COVID-19 Genomics UK (COG-UK) Consortium

Report #10 - 11th August 2020

This report is provided at the request of SAGE and includes information on the ongoing state of the research being carried out. It should not be considered formal or informal advice. The conclusions of the ongoing scientific studies may be subject to change as further evidence becomes available and as such any firm conclusions would be premature.

Executive Summary

- In just 5 months, COG-UK has sequenced and analysed more than 40K SARS-CoV genomes, accounting for ~50% of the global total.

- Beyond generating this unprecedented viral genomic dataset, COG-UK has already had a substantial impact on national and global efforts to understand and tackle the SARS-CoV-2 pandemic, as demonstrated by the suite of dedicated tools developed by COG-UK researchers, the growing list of ground-breaking publications using COG-UK data and tools, and by the increasing focus on integrating genomic insights into infection control decisions.

- Genomic analyses have demonstrated that cases of shock and multisystem inflammation in children positive for SARS-CoV-2 (known formally as paediatric inflammatory multisystem syndrome temporally associated with SARS-CoV-2 (PIMS-TS)) are not associated with specific polymorphisms in any viral gene.

COG-UK update

Across 19 sequencing sites, the total number of SARS-CoV-2 genomes now sequenced by COG-UK stands at 40,035 (Figure 1a) which constitutes ~50% of the global total number of SARS-CoV-2 genomes sequenced (Figure 1b).

During 5 months since the inception of COG-UK, not only have an unprecedented number of viral genomes been sequenced and analysed, but the organisational, regulatory, laboratory and bioinformatic workflows, pipelines, tools and documentation underpinning this massive effort have been established, adapted or updated. Akin to assembling an aeroplane from parts while already in the process of taking off, hundreds of COG-UK researchers and individuals in support roles have maintained a furious pace to achieve milestones that under different circumstances would have been several years in the making. While COG-

UK sequencing and analyses continues unabated, the advent of summer has allowed individual consortium members to take some time to recharge ahead of the next stages to come in the autumn and winter.
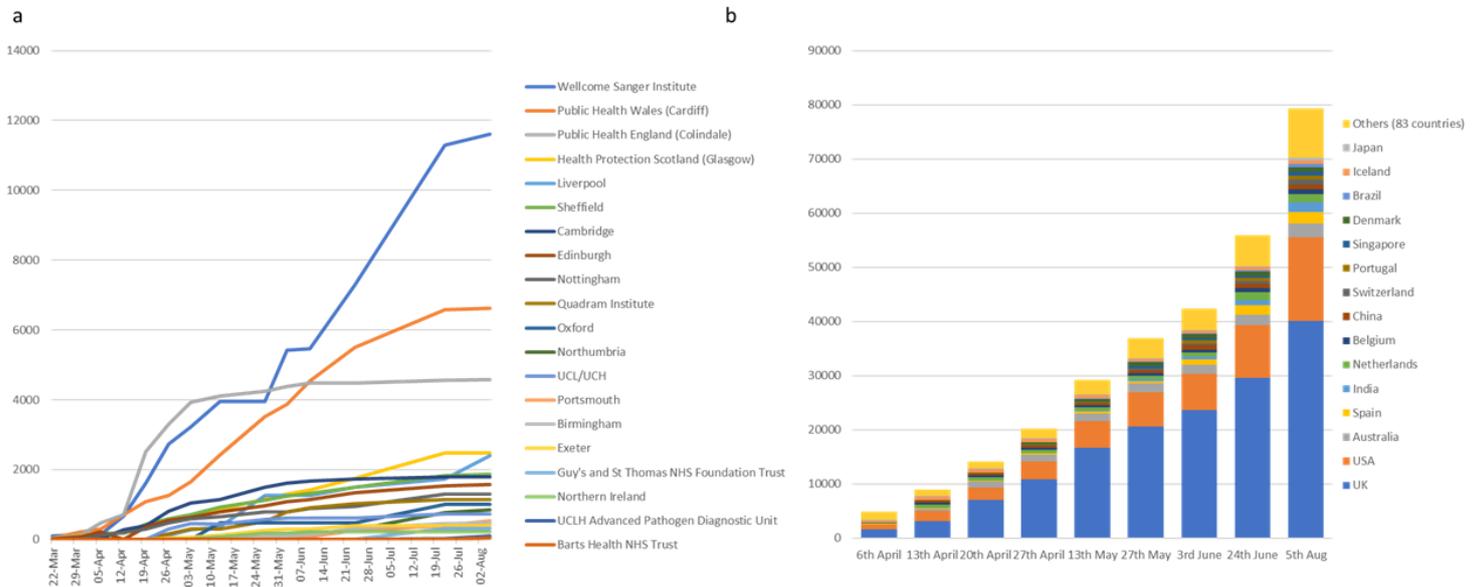

