COVID-19 Genomics UK (COG-UK) Consortium

Weekly Report #1   23rd March 2020

Executive Summary

- The COG-UK consortium has rapidly established a multi-agency national network to deliver coordinated large-scale sequencing and analysis of SARS-CoV-2 genomes.
- In the space of 12 days, the COG-UK centres already online have sequenced and analysed 260 SARS-CoV-2 genomes.
- Initial analyses indicate a large number of independent SARS-CoV-2 introductions to the UK, from multiple locations around the world.
- Enabling access to patient electronic health records to allow a detailed epidemiological analysis of the early spread of cases in the UK COVID-19 epidemic is a crucial next step.

About the COG-UK Consortium

The COVID-19 Genomics UK Consortium (COG-UK) – composed of the NHS, UK Public Health Agencies, multiple UK Universities, and the Wellcome Sanger Institute – has been established to deliver large scale, rapid sequencing of SARS-CoV-2 genomes.

In just 12 days, COG-UK has assembled a truly national network of sequencing centres and analysis groups located in Belfast, Birmingham, Cambridge, Cardiff, Edinburgh, Exeter, Glasgow, Liverpool, London, Norwich, Nottingham, Oxford and Sheffield. These initial locations are based on sequencing capacity that is immediately available; we envisage incorporating sequencing capacity at other locations in future.

A national operations hub and large-scale sequencing facility has been established at the Wellcome Sanger Institute and the University of Cambridge. It currently includes approximately 30 people drawn from Sanger core staff and has the potential to grow to meet demand.
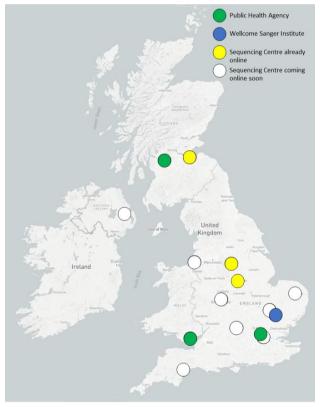

Figure 1: Location of current COG-UK sites.

The coordinated sequencing of samples from patients with confirmed cases of COVID-19, and subsequent analysis of genomic data by the COG-UK consortium will enable clinicians and public health teams to rapidly investigate clusters of cases, to understand how the virus is spread and to implement appropriate infection control measures.

Achievements to date

- A meeting at the Wellcome Trust in London on 11th March brought together key stakeholders to explore the opportunity to harness national capabilities to support genomic surveillance of COVID-19.

- Proposal submitted to the UK government on 16th March outlining the COG-UK consortium plan to deliver a large scale and rapid SARS-CoV-2 sequencing capacity to local NHS centres and the UK government at pace.

- By the data cut-off for this report on 20th March, COG-UK has sequenced, assembled, collated and analysed 260 complete SARS-Cov-2 genomes from 7 sequencing centres in the network. To place this in context, in just 11 days COG-UK generated more virus genomes than reported by any other country, other than China, during the entire epidemic to date.

Sequencing Centre	Number of SARS-CoV-2 genomes*
Edinburgh	32
Glasgow	27
Nottingham	1
Public Health England (Colindale)	44
Public Health Wales (Cardiff)	48
Wellcome Sanger Institute	95
Sheffield	13
TOTAL	260

Table 1: Number of SARS-CoV-2 genomes sequenced and analysed by the COG-UK centres already online. *by the data cut-off on 20/03/20

Country	Number of SARS-CoV-2 genomes by 11/03/20	Number of SARS-CoV-2 genomes by 22/03/20
China	206	280
UK	24	260
Netherlands	28	190
USA	70	182
France	10	61
Australia	18	30
Canada	3	17
Germany	3	15
Italy	5	5
Spain	0	3

Table 2: Number of SARS-CoV-2 genomes sequences by country on 11/03/20 and 22/03/20. Ten highest countries shown.

