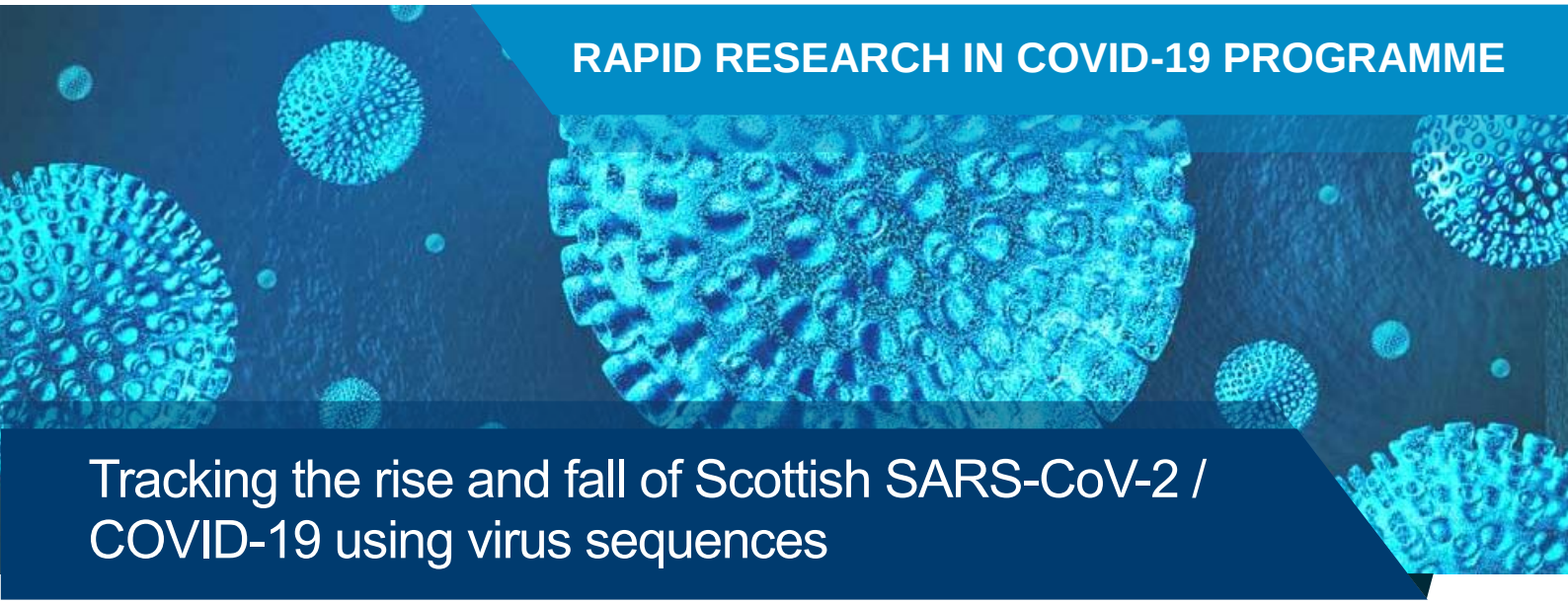
CHIEF SCIENTIST OFFICE

Scottish Government
Riaghaltas na h-Alba
gov.scot

THE UNIVERSITY of EDINBURGH

RAPID RESEARCH IN COVID-19 PROGRAMME



Tracking the rise and fall of Scottish SARS-CoV-2 / COVID-19 using virus sequences

AIMS

SARS-CoV-2 variants from COVID-19 cases in the UK are being sequenced as part of a national effort via the COVID-19 Genomics UK (COG-UK) Consortium . Since virus sequence data accumulates mutations over time, even over the time scale of a few weeks there are sufficient differences between sequences to infer clustering and transmission patterns across the UK. This project used the virus sequence data from Scotland in models to provide a genomic view on the origins of the multiple introductions of SARS-CoV-2 into Scotland and how the epidemic was spreading within regions of Scotland.

