Tracking the international spread of SARS-CoV-2 lineages B.1.1.7 and B.1.351/501Y-V2 [version 1; peer review: awaiting peer review]

Dutch National SARS-CoV-2 surveillance program, Division of Emerging Infectious Diseases (KDCA), Tulio de Oliveira\textsuperscript{8}, Nuno Faria\textsuperscript{2,49}, Andrew Rambaut\textsuperscript{1}, Moritz U. G. Kraemer\textsuperscript{2}

\textsuperscript{1}Institute of Evolutionary Biology, University of Edinburgh, Edinburgh, UK
\textsuperscript{2}Department of Zoology, University of Oxford, Oxford, UK
\textsuperscript{3}Li Ka Shing Knowledge Institute, St. Michael's Hospital, Toronto, Canada
\textsuperscript{4}BlueDot, Toronto, Canada
\textsuperscript{5}Department of Medicine, University of Toronto, Toronto, Canada
\textsuperscript{6}Divisions of General Internal Medicine and Infectious Diseases, University Health Network, Toronto, Canada
\textsuperscript{7}Department of Geography, University of Oxford, Oxford, UK
\textsuperscript{8}KwaZulu-Natal Research Innovation and Sequencing Platform (KRISP), Nelson R Mandela School of Medicine, University of KwaZulu-Natal, Durban, South Africa
\textsuperscript{9}Botswana Institute for Technology Research and Innovation, Gaborone, Botswana
\textsuperscript{10}National Health Laboratory, Gaborone, Botswana
\textsuperscript{11}Ministry of Health and Wellness, Gaborone, Botswana
\textsuperscript{12}Department of Natural Sciences, Lebanese American University, Beirut, Lebanon
\textsuperscript{13}Faculty of Public Health, Lebanese University, Beirut, Lebanon
\textsuperscript{14}Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand
\textsuperscript{15}Institute of Environmental Science and Research, Wellington, New Zealand
\textsuperscript{16}School of Natural and Computational Sciences, Massey University, Auckland, New Zealand
\textsuperscript{17}Department of Neurovirology, National Institute of Mental Health and Neurosciences, Bengaluru, India
\textsuperscript{18}Institute of Virology, Hannover Medical School, Hannover, Germany
\textsuperscript{19}Department of Biosystems Science and Engineering, ETH Zürich, Zurich, Switzerland
\textsuperscript{20}Istituto Zooprofittato dall'Aversità degli Studi di Modena e Reggio Emilia, Modena, Italy
\textsuperscript{21}Hospital General Universitario Gregorio Marañón; Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain
\textsuperscript{22}CIBER Enfermedades Respiratorias CIBERES, Madrid, Spain
\textsuperscript{23}Bioinformatics Unit, Department of Infectious Diseases, National Institute of Health Doutor Ricardo Jorge (INS), Lisbon, Portugal
\textsuperscript{24}Innovation and Technology Unit, Department of Human Genetics, National Institute of Health Doutor Ricardo Jorge (INS), Lisbon, Portugal
\textsuperscript{25}Central Virology Laboratory, Israel Ministry of Health, Sheba Medical Center, Ramat Gan, Israel
\textsuperscript{26}Microbiological Diagnostic Unit Public Health Laboratory, Department of Microbiology & Immunology, University of Melbourne at the Peter Doherty Institute for Infection & Immunity, Melbourne, Australia