Figure 1: **a)** Number of SARS-CoV-2 genomes sequenced and analysed by the COG-UK centres by date. A total of 40,035 genomes have been sequenced across all COG-UK centres. Note that each sequencing centre has a different quota. **b)** Number of SARS-CoV-2 genomes sequences reported (in MRC CLIMB and GISAID). Data shown up to 5th August.

Recent improvements in sample handling and sequencing pipelines have enabled all positive Lighthouse samples to be processed as they arrive on site at the Wellcome Sanger Institute, which is a major milestone in our strategy for achieving prospective outbreak identification. Accordingly the use of COG-UK data is increasingly being integrated into the decision making process adopted by local and regional infection control teams. Work will continue to further reduce the time taken from a positive test at a diagnostic lab to QC genome data being made available for analysis and to better support infection control teams in generating and using genomic data.

A suite of dedicated bioinformatic platforms, pipelines and tools have been developed by COG-UK researchers (see 'Summary of tools and pipelines developed by COG-UK' below). In addition to forming the basis through which COG-UK viral genome data is analysed in the UK, many of these tools have become the gold standard approaches adopted by individual researchers and genomic consortia in countries around the world.

COG-UK researchers lead the world in applying genomic surveillance to understand the dynamics of the SARS-CoV-2 pandemic and to identify opportunities for intervention, resulting in the publication (or pre-printing) of 10 studies so far, with many more in the pipeline (see 'Summary of COG-UK publications' below).

In addition to these and other studies being undertaken, a major COG-UK initiative, the Hospital Onset Covid-19 Infection (HOCI) study, is now underway. The goal of HOCI is to determine empirically the impact that integrating real-time genomic sequencing can have on decision making by infection control teams (see 'Progress update on COG-UK Hospital Onset Covid-19 Infection (HOCI) Study' below).

Highlighted findings with public health implications

- Despite speculation to the contrary, a study of SARS-CoV-2 genomes from 61 children hospitalized for COVID-19 in London between late-March and mid-May 2020 found no evidence for specific single nucleotide polymorphisms to be associated with cases of paediatric inflammatory multisystem syndrome temporally associated with SARS-CoV-2 (PIMS-TS).

