



## GENOME WATCH

# Genomics and epidemiological surveillance

Stephanie W. Lo and Dorota Jamroz

This month's Genome Watch highlights how genomic surveillance can provide important information for identifying and tracking emerging pathogens such as SARS-CoV-2.

The timely detection and surveillance of infectious diseases and responses to pandemics are crucial but challenging. Whole-genome sequencing (WGS) is a common tool for pathogen identification and tracking, establishing transmission routes and outbreak control.

At the turn of 2019/20, Wu et al<sup>1</sup> used metagenomic RNA sequencing to identify the aetiology of an at this point unknown respiratory disease in a single patient in Wuhan, China, where several cases of severe respiratory infections have been reported. The authors identified the potential causative pathogen as a new coronavirus by reconstructing the viral genome from the bronchoalveolar lavage fluid sample of the patient. In early January 2020, the viral genome sequence was released, which facilitated the development of rapid molecular diagnostics assays worldwide. Subsequently, the virus (now known as SARS-CoV-2, which causes the ongoing coronavirus disease 2019 (COVID-19) pandemic) rapidly spread globally, and there has been an immediate effort to study viral transmission and evolution using WGS. For example, SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology and Surveillance (SPHERES) in the United States and the COVID-19 Genomics UK<sup>2</sup> (COG-UK) in the United Kingdom.

The latter consortium was launched in March 2020 as a nationwide genomic surveillance network that aims to track

viral transmission, identify viral mutations and integrate viral data with health data<sup>3</sup>. By June 2020, the consortium sequenced >20,000 SARS-CoV-2 genomes and defined transmission lineages based on phylogeny. Open data sharing and standardized lineage definitions (Global Initiative on Sharing All Influenza Data (GISAID)) were established to enable global efforts in detecting emerging lineages and mutations that are relevant for outbreak control and vaccine development on an international level<sup>3</sup>. By the end of June 2020, >57,000 SARS-CoV-2 genomes from around 100 different countries have been deposited in the GISAID database. To overcome the challenge in data analysis and interpretation, user-friendly web-based applications were designed for lineage assignment (Pangolin COVID-19 Lineage Assigner) and to interactively visualize the circulating lineages on national and international scales (for example, Microreact and Nextstrain).

The international effort towards open data sharing is of major scientific benefit, enabling monitoring of SARS-CoV-2 evolution in nearly real time and on a global level. Korber et al.<sup>4</sup> developed a bioinformatics pipeline to track changes in the SARS-CoV-2 spike glycoprotein, which mediates host cell entry and is a key vaccine target. The pipeline monitors changes in the amino acid sequence of spike over time to identify variants that are concomitantly increasing in frequency in different geographic locations. The analysis, which was enabled by data from GISAID, suggested that a SARS-CoV-2 variant carrying a particular spike mutation (D614G) became globally dominant over a period of one month. Comparison of different regions revealed consistent patterns of the G614 variant replacing a previously established

D614 variant, which might be indicative of potential positive selection. The viral genome data were linked with patient clinical information, which showed that the G614 variant might be associated with potentially higher viral loads but not with disease severity. Updated data and current global counts of the spike 614 variants are available in the COVID-19 Viral Genome Analysis Pipeline.

Genomic surveillance can generate a rich source of information for tracking pathogen transmission and evolution on both national and international levels. More importantly, the recent application of genomics in surveillance of COVID-19 highlighted its usefulness in the nearly real-time response to a public health crisis.

Stephanie W. Lo<sup>✉</sup> and Dorota Jamroz<sup>✉</sup>  
Parasites and Microbes, Wellcome Sanger Institute,  
Wellcome Genome Campus, Hinxton, UK.  
✉e-mail: [microbes@sanger.ac.uk](mailto:microbes@sanger.ac.uk)

<https://doi.org/10.1038/s41579-020-0421-0>

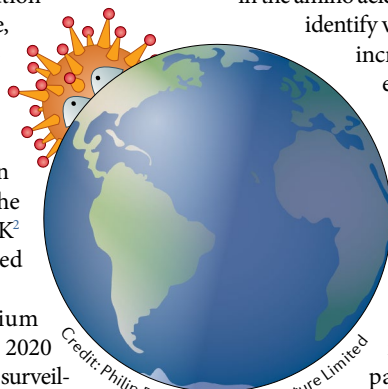
1. Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **579**, 265–269 (2020).
2. The COVID-19 Genomics UK (COG-UK) consortium. An integrated national scale SARS-CoV-2 genomic surveillance network. *Lancet Microbe* **1**, e99–e100 (2020).
3. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* <https://doi.org/10.1038/s41564-020-0770-5> (2020).
4. Korber, B. et al. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* <https://doi.org/10.1016/j.cell.2020.06.043> (2020).

### Competing interests

The authors declare no competing interests.

### RELATED LINKS

COG-UK: <https://www.cogconsortium.uk>  
 COVID-19 Viral Genome Analysis pipeline: <https://cov.lanl.gov/content/index>  
 GISAID: <https://www.gisaid.org/>  
 Microreact: <https://microreact.org/showcase>  
 Nextstrain: <https://nextstrain.org>  
 Phylogenetic Assignment of Named Global Outbreak Lineages COVID-19 Lineage Assigner: <https://pangolin.cog-uk.io/>  
 SPHERES: <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/spheres.html>



Credit: Philip Patenall/Springer Nature Limited