Divergent avian strains drive an off-season influenza A peak in municipal wastewater

Alexander L. Jaffe[1]*, Alessandro Zulli[1]*, Dorothea Duong[2], Bridgette Shelden[2], Miriam Goldman[2], Miles Richardson[2], Marlene K. Wolfe[3], Alexandria B. Boehm[1†]

*these authors contributed equally.
[†]correspondence should be addressed to aboehm@stanford.edu.
[1]Department of Civil and Environmental Engineering, Stanford University
[2]Verily Life Sciences, South San Francisco, CA
[3]Gangarosa Department of Environmental Health, Rollins School of Public Health, Emory University

Summary

Wastewater sequencing is an increasingly valuable tool in tracking the spread of infectious disease agents across space and time in areas of dense human settlement. Among pathogens that can be readily detected by this approach is influenza A, which follows predictable patterns of prevalence through the winter months in North America. Here, we leverage routine surveillance of a municipal wastewater treatment plant in Northern California to describe an atypical, off-season spike in influenza A concentrations that rivals that of the winter respiratory virus season. Drawing upon metagenomic data generated through hybrid-capture sequencing, we assemble and subsequently characterize fragments of divergent influenza genomes that appear to derive predominantly from the avian H16 clade. These strains exhibit close evolutionary relationships to influenza isolated from migratory shorebirds, hinting at potential host species and mechanisms of geographic spread. Analysis of read abundances suggest that these avian strains dominate the pool of influenza circulating during the summer months, when typical human-infecting strains are essentially absent. Together, our results expand the value of wastewater sequencing to encompass sensitive tracking of outbreaks within animals in interface regions where human settlement abuts wildlands, increasing overall pandemic preparedness.

Introduction

Influenza A virus (IAV) is one of the leading causes of morbidity and mortality worldwide, and one of the major respiratory illnesses impacting human health.[1,2] The World Health Organization currently estimates that IAV is responsible for up to 600,000 deaths and 1 billion cases each year.[1] Influenza A is seasonal, with most cases occurring in each hemisphere's winter season, and there is a vaccine available, though both immune history and antigenic mismatch can lower vaccine effectiveness.[3,4] The majority of IAV cases each year are caused by commonly circulated IAV subtypes such as H3N2 and H1N1. However, IAV can undergo genetic reassortment and has broad host tropism[5], allowing new subtypes and variants to emerge, sometimes with pandemic potential.

The most recent IAV pandemic was in 2009, with a quadruple reassorted swine influenza H1N1 infecting an estimated 11-21% of the global population and causing an estimated 575,400 deaths.[6] This recent flu pandemic highlights how new variants emerge through zoonoses and

spillover and emphasizes the need for surveillance of influenza evolution.[6,7] Studying and predicting the evolution of influenza as it undergoes genetic reassortment through multiple hosts is key to preventing future pandemics, allowing researchers to rapidly develop vaccines and other countermeasures.[8,9] Significant advances have been made in predicting IAV evolution, including using deep-learning and transformer based-models to predict viral fitness based on predicted sequences and protein conformation.[9,10] These efforts require massive amounts of data, and to that end, a multitude of surveillance systems have emerged both nationally and internationally for sequencing and disseminating genome sequences, ranging from clinical (Center for Disease Control, Global Initiative on Sharing All Influenza Data) to environmental (Animal and Plant Health Inspection Service).[11–13] These are all capital and time intensive undertakings that could be complemented by wastewater-based surveillance.

Recently, there has been significant progress in using wastewater surveillance as an integral tool for studying IAV epidemiology.[14–16] These have focused on quantitative approaches like quantitative polymerase chain reactions (qPCR) or digital droplet polymerase chain reactions (ddPCR) that are limited to identifying either all types of IAV simultaneously, or specific types individually. Despite this limitation, this approach has been successfully used to track both human IAV and emerging avian strains such as H5N1.[17] Since 2024, wastewater has been used in the United States to track H5N1 due to outbreaks in cattle and the resulting risk of spillover into human populations.[18,19] Recently, a newer approach to wastewater-based surveillance of IAV, probe-based hybridization and sequencing, has demonstrated promising results for quantitative IAV surveillance while also allowing for simultaneous strain subtyping.[20]

In this study, we present the results of an investigation into an atypical peak in IAV wastewater concentrations comparable to the winter season in Palo Alto, California during the summer of 2025. We combine reverse-transcription ddPCR (RT-ddPCR) results with probe-based hybridization sequencing to systematically exclude common circulating IAV subtypes from humans, and identify a set of avian strains likely associated with migratory shorebirds as the main source of this viral material. This study demonstrates the value of metagenomic sequencing of wastewater samples to complement existing PCR-based surveillance systems, even for pathogens of non-human animals.

Methods

Sample collection and storage

One hundred and forty-seven wastewater treatment plants (WWTPs) located in 40 US states provided either “grab” samples from primary clarifiers or 24-hour composite samples of influent from the headworks (Table S1). Depending on WWTP location, samples were collected between 3-7 times a week between September 1, 2024 and November 1, 2025. Additional details of sampling as well as the location of each WWTP are available in Boehm et al.[21,22]

After collection, samples were then stored at 4°C, shipped on ice to the laboratory, and processed immediately for RT-ddPCR analysis. Aliquots of samples from Palo Alto WWTP were taken and stored at 4°C for ~0.5-10 weeks (depending on sample) for later metagenomic sequencing. Storage temperature was informed by our previous work indicating that wastewater storage at 4°C is preferable over freezing, as freeze-thaw cycles can increase the degradation of nucleic acids.[23]

