

Executive Summary

- COG-UK has brought online an additional sequencing centre (Exeter) and increased the number SARS-CoV-2 genomes sequenced and analysed to a total of 1679 (up from 806 on the 31st of March). The UK has now reported the largest number of genomes of any individual country in the pandemic to date, accounting for around one third of the global total.
- Roadblocks remain to access and integrate epidemiological data with viral genome sequence data for some regions of the UK, limiting the ability of COG-UK to realise the full potential of genomic surveillance to impact the course of the COVID-19 pandemic.
- Analyses are underway to develop weekly phylodynamic situation reports with outputs, such as reproduction number estimates, at the level of cities or local authorities, and to integrate aggregate mobility data with SARS-CoV-2 genomic data.

COG-UK update

In the past week, an additional sequencing centre at the University of Exeter has been brought on line, bringing the total number of already active COG-UK sites to nine. A further seven sequencing centres are expected to come online shortly (Figure 1).

Online sequencing capacity is sufficient to sequence ~9K samples per week, which currently exceeds ability to access positive samples.

The pipeline for downloading latest datasets from GISAID, quality control, integration with the latest UK data and installation on CLIMB is now essentially automated. With a firm cut off on Friday afternoon, the data and phylogenetic trees should now be available for annotation and subsequent analysis from Monday morning each week.

By the data cut-off for this report, the total number of viral genomes now stands at 1679 (Table 1). To place this in context, in the month that the consortium has been active to date, COG-UK has generated more genomes than reported by any other country during the entire epidemic to date, accounting for approximately one third of the total number of SARS-CoV-2 genomes reported (Table 2).

COG-UK has proposed (and adopted) a workable and practical nomenclature system for SARS-CoV-2 to describe virus lineages and to facilitate real-time genomic epidemiology by providing commonly-agreed labels to refer to viruses circulating in different parts of the world (Full report [here](#))

From the genomes available, we have identified at least 39 viral lineages that have been or are in circulation within the UK. Figure 2 below illustrates the duration, size, and geographic distribution of these lineages. As noted in previously, this is likely a substantial underestimate of the number of independent virus introductions and active transmission chains in the country

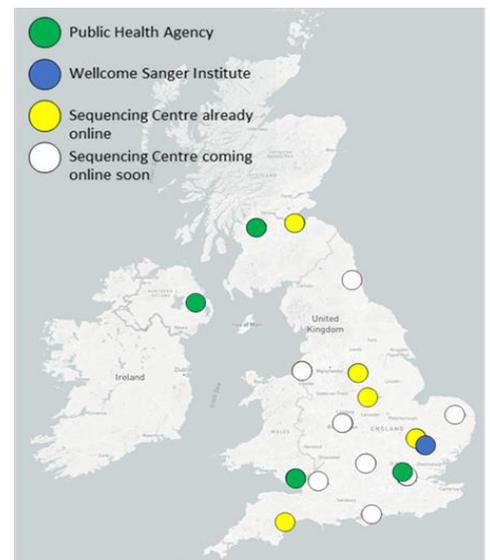


Figure 1: Location of current COG-UK sites.

Using data generated through COG-UK, we are working towards establishing a system for generating weekly phylodynamic (i.e. how epidemiological, immunological, and evolutionary processes act and interact to shape viral phylogenies) situation reports with outputs, such as reproduction number estimates, at the level of cities or local authorities.

Access to aggregated UK mobility data from 20th March onwards from the mobile phone operator O2 has been secured and work is underway to integrate this mobility data with SARS-CoV-2 genomic data to help discriminate among different scenarios of spatial spread when the genomic data are not sufficiently informative. An initial trial on a location with a high density of sampling and good metadata for this type of analysis is being considered.

Sequencing Centre	Number of SARS-CoV-2 genomes		
	22/03	31/03	06/04
Cambridge	-	33	33*
Edinburgh	32	86	223
Exeter	-	-	4
Glasgow	27	102	223
Nottingham	1	28	116
Public Health England (Colindale)	44	147	469
Public Health Wales (Cardiff)	48	218	298
Wellcome Sanger Institute	95	143	143*
Sheffield	13	49	190
TOTAL	260	806	1679

Table 1: Number of SARS-CoV-2 genomes sequenced and analysed by the COG-UK centres. *New genomes sequenced at these centres during this period were not reported ahead of the data cutoff, so will be included in the following week's numbers

Country	Number of SARS-CoV-2 genomes			
	11/03	22/03	31/03	06/04
UK	24	260	806	1679
USA	70	182	744	793
Australia	18	30	71	385
China	206	280	296	344
Iceland	-	-	343	343
Belgium	-	-	-	255
Netherlands	28	190	190	190
France	10	61	119	167
Canada	3	17	129	129
Japan	10	12	84	101
Hong Kong	-	-	-	89
Spain	-	3	42	89
Germany	3	15	27	51
Portugal	-	-	-	46
Finland	-	-	-	40

Table 2: Number of SARS-CoV-2 genomes sequences reported in GISAID. Fifteen highest countries shown.