\textsuperscript{27}New South Wales Health Pathology - Institute of Clinical Pathology and Medical Research, Sydney, Australia
\textsuperscript{28}New South Wales Health Pathology Randwick, Prince of Wales Hospital, Sydney, Australia
\textsuperscript{29}Kinghorn Centre for Clinical Genomics, Sydney, Australia
\textsuperscript{30}Queensland Reference Centre for Microbial and Public Health Genomics, Forensic and Scientific Services, Health Support Queensland, Queensland Health South Australia Pathology, Adelaide, Australia
\textsuperscript{31}South Australia Pathology, Adelaide, Australia
\textsuperscript{32}Department of Biomedical Sciences and Human Oncology, University of Bari, Bari, Italy
\textsuperscript{33}Centre National de Référence des virus des infections respiratoires, Hospices Civils de Lyon, Lyon, France
\textsuperscript{34}University of Sydney, Sydney, Australia
\textsuperscript{35}Wadsworth Center, New York State Department of Health, Albany, New York, USA
\textsuperscript{36}ErasmusMC, Department of Viroscience, WHO collaborating centre for arbovirus and viral hemorrhagic fever Reference and Research, Rotterdam, The Netherlands
\textsuperscript{37}The Public Health Agency of Sweden, Department of Microbiology, Solna, Sweden
\textsuperscript{38}Upgraded Department of Microbiology, Osmania Medical College, Hyderabad, Telangana, India
\textsuperscript{39}Department of Virology, University of Helsinki, Helsinki, Finland
\textsuperscript{40}HUS Diagnostic Center, HUSLAB, Clinical Microbiology, University of Helsinki and Helsinki University Hospital, Helsinki, Finland
\textsuperscript{41}Department of Virology, National Institute of Health, Islamabad, Pakistan
\textsuperscript{42}National Institute for Infectious Diseases “L. Spallanzani”, Rome, Italy
\textsuperscript{43}Institute for Virology, Charité Universitätsmedizin, Berlin, Germany
\textsuperscript{44}Robert Koch-Institut, , Head, Unit 17, Influenza and other Respiratory Viruses, Seestr. 10, Berlin, Germany
\textsuperscript{45}WHO COVID-19 reference laboratory, Centre for Infectious Disease Control-National Institute for Public Health and the Environment, Bilthoven, The Netherlands
Abstract
Late in 2020, two genetically-distinct clusters of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) with mutations of biological concern were reported, one in the United Kingdom and one in South Africa. Using a combination of data from routine surveillance, genomic sequencing and international travel we track the international dispersal of lineages B.1.1.7 and B.1.351 (variant S01Y-V2). We account for potential biases in genomic surveillance efforts by including passenger volumes from location of where the lineage was first reported, London and South Africa respectively. Using the software tool grinch (global report investigating novel coronavirus haplotypes), we track the international spread of lineages of concern with automated daily reports. Further, we have built a custom tracking website (cov-lineages.org/global_report.html) which hosts this daily report and will continue to include novel SARS-CoV-2 lineages of concern as they are detected.

