https://doi.org/10.1038/s41587-025-02636-6

Lengthy delays in H5N1 genome submissions to GISAID

Real-time surveillance of viral genomes enables the detection of new variants, the assessment of their impact on infectivity and disease severity, the estimation of rates of spread and routes of transmission using new phylogenetic tools, and evidence-based decision making by public health authorities, but it requires access to timely genomic information. Up-to-date genomic information also facilitates rapid, collaborative and interdisciplinary research and evidence-based public health responses (for example, vaccine development and deployment). Viral genetic changes have enabled transmission of the highly pathogenic avian influenza H5N1 to hundreds of species of birds and mammals[1–3], leading to repeated animal-to-human transmission events, including 72 reported to the World Health Organization in 2024 (ref. 4). Two recent reports describe patients with severe respiratory infections in Canada[5] and the United States[6] with viruses that were polymorphic for genetic changes previously predicted by deep mutational scanning to improve binding to human cells[7]. While these changes may have facilitated within-host viral replication, efficient human-to-human transmission of H5N1 has yet to be observed. However, this may change at any time. Real-time reporting of current H5N1 genomes is crucial, yet we find extensive delays of 7.5 months between H5N1 sample collection and submission to the Global Initiative on Sharing All Influenza Data (GISAID) repository for virus data and associated metadata[8,9].

Early on in the COVID-19 pandemic, a previous study[10] highlighted lengthy delays in SARS-CoV-2 sequence submissions to GISAID, with an average of 48 days between sample collection and submission. This global analysis highlighted countries with rapid data sharing practices and pointed out others that lagged behind. Many countries subsequently improved pipelines for data submission, with collection-to-submission times (CST) now down to 30 days for samples submitted to GISAID in 2024. As an example, Canada had a CST of 88 days early in the pandemic[10], but now has a median CST for SARS-CoV-2 sequences of only 16 days. Such dramatic improvements aided global efforts to track variants and to monitor the spread and public health impacts of COVID-19 over the last few years.

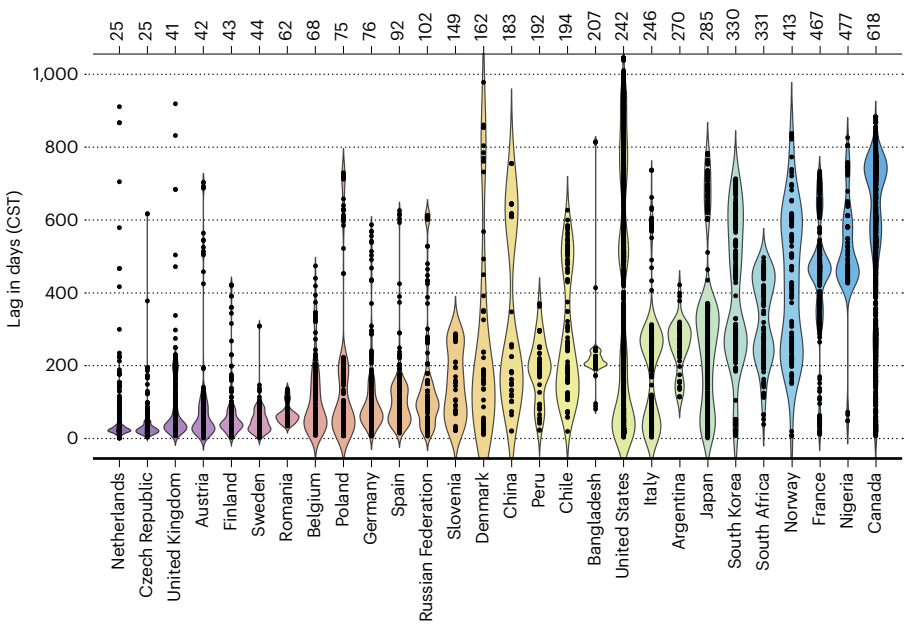

Fig. 1 | Collection-to-submission times by country for highly pathogenic avian influenza (H5N1). Violin plots show the distribution of days from sample collection to submission for all countries that had submitted at least 50 accessions between 2021 and 2024, ordered by median CST in days (numbers at top of graph). See Supplementary Tables for numbers of sequences, means, s.d. and 95% quantiles by country (Supplementary Table 1) and host information (Supplementary Table 2). Data were accessed from GISAID on 26 February 2025 and analyzed in Mathematica (code available as Supplementary Code).