Analysis updates

Microreact visualization of SARS-CoV-2 lineage distribution

The latest data release into Microreact (Figure 2) shows 32,701 UK SARS-CoV-2 genomes highlighted on a tree of 69,358 global SARS-CoV-2 genomes. Lineage distributions are highlighted as pie charts on the map. Also shown is the global distribution on the spike protein residue 614 variants.
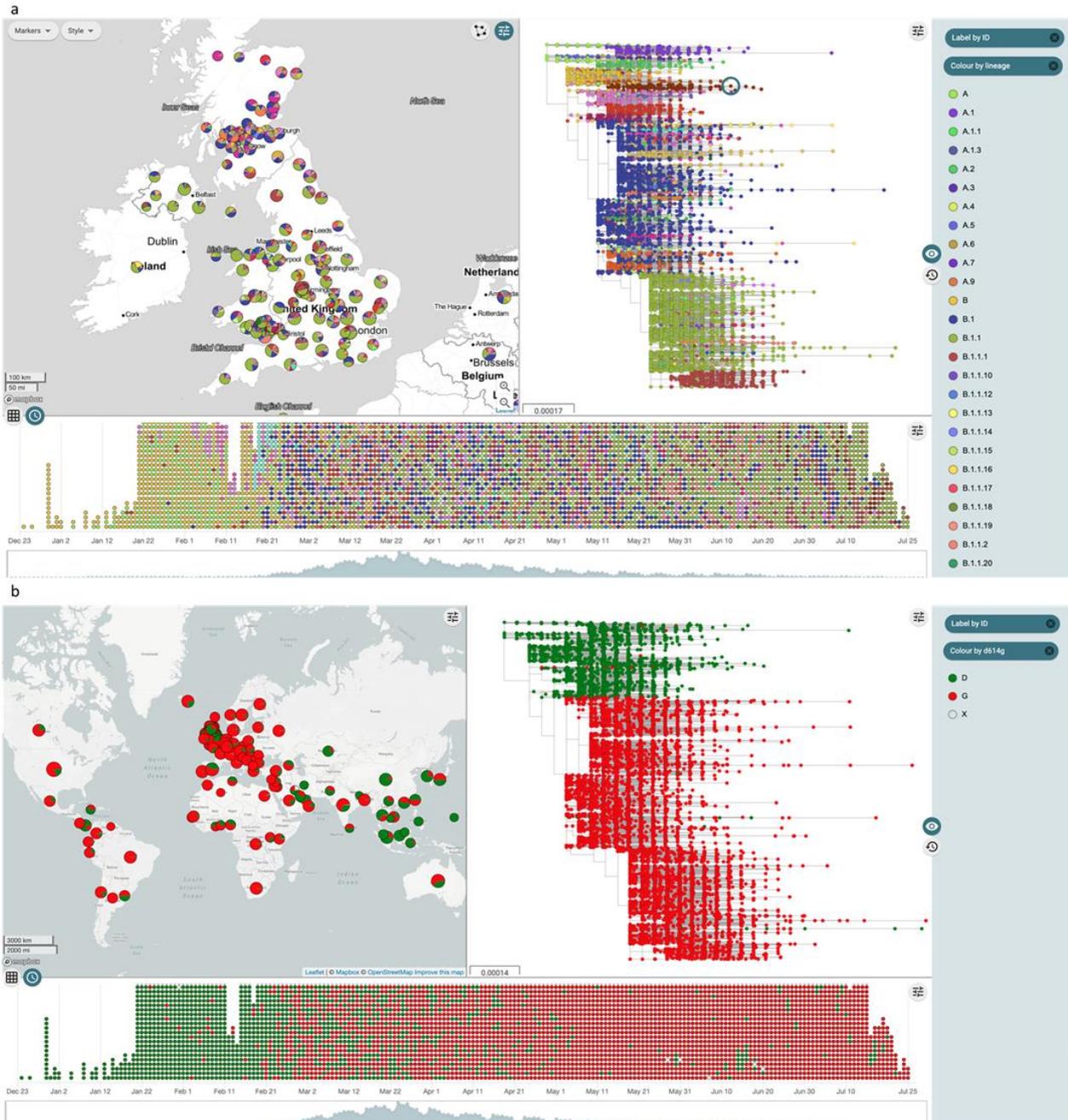

Figure 2: Data linked and delivered through Microreact. a) Distribution of lineages are indicated by location and UK lineages are contextualised within the global phylogenetic tree. The lower timeline can be used to investigate spread and location of lineages over time. b) Global distribution of the spike protein 614D (green) and 614G (red) variants from the first to last genome within the global dataset. https://microreact.org/project/cogconsortium.

No evidence of viral polymorphisms associated with Paediatric Inflammatory Multisystem Syndrome Temporally Associated With SARS-CoV-2 (PIMS-TS).

Study leads

Judith Breuer, Juanita Pang, Florencia A.T. Boshier (UCL Great Ormond Street Institute of Child Health, University College London), Nele Alders, Garth Dixon (Great Ormond Street Hospital for Children NHS Foundation Trust),

Question addressed

To determine whether cases of shock and multisystem inflammation in children positive for SARS-CoV-2 (known formally as paediatric inflammatory multisystem syndrome temporally associated with SARS-CoV-2 (PIMS-TS)) are associated with specific polymorphisms in any viral gene.

Methodology

SARS-CoV-2 genomes were sequenced from 61 children hospitalized for COVID-19 in London between late-March and mid-May 2020, 36 of which were diagnosed with PIMS-TS, 11 of which were positive for SARS-CoV-2 viral RNA. Reads were quality checked and mapped to the SARS-CoV-2 reference genome prior to phylogenetic analysis.

Findings

A maximum likelihood phylogeny constructed using the paediatric COVID-19 cohort together with 130 SARS-CoV-2 sequences from community cases across North London revealed no clustering of viral sequences from PIMS-TS patients or non-PIMS-TS patients.

No single nucleotide polymorphisms (SNPs) were observed to be unique to genomes from PIMS-TS or other childhood cases and there was no difference in SNP distribution in the PIMS-TS, non-PIMS-TS or community cases. Looking at SARS-CoV-2 viral spike (S) protein in particular, which has previously been suggested as having the potential to drive the development of PIMS-TS, all genomes carried the D839 and A831 variants, while the majority of PIMS-ST (3/5), non PIMS-ST (6/8) and community cases (118/130) were 614G positive.