KEY FINDINGS

- This study has generated new sequence analysis pipelines which show how sequence data can be used to assess the epidemic at a regional population scale.
- Sequence data was used to trace the timing and geographic origins of groups or chains of related infections, and to assess growth and decline of these infections within Scotland.
- Infections in Scotland in the first wave had origins traced predominantly to both England and mainland Europe.
- During the first wave restrictions, the number of new imports into Scotland declined sharply and the spread within Scotland also declined to low levels, with little persistence of wave one viruses over the summer.
- When the restrictions were relaxed over the summer the initial new introductions of infections again had their origins traced predominantly to England and Europe. By September SARS-CoV-2 was circulating and increasing in numbers within Scotland once again.
- Using sequence data shows that lockdown restrictions can be effective in controlling the epidemic within Scotland, but preventing additional waves of infection also requires the control of new incursions of infection originating from outside of Scotland.

WHAT DID THE STUDY INVOLVE?

This study involved the modelling and analysis of virus genome sequences from cases in Scotland. Samples were collected for sequencing and the sequencing was performed via the COVID-19 Genomics UK (COG-UK) consortium. The Scottish samples were sequenced by the MRC-University of Glasgow Centre for Virus Research (CVR), Virology Department, Royal Infirmary of Edinburgh, NHS Lothian and Wellcome Sanger Institute.

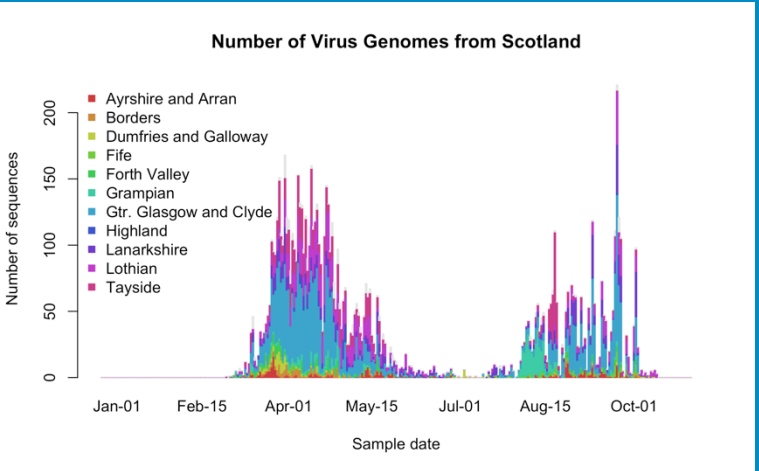
Advanced computational methods for analysing outbreak sequence data have been under development since the 2009 H1N1 pandemic, however these have only been suitable for relatively small bounded data sets. Therefore in this project we developed algorithms and an analysis pipeline to make them suitable for the Scottish data embedded within the larger UK and global data sets.

Specifically making use of the division of the UK data into groups of related sequences (lineages), we used fast approximations to estimate the origin time and location, and to indicate the growth and persistence properties of each lineage which had some representatives in Scotland. Sequences were further grouped into NHS Health Boards, to enable within Scotland spread to be estimated as a function of time. Using the initial estimates from the data as starting points, we developed simulations of how the infection might spread through Scotland between Health Boards. Additionally, we developed an online webtool to show the results of the analysis.

WHAT WERE THE RESULTS AND WHAT DO THEY MEAN?

More than 8000 virus samples from cases in Scotland were sequenced and analysed in this study. These represent approximately 20% of the cases which tested positive in Scotland. Sequence data sets are updated with the latest UK and global data weekly as part of COG-UK, and we have been further analysing the data from Scotland. Results from the analysis can be explored in an online webtool we developed - http://phylodynamics.lycett.roslin.ed.ac.uk:3838/RiseFallScotCOVID/

