HOSPITAL OUTBREAKS

Numerous studies are underway on hospital outbreaks across COG-UK, and sequence data is being used in close to real time in numerous hospitals. The study described here is one of the first to be published.

A Rapid implementation of SARS-CoV-2 sequencing to investigate healthcare-associated COVID-19 infections at Cambridge University Hospitals has been accepted for publication in the Lancet Infectious Diseases. A pre-print of the article is available here:

<https://www.medrxiv.org/content/10.1101/2020.05.08.20095687v1>

Rapid SARS-CoV-2 nanopore sequencing was performed on PCR-positive diagnostic samples collected from Cambridge University Hospitals (CUH) and a random selection from the East of England (EoE), enabling sample-to-sequence in under 24 hours. A weekly review process was established with integration of genomic and epidemiological data to investigate suspected healthcare-associated COVID-19 cases, led by infection clinicians and supported by the local PHE Field Service.

Between 13 March and 24 April 2020, clinical data and samples were collected from 5,613 confirmed COVID-19 patients from across EoE. Sequencing was attempted on 1,000 samples, producing 747 high-quality genomes. Combined epidemiological and genomic

analysis was performed on the 299 cases from CUH (253 patients, 46 healthcare workers (HCWs)); 35 clusters of identical viruses were identified involving 159 cases, ranging in size from 2 to 18 cases. 92 cases (57.9%) had strong epidemiological links to support recent transmission and 32 cases (20.1%) had plausible epidemiological links. Transmission events were identified in 12 hospital wards and an outpatient dialysis unit. 9 of the hospital wards were "green" (i.e. non-COVID-19) at the onset of cluster cases, and 3 were "red" (i.e. confirmed or highly suspected COVID-19). Additionally, community transmission events were identified in 3 care homes (in both residents and carers), hostel accommodation and a number of households. Transmission events involving HCWs were identified in both hospital and community settings; for example, one cluster included several paramedics and CUH HCWs who live in the same accommodation. In addition, a number of cases previously thought to have been community acquired were linked to an outpatient dialysis unit where shared transport to and from the clinic was identified as a likely contributing factor. These results were fed back to clinical, infection control and hospital management teams, resulting in infection control interventions and informing patient safety reporting.

Sequencing provided greater understanding of transmission networks involving patients and HCWs, including the ability to rule out transmission events, and supported focused efforts at a time of unprecedented demand on infection control teams. Rapid turnaround of sequencing enabled its use in real-time infection control investigations. This approach enabled the detection of cryptic transmission events and identified opportunities to target infection control interventions to reduce further healthcare-associated infections. Sequencing has informed patient safety review processes, including investigations of possible hospital-onset COVID-19 where the patient became symptomatic within 14 days of admission and it was unclear if the virus was acquired in hospital or the community. Genomic data informed reviews of patient placement and isolation procedures, assessment of PPE use and staff working patterns and break arrangements.

CARE HOMES

One of the first studies to be completed in the UK by *Shamez Ladhani, Maria Zambon and others* describes an investigation in six London care homes reporting suspected COVID-19 outbreaks during April 2020. Residents and staff had nasal swabs taken for SARS CoV-2 testing using RT-PCR and were followed-up for 14 days.

They found that 107/268 (39.9%) residents and 51/250 (20.4%) staff were SARS CoV-2 positive. The 158 PCR positive samples were sequenced and 99 (68 residents, 31 staff) distributed across all the care homes yielded sufficient sequence data for analysis. This identified multiple independent introductions into each care home, rather than a single introduction followed by within-care home transmission. Several introduction events were followed by considerable within-care home transmission, although there were numerous instances where an introduction event was not followed by any detected forward transmission. Clusters commonly contained isolates from both residents and staff members, although it is not possible to infer directionality. There was no instance where clusters contained isolates from different care homes, refuting the suggestion that staff working in multiple care homes had transmitted SARS CoV-2 between different homes in this investigation.

EXAMPLES OF COG-UK INTEGRATION INTO NATIONAL STUDIES

COG-UK will sequence PCR-positive samples identified by the following national studies, the purpose of which is to identify points of failure in infection control that require changes in practice and policy in different settings.

1. VIVALDI, a study of more than 6500 care home staff and more than 4000 residents across >100 care homes in England (May 2020 – April 2021). Intensive PCR testing will be conducted in outbreak and non-outbreak settings. A combined epidemiological

and genomic analysis will identify patterns of transmission, and sequence data from patients in neighbouring hospitals and the surrounding community will indicate potential sources and connectedness of transmission pathways.

2. A study of the prevalence and incidence of SARS-CoV-2 infection in prisons in England. Ten thousand staff and 20,000 prisoners will be sampled every three weeks, between July and September. A combined epidemiological and genomic analysis will identify what activities are associated with introduction events, the effectiveness of cohorting strategies, and the effect of increasing flow of prison populations as courts become open, on transmission chains.
3. SIREN, which aims to determine if prior SARS-CoV-2 infection in health care workers confers future immunity to re-infection. Conducted in 40 NHS Trusts and involving 100,000 staff, the primary focus of the study is serological testing, but PCR-positive samples based on healthcare worker screening will be sequenced. This will provide evidence for transmission between staff.
4. Track and Trace. We are initiating a retrospective pilot study of people contacted through the Track and Trace programme, in which we will sequence positive samples and compare these with other cases and people in the locality who have been diagnosed elsewhere. This will provide insights into the accuracy of backward contact tracing.

FUTURE PLANS

We hypothesise that routine surveillance sequencing of all available SARS-CoV-2-positive samples from people with COVID-19 has the potential to detect a cluster or hotspot, possibly even before it has been detected by other means.

Cluster detection is vital, since suppression of clusters may have a significant impact on overall transmission, particularly if the severity of a superspreading event can be rapidly curtailed. The current approach for their detection is for public health agencies to monitor for clusters of cases, which is most readily achieved when this occurs in a specific location such as a health care facility or factory. However, genomic epidemiology represents a method of sentinel surveillance with the potential to detect clusters and transmission at the earliest opportunity. Our future plan is to evaluate the value of comprehensive, prospective sequencing of all available SARS-CoV-2 from affected people to support genomic epidemiological surveillance.

Detecting clusters (and excluding transmission when cases have come together by chance) represents an important way to prioritise and refine outbreak investigations, enabling these to concentrate effort (contact tracing and other interventions) on situations where they can make maximum impact in preventing the spread of the virus.

Combined epidemiology and genomic analyses are currently labour intensive and require scientific expertise. However, there is potential for greater automation of sequencing and analysis pipelines and their integration with hospital data and public health systems. We will also need to achieve efficient and rapid integrating of genomic, epidemiological and health informatics data.