Key conclusions

Overall, the data suggest that viruses causing PIMS-TS are representative of locally circulating SARS-CoV-2 and that there is no evidence for an association of PIMS-TS and the presence of new or unusual sequence polymorphisms.

Progress update on COG-UK Hospital Onset Covid-19 Infection (HOCI) Study

Study leads
Judith Breuer (UCL Great Ormond Street Institute of Child Health, University College London)

Question addressed
Hospitals and healthcare settings have been shown to have a major role in the spread of SARS-CoV-2, including into the community. While rapid and repeated testing of patients and staff is key to identifying potential hospital outbreaks, it cannot tell us which infections are actually linked and therefore where to focus infection prevention and control (IPC) measures. In particular where the number of infections is also high in the community, it is difficult to distinguish hospital and community acquired infections using standard IPC[1-3]. Sequencing of pathogen genomes to identify closely matched sequences has been repeatedly shown to better identify hospital outbreaks than standard IPC measures alone[1-3]. Most studies are retrospective and therefore cannot easily interrupt evolving outbreaks. Recently however, rapid SARS-CoV-2 sequencing has been reported as able to identify sources of hospital outbreaks fast enough to change ongoing IPC practice and reduce further spread[4]. COG-UK Hospital Onset Covid Infection (COG-HOCI), a phase three interventional clinical trial aims to quantify just how useful rapid pathogen sequencing can be for IPC practice, by measuring changes in IPC, if any, prompted by the receipt of rapid SARS-CoV-2 sequencing results and determining whether this in turn reduces nosocomial spread of infection. The study is Urgent Public Health adopted.

Methodology
COG-UK HOCI will involve over 20 NHS hospitals across the UK. All hospitals are linked to a COG-UK sequencing hub and will open as trial sites when evidence of increasing SARS-CoV-2 infection is reported. The clinical trial, which is supported by the UCL Comprehensive Clinical Trials Unit (CCTU) will use standardised Case Report Forms to measure the impact of delivery of a genomic sequencing data report to the infection prevention and control team, either within 48 hours or >4 days (to simulate a centralised sequencing facility) of sampling. The primary outcome measures will be:

1. Whether rapid (<48 hours) availability of SARS-CoV-2 sequence data reduces the incidence of IPC-defined HOCIs compared with delayed (>4 days) or standard of care (i.e. no sequence data).
2. Whether rapid genome data can identify previously undetected nosocomial transmission and how this compares with delayed (>4 days) or no sequence data.

In addition we will use qualitative research methods to analyse the acceptability and feasibility of having SARS-CoV-2 sequence data for routine IPC practice and what would be required if routine sequencing for IPC were to be introduced across the NHS. Finally, COG-UK HOCI will carry out a health economic evaluation of sequencing for SARS-CoV-2 IPC.

Progress
- Full ethical approval.
- Urgent Public Health status.
- Completion of 7 case report forms.
- Completion of a sequence reporting tool (separately described, below).
- Completion of an implementation guide based on Behaviour Change Techniques, highlighting how best to ensure viral sequencing and the SRT delivers positive outcomes in relation to reducing COVID-19 other outbreaks.
- Completion of site set up protocols.
- Ongoing setup of hospitals and sequencing hubs across the UK*.
- Site initiation visits booked at lead sites in Sheffield Glasgow, London (IHT &GSTT).

*Portsmouth, Southampton, Cardiff, Swansea, Glasgow, Edinburgh, Liverpool, Leeds, Sheffield, Manchester, Stoke, Nottingham, Birmingham (2 sites), London (Imperial Healthcare Trust, Guys and St Thomas trust, UCLH, Royal Free hospital Trust, Barts Healthcare Trust, St George's Hospital Trust), Exeter.

Development of a Sequence Reporting Tool

Authors
Oliver Stirrup* (University College London), Josh Singer* (CVR Glasgow), Joseph Hughes (CVR Glasgow), Matt Parker (The University of Sheffield), David Partridge (Sheffield University Trust), Asif Tamuri (University College London), James Blackstone (University College London), Thushan de Silva (The University of Sheffield), Emma Thomson (The University of Glasgow), Judy Breuer (UCL Great Ormond Street Institute of Child Health, University College London).