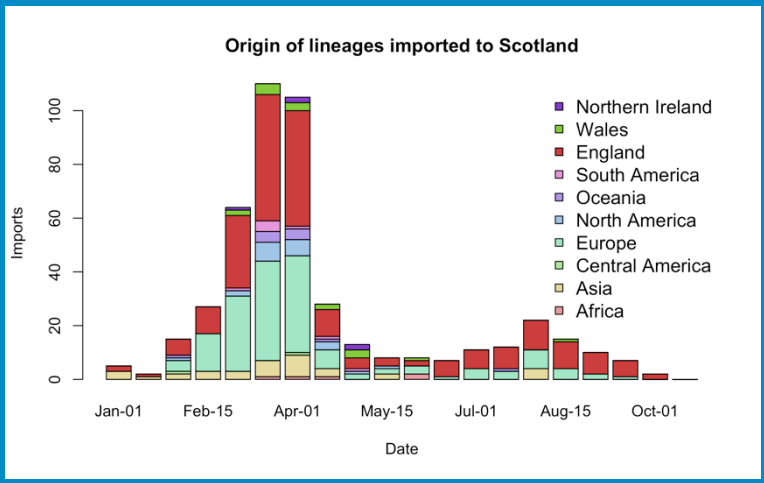
Figure 1: Number of virus genomes from samples from Scotland analysed in this study. The sequences are grouped by Health Board (sequences numbers from Orkney, Shetland and the Western Isles are too low to display). This data contains sequences to October 2020.



The sequences can be classified into detailed strains using a global lineage and/or UK specific lineage naming convention (Rambaut et al. 2020). Of the diversity present in the first wave, a wide sample of global diversity was found in Scotland (55 of the 164 named global lineages to August 2020); and although an even wider sample was found in the UK (101 named lineages) – the diversity in Scotland was not just a subset, as there were 14 lineages with a much higher proportion of sequences (>=90%) in Scotland than the rest of the UK; most likely reflecting different travel patterns (see da Silva Filipe et al. 2021)). Furthermore, there is diversity within Scotland and different Health Boards had different distributions of global lineages.

Using data to August 2020 (5000+ whole genomes from Scotland), we estimated that there were >1000 introductions into the whole of the UK. About 300 introductions seeded lineages in Scotland, but most of these were represented by five sequences or less, and only around 25% had more than 5 sequences. We estimated the approximate time and place of origin of these individual lineages, and found that the majority had originated in England or Europe, from before the lock-down. Figure 2 shows the estimated number and origin of lineages imported to Scotland with time using data to October 2020 (8000+ whole genomes from Scotland).

Figure 2: Estimation of the origin time and location of lineages imported to Scotland. The majority of the imports are from England (red) and Europe (pale green). This data contains sequences to October 2020.



Within Scotland, the diversity of the UK lineages circulating in a Health Board, and how imports versus exports to other regions within Scotland, and also from/to England, Wales and Northern Ireland changed through time was inferred from the UK lineage data. Using data to October 2020, the imports and exports between Health Boards over time are represented by arrows on the maps in figures 3A and B (please see online tool for movie). These results show the virus spreading rapidly around Scotland after the first introductions; a decrease in between Health Board transmission after lockdown, but a subsequent increase in transmission through August and September.