Keywords
genomic surveillance, air travel, SARS-CoV-2, genomics, genome sequencing, virus, surveillance, pandemic, B.1.1.7, B.1.351, NS01Y, coronavirus, sequencing, genomic epidemiology

This article is included in the Coronavirus (COVID-19) collection.
**Introduction**

In December 2020, routine genomic surveillance in the United Kingdom (UK) reported a new and genetically distinct phylogenetic cluster of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (variant VOC202012/01, lineage B.1.1.7). Preliminary analysis suggests that this lineage carries an unusually large number of genetic changes. The earliest known cases of B.1.1.7 were sampled in southern England in late September 2020, and by December the lineage had spread to most UK regions and was growing rapidly. In October 2020, a separate SARS-CoV-2 cluster (variant 501Y.V2, lineage B.1.351), which carried a different constellation of genetic changes, was detected by the Network for Genomic Surveillance in South Africa. Both lineages carry mutations, especially in the virus spike protein, that may affect virus function, and both appear to have grown rapidly in relative frequency since their discovery. Early analyses of the spatial spread of SARS-CoV-2 highlights the potential for rapid virus dissemination through national and international travel. Therefore continued genomic monitoring of lineages of concern is required.

**Methods**

To better characterise the international distribution of lineages B.1.1.7 and B.1.351 we collated SARS-CoV-2 sequences from GISAID and assigned lineages using pangolin (v2.1.6, https://github.com/cov-lineages/pangolin), which implements the nomenclature scheme described in Rambaut et al.,26. Genomes are assigned lineage B.1.1.7 if they exhibit at least 5 of the 17 mutations inferred to have arisen on the phylogenetic branch immediately ancestral to the cluster (Table 1); or to B.1.351 if they exhibit at least 5 of 9 lineage-associated mutations (Table 1). Lineage count and frequency data have been calculated daily using grinch. Using Air Transport Association (IATA) travel data from October 2020, available through bluedot.global, we aggregated and collated the passenger volumes from international airports in London and South Africa to international destinations on same booking. Destinations with more than 5,000 passengers from London and more than 300 passengers from South Africa during the month of October are displayed on the cov-lineages.org website and in the underlying data for this publication.

**Implementation**

All of the code underlying this daily lineage tracking web-report can be found at GitHub and Zenodo. grinch is a python-based tool, the analysis pipeline of which is built on a snakemake backbone. Every 24 hours a scheduled crontab task runs on our local servers. We download the latest data from GISAID and deduplicate based on sequence names. The sequences are assigned their most likely lineage using pangolin’s latest version and model files. All processed metadata is available and maintained on the cov-lineages.org GitHub repository. To run grinch, the user must have access to a GISAID direct download key and a password and provide these within a configuration file for use. The command used to run grinch is grinch -i grinch_config.yaml, using the config file provided at doi:10.5281/zenodo.4640379.

**Operation**

Most users will not run grinch themselves, instead all information and useful descriptive figures are provided daily on the web report. Users can navigate to cov-lineages.org in a web browser of choice to view the latest daily report.

**Results and discussion**

As of 7th Jan 2021, 45 countries had reported the presence of B.1.1.7 and 13 countries had reported B.1.351/501Y.V2. B.1.1.7 and B.1.351 genome sequences were available for 28 and 8 countries, respectively (Figure 1a, b, c). Although some countries report increases in the relative frequency of B.1.1.7, genome sequencing efforts vary considerably. Potential targeting of sequencing towards travelers from the UK could bias frequency estimates upwards (Figure 1b, c) and differing genome sharing policies and delays may also skew reporting estimates. The time between the initial collection date of a new variant sample in a country and the first availability of a corresponding virus genome on GISAID was, on average, 12 days (range 1–71).

The number of B.1.1.7 and B.1.351/501Y.V2 genome sequences reported in each country is a consequence of (i) the intensity of local genomic surveillance; (ii) the level of concern about new variant introductions; (iii) the volume of international travel among affected countries, and (iv) the amount of local transmission following the introduction of lineage from elsewhere. To explore these factors, we analysed the most recent available International Air Transport Association (IATA) travel data (October 2020). We collated the total number of origin-to-destination air journeys between major London international airports and each country. The calculation was repeated for journeys originating in all international South African airports. We focussed on London and South Africa as they are the locations with the first reports and highest reported prevalence of lineages

<table>
<thead>
<tr>
<th>Lineage</th>
<th>Defining mutations</th>
</tr>
</thead>
<tbody>
<tr>
<td>B.1.1.7</td>
<td>orf1ab:T1001I; orf1ab:A1708D; orf1ab:L2230T; del:11288:9; del:21765:6; del:21991:3; S:N501Y; S:A570D; S:P681H; S:T716I; S:S982A; S:D1118H; Orf8:Q274; Orf8:R521; Orf8:Y73C; N:D3L; N: S235F</td>
</tr>
<tr>
<td>B.1.351/501Y.V2</td>
<td>E:P71L; N:T205I; orf1a:K1655N; S:D80A; S:D215G; S:K417N; S:E484K; S:N501Y; S:E484K</td>
</tr>
</tbody>
</table>
Figure 1. a) The cumulative number of countries with reports of lineage B.1.1.7 (grey line) and cumulative number of genomes of B.1.1.7 deposited in GISAID. b) Rolling seven-day average of the proportion of B.1.1.7 genomes in countries with more than ten sequences of the variant, and with more than ten days between the first B.1.1.7 sequence and the most recent one compared to all sampled genomes in that country. c) Number of sequences (log10) per country. Colour indicates the proportion of sequences that are classified as lineage B.1.1.7. d) Number of air travellers from major international London airports (Heathrow, Gatwick, Luton, City, Stansted, Southend) during October 2020. Colour indicates the number of sampled genomes of lineage B.1.1.7. e) Map of international flights from major international London airports to countries with B.1.1.7 sequences. Colours indicate the date of earliest detection of B.1.1.7 in each country. The width of the lines indicates the number of flights. International Air Transport Association data used here account for ~90% of passenger travel itineraries on commercial flights, excluding transportation via unscheduled charter flights (the remainder is modelled using market intelligence). Data shown represents origin-destination journeys during October 2020. Routes to countries that have not yet detected B.1.1.7 and deposited data on GISAID are not included.