Previous studies where sequence data have been used to inform IPC activity have relied on identifying putatively linked sequences from phylogenetic trees and then gathering metadata on the relevant patients to determine if it supports the observed linkage[4]. This process can delay full interpretation of sequence data for several days. In addition, the low substitution rate of SARS-CoV-2 means that identical sequences may not represent recent transmissions or hospital acquired infections, particularly if a sequence genotype in question is common in the community and or hospital. To overcome these barriers, we have developed a Sequence Reporting Tool (SRT).

The HOCI SRT uses a probability model to integrate viral sequences, patient metadata and sequence data on other patients in the ward, hospital and community, providing a likelihood that the "focus" HOCI patient acquired the virus on their current ward or elsewhere in the hospital. The HOCI SRT computes a report that can be generated as soon as the sequence data are available within 48 hours. The HOCI-SRT will enable rapid:

1. Refuting of IPC-identified ward outbreaks where the HOCI patient and other patient sequences differ.

2. Confirmation that a HOCI is part of an ward outbreak where HOCI patient's sequence is closely matched to others on the ward.

3. Identification of healthcare workers potentially linked to the HOCI focus case and other closely matched cases.

4. Identification of other patients in the hospital potentially linked to the HOCI focus case.

5. Identification of any other patients and staff with closely matched sequences who have been collocated with the HOCI focus case at any time within the previous three weeks.

The report generated by the SRT will provide a concise and easily interpreted summary of salient data regarding other sequences that could plausibly form an outbreak cluster within the ward or hospital that includes the focus case. The report will allow IPC decisions to be implemented quickly and thus maximise opportunities to use the sequence data to interrupt spread of infection. Preliminary results from a pilot of the HOCI-SRT using data from Glasgow and Sheffield has confirmed its favourable performance in relation to standard methods (paper in preparation).

Linking of the HOCI SRT to electronic patient records to incorporate detail of patient locations and patient and staff movements will further improve the probability readouts. Modification of the HOCI SRT for early identification of outbreaks in care homes and the community is planned.

1. Roy S, Hartley J, Dunn H, Williams R, Williams CA, Breuer J. Whole-genome Sequencing Provides Data for Stratifying Infection Prevention and Control Management of Nosocomial Influenza A. *Clin Infect Dis.* 2019;69(10):1649-1656.

2. Brown JR, Roy S, Shah D, et al. Norovirus Transmission Dynamics in a Pediatric Hospital Using Full Genome Sequences. *Clin Infect Dis.* 2019;68(2):222-228.

3. Houldcroft CJ, Roy S, Morfopoulou S, et al. Use of Whole-Genome Sequencing of Adenovirus in Immunocompromised Pediatric Patients to Identify Nosocomial Transmission and Mixed-Genotype Infection. *J Infect Dis.* 2018;218(8):1261-1271.

4. Meredith LW, Hamilton WL, Warne B, 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.* 2020.

Summary of major tools and pipelines developed by COG-UK

In 5 months, COG-UK has sequenced more than 40,000 SARS-CoV-2 genomes, more than half of the global total. By comparison, the largest previous dataset for real-time virus epidemiology was ~1500 genomes from the West African Ebola outbreak, which were sequenced over the course of several years. Analysing genomic surveillance data and metadata on this scale and using it in real-time to inform disease control interventions is highly complex and entirely unprecedented. The task of grappling with this volume of genomic data has required COG-UK researchers to rapidly retool existing data management platforms and analytic approaches, as well as to develop entirely new pipelines and methods.

Below is a summary of the major data pipelines and analytical tools developed (or adapted) as part of COG-UK to date (See Figure 3 for a summary diagram of the relationship between the described tools). All of these tools are open source, in line with COG-UK's commitment to open science, and sharing all data that we can as rapidly as possible. Most of the tools developed are currently hosted on MRC-CLIMB (Cloud Infrastructure for Microbial Bioinformatics).

COG-UK data and tools are being increasingly used in the UK and globally both for retrospective academic studies and to enable real-time genomic epidemiology to inform infection control strategies.

Majora (Malleable All-seeing Journal Of Research Artifacts)

https://majora.covid19.climb.ac.uk/public/dashboard

https://github.com/SamStudio8/majora

Sam Nicholls (University of Birmingham)

Majora is a laboratory information management system (LIMS) developed as part of COG-UK. Majora enables information on samples and digital files to be stored together and can reconstruct the journey that a sample has taken from tube check-in at the lab through to data upload to a public database. The system uses a polymorphic artifact and process model, allowing for flexibility to store almost any metadata about any artifact. The initial version of Majora was online within 3 days, within 3 weeks it became the central repository for information about any sample within COG. Users upload metadata about their samples and sequencing to Majora which then makes that information available to all users and analysts in the consortium. The pipelines responsible for inbound and outbound data distribution link into Majora to find out what new samples to process and release.