Rapid sharing of the genomic data needed to monitor pathogens with pandemic potential has not extended to highly pathogenic avian influenza. Here, we analyzed the metadata for H5N1 (clade 2.3.4.4b) sequences submitted to GISAID between 1 January 2021 and 31 December 2024 with complete collection dates. To avoid analyses of historical data and an undue influence of outlier data with long CSTs, we removed data collected more than 1 year before this range, leaving $n = 18{,}817$ accessions from 84 countries. Only 14 sequences with complete date information were collected earlier (some as early as 1905); all were dropped to avoid skewing the data. Overall, the median CST delay was 228 days for H5N1 data (2.5%–97.5% quantiles of 14 to 932 days). For countries submitting at least 50 sequences, median delays ranged from a rapid 25 days for the Czech Republic and the Netherlands to 618 days for Canada (Fig. 1 and Supplementary Table 1). While a handful of countries are quick to share data, half of the countries had median CSTs over 6 months.

Furthermore, influenza metadata on GISAID were vague or missing in most cases. When reporting the host species, 66% of submissions lacked sufficient detail to determine the genus of the host (Supplementary Table 2), while 90% lacked information about whether the animal was domestic or wild, and 87% lacked animal health status. Some data sources, particularly the US Department of Agriculture (USDA), provide incomplete information when first submitting sequences, adding details only later. Having accessed the GISAID data on both 10 January 2025 and 26 February 2025, we were able to document back-filling

of metadata. For example, collection day was initially missing for 1,763 sequences but later completed for 799 of these (all by the USDA). Similarly, the US state was initially missing for 1,740 sequences, with 805 of these later included.

As a comparison, data delays were similar during the H1N1 'swine flu' pandemic a decade earlier (Supplementary Fig. 1). To see this, we matched the criteria used for H5N1; we considered submission dates to GISAID from 1 January 2010 to 31 December 2013, for samples collected at most 1 year earlier (n = 21,387 accessions from 150 countries). The median CST was similarly large: 289 days for H1N1 (2.5%–97.5% quantiles of 32 to 1,115 days).

We considered factors that might account for such delays. Among the countries included in our analyses, gross domestic product (GDP) per capita as reported by the International Monetary Fund was not significantly related to CST, with long delays in several high-GDP countries such as Norway, France and Canada (413, 467 and 618 days, respectively). Meanwhile, countries such as Peru and Bangladesh, with a lower GDP per capita, submitted data 2–3 times faster. Thus, GDP limitations do not provide a good explanation for the lengthy delays reported here.

One factor that does contribute is the host species from which viruses were sampled. For H5N1, CST was only 15 days for the few samples sequenced from humans (n = 65), as compared to 230 days when sampling from other mammals or birds (n = 18,592, excluding environmental and non-specific sources). Delays for wildlife were particularly long (Supplementary Table 2). Similarly, for the earlier swine flu dataset, longer delays were observed for H1N1 sampled from animal hosts (median CST of 422 days for animal samples (n = 3,782) versus 244 days for human samples (n=17,600)). Longer delays for animal-collected data are also observed among H1N1 data submitted to GISAID in 2024 (median CST of 127 days for animal samples (n = 922) versus 60 days for human samples (n = 35,681)).

Animal samples also show longer CSTs for SARS-CoV-2 data. For data submitted to GISAID in 2024, viruses collected from animals took an average of 787 days to submit (n = 654) compared to 30 days for human (n = 693,824).

Viruses collected from animals fall under the jurisdiction of different agencies than human data (for example, USDA or Canadian Food Inspection Agency), with different policies regarding sample processing and data sharing. Delays may be added when multiple groups are involved in the collection, storage and sequencing of animal specimens, compounded by a lack of resources budgeted for timely processing. H5N1 data are often uploaded in bulk (Supplementary Fig. 2), potentially reflecting the withholding of data until release is required for publication. Samples collected from animals may also be considered lower priority for sequence submission to global databases, but evaluating the risk of spillover into humans requires access to timely data from these animal samples.