Recommendations

- Building the infrastructure and embedding COG-UK personnel to access patient electronic health records and obtain the metadata that will enable our analyses to provide actionable insights remains a roadblock in some areas. To realise the full potential of COG-UK genomic surveillance and to begin to address pressing epidemiological questions, it is imperative that these barriers are surmounted with a high priority.
- Sequencing capacity now outstrips the ability to access positive samples. Further work to enable the smooth transfer of samples from NHS laboratories to COG-UK sequencing centres (for instance, direction that material transfer agreements are not essential) needs to be prioritized.

Analysis Summaries

This weekly report focuses on providing an update on COG-UK progress, reporting key numbers and statistics, and highlighting specific bespoke analyses that are being undertaken. This section will provide brief summaries of these analyses. In-depth reports will be submitted separately, where appropriate.

Visualising transmission chains

As more genomic data becomes available, our analyses are able to map the duration, size, and geographic distribution of the SARS-CoV-2 lineages circulating in the UK and to reconstruct transmission chains. As can be seen in Figure 2, we can identify at least 39 lineages that have been or are currently circulating. Some of these lineages are probably no longer in circulation (as demonstrated by the lack of genomes from later time points), although additional data in the coming weeks will be required to confirm this.

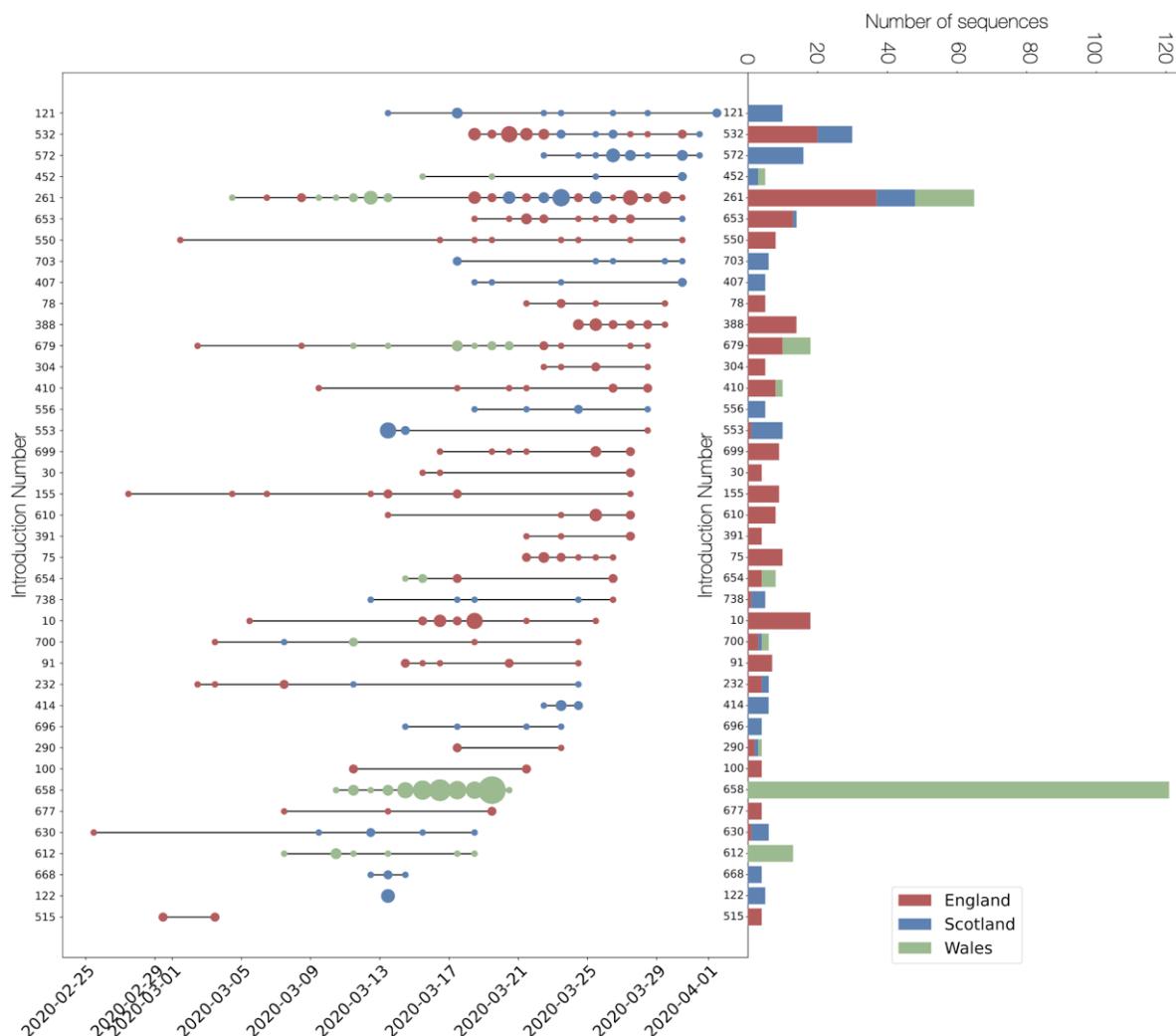


Figure 2 | Reconstructed UK transmission lineages with at least 5 viral genomes. Each row represents a single introduction and subsequent UK spread reconstructed from the virus genomes captured in the COG-UK sample in combination with global genome data. The diameter of the circles represent the number of virus genomes for the day of sampling. The bars denote the number of virus genomes in that lineage.

Appendix

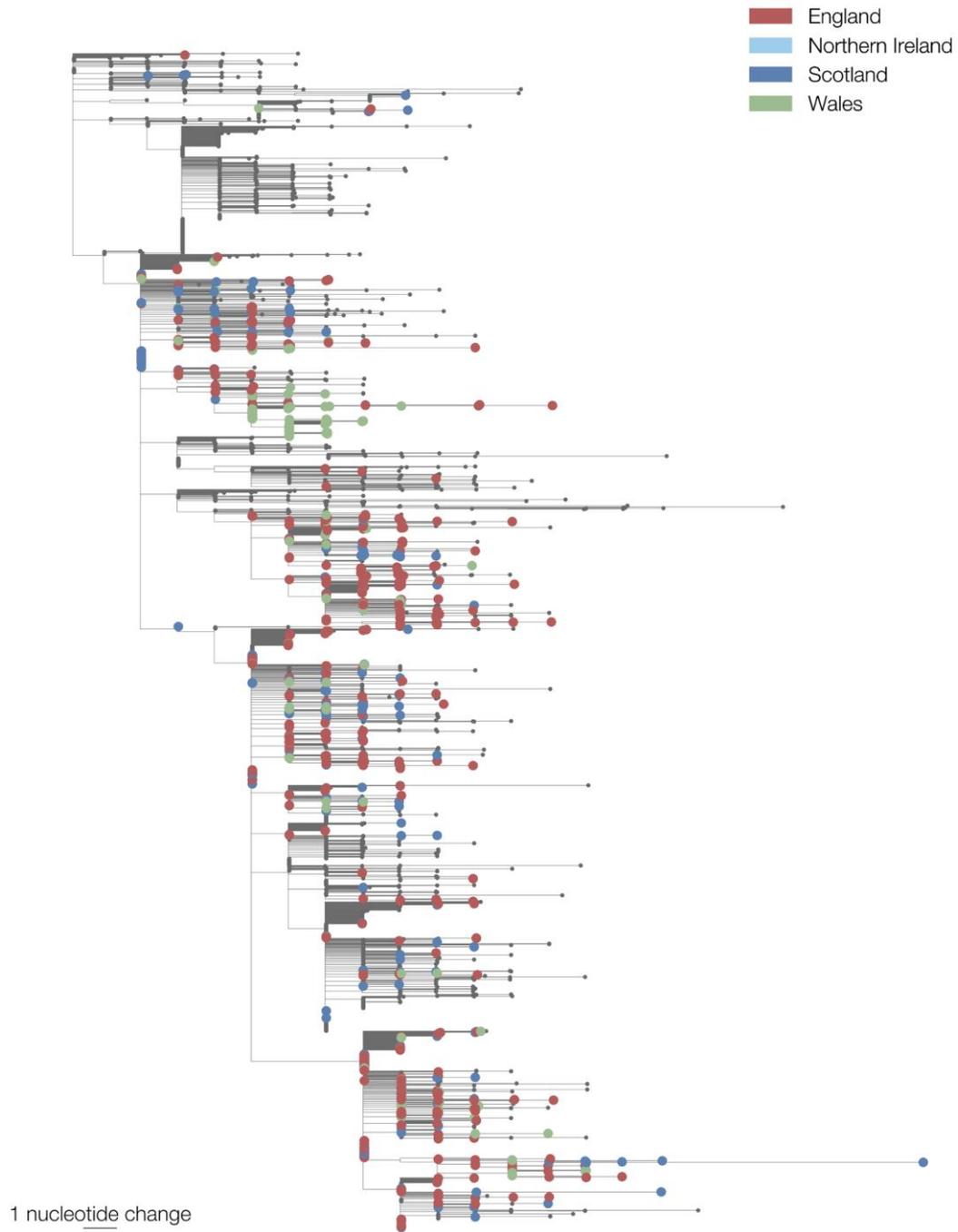


Figure S1 | Phylogenetic tree of the first 1679 UK genome sequences in the context of all global data. Larger circles denote cases from England (red), Northern Ireland (pink), Scotland (blue) and Wales (green). The distribution of UK cases across the entire global diversity reveals the many imports of 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 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.

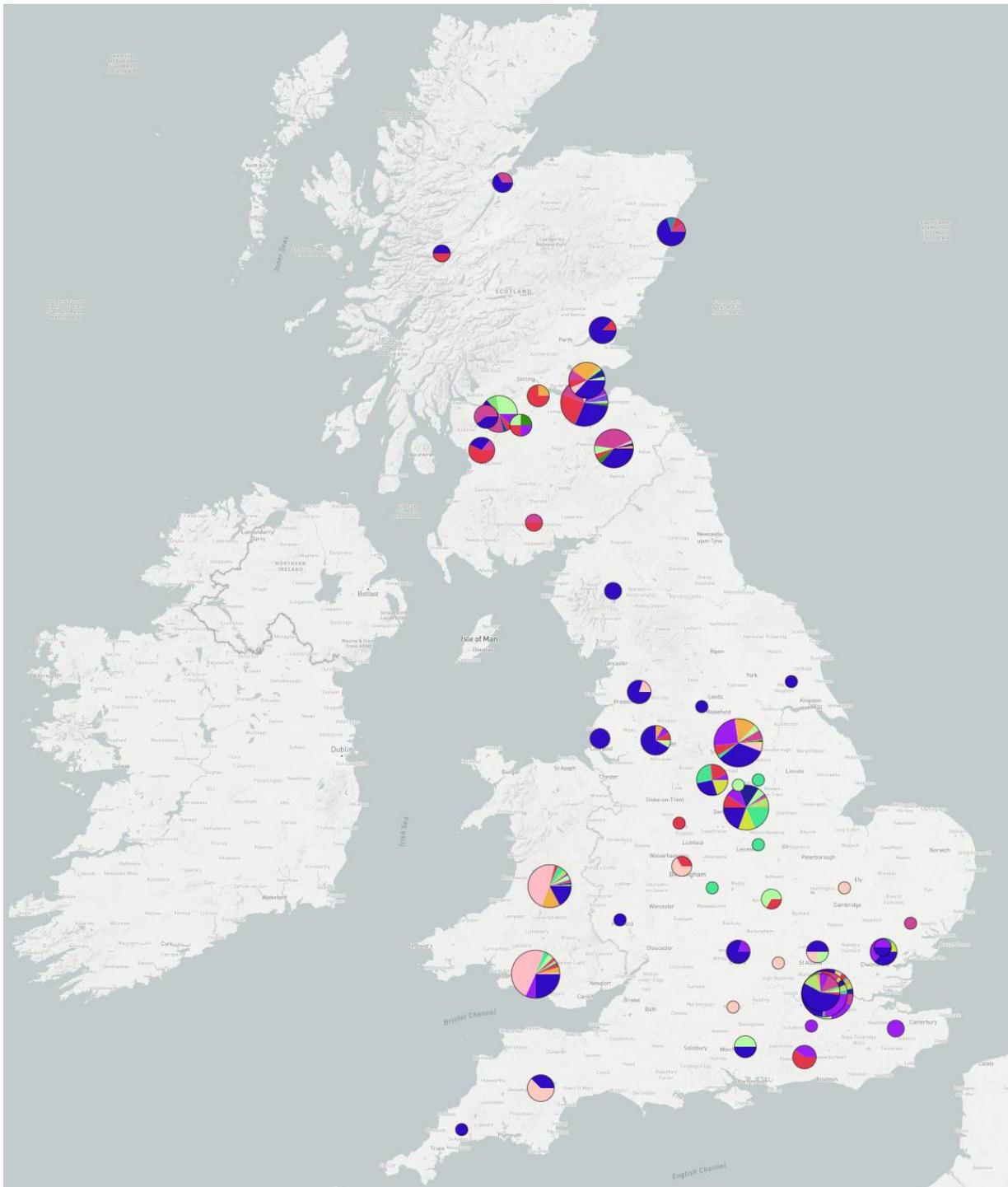


Figure S2 | Geographic distribution of known SARS-CoV-2 lineages in the UK - Pie charts summarise distribution of lineages at each location.

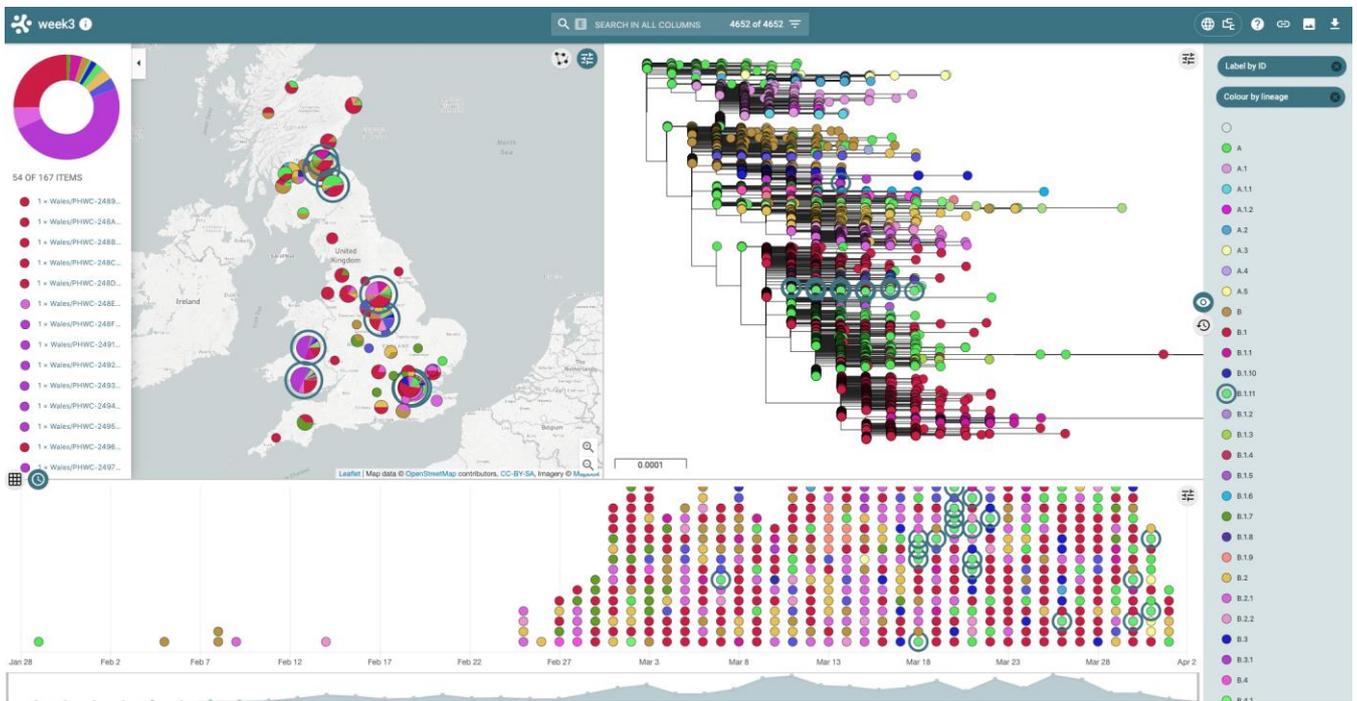


Figure S3 | Current live view of data linked and visualised using Microreact - geographic and temporal patterns in lineages can be shared and investigated. Figure above highlights geographic and temporal distribution of Lineage B.1.11