Since March, Majora has been populated with over 40,000 samples and has served nearly 750,000 API requests within the consortium.

Elan

https://github.com/SamStudio8/elan-nextflow

Sam Nicholls (University of Birmingham)

Elan is a reproducible workflow for enumerating and incorporating SARS-CoV-2 samples to ensure that files uploaded from across the consortium are valid. Elan is also responsible for conducting quality control and organising valid "artifacts" (consensus viral sequences and aligned reads) for use by analysts in downstream pipelines.

Metadata uploader

http://metadata.cog-uk.io/

Anthony Underwood, Ben Taylor, Khalil Abudahab, David Aanensen (The Centre for Genomic Pathogen Surveillance, University of Oxford)

The metadata uploader is a standalone web application that allows COG-UK members to easily populate Majora with data, through dragging and dropping metadata files containing information about sequencing and metadata. Every sample sequenced and analysed by COG-UK has to go through Majora, Elan and the Metadata uploader, making them central to COG-UK data handling.

CoV-GLUE

http://cov-glue.cvr.gla.ac.uk/

Josh Singer, David Robertson (MRC-University of Glasgow Centre for Virus Research)

CoV-GLUE is a publicly-accessible web application for the interpretation and analysis of SARS-CoV-2 genome sequences. CoV-GLUE is based on the GLUE software framework and is enabled by data from GISAID. It allows users to browse a database of amino acid replacements and coding region insertions and deletions observed in SARS-CoV-2 genome sequences from the pandemic. CoV-GLUE also allows users to analyse their own SARS-CoV-2 sequences by submitting them to the web application to receive an interactive report.
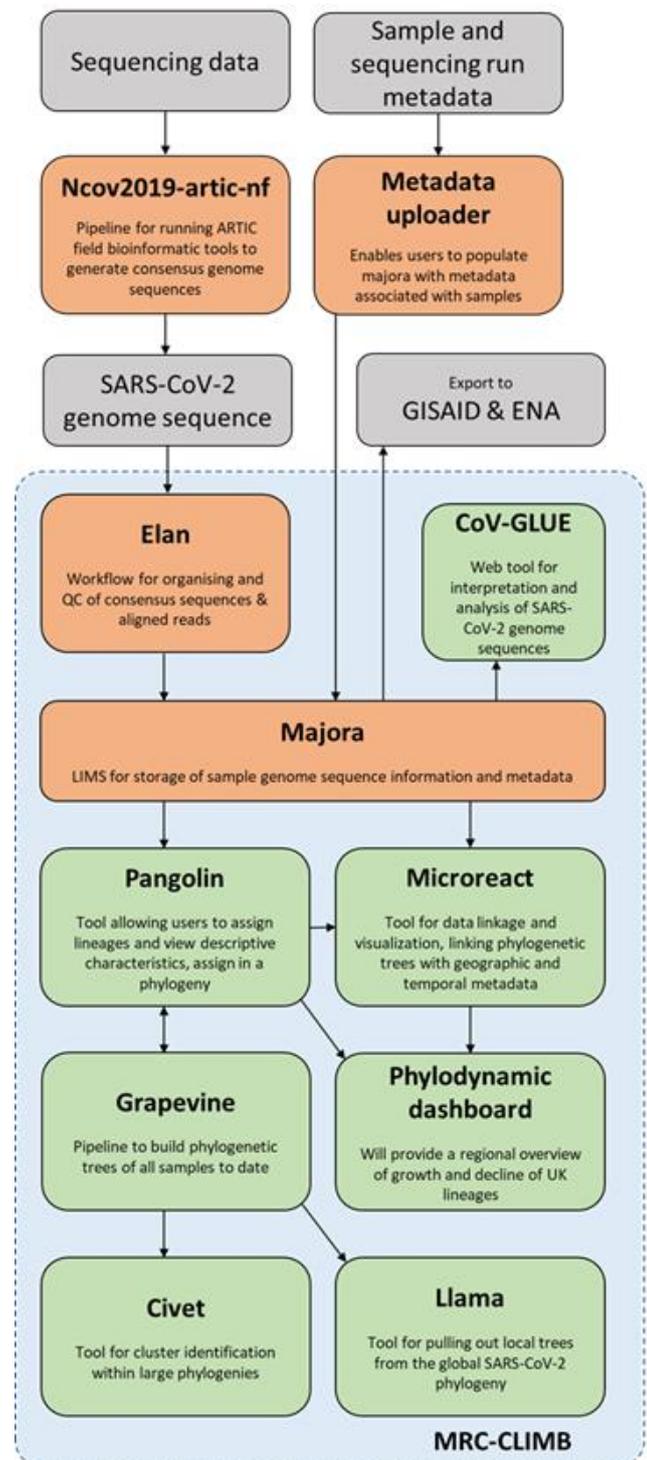