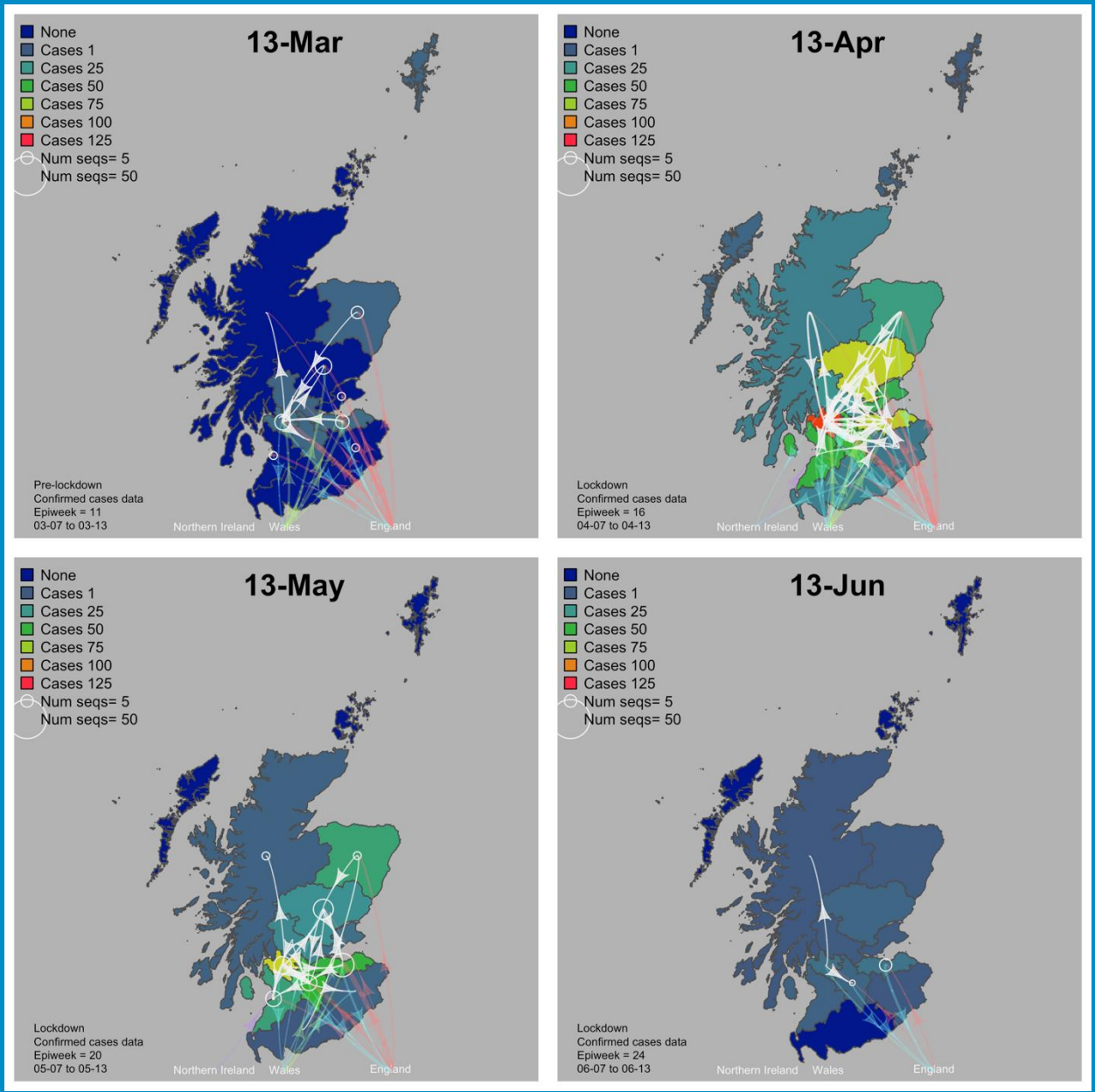


Figure 3A: Maps showing the import and export relationships between the NHS Health Boards within Scotland at different timepoints from March to June.

The width of the arrows on the map represent the number of imports/exports in a week. White arrows are between Health Board transmissions; from England (pink); from Wales (green); from Northern Ireland (purple), from Scotland to the other nation states (blue) . [Legend continues next panel].