RT-ddPCR quantification

Nucleic acids were extracted from wastewater solids and used neat as template in RT-ddPCR to quantify IAV genomes using commercial kits, as previously described.[17,21,22] Briefly, wastewater solids were centrifuged at 24,000×g and resuspended at 75 mg/ml in a buffer containing bovine coronavirus (BCoV) nucleic acid recovery control. After homogenization and nucleic acid extraction from 6-10 replicate aliquots, RT-ddPCR was performed using established assays targeting the conserved M gene of IAV, as well as hemagglutinin (HA) gene subtypes H1, H3, and H5.[21] IAV H3 concentrations were measured after March 2025 only.[21] The 6-10 nucleic acid extract replicates were run in their own wells. Results required ≥3 positive droplets across merged wells for detection, corresponding to ~500-1000 copies/gram. Full nucleic-acid extraction and quantification methods are provided in previous data descriptors.[21,22] Methods followed EMMI and MIQE guidelines and thus included negative and positive controls, and replication as explained in Boehm et al.[21,24,25]

Concentrations of M, H1, H3, and H5 markers were reported as population weighted concentrations as described in Chan et al.[26,27] Five-day trimmed averages were also calculated and plotted as described elsewhere.[26,27]

Sample preparation for metagenomic sequencing

Biobanked Palo Alto wastewater samples were warmed at room temperature for 30-45 minutes before processing. Samples were spun down at 4200xg for 5 minutes, and 50 mL of resulting supernatant was vacuum filtered through 0.45 µm pore size polyethersulfone (PES) filters (Thermo Scientific #165-0045, Waltham, Massachusetts, USA). Viral particles were then concentrated from 10 mL of the filtrate using Ceres Nanotrap Microbiome A particles (Manassas, Virginia, USA). Total nucleic acids were then extracted using the MagMAX Viral/Pathogen Kit (Waltham, Massachusetts, USA) on an automated KingFisher instrument, followed by PCR inhibitor removal using the Zymo OneStep Kit. Nucleic acids were then diluted 1:10 with nuclease-free water for all samples except wwd006407 and wwd006408 (January 2026), which were prepared without dilution. cDNA synthesis was performed using a ProtoScript II First Strand cDNA Synthesis Kit (New England Biolabs #E6560L, Ipswich, Massachusetts, USA). cDNA was subsequently purified using Kapa HyperPure Beads, and samples were normalized to a concentration of 1 ng/µL using nuclease-free water for library preparation. Enzymatic fragmentation, adapter ligation, and PCR amplification were accomplished using the Twist Library Preparation Kit (Twist Biosciences, #103548). Libraries were additionally prepared with the Twist Unique Molecular Identifiers (UMI) adapter system.

Finally, libraries were enriched using the Twist Comprehensive Viral Research Panel (Twist Biosciences, #103548), which targets influenza A genomes among more than 3000 viral genomes, through hybridization at 70°C for 16 hours. Ribosomal RNA was depleted using CRISPR-Cas9 technology (Jumpcode Genomics, #KIT1014). Prepared libraries were ultimately sequenced on an Illumina NovaSeq X Plus (150 or 250 bp paired-end reads), alongside negative extraction/sequencing controls and a positive control, consisting of ATCC virome nucleic acid mix (MSA-1008) at a concentration of 1000 genome copies per reaction mixed with 10 ng *E. coli* genomic DNA (Zymo D5016) and 10 ng *E. coli* total RNA (Invitrogen AM7940). Full library preparation details are provided on protocols.io.[28]

Sequence read processing and assembly

Sequencing reads were screened for those of human origin using a three-step, iterative process employing Kraken2 (v2.1.3), bowtie2 (v2.5.2) with reference GRCh38.p14, HRRT (v2.2.1) with the human_filter.db.20250916v2 database. Reads were next trimmed of sequencing adapters and UMIs using fastp (v0.23.4), and de-duplicated using HUMID (https://github.com/jfjlaros/HUMID).[29] Filtered and de-duplicated reads were subjected to initial classification with EsViritu v1.1.0.[30,31] Read profiling was complemented by *de novo* assembly of full hybrid capture metagenomes. Specifically, each set of sequenced reads were subjected to assembly by SPAdes in RNA viral mode (default parameters).[32,33] To identify assembled influenza segments, assembled contigs were BLASTed against all influenza segments contained within the EsViritu database (database v3.2.2). Any contig that attained ≥50% coverage when aligned against a reference segment was retained.

Putative influenza segments were combined with reference influenza genomes from the EsViritu database as well as H16 genomes drawn from the NCBI (-nt database) and GISAID (https://gisaid.org/), including genomes from several recent studies.[34–37] A full table of the metadata of sequences from GISAID is available in the Supporting Information (Table S2). Each segment was stored as a separate FASTA record, and the total sequence set was clustered using USEARCH at 95% and 99% identity (https://www.drive5.com/usearch/) (*-cluster_fast).* Clusters were filtered to those corresponding to the HA and neuraminidase (NA) segments (segments 4 and 6, respectively), and the longest contig within each cluster was selected as a representative. Where clusters contained a newly-assembled wastewater fragment, the longest wastewater fragment was chosen as a cluster representative. To ensure that representative fragments did not contain assembly errors, de-duplicated sequencing reads were aligned to representative fragments using bowtie2 and assessed manually for regions without read support. One such alignment is visualized in Fig. S1.

Phylogenetic reconstruction of assembled sequences

To reconstruct the identity and evolutionary history of newly-assembled flu fragments, both protein and nucleotide trees were constructed for the HA and NA segments, respectively. For protein trees, cluster representatives at 95% nucleotide identity were concatenated, and

encoded protein sequences were predicted using Prodigal (*-p meta*).[38] Proteins were aligned using MAFFT (*--auto*) if they exceeded 100 amino acids in length, achieved ≥50% of the total alignment length, and contained fewer than 10 ambiguous bases.[39] This filtering process removed one HA sequence from consideration (WWD005552_RVS_scaffold_2984), possibly due to ambiguities in sequence assembly or errors in the prediction of open reading frames. Alignments were subsequently trimmed using trimal (*-gt 0.1*) to remove uninformative positions.[40] Finally, the protein tree was inferred using IQTree2 with a model selection step (*-bnni -m TEST -st AA -bb 1000 -nt AUTO),* which selected FLU+I+G4 as the best-fitted model.[41]