Figure 3: Major pipelines and tools developed by COG-UK for import and storage (orange) as well as analysis and visualization (green) of SARS-CoV-2 sequence data and metadata (grey). All tools are open source, many sit within the MRC-CLIMB infrastructure.

Pangolin (Phylogenetic Assignment of Named Global Outbreak LINeages)

https://pangolin.cog-uk.io/

Áine O'Toole, Verity Hill, JT McCrone, Emily Scher, Ben Jackson Andrew Rambaut (University of Edinburgh), Khali Abu-Dahab, Ben Taylor, Anthony Underwood, Corin Yeats and David Aanensen (The Centre for Genomic Pathogen Surveillance).

Pangolin is an open-source tool and web application developed to make it as easy as possible for researchers, public health workers and clinicians to obtain useful information from genome sequencing of SARS-CoV-2 by allowing them to assign lineages to genome sequences, view descriptive characteristics of the assigned lineages, view placement of the lineage in a global phylogeny and view the temporal and geographic distribution of the assigned lineages. Pangolin enables user samples to be contextualised within the global context by linking to Microreact (see below), which can visualize where and when sequenced samples of the same lineage have been observed. Pangolin is being used by researchers around the world, and as of the end of July, has assigned ~150,000 unique sequences globally.

Microreact

https://microreact.org/project/cogconsortium

Khalil Abudahab, Ben Taylor, Anthony Underwood, David Aanensen (The Centre for Genomic Pathogen Surveillance, University of Oxford)

Microreact is a web application that provides a simple, yet powerful, data linkage and visualization method for linking genomics to epidemiology, By linking phylogenetic trees together with geographic, temporal or other associated metadata research and public health audiences can easily interpret data. Microreact also encourages the open sharing of data. Within the Microreact COG-UK project, global SARS-CoV-2 lineage distributions (as defined by Pangolin) can be visualised together with, minimal metadata associated with genomes and a global tree indicating genome similarity. These data are currently updated when new genomes are processed and further automation will move data updates to close to real-time to enable the monitoring of trends in lineage distribution and movement. Microreact is used globally with several countries creating bespoke country-specific data views. Furthermore, local instances within PHW and PHS are enabling linkage of local sensitive data to genomic outputs in real-time.

Grapevine

https://github.com/COG-UK/grapevine

Ben Jackson, Verity Hill, Rachel Colquhoun, Andrew Rambaut (University of Edinburgh)

Grapevine is a phylogenetics pipeline that operates (currently twice-weekly) on the UK SARS-CoV-2 genome sequences produced by Elan, to which it adds a dataset of sequences from the rest of the world, with the central aim of building a phylogenetic tree that captures the evolutionary relationship between all viruses sampled to date. This adds evolutionary context to samples' epidemiological metadata, which

together can be used to understand aspects such as transmission chains and introduction events. Grapevine defines and extracts UK SARS-CoV-2 clusters, assays sequences for genetic variants of interest (such as the D614G spike protein mutation), and assigns all sequences to a Pangolin global SARS-CoV-2 lineage. It produces the global tree and metadata used by Civet (below) to facilitate local cluster investigation, and by Microreact (below) to visualise the COG-UK data. For each run it automatically generates reports at UK, constituent nation, and regional levels, which summarise the geographic and genetic distribution of SARS-CoV-2 genomic samples in the UK.