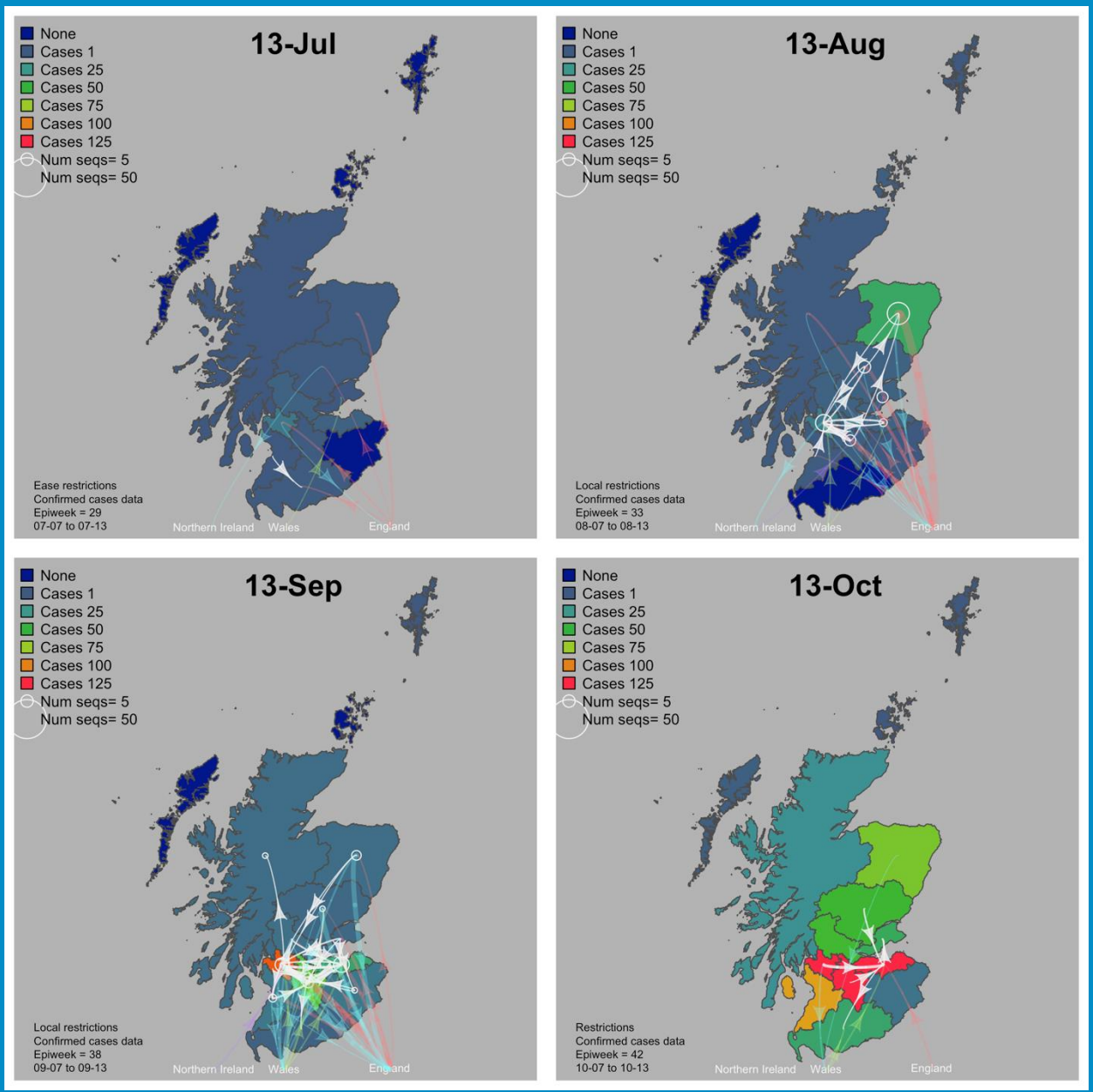


Figure 3B: Maps showing the import and export relationships between the NHS Health Boards within Scotland at different timepoints July to October.

White circle size: maximum number of sampled sequences in 7 days. Note that the transmissions are inferred so may pre-date the sampled sequences. Background colour of the Health Boards: maximum number of new confirmed cases detected in 7 days.

Analysis of the individual lineages show that there were many lineages which have sequences in Scotland before mid-July but die out afterwards (388 before 19th July 2020, of any size and 66 with > 5 sequences in Scotland), and a resurgence of different lineages at the end of August (30th August 2020). There however were some lineages (9 with > 5 sequences in Scotland) which were detected both before 19th July and after 30th August in Scotland – these lineages may have died out in Scotland in the summer and been re-introduced (most likely from England since these lineages were present in England also) or may just not have been detected. From the end of August onwards, there were 30 new lineages with sequences in Scotland. The time-scaled phylogenetic analyses shows that several of the new lineages in Scotland were seeded by European imports and some by English imports over the summer period, and that this largely agrees with independent epidemiological investigations into the travel history of cases.

An estimate of the relative contribution of the imported lineages to the subsequent circulation in Scotland was made by considering which UK lineages were imported in the summer period (17th July - 30th August) and autumn (30th August onwards). It was estimated that imported lineages over the summer contributed to 26% of sequences from the summer onwards. But in the autumn, although there were more sequences, these were mostly from lineages growing within Scotland, rather than from new lineages starting from elsewhere, so overall it is estimated that imported lineages over the autumn only contributed to 1% of sequences from the autumn onwards. Of this 1% the majority of the sequences in these lineages had an English origin.

WHAT IMPACT COULD THE FINDINGS HAVE?

- The sequence analysis is at the Health Board level, and enables an overview of the origin, growth and decline of infections and lineages in that region. This information is influenced by which samples are sequenced, but the analysis can be performed on five or more sequences.
- The number of lineages seeded at different points in time is a difficult quantity to estimate without sequence data, and results from this analysis can be used in further simulation modelling to help generate possible epidemic trajectories.
- The detailed outcomes of this study can provide virus centric information to policy makers to help inform decisions about when and where restrictions could be tightened or eased.

HOW WILL THE OUTCOMES BE DISSEMINATED?

A report written together with several from COG-UK has been written and sent to SAGE “Epidemic waves of COVID-19 in Scotland: a genomic perspective on the impact of the introduction and relaxation of lockdown on SARS-CoV-2” Samantha J Lycett, David L Robertson et al. (COVID-19 Genomics UK (COG-UK) Consortium), 28/10/2020. Additionally the report, converted to the form of an academic paper, is available as a pre-print on medRxiv at https://doi.org/10.1101/2021.01.08.20248677 and will be submitted for publication.

The results of the analyses including a between Scottish health board spread movie are displayed in an on-line tool, this is accessible via http://phylodynamics.lycett.roslin.ed.ac.uk:3838/RiseFallScotCOVID/

CONCLUSION

Analysis of whole genome viral sequence data has revealed the pattern of introductions and spread of SARS-CoV-2 into and within Scotland. Hundreds of introductions resulted in lineages being established in and spread around Scotland, with the majority of imports from England and mainland Europe. Movement and other restrictions placed on the population in March 2020 reduced the spread of the virus into and around Scotland, and extinguished many lineages, with only a few persisting through the summer. However, the virus returned in the autumn when restrictions were eased.

The sequence data from Scotland shows that it is important to continue to control chains and clusters of infections arising from new imports from elsewhere in the UK and globally, and to control the growth of lineages within Scotland. This work summarises the rise, fall and resurgence of SARS-CoV-2 in Scotland from a virus genomic perspective and indicates that restrictions are effective in controlling the virus in Scotland, but that they will probably be needed again in 2021 until effective vaccines or drugs are available widely in the population.

RESEARCH TEAM & CONTACT

Dr Samantha Lycett	samantha.lycett@ed.ac.uk
Infection & Immunity Division Roslin Institute, University of Edinburgh Easter Bush Campus Midlothian, EH25 9RG	0131 651 9100 (Roslin reception)

ADDITIONAL INFORMATION

Project details: COV/EDI/20/11 Duration: 1 May - 31 Oct 2020 Award amount: £62k

Project team members and associates: Samantha Lycett, Thomas Doherty, Gianluigi Rossi, Daniel Balaz, Rowland Kao, Lu Lu, Rhys Inward, Amy Shepherd