- Working groups have been established in five key areas:
  - Sample collection : finding effective ways of working with NHS and public health labs at both the local and national level.
  - Methodology: sharing and standardising protocols for sample processing and genome sequencing.
  - Metadata: : developing protocols and informatics solutions to link genomic data to clinical, epidemiological and biological data collected by the NHS, public health agencies and academic groups.
  - Analysis: inferring epidemiological processes from genome sequence data using phylogenetic trees and other computational methods.
  - Report writing: producing weekly reports based on the analytical findings and tailored for the needs of the NHS, public health agencies and government.

- A coordinated naming system for all sequenced UK samples has been developed and agreed.

- A digital analysis and data sharing environment has been established on MRC CLIMB (Cloud Infrastructure for Microbial Bioinformatics), where all samples were collated and analysed.

- A system for sharing visualizations of the relationships and distribution of samples has been deployed, using Microreact (Appendix Fig S3).

- Plans for the collation of virus genome sequence and linked, anonymised eHealth data have been initiated with HDR-UK.

- An initial outline plan for the types of required metadata has been developed and circulated to NHS partners for approval.

- A system for metadata submission and collection from the diverse range of organisations involved has been created using Epicollect.

- A comprehensive system for collaborative online working and teleconferencing has been established.

Questions COG-UK will address

We have identified multiple public health applications of large-scale SARS-CoV-2 genome sequencing:

1. **Enhance our understanding of epidemiology and transmission of SARS-CoV-2.**
   - Infer the relative contribution of local transmission versus imported cases.
   - Distinguish individual chains of transmission at a local level.
   - Estimate rates of epidemic growth, unreported cases and rate of sampling.
   - Determine patterns of within-country virus spread among locations and population sub-groups.
   - Provide insights into mechanisms and drivers of long-distance dispersal.

2. **Enable monitoring of interventions and treatment**
   - Monitor the effect of non-pharmacological interventions (such as travel restrictions, social distancing and isolation) on SARS-CoV-2 spread, epidemiology and biology.
   - Monitor changes in virus antigenicity and the emergence of resistance mutations in response to the introduction of vaccines and drugs.
   - Identify virus genetic markers associated with clinical severity to prioritize for functional investigation.
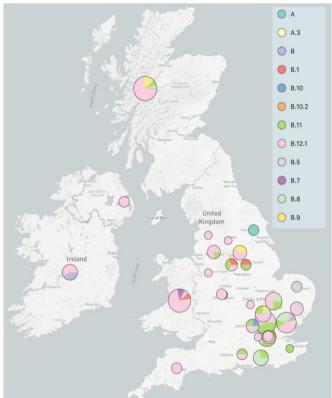
3. **Expand biological understanding and further research**
   - Assess the functional and phenotypic relevance of mutations.

Initial analytical findings

Only 12 days since the initial conception of the COG-UK consortium, the genomic data being generated and analysed have already allowed us to begin to address important questions regarding the spread of COVID-19 in the UK and beyond.

- We have identified 12 viral lineages in the 260 UK SARS-CoV-2 genomes sequenced to date. However, under-sampling in the UK and elsewhere means the number of independent introductions of SARS-CoV-2 is very likely substantially higher.

- The data are consistent with a large number of independent SARS-CoV-2 introductions to the UK, from multiple locations around the world.

- Major import to the UK seems to have occurred from locations with large epidemics and high travel volumes, notably Italy and other parts of Europe.