While submission delays for influenza data from animals have improved, shifting from a median of ~14 months from the H1N1 swine flu pandemic (submissions in 2010–2013) to ~7.5 months for the current H5N1 outbreak (submissions in 2021–2024), the delays remain substantial and prevent real-time evaluation of what is happening on the ground. In addition, data with very long delays have not yet been submitted or may never be submitted (right truncation), widening the data gap even further than what we report here.

Advances in evolutionary biology and genomics have vastly improved the power to predict attributes of viruses from sequence data, allowing early assessments of the risks posed by viruses as they evolve. From deep mutational scans of H5N1, we know which sites in the hemagglutinin protein shift cell-binding preferences away from the sialic acid receptors common in birds (α2,3) to those common in mammals (α2,6)[7]. This work also indicates which sites affect H5N1 pseudovirus cell entry speed, immune evasiveness, and protein stability at low pH, which is thought to be a key trait for aerosol transmission. Combining genomic strategies with current data provides a promising early-warning system by monitoring risks of circulating viruses and assessing their pandemic potential — but timely genomic data submission is needed to serve this purpose.

Globally and collectively, we are hampered in our ability to determine how H5N1 is evolving because the data made available are months old. We hope that drawing attention to these delays for viruses with pandemic potential collected from animals will improve resource allocation for surveillance and incentivize better data-sharing practices, enabling more rapid detection and efficient response to evolutionary changes that risk the health of both animals and humans.

Data availability

The findings of this study are based on metadata accessed from GISAID on 26 February 2025 and includes H5N1 sequences up to 31 December 2024 (list of EPI_SET IDs at https://doi.org/10.55876/gis8.250226wp), H1N1 sequences up to 31 December 2013 (EPI_SET IDs at https://doi.org/10.55876/gis8.250226ad) and H1N1 sequences sampled in 2024 (EPI_SET IDs at https://doi.org/10.55876/gis8.250226zt). Summary data for SARS-CoV-2 were obtained from GISAID's Dates + Location summary file, limited to submissions in 2024 (EPI_SET IDs at https://doi.org/10.55876/gis8.250226hm).

Sarah P. Otto ✉ **& Sean Vidal Edgerton**
Department of Zoology & Biodiversity Research Centre, The University of British Columbia, Vancouver, British Columbia, Canada.
✉e-mail: otto@zoology.ubc.ca

Published online: 25 March 2025

References

1. Youk, S. et al. *Virology* **587**, 109860 (2023).
2. Peacock, T. P. et al. *Nature* **637**, 304–313 (2025).
3. Animal and Plant Health Inspection Service. Detections of highly pathogenic avian influenza. https://www.aphis.usda.gov/livestock-poultry-disease/avian/avian-influenza/hpai-detections (20 March 2025).
4. World Health Organization. Cumulative number of confirmed human cases for avian influenza A(H5N1) reported to WHO, 2003-2024. https://www.who.int/publications/m/item/cumulative-number-of-confirmed-human-cases-for-avian-influenza-a(h5n1)-reported-to-who--2003-2024--20-december-2024 (12 December 2024).
5. Jassem, A. N. et al. *N. Engl. J. Med.* **392**, 927–929 (2025).
6. Centers for Disease Control and Prevention. Genetic sequences of highly pathogenic avian influenza A(H5N1) viruses identified in a person in Louisiana. *Avian Influenza (Bird Flu)* https://www.cdc.gov/bird-flu/spotlights/h5n1-response-12232024.html (2024).
7. Dadonaite, B. et al. *PLoS Biol.* **22**, e3002916 (2024).
8. Shu, Y. & McCauley, J. *Euro Surveill.* **22**, 30494 (2017).
9. Elbe, S. & Buckland-Merrett, G. *Glob. Chall.* **1**, 33–46 (2017).
10. Kalia, K., Saberwal, G. & Sharma, G. *Nat. Biotechnol.* **39**, 1058–1060 (2021).

Acknowledgements

We gratefully acknowledge all data contributors, that is, the authors and their originating laboratories responsible for obtaining the specimens, and their submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this study is based. We would like to thank Florence Débarre for comments on an earlier draft. We acknowledge the support of the Natural Sciences and Engineering Research Council of Canada (NSERC RGPIN-2022-03726).

Author contributions

S.P.O. and S.V.E. co-designed the study, planned the manuscript and revised the manuscript; S.P.O. analyzed the data and drafted the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41587-025-02636-6.