ACKNOWLEDGEMENTS

All collaborators: Samantha J Lycett[1], Joseph Hughes[2], Martin P McHugh[3,11], Ana da Silva Filipe[2], Rebecca Dewar[3], Lu Lu[4], Thomas Doherty[1], Amy Shepherd[1], Rhys Inward[1], Gianluigi Rossi[1], Daniel Balaz[1], Rowland Kao[1], Stefan Rooke[5], Seb Cotton[3], Michael D Gallagher[3], Carlos Balcazar Lopez[3], Áine O'Toole[5], Emily Scher[5], Verity Hill[5], John T McCrone[5], Rachel M Colquhoun[5], Ben Jackson[5], Thomas C Williams[6], Kathleen A Williamson[6], Natasha Johnson[2], Katherine Smollett[2], Daniel Mair[2], Stephen Carmichael[2], Lily Tong[2], Jenna Nichols[2], Kirstyn Brunker[7], James G Shepherd[2], Kathy Li[2], Elihu Aranday-Cortes[2], Yasmin A Parr[2], Alice Broos[2], Kyriaki Nomikou[2]; Sarah E McDonald[2], Marc Niebel[2], Patawee Asamaphan[2], Igor Starinskij[8], Natasha Jesudason[2], Rajiv Shah[2], Vattipally B Sreenu[2], Tom Stanton[9], Sharif Shaaban[10], Alasdair MacLean[8], The COVID-19 Genomics UK (COG-UK) consortium[12], Mark Woolhouse[4], Rory Gunson[8], Kate Templeton[3], Emma C Thomson[2], Andrew Rambaut[5], Matthew T.G. Holden[9,11], David L Robertson[2]

[1]Roslin Institute, University of Edinburgh; [2]MRC-University of Glasgow Centre for Virus Research (CVR); [3]Virology Department, Royal Infirmary of Edinburgh, NHS Lothian; [4]Usher Institute, University of Edinburgh; [5]Institute of Evolutionary Biology, University of Edinburgh; [6]MRC Institute of Genetics and Molecular Medicine, University of Edinburgh; [7]Institute of Biodiversity, Animal Health and Comparative Medicine, Boyd Orr Centre for Population and Ecosystem Health, University of Glasgow; [8]West of Scotland Specialist Virology Centre, NHSGGC and University of Edinburgh; [9]Edinburgh Medical School, University of Edinburgh; [10]Public Health Scotland; [11]School of Medicine, University of St Andrews; [12]https://www.cogconsortium.uk

We thank the Scottish NHS virology laboratories who provided samples for sequencing, the authors who have shared genome data on GISAID, https://www.gisaid.org. We thank Scott Arkison for HPC upgrades and systems administration. Funding: MRC (MC UU 1201412), Wellcome Trust Collaborator Award (206298/Z/17/Z – ARTIC Network), TCW Wellcome Trust Award 204802/Z/16/Z, Chief Scientist Office Project (COV/EDI/20/11), MRC EAVE II (MR/R008345/1), UKRI Industrial Strategy Challenge Fund BREATHE (MC_PC_19004), Scottish Government DG Health and Social Care, and Tommy's Charity (1060508, SC03928). COG-UK is supported by funding from the Medical Research Council (MRC) part of UK Research & Innovation (UKRI), the National Institute of Health Research (NIHR) and Genome Research Limited, operating as the Wellcome Sanger Institute. CLIMB is funded by the Medical Research Council (MRC) through grant MR/L015080/1. SL additionally received support from Scottish Government Rural and Environment Science and Analytical Services Division as part of Centre of Expertise on Animal Disease Outbreaks (EPIC), and the BBSRC Institute Strategic Programme grant to Roslin Institute (BB/J004235/1).

REFERENCES

Rambaut, A., Holmes, E.C., O'Toole, Á. et al (2020) "A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology". Nat Microbiol 5, 1403–1407 (2020). https://doi.org/10.1038/s41564-020-0770-5

da Silva Filipe, A., Shepherd, J.G., Williams, T. et al. (2021) "Genomic epidemiology reveals multiple introductions of SARS-CoV-2 from mainland Europe into Scotland" Nat Microbiol 6, 112–122 (2021) https://doi.org/10.1038/s41564-020-00838-z