

## Perspectives

## GISAID's Role in Pandemic Response

Shruti Khare<sup>1,2</sup>; Céline Gurry<sup>1</sup>; Lucas Freitas<sup>1,3</sup>; Mark B Schultz<sup>1</sup>; Gunter Bach<sup>1</sup>; Amadou Diallo<sup>1,4</sup>; Nancy Akite<sup>1</sup>; Joses Ho<sup>1,2</sup>; Raphael TC Lee<sup>1,2</sup>; Winston Yeo<sup>1,2</sup>; GISAID Core Curation Team<sup>1,2,3,5,6,7,8</sup>; Sebastian Maurer-Stroh<sup>1,2,9,10,11,#</sup>

GISAID is a global data science initiative and the primary source of genomic and associated metadata of all influenza viruses, Respiratory Syncytial Virus (RSV) and severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the pandemic coronavirus causing coronavirus disease 2019 (COVID-19). GISAID's publicly accessible data sharing platform enables collaboration of over 42,000 participating researchers from 198 nations and data generators from over 3,500 institutions across the globe. Since the first whole-genome sequences were made available by China CDC through GISAID on January 10, 2020, over 5 million genetic sequences of SARS-CoV-2 from 194 countries and territories have been made publicly available through GISAID's EpiCoV database as of November 9, 2021. This high-quality, curated data enabled the rapid development of diagnostic and prophylactic measures against SARS-CoV-2 including the first diagnostic tests and the first vaccines to combat COVID-19 as well as continuous monitoring of emerging variants in near real-time.

### GISAID'S MISSION AND BACKGROUND

GISAID was launched in 2008 with the support of many governments and in partnership with public health and scientific institutions, including the Chinese Academy of Sciences, to respond to an increased reluctance of countries, and scientists around the world, to share their data during disease outbreaks in a timely manner.

Access to the latest genomic data for the highly pathogenic avian influenza (H5N1) was limited, in part due to the hesitancy by WHO Member States to share their virus genomes. In addition, the scientific community's reticence to share data pre-publication (fear of being scooped) delayed sharing. Public-domain archives offer no protection of data providers' interests, nor provide transparency on the use of data as the access and use of data take place anonymously. This limits the incentive to share data voluntarily.

By introducing a new data sharing mechanism, that recognizes the contributions and interests of data providers and users alike, GISAID successfully overcomes the reluctance for data sharing by providing an option to share data with the public. GISAID's sharing mechanism incentivizes and encourages data generators to make their data publicly accessible by guaranteeing that researchers using the data will acknowledge the contributions of, and make efforts to collaborate with, data generators.

The GISAID Initiative is an independent, non-profit, public-private partnership that involves various governments with contributions from Brazil, China, France, Germany, Senegal, Singapore, the United Kingdom, and the United States. Furthermore, GISAID receives grants from the WHO and public donors, including the Rockefeller Foundation in addition to donations from private philanthropy. GISAID is an essential asset for the Global Influenza Surveillance and Response System (GISRS) and for post-regulatory quality control of manufacturer seed viruses relative to candidate vaccine viruses.

### SARS-CoV-2 DATA SHARING THROUGH GISAID

Not long after a previously unknown human coronavirus was detected in late 2019 in patients in the City of Wuhan, who suffered from respiratory illnesses including atypical pneumonia (Pneumonia of unknown Etiology, PUE), the human coronavirus disease, later named COVID-19, was identified as a newly emerging viral respiratory disease (*1*). Researchers at China CDC looked to GISAID for its expertise to facilitate the rapid sharing of the first whole genome sequences of the earliest collected samples, thus setting in motion an unparalleled global response. GISAID has gained much experience during previous, significant outbreaks, including the 2009 swine influenza pandemic (H1N1) and the 2013 avian influenza outbreak (H7N9) in China (*2–3*). Building over its extensive expertise in influenza data sharing and its extensive collaboration network, GISAID was

well positioned to respond to what amounted for GISAID as “Disease X.” Thanks to GISAID’s longstanding partnership with China CDC, the global scientific community was able to access whole genomes within 48 hours following the identification of the pathogen. Since then, the world continues to witness an unprecedented increase in data submissions to GISAID’s EpiCoV database. The World Health Organization’s (WHO’s) Chief Scientist called GISAID a “game-changer” (4). This high-quality, curated initial set of genomes enabled the rapid development of diagnostic and prophylactic measures against SARS-CoV-2 including the first diagnostic tests (5) and the first vaccines (6) to combat COVID-19.

## SALIENT FEATURES OF GISAID

Data contributors rely on sharing their genomic data via GISAID because all submitted data are reviewed and curated in real-time and annotated by a global team of curators (7), prior to release. These curated data are enhanced with computed results and delivered downstream for analyses by countless public health and research institutions, via customized data feeds using an Application Programming Interface (API).