- Community transmission of multiple lineages/clusters following introduction is likely, as revealed by cases where genomes have come from patients with no known contact with an international source.
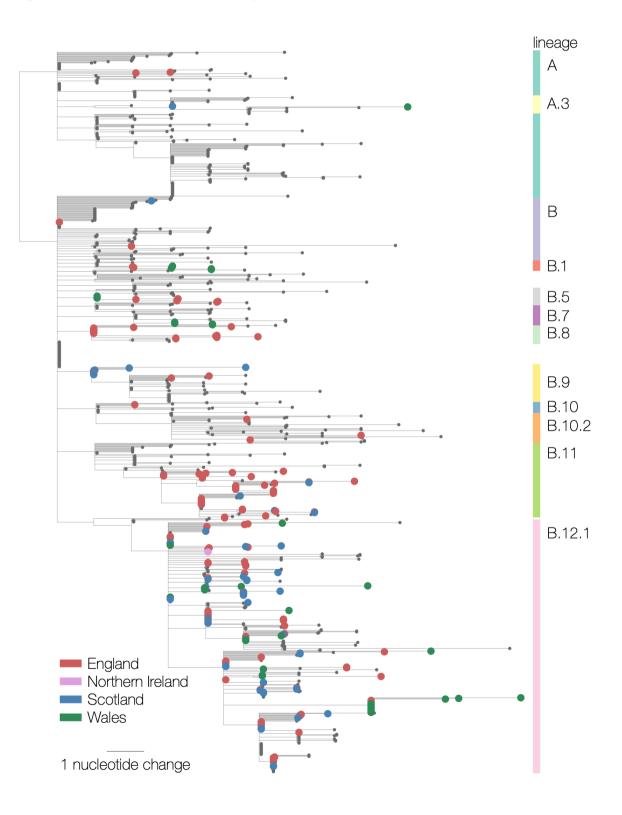

Figure 2: Geographic distribution of known SARS-CoV-2 lineages in the UK.
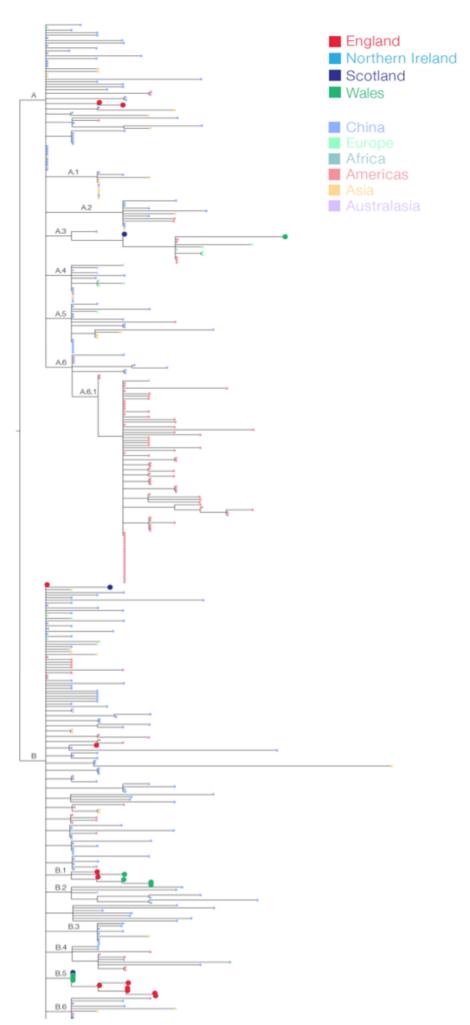
Recommendations

- A crucial next step for COG-UK is to access patient electronic health records to allow a detailed epidemiological analysis of the early spread of cases in the UK COVID-19 epidemic. This analysis will potentially provide important insight to help design future early interventions of viral introductions and transmission.