B.1.1.7 and B.1.351 respectively. However, due to low SARS-CoV-2 genomic surveillance in many locations, we cannot reject the hypotheses that these lineages initially originated elsewhere. Figure 1d shows destinations receiving >5,000 travellers in October 2020 from the UK (Figure 2 shows destinations receiving >300 travellers from South Africa).

Of the countries that receive >5,000 travellers from London, 16 have sequenced B.1.1.7. Of the 45 countries that have identified B.1.1.7 (32 in travellers and 13 with local onward transmission), only 6 perform real-time routine genomic surveillance (Denmark, UK, Iceland, The Netherlands, Australia, Sweden), 3 have prioritised sequencing based on S-gene target failure tests, 30 primarily targeted sequencing towards arriving travellers from the UK, and there was no information available for 10 (details at https://github.com/cov-lineages/lineages-website/blob/master/_data/). Of the 13 countries that have identified B.1.351 (four with local onward transmission including South Africa), 4 perform routine sequencing (South Africa, UK, Botswana, Australia), 6 target sequencing of travellers, and there was no information available for 3. Consequently, the number of sequences reported does not correlate with flight numbers, but rather reflects current genomic surveillance effort. For example, in September, the UK sequenced ~13% of its reported cases and Denmark sequenced ~21%. In comparison, Israel sequenced ~0.002% of its cases during the same period.
Our study has several limitations. The passenger flight data do not include recent changes to holiday travel, and recent restrictions on travel from the UK and South Africa is not reflected in the mobility data. Further, flight data may not accurately reflect the final destination if multiple tickets are purchased.

The discovery and rapid spread of B.1.1.7 and B.1.351/501Y.V2 highlights the importance of real-time and open data for tracking the spread of SARS-CoV-2 and for informing future public health interventions and travel advice.

Data availability

Underlying data


This project contains the following underlying data:
- Accession IDs of B.1.1.7 and B.1.351 genome sequences included in report up until January 7th, 2021. All accession IDs link to data on the GISAID repository, http://doi.org/10.17616/R3Q59F. These data are available under the terms of the GISAID EpiFlu™ Database Access Agreement.


This project contains the following underlying data:
- Website data archived at time of publication

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).
Extended data

This project contains the following extended data:
- Supplementary materials with group authorship affiliations and full acknowledgements.

Data are available under the terms of the Creative Commons Attribution 4.0 International license (CC-BY 4.0).

Software availability
- Source code available from: https://github.com/cov-lineages/grinch
- Archived source code at time of publication: https://doi.org/10.5281/zenodo.46403715
- License: GNU General Public License v3.0; Creative Commons Attribution 4.0 International license (CC-BY 4.0).

References

Acknowledgements
An earlier version of this article can be found on Virological (url: https://virological.org/b/trackingsarco2lineagesb117andb1351501yv2/592).

We thank Norelle Sherry, Benjamin Howden and Michelle Sait for their contribution to sequencing in Australia. We also include full acknowledgements and details of group authorships at https://doi.org/10.5281/zenodo.47044719. We would also like to extend our gratitude to everyone involved in the global sequencing effort.