GISAID is known for high-quality data standards and being a driver for innovative technology. GISAID facilitates high-throughput submissions by employing Command Line Interface technology (CLI) and API interconnectivity enabling downstream analysis for public health surveillance as well as research and development. GISAID also collaborates with developers and manufacturers of vaccines and therapeutics to facilitate the collection of viral genetic sequence and metadata from clinical trial specimens.

## SARS-CoV-2 DATA ANALYSES THROUGH GISAID

GISAID enables real-time monitoring of SARS-CoV-2 genomic data. The submission tracker provides country-wise submission statistics (Figure 1A). Tracking the distribution of emerging variants like the Variant of Concern (VOC) Delta (B.1.617.2 and AY lineages) across the globe along with estimation of country-wise prevalence (Figure 1B) are made possible via GISAID. Other variants that could become relevant are also monitored for signs of increased spread estimated primarily by change in number of locations and other critical factors. A global phylogenetic tree comprising of all high-quality

sequences is available to all GISAID users (Figure 1C). The CoVsurfer tool performs sequence alignments and annotations highlighting phenotypically or epidemiologically interesting candidate amino acid changes (Figure 1D) along with 3D structural mapping. GISAID’s high throughput data sharing provisions enable numerous web applications to facilitate near real-time mutation analysis and genomic epidemiology. GISAID issues analysis updates twice a week that provide comprehensive analyses including time course of variant distribution and receptor binding surveillance to ensure that decision makers are well-informed of the emerging trends in viral spread.

## OTHER PRIORITY PATHOGENS

Since its handling of the COVID-19 pandemic, GISAID is considered uniquely positioned to follow the call by WHO Member States and public health authorities to make available its data sharing mechanism to other pathogens. These calls are likely to result in GISAID to host other priority pathogens on its platform, i.e., those with the potential of a significant global outbreak and part of the WHO R&D Blueprint. GISAID may also make its sharing mechanism available to provide access to existing data repositories that are currently not accessible to the public. GISAID’s EpiFlu database was launched in May 2008, its EpiCoV database in January 2020 and its EpiRSV database in June 2021.

GISAID continues to adhere to high quality standards and offers a trusted framework for sharing data.

**GISAID Core Curation Team:** Yi Hong Chew, Meera Makheja, Priscila Born, Gabriela Calegario, Constanza Schiavina, Sofia Romano, Juan Finello, Ya Ni Xu, Suma Tiruvayipati, Shilpa Yadahalli, Lina Wang, Xiaofeng Wei, Mikhail Bakaev, and Motharasan Manogaran.

doi: [10.46234/ccdcw2021.255](https://doi.org/10.46234/ccdcw2021.255)

# Corresponding author: Sebastian Maurer-Stroh, [sebastianms@gisaid.org](mailto:sebastianms@gisaid.org).

<sup>1</sup> GISAID Global Data Science Initiative (GISAID), Munich, Germany; <sup>2</sup> Bioinformatics Institute, Agency for Science Technology and Research, Singapore; <sup>3</sup> Oswaldo Cruz Foundation (FIOCRUZ), Rio de Janeiro, Brazil; <sup>4</sup> Institut Pasteur de Dakar, Dakar, Senegal; <sup>5</sup> National Institutes of Biotechnology Malaysia, Selangor, Malaysia; <sup>6</sup> Smorodintsev Research Institute of Influenza, St. Petersburg, Russia; <sup>7</sup> Genome Institute of Singapore, Agency for Science Technology and Research, Singapore; <sup>8</sup> China National GeneBank, Shenzhen, China; <sup>9</sup> A\*STAR Infectious Disease Labs (ID Labs), Singapore; <sup>10</sup> National Public Health Laboratory, National Centre for Infectious Diseases, Ministry of Health, Singapore; <sup>11</sup> Department of Biological Sciences,