- Secondly, access to electronic health records will enable COG-UK to investigate any association between viral lineage, genotype and severity of disease.
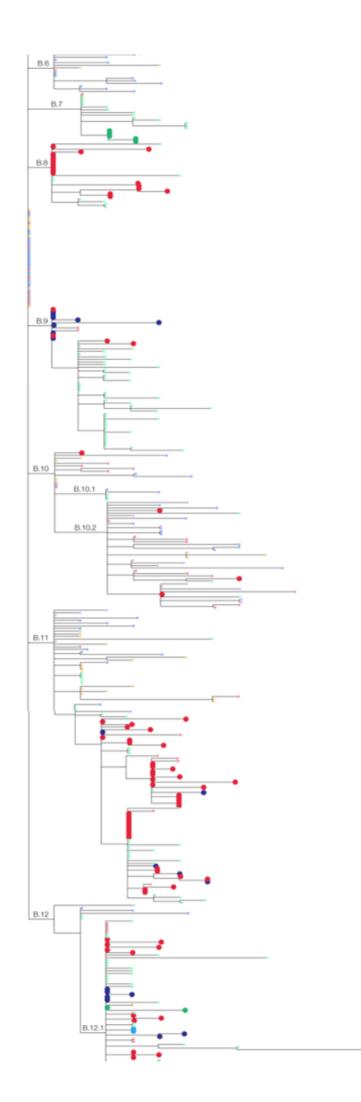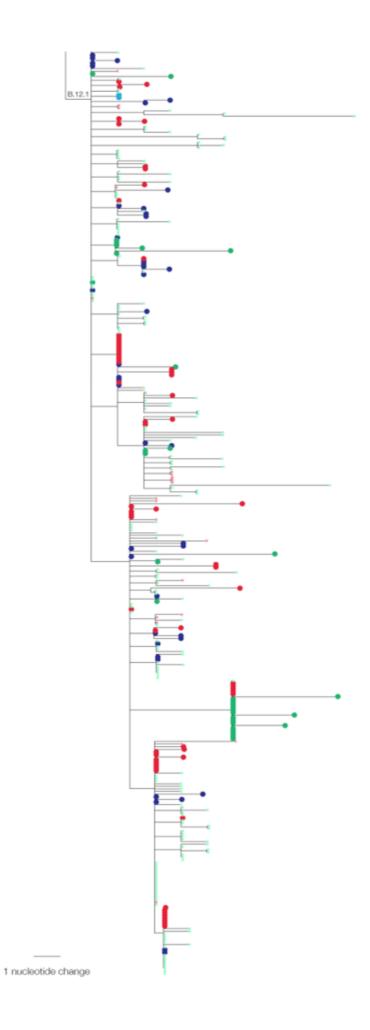
Appendix

Figure S1 | Phylogenetic tree of the first 260 UK genome sequences in the context of all global data. Larger circles denote cases from England (red), Northern Ireland (pink), Scotland (blue) and Wales (green). The distribution of UK cases across the entire global diversity reveals the many imports of the virus from across the world. Bars on the right denote the lineages with cases in the UK and correspond to those in Figure 2. The lineages B.11 and B.12.1 at the bottom of the tree are the predominant lineages circulating in Italy and continental Europe. Clustering of UK genomes together may be indications of community spread but this must be interpreted with caution as such groupings would also be expected as a result of travellers returning from common destinations.

Figure S2 | Detailed phylogenetic tree of all global SARS-CoV-2 genomes from GISAID as of 21-Mar-2020 combined with 260 UK virus genomes sequenced by COG-UK. Viruses are coloured by location of sampling and lineages are labelled using the same scheme as Figures 2 and S1.
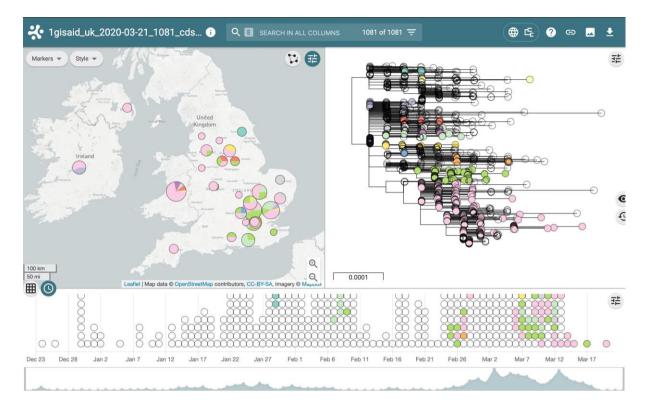
B.6

B.7

B.8

B.9

B.10

B.10.1

B.10.2

B.11

B.12

B.12.1

B.12.1

1 nucleotide change

Figure S3 | Data are live linked and visualised using Microreact (web application) enabling ongoing monitoring of trends across the consortium.