Civet (Cluster Investigation & Virus Epidemiology Tool)

https://github.com/COG-UK/civet

Áine O'Toole, Verity Hill, JT McCrone, Ben Jackson, Andrew Rambaut (University of Edinburgh)

Civet is an open source tool for cluster identification developed with 'real-time' genomics in mind. With the large phylogeny available through the COG-UK infrastructure on CLIMB, civet generates reports for sets of sequences of interest i.e. outbreak investigations. If the sequences are already on CLIMB and part of the large tree, civet will pull out the local context of those sequences, merging the smaller local trees as appropriate. If sequences haven't yet been uploaded to CLIMB, for instance if they have just been sequenced, civet will find the closest sequence in the COG-UK database on climb, pull the local tree of that sequence out and add new sequences in. The local trees then get collapsed to display in detail only the sequences of interest so as not to inform investigations beyond what was suggested by epidemiological data. A report summarising the query sequences and rendering the collapsed trees is generated. The tips of these trees can be coloured by any categorical trait present in the input csv, and additional fields added to the tip labels. Optional figures may be added to describe the local background of UK lineages and to map the query sequences using coordinates, again colourable by a custom trait. Civet is a tool particularly suited to investigating outbreaks and reporting on new sequences produced across the UK.

Llama (Local lineage and monophyly assessment)

https://cov-lineages.org/llama.html

Áine O'Toole, Verity Hill, JT McCrone, Andrew Rambaut (University of Edinburgh)

Llama is an open source tool for pulling out local phylogenetic trees from a large tree (e.g. the global SARS-CoV-2 phylogeny) and enables the addition of new sequences directly to local trees.

ncov2019-artic-nf

https://github.com/connor-lab/ncov2019-artic-nf

Matt Bull (Public Health Wales)

ncov2019-artic-nf is a Nextflow pipeline for running the ARTIC network's field bioinformatics tools (https://artic.network/ncov-2019/ncov2019-bioinformatics-sop.html) to take sequencing data (Illumina or

Nanopore) and generate consensus genome sequences. The pipeline includes steps for basecalling, de-multiplexing, mapping, polishing and consensus generation.

Phylodynamics dashboard

Rob Johnson, Erik Volz (Imperial College London)

The phylodynamics dashboard is currently under development. The dashboard will provide an overview of growth and decline of SARS-CoV-2 lineages circulating in the UK. The dashboard will also facilitate exploration of data and visualization of trends over time and differences between regions.

Summary of COG-UK publications

Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity
Volz *et al.* medRxiv [preprint]
https://www.medrxiv.org/content/10.1101/2020.07.31.20166082v1

Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study
Meredith *et al*. *Lancet Infect. Dis*. 2020 Jul 14;S1473-3099(20)30562-4
https://www.thelancet.com/journals/laninf/article/PIIS1473-3099(20)30562-4/fulltext

No evidence of viral polymorphisms associated with Paediatric Inflammatory Multisystem Syndrome Temporally Associated With SARS-CoV-2 (PIMS-TS).
Pang *et al*. medRxiv [preprint]
https://www.medrxiv.org/content/10.1101/2020.07.07.20148213v1

periscope: sub-genomic RNA identification in SARS-CoV-2 ARTIC Network Nanopore Sequencing Data
Parker *et al*. bioRxiv [preprint]
https://www.biorxiv.org/content/10.1101/2020.07.01.181867v1

CoronaHiT: large scale multiplexing of SARS-CoV-2 genomes using Nanopore sequencing
Baker *et al.* bioRxiv [preprint]
https://www.biorxiv.org/content/10.1101/2020.06.24.162156v1

Genomic epidemiology of SARS-CoV-2 spread in Scotland highlights the role of European travel in COVID-19 emergence

Da Silva Filipe *et al.* medRxiv [preprint]
https://www.medrxiv.org/content/10.1101/2020.06.08.20124834v1

An integrated national scale SARS-CoV-2 genomic surveillance network
The COVID-19 Genomics UK (COG-UK) consortium. Lancet Microbe. 2020 Jul; 1(3): e99–e100
https://www.thelancet.com/journals/lanmic/article/PIIS2666-5247(20)30054-9/fulltext

Shared SARS-CoV-2 diversity suggests localised transmission of minority variants
Lythgoe *et al.* bioRxiv [preprint]
https://www.biorxiv.org/content/10.1101/2020.05.28.118992v3

Screening of healthcare workers for SARS-CoV-2 highlights the role of asymptomatic carriage in COVID-19 transmission
Rivett *et al*. *eLife*. 2020 May 11;9:e58728.
https://elifesciences.org/articles/58728

Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2
Korber *et al*. bioRxiv [preprint]
https://www.biorxiv.org/content/10.1101/2020.04.29.069054v2