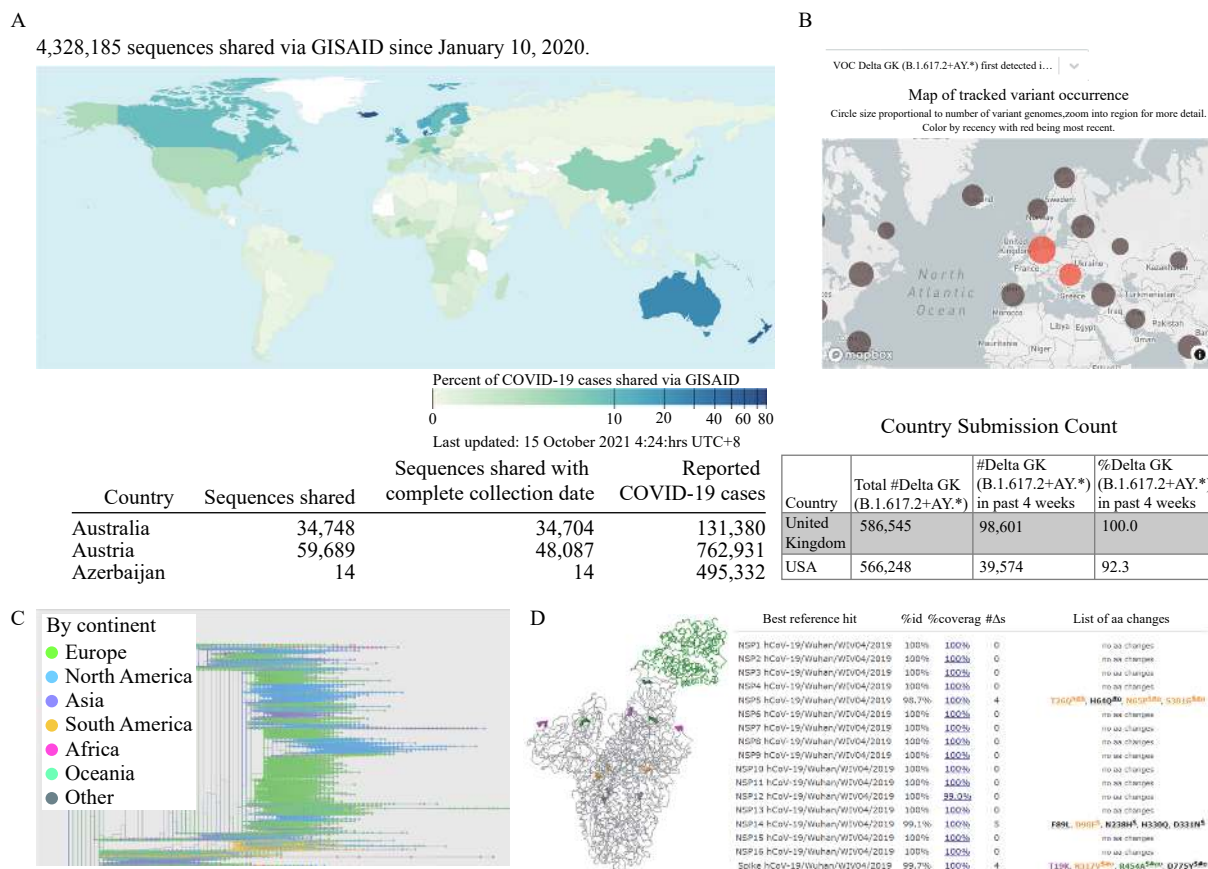


FIGURE 1. Real-time monitoring SARS-CoV-2 genomic data. (A) Submission tracker (source: <https://www.gisaid.org/index.php?id=208>); (B) Global distribution and country-wise submission statistics of tracked variants, e.g., VOC Delta (source: <https://www.gisaid.org/hcov19-variants/>); (C) Global phylogenetic tree comprising of all high-quality sequences (source: <https://www.epicov.org/epi3/cfrontend#19688e>); (D) Monitoring of nucleotide and amino acid variations and 3D structural mapping (source: <https://www.gisaid.org/epiflu-applications/covsurver-mutations-app/>). Abbreviations: SARS-CoV-2=severe acute respiratory syndrome coronavirus 2; VOC=variant of concern.

National University of Singapore, Singapore.

Submitted: November 10, 2021; Accepted: November 17, 2021

## REFERENCES

1. Tan WJ, Zhao X, Ma XJ, Wang WL, Niu PH, Xu WB, et al. A novel coronavirus genome identified in a cluster of pneumonia cases—Wuhan, China 2019–2020. *China CDC Wkly* 2020;2(4):61–2. <http://dx.doi.org/10.46234/ccdcw2020.017>.
2. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* 2017;1(1):33–46. <http://dx.doi.org/10.1002/GCH2.1018>.
3. Shu YL, McCauley J. GISAID: global initiative on sharing all influenza data—from vision to reality. *Eurosurveillance* 2017;22(13):30494. <http://dx.doi.org/10.2807/1560-7917.ES.2017.22.13.30494>.
4. Swaminathan S. The WHO's chief scientist on a year of loss and learning. *Nature* 2020;588(7839):583–5. <http://dx.doi.org/10.1038/d41586-020-03556-y>.
5. Bohn MK, Mancini N, Loh TP, Wang CB, Grimmer M, Gramegna M, et al. IFCC interim guidelines on molecular testing of SARS-CoV-2 infection. *Clin Chem Lab Med* 2020;58(12):1993–2000. <http://dx.doi.org/10.1515/cclm-2020-1412>.
6. Polack FP, Thomas SJ, Kitchin N, Absalon J, Gurtman A, Lockhart S, et al. Safety and efficacy of the BNT162b2 mRNA covid-19 vaccine. *N Engl J Med* 2020;383(27):2603–15. <http://dx.doi.org/10.1056/NEJMoa2034577>.
7. GISAID. EpiCoV Data Curation Team. 2021. <https://www.gisaid.org/about-us/acknowledgements/data-curation/>. [2021-10-21].