For the HA segment, a more detailed nucleotide tree was then constructed using all sequence clusters in the immediate phylogenetic context of wastewater contigs, as well as a nearby outgroup. For this tree, formerly redundant sequences were restored for greater phylogenetic resolution, as long as these sequences obtained at least 300 nucleotides in length. Otherwise, the same bioinformatic steps were employed. Host type, geographic location, and year of collection were derived from each database, manually curated where missing, and visualized alongside the tree using iTol.[42]

We also undertook additional analyses of sequence similarity for the HA segment. All redundant sequences from clusters that were phylogenetically proximal to wastewater contigs were included, provided that these contigs attained 1000 bp in length or greater. Comparisons were then made between these sequences and wastewater contigs by first predicting nucleotide gene sequences with prodigal, aligning and alignment trimming (as above), and calculation of pairwise percent identity across the trimmed alignment at all non-gapped sites. Preliminary subtype classifications for segments other than HA/NA were made by subjecting reconstructed nucleic acid to BLASTn (core_nt database) and selecting the top hit by total bitscore.

Temporal dynamics of influenza subtypes

To study abundance dynamics of influenza subtypes over time, we examined the recruitment of metagenomic reads by representative HA and NA sequences from both our *de novo* assemblies as well as public data repositories. Cluster representatives (95% identity) were concatenated and used to construct a bowtie2 index. Next, trimmed and de-duplicated metagenomic reads were competitively aligned to this index, also using bowtie2.[43] For each alignment file, reads aligned to representative sequences were identified using the fetch function in pysam (https://github.com/pysam-developers/pysam). Individual reads (forward or reverse) were classified as 'stringently mapped' if the number of mismatches between read and reference was less than or equal to 5% of the read length after trimming. Otherwise, reads were classified as 'loosely mapped'. The proportion of reads mapping both stringently and loosely to each representative sequence was then visualized across the time series as a function of either HA or NA clade. Finally, read counts for each segment were summed as a function of collection date and normalized by the total number of trimmed reads in the sample to calculate effective sequencing depth in reads per million (RPM). All numerical data were plotted and analyzed using RStudio (v4.5.1) and the seaborn library in Python.

Results and Discussion

An off-season peak of Influenza A in Palo Alto municipal wastewater

RT-ddPCR results showed a large influenza A spike in measurements of the M gene beginning in July of 2025. This spike was unique to Palo Alto and not present nationwide (146 other WWTPs). **Figure 1** shows how atypical these measurements were, with very few detections nationwide during that period, while Palo Alto showed sustained high detections ≥3,000 gene copies per gram dry weight (Table S3, S4). Investigating these detections using HA subtype specific RT-ddPCR assays to detect circulating H1, H3 and H5 sequences yielded only sporadic detections which did not adequately explain the high levels of IAV M gene.[22] Notably, H1, H3, and H5 were also not detected at substantial concentrations at other WWTPs in the study (Figure 1). All assays were confirmed to be sensitive and specific to contemporary strains based on *in silico* analyses (Supplementary Methods) .

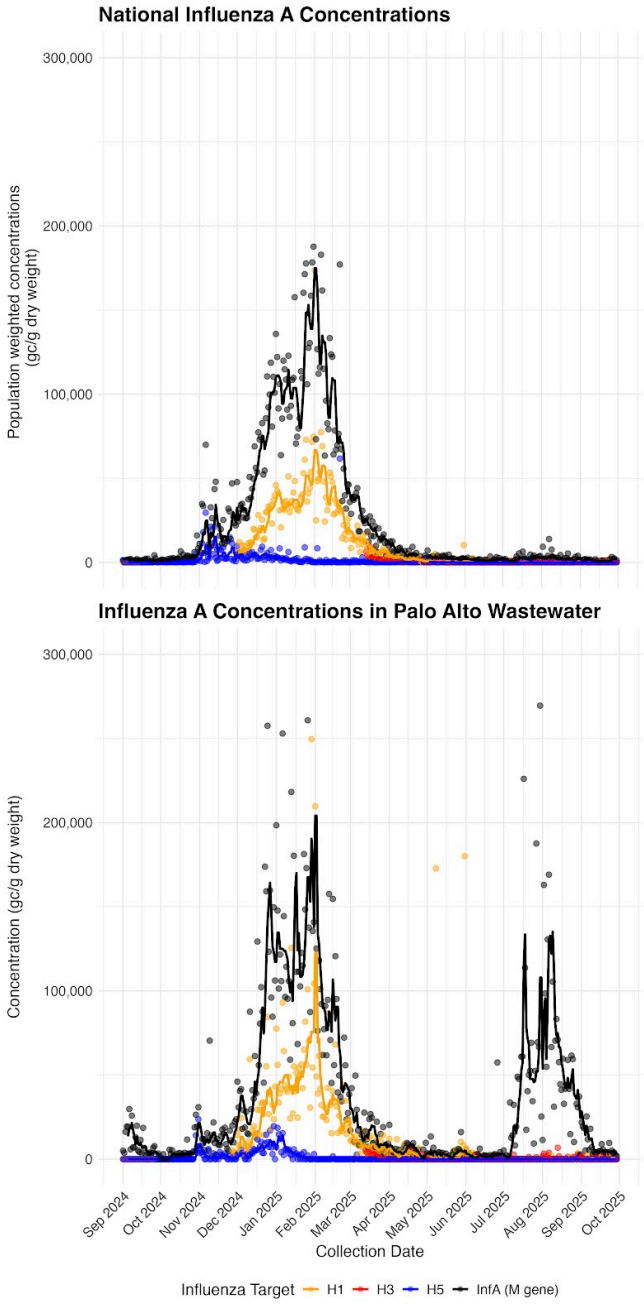

Figure 1. (Upper panel) Population weighted average national concentrations of influenza A M, H1, H3, and H5 genes in wastewater across 147 WWTPs, including Palo Alto, in 40 states between September 2024 and November 2025. Dots represent population weighted concentration in gene copies per gram of dry weight while lines represent 5-day trimmed averages. (Lower panel) Concentrations of influenza A M, H1, H3, and H5 genes in wastewater in Palo Alto, California. Dots represent concentration in gene copies per gram of dry weight while lines represent 5-day trimmed averages.

Influenza A sequences assembled from wastewater likely derive from shorebirds

To more deeply investigate the biological basis of this pattern, we performed hybrid capture sequencing targeting a wide variety of animal viruses (including influenza A) on 14 daily samples of wastewater solids from the Palo Alto WWTP, corresponding to the highest IAV concentrations observed during the summer peak (Methods). Resulting sequencing reads were initially analyzed using a read-based profiler (Methods), which indicated that a mixture of influenza subtypes - including H13N6, H7N7, H4N8, and H6N5 - was detected based on read mapping to reference genomes (Table S5). However, recruited reads only attained ~94.9% identity to reference sequences on average, suggesting the potential presence of divergent strain(s) (Table S5). Read profiling of negative extraction and sequencing controls yielded only detections of human adenovirus and papillomavirus (Table S5).

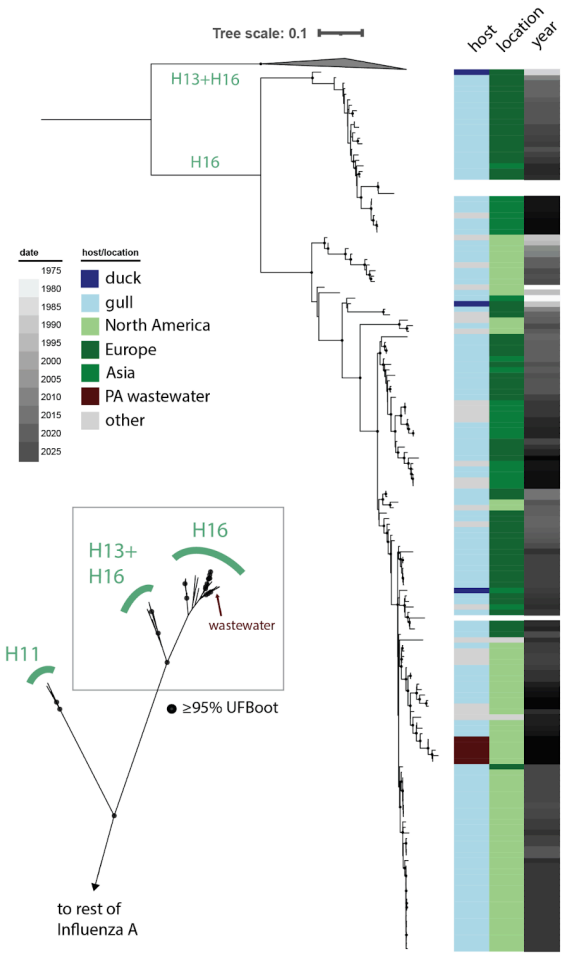

Figure 2. Phylogenetic trees depicting evolutionary relationships of wastewater strains to reference sequences based on HA protein sequence (inset) and HA nucleotide sequence (main tree). In both trees, black dots represent nodes supported by ≥95% ultra-fast (UF) bootstrap.

To evaluate the potential presence of divergent influenza strains, we performed *de novo* assembly of each set of metagenomic reads, utilizing an algorithm tailored for RNA viruses (Methods). Despite low sequencing depth, we successfully assembled contigs that appeared to derive from all eight segments of the influenza genome, including the 'typing' segments encoding the hemagglutinin and neuraminidase proteins (Table S6). Clustering of assembled contigs at 95% identity indicated that between 3 and 5 distinct clusters were assembled per segment, suggesting that some degree of sequence diversity was present among influenza strains (Table S6). We next focused on the HA (max length 1,743 bp) and NA segments (max length 1,478 bp), predicting their encoded proteins and placing them within a phylogenetic tree alongside sequences from diverse reference strains (Fig. 2a, Fig. S2). Intriguingly, assembled HA proteins fell within a clade derived from H16 (primarily H16N3) – an avian subtype primarily hosted by gulls and terns – in contrast with our initial read-based analysis.[36,44,45] A more detailed tree based on HA gene nucleotide sequence confirmed this placement and further revealed that assembled sequences were closely related to one another, forming a monophyletic group within a larger clade of strains derived primarily from North American gulls (Fig. 2a).

The nucleotide tree topology also revealed that an H16N3 strain sampled from an Alaskan gull in 2021 (EPI_ISL_16641800, A/gull/Alaska/21MB00111/2021) was the closest relative of our newly-reconstructed HA genes. Analysis based solely on sequence similarity, rather than evolutionary reconstruction, suggested that this gene obtained ~96.1% identity to one of our newly-assembled genes, the most similar of any comparison made with reference sequences (Table S7). Other newly-assembled HA genes were substantially more divergent from reference sequences, attaining only 74.4%-82.9% identity at most (Table S7). To our knowledge, no more closely related strains have been sequenced to date, possibly due to lack of sampling effort on the U.S. western coast after 2021. We anticipate that sequencing of historical samples from these bird populations would reveal currently missing links between the wastewater strains presented here and their closest relatives in the public record, which emerged 4-5 years prior.

Phylogenetic placement of assembled NA proteins, on the other hand, revealed two sequences falling in with N3 (again, mostly H16N3) references (Fig. S2). We also recovered one sequence that fell within the N8 clade, which was most closely related to diverse avian and swine N8 strains. With the methods employed here, it remains unclear whether this recovered N8 fragment was associated with an HA segment from the H16 clade, or another that could not be assembled. However, given the repeated assembly of HA segments from the H16 clade, as well as NA segments from the N3 clade, we consider the presence of at least one H16N3 strain in sampled wastewater to be highly likely. This inference is consistent with patterns of sequence similarity observed for other segments assembled but not subjected to phylogenetic reconstruction, which mostly resembled H16N3 strains (Table S6).

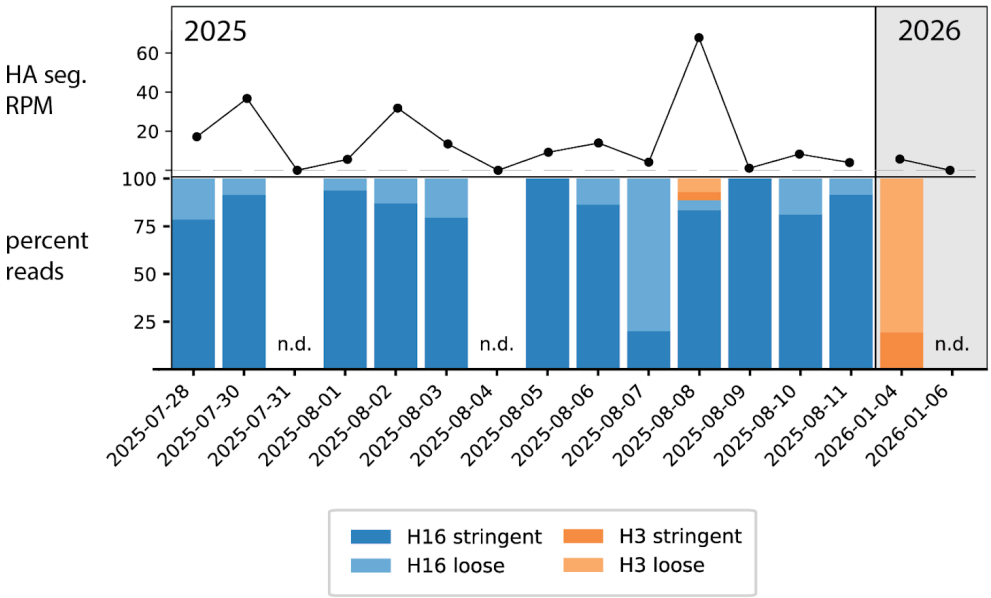

Figure 3. Relative abundance dynamics of wastewater strains across the time series, based on reads mapped to a database of HA segments, including newly-assembled sequences. Reads mapping stringently to HA segments, as well as those mapping loosely (Methods) are depicted in different shades. Upper panel indicates the total number of HA reads considered at each time point, expressed in reads per million (RPM). Abbreviations: HA, hemagglutinin; n.d., not detected.

Abundance dynamics and potential host range of wastewater influenza strains

Finally, to assess the extent to which the identified strains could drive the unusual influenza peak observed in the RT-ddPCR data (Fig. 1), we examined abundance patterns of strains across the time series. By competitively recruiting reads to the set of HA contigs drawn from our assemblies as well as existing references from diverse influenza subtypes, we observed that H16 sequences were dominant through the late July/August peak (Fig. 3, Table S8). Newly-assembled H16 segments recruited reads at all time points where HA was detected. Furthermore, most reads were mapped stringently (Methods), suggesting that mapping targets were generally representative of the sequence diversity present. In one instance, where effective sequencing depth was greater, we also detected low relative abundance of two avian H3 strains (Table S8). Supporting the robustness of our method, application of the same approach to additional samples from January 2026 revealed a read population entirely attributable to mammalian H3 strains (Fig. S3, Table S8). These findings are consistent with clinical reports showing that H3N2 was the dominant IAV subtype circulating among humans in the winter 2025-2026 season.[46] In several instances, no HA reads were detected in wastewater samples, likely due to variable efficiency of hybrid capture ('n.d.').

We applied this same approach to study strain dynamics based on the NA segment (Figure S2). There, we observed that N3-type genes accounted for most reads over time, whereas N8 genes

only appeared to be numerically relevant at one time point (Fig. S2, Table S8). Overall, we infer that H16 and N3 type strains are primarily responsible for the observed wastewater peak, though we cannot definitively rule out the presence of other, low-abundance strains that were unable to be assembled here or that are otherwise absent from reference databases.

Regardless of the precise subtypes in circulation, our combined results raise intriguing questions about the sources of viral material sampled from wastewater. Palo Alto WWTP is situated less than 2 kilometers away from a large nature preserve (Fig. S3) comprising nearly 2,000 acres of pristine marshland and hosting large numbers of both resident and migratory shorebirds.[47] While input of environmental runoff is theoretically rare in 'closed' WWTP systems like this one, we speculate that standing water from the preserve may enter into the system through a currently unknown mechanism; or, alternatively, fecal material may directly enter the system via recirculation of effluent from open-air settling tanks, where shorebirds can congregate.[48] One factor to consider in evaluating these hypotheses is that many infected individuals would likely be required to achieve the influenza concentrations observed by RT-ddPCR, although these estimates are sensitive to shedding rates used for modeling.[48]

Of course, the composite nature of wastewater does not permit definitive identification of host species for the strains analyzed here. However, our molecular evidence points firmly to a closely related set of H16Nx strains from a larger clade hosted primarily by gulls (Fig. 2). Furthermore, the timing of the influenza peak observed in Palo Alto wastewater is highly consistent with previous observations that IAV prevalence is highest among gulls during the late summer and early fall, potentially coinciding with the end of the breeding season for some Northern hemisphere species.[45] During this time, both adults and immunologically naive juveniles aggregate into dense colonies that may greatly favor the spread of viruses.[45] Thus, the wastewater peak observed in our study could reflect the dynamics of an IAV outbreak among avian hosts, rather than a temporary incursion of fecal material into wastewater streams. In the future, sampling of birds during similar anomalies will both allow for more clear delineation of host range as well as the reconstruction of genome sequences for individual strains, complementing the population-level consensuses generated here with metagenomics.

Finally, the finding that wastewater IAV sequences were most closely related to a strain detected nearly 4-5 years prior in Alaska suggests a role for host migration in the spread of these IAVs. Indeed, it is well established that migratory patterns of wild birds contribute substantially to the distribution of the IAV viruses (including high pathogenicity strains like H5) over vast geographic spans.[49–51] The spread and genetic reassortment of these strains is facilitated by the tendency of multiple migratory shorebirds to congregate in high densities in shared 'staging' areas – like Alaska and Iceland – between summer breeding grounds in arctic/subarctic regions and southern wintering areas.[45,51–53] It is increasingly recognized that targeting surveillance efforts at these hotspots, as well as the migratory routes that stem from them, may be advantageous in tracking viral spread/evolution at the global scale.[53,54] One intriguing outcome of our study is the notion that passive wastewater surveillance in locations where wildlands abut human settlement might offer this opportunity, increasing overall pandemic preparedness without the need for additional field sampling. However, future work characterizing the specific mechanisms by which

wildlife contributes to wastewater, as well as the nature of these contributions, will be necessary to most accurately interpret signals like the ones presented in this study.

Data Availability

The hybrid-capture metagenomes analyzed here will be made available through the NCBI under BioProject PRJNA1438722. Nucleotide sequences for the assembled influenza segments are available in Table S6. Custom code used to perform the described bioinformatic analyses is available on GitHub: https://github.com/alexanderjaffe/flu-genomics.

Acknowledgments

We thank Mike Tisza, Alex Crits-Christoph, Marty Freeland, Maya Xu, Katie LaBarbera, and Barbara Mühlemann for helpful discussions and comments on the manuscript. We thank Amanda Bidwell for contributing to the creation of supplementary material. We gratefully acknowledge all data contributors, i.e., 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 research is based. Finally, we thank the Palo Alto Regional Water Quality Control Plant, the Santa Clara County Public Health Department, and the California Department of Public Health.

Funding

This project was supported by a gift from the Sergey Brin Family Foundation to A.B.B.

References

(1) *Influenza (Seasonal)*. https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal) (accessed 2024-08-30).
(2) Al Hajjar, S.; McIntosh, K. The First Influenza Pandemic of the 21st Century. *Ann. Saudi Med.* **2010**, *30* (1), 1–10. https://doi.org/10.4103/0256-4947.59365.
(3) Hope-Simpson, R. E. The Role of Season in the Epidemiology of Influenza. *Epidemiol. Infect.* **1981**, *86* (1), 35–47. https://doi.org/10.1017/S0022172400068728.
(4) Lewnard, J. A.; Cobey, S. Immune History and Influenza Vaccine Effectiveness. *Vaccines* **2018**, *6* (2), 28. https://doi.org/10.3390/vaccines6020028.
(5) Medina, R. A.; García-Sastre, A. Influenza A Viruses: New Research Developments. *Nat. Rev. Microbiol.* **2011**, *9* (8), 590–603. https://doi.org/10.1038/nrmicro2613.
(6) CDC. *2009 H1N1 Pandemic*. Centers for Disease Control and Prevention. https://archive.cdc.gov/flu/pandemic-resources/2009-h1n1-pandemic.html (accessed 2025-09-07).
(7) Kelly, H.; Peck, H. A.; Laurie, K. L.; Wu, P.; Nishiura, H.; Cowling, B. J. The Age-Specific Cumulative Incidence of Infection with Pandemic Influenza H1N1 2009 Was Similar in Various Countries Prior to Vaccination. *PLOS ONE* **2011**, *6* (8), e21828. https://doi.org/10.1371/journal.pone.0021828.
(8) Lowen, A. C. Constraints, Drivers, and Implications of Influenza A Virus Reassortment. *Annu. Rev. Virol.* **2017**, *4* (Volume 4, 2017), 105–121. https://doi.org/10.1146/annurev-virology-101416-041726.

(9) Thadani, N. N.; Gurev, S.; Notin, P.; Youssef, N.; Rollins, N. J.; Ritter, D.; Sander, C.; Gal, Y.; Marks, D. S. Learning from Prepandemic Data to Forecast Viral Escape. *Nature* **2023**, *622* (7984), 818–825. https://doi.org/10.1038/s41586-023-06617-0.

(10) Notin, P.; Dias, M.; Frazer, J.; Marchena-Hurtado, J.; Gomez, A.; Marks, D. S.; Gal, Y. Tranception: Protein Fitness Prediction with Autoregressive Transformers and Inference-Time Retrieval. arXiv May 27, 2022. https://doi.org/10.48550/arXiv.2205.13760.

(11) *GISAID Initiative*. https://platform.epicov.org/epi3/frontend#2fb84f (accessed 2025-09-07).

(12) *Wild Bird Avian Influenza Surveillance Dashboard*. https://www.aphis.usda.gov/livestock-poultry-disease/avian/avian-influenza/wild-bird-surveillance-dashboard (accessed 2025-09-07).

(13) CDC. *Influenza Virus Genome Sequencing and Genetic Characterization*. Influenza (Flu). https://www.cdc.gov/flu/php/viruses/genetic-characterization.html (accessed 2025-09-07).

(14) Schoen, M. E.; Bidwell, A. L.; Wolfe, M. K.; Boehm, A. B. United States Influenza 2022–2023 Season Characteristics as Inferred from Wastewater Solids, Influenza Hospitalization, and Syndromic Data. *Environ. Sci. Technol.* **2023**, *57* (49), 20542–20550. https://doi.org/10.1021/acs.est.3c07526.

(15) Boehm, A. B.; Hughes, B.; Duong, D.; Chan-Herur, V.; Buchman, A.; Wolfe, M. K.; White, B. J. Wastewater Concentrations of Human Influenza, Metapneumovirus, Parainfluenza, Respiratory Syncytial Virus, Rhinovirus, and Seasonal Coronavirus Nucleic-Acids during the COVID-19 Pandemic: A Surveillance Study. *Lancet Microbe* **2023**, *4* (5). https://doi.org/10.1016/S2666-5247(22)00386-X.

(16) CDC. *Wastewater COVID-19 National and Regional Trends*. Centers for Disease Control and Prevention. https://www.cdc.gov/nwss/rv/InfluenzaA-national-data.html (accessed 2025-09-07).

(17) Wolfe, M. K.; Duong, D.; Shelden, B.; Chan, E. M. G.; Chan-Herur, V.; Hilton, S.; Paulos, A. H.; Xu, X.-R. S.; Zulli, A.; White, B. J.; Boehm, A. B. Detection of Hemagglutinin H5 Influenza A Virus Sequence in Municipal Wastewater Solids at Wastewater Treatment Plants with Increases in Influenza A in Spring, 2024. *Environ. Sci. Technol. Lett.* **2024**, *11* (6), 526–532. https://doi.org/10.1021/acs.estlett.4c00331.

(18) Tisza Michael J.; Hanson Blake M.; Clark Justin R.; Wang Li; Payne Katelyn; Ross Matthew C.; Mena Kristina D.; Gitter Anna; Javornik Cregeen Sara J.; Cormier Juwan; Avadhanula Vasanthi; Terwilliger Austen; Balliew John; Wu Fuqing; Rios Janelle; Deegan Jennifer; Piedra Pedro A.; Petrosino Joseph F.; Boerwinkle Eric; Maresso Anthony W. Sequencing-Based Detection of Avian Influenza A(H5N1) Virus in Wastewater in Ten Cities. *N. Engl. J. Med.* **2024**, *391* (12), 1157–1159. https://doi.org/10.1056/NEJMc2405937.

(19) Wolfe, M. K.; Duong, D.; Shelden, B.; Chan, E. M. G.; Chan-Herur, V.; Hilton, S.; Paulos, A. H.; Xu, X.-R. S.; Zulli, A.; White, B. J.; Boehm, A. B. Detection of Hemagglutinin H5 Influenza A Virus Sequence in Municipal Wastewater Solids at Wastewater Treatment Plants with Increases in Influenza A in Spring, 2024. *Environ. Sci. Technol. Lett.* **2024**, *11* (6), 526–532. https://doi.org/10.1021/acs.estlett.4c00331.

(20) Zulli, A.; Bowie, K. R.; Ruedaflores, M.; Sahly, J.; Angel, D.; Rohatgi, K.; Malik, W. S.; Hao, R. M.; Shepherd, J.; Peccia, J. Probe-Based Enrichment Sequencing Applied to Wastewater Surveillance Accurately Tracks Multiple Viral Respiratory Outbreaks. *ACS EST Water* **2025**, *5* (8), 4575–4583. https://doi.org/10.1021/acsestwater.5c00262.

(21) Boehm, A. B.; Wolfe, M. K.; Bidwell, A. L.; Zulli, A.; White, B. J.; Shelden, B.; Duong, D. Pathogen Nucleic Acids Data in Wastewater Solids from 147 Treatment Plants in the United States: 2024–2025. *Data in Brief* **2026**, *65*, 112503. https://doi.org/10.1016/j.dib.2026.112503.

(22) Boehm, A. B.; Wolfe, M. K.; Bidwell, A. L.; Zulli, A.; Chan-Herur, V.; White, B. J.; Shelden, B.; Duong, D. Human Pathogen Nucleic Acids in Wastewater Solids from 191

Wastewater Treatment Plants in the United States. *Sci. Data* **2024**, *11* (1), 1141. https://doi.org/10.1038/s41597-024-03969-8.

(23) Simpson, A.; Topol, A.; White, B. J.; Wolfe, M. K.; Wigginton, K. R.; Boehm, A. B. Effect of Storage Conditions on SARS-CoV-2 RNA Quantification in Wastewater Solids. *PeerJ* **2021**, *9*, e11933. https://doi.org/10.7717/peerj.11933.

(24) Borchardt, M. A.; Boehm, A. B.; Salit, M.; Spencer, S. K.; Wigginton, K. R.; Noble, R. T. The Environmental Microbiology Minimum Information (EMMI) Guidelines: qPCR and dPCR Quality and Reporting for Environmental Microbiology. *Environ. Sci. Technol.* **2021**, *55* (15), 10210–10223. https://doi.org/10.1021/acs.est.1c01767.

(25) dMIQE Group; Huggett, J. F. The Digital MIQE Guidelines Update: Minimum Information for Publication of Quantitative Digital PCR Experiments for 2020. *Clin. Chem.* **2020**, *66* (8), 1012–1029. https://doi.org/10.1093/clinchem/hvaa125.

(26) Chan, E. M. G.; Burnor, E.; Zulli, A.; Lu, C.; Yu, A. T.; Boehm, A. B. Spatial Aggregation Methods for Interpreting Wastewater Concentrations at Jurisdictional Scales: Insights from Two SARS-CoV-2 Monitoring Programs. Research Square December 3, 2025. https://doi.org/10.21203/rs.3.rs-8205614/v1.

(27) Zulli, A.; Varkila, M. R. J.; Parsonnet, J.; Wolfe, M. K.; Boehm, A. B. Observations of Respiratory Syncytial Virus (RSV) Nucleic Acids in Wastewater Solids Across the United States in the 2022–2023 Season: Relationships with RSV Infection Positivity and Hospitalization Rates. *ACS EST Water* **2024**, *4* (4), 1657–1667. https://doi.org/10.1021/acsestwater.3c00725.

(28) *WWSCAN Probe-based hybridization of wastewater for metagenomic analysis of viruses*. protocols.io. https://www.protocols.io/view/wwscan-probe-based-hybridization-of-wastewater-for-g8qabzvsf (accessed 2025-09-07).

(29) Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34* (17), i884–i890. https://doi.org/10.1093/bioinformatics/bty560.

(30) Cmmr/EsViritu, 2025. https://github.com/cmmr/EsViritu (accessed 2025-09-12).

(31) Tisza, M.; Javornik Cregeen, S.; Avadhanula, V.; Zhang, P.; Ayvaz, T.; Feliz, K.; Hoffman, K. L.; Clark, J. R.; Terwilliger, A.; Ross, M. C.; Cormier, J.; Moreno, H.; Wang, L.; Payne, K.; Henke, D.; Troisi, C.; Wu, F.; Rios, J.; Deegan, J.; Hansen, B.; Balliew, J.; Gitter, A.; Zhang, K.; Li, R.; Bauer, C. X.; Mena, K. D.; Piedra, P. A.; Petrosino, J. F.; Boerwinkle, E.; Maresso, A. W. Wastewater Sequencing Reveals Community and Variant Dynamics of the Collective Human Virome. *Nat. Commun.* **2023**, *14* (1), 6878. https://doi.org/10.1038/s41467-023-42064-1.

(32) *SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing - PubMed*. https://pubmed.ncbi.nlm.nih.gov/22506599/ (accessed 2024-04-18).

(33) Prjibelski, A.; Antipov, D.; Meleshko, D.; Lapidus, A.; Korobeynikov, A. Using SPAdes De Novo Assembler. *Curr. Protoc. Bioinforma.* **2020**, *70* (1), e102. https://doi.org/10.1002/cpbi.102.

(34) Khare, S.; Gurry, C.; Freitas, L.; Schultz, M. B.; Bach, G.; Diallo, A.; Akite, N.; Ho, J.; Lee, R. T.; Yeo, W.; Team, G. C. C.; Maurer-Stroh, S. GISAID's Role in Pandemic Response. *China CDC Wkly.* **2021**, *3* (49), 1049–1051. https://doi.org/10.46234/ccdcw2021.255.

(35) Shu, Y.; McCauley, J. GISAID: Global Initiative on Sharing All Influenza Data – from Vision to Reality. *Eurosurveillance* **2017**, *22* (13), 30494. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494.

(36) Peng, P.; Shen, J.; Shi, W.; Guo, J.; Wang, M.; Li, W.; Yue, Z.; Sun, X.; Guan, M.; Liu, L.; Xu, H.; Xie, Y.; Ren, A.; Liu, M.; Liu, W.; Zhang, Z.; Xiao, Z.; Li, X. Novel H16N3 Avian

Influenza Viruses Isolated from Migratory Gulls in China in 2023. *Front. Microbiol.* **2025**, *15*. https://doi.org/10.3389/fmicb.2024.1543338.

(37) Wang, Y.; Zhang, H.; Wang, M.; Guo, J.; Zhao, C.; Cui, P.; Deng, G.; Chu, D.; Shen, J.; Sun, X.; Gao, X.; Li, Y.; Liu, W.; Peng, P.; Li, X. Genetic Analysis of a Novel H16N3 Virus Isolated from a Migratory Gull in China in 2021 and Animal Studies of Infection. *Microbiol. Spectr.* **2022**, *10* (6), e02484-22. https://doi.org/10.1128/spectrum.02484-22.

(38) Hyatt, D.; Chen, G.-L.; LoCascio, P. F.; Land, M. L.; Larimer, F. W.; Hauser, L. J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinformatics* **2010**, *11* (1), 119. https://doi.org/10.1186/1471-2105-11-119.

(39) Katoh, K.; Asimenos, G.; Toh, H. Multiple Alignment of DNA Sequences with MAFFT. In *Bioinformatics for DNA Sequence Analysis*; Posada, D., Ed.; Humana Press: Totowa, NJ, 2009; pp 39–64. https://doi.org/10.1007/978-1-59745-251-9_3.

(40) Capella-Gutiérrez, S.; Silla-Martínez, J. M.; Gabaldón, T. trimAl: A Tool for Automated Alignment Trimming in Large-Scale Phylogenetic Analyses. *Bioinformatics* **2009**, *25* (15), 1972–1973. https://doi.org/10.1093/bioinformatics/btp348.

(41) Minh, B. Q.; Schmidt, H. A.; Chernomor, O.; Schrempf, D.; Woodhams, M. D.; von Haeseler, A.; Lanfear, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol. Biol. Evol.* **2020**, *37* (5), 1530–1534. https://doi.org/10.1093/molbev/msaa015.

(42) Letunic, I.; Bork, P. Interactive Tree of Life (iTOL) v6: Recent Updates to the Phylogenetic Tree Display and Annotation Tool. *Nucleic Acids Res.* **2024**, *52* (W1), W78–W82. https://doi.org/10.1093/nar/gkae268.

(43) Langmead, B.; Salzberg, S. L. Fast Gapped-Read Alignment with Bowtie 2. *Nat. Methods* **2012**, *9* (4), 357–359. https://doi.org/10.1038/nmeth.1923.

(44) Fouchier, R. A. M.; Munster, V.; Wallensten, A.; Bestebroer, T. M.; Herfst, S.; Smith, D.; Rimmelzwaan, G. F.; Olsen, B.; Osterhaus, A. D. M. E. Characterization of a Novel Influenza A Virus Hemagglutinin Subtype (H16) Obtained from Black-Headed Gulls. *J. Virol.* **2005**, *79* (5), 2814–2822. https://doi.org/10.1128/jvi.79.5.2814-2822.2005.

(45) Verhagen, J. H. Ecology of Avian Viruses. In *Studies in Viral Ecology*; John Wiley & Sons, Ltd, 2021; pp 511–539. https://doi.org/10.1002/9781119608370.ch14.

(46) CDC. *FluView*. FluView. https://www.cdc.gov/fluview/index.html (accessed 2026-03-16).

(47) *Baylands Nature Preserve*. https://www.paloalto.gov/Departments/Community-Services/Parks-Open-Space-Golf-Division/Neighborhood-Parks/Baylands-Nature-Preserve (accessed 2026-03-16).

(48) Paulos, A. P.; Hilton, S. P.; Shelden, B.; Duong, D.; Boehm, A. B.; Wolfe, M. K. Detection of Hemagglutinin H5 Influenza A Virus RNA and Model of Potential Inputs in an Urban California Sewershed. *Environ. Sci. Technol.* **2025**, *59* (31), 16168–16179. https://doi.org/10.1021/acs.est.4c14792.

(49) Yang, Q.; Wang, B.; Lemey, P.; Dong, L.; Mu, T.; Wiebe, R. A.; Guo, F.; Trovão, N. S.; Park, S. W.; Lewis, N.; Tsui, J. L.-H.; Bajaj, S.; Cheng, Y.; Yang, L.; Haba, Y.; Li, B.; Zhang, G.; Pybus, O. G.; Tian, H.; Grenfell, B. Synchrony of Bird Migration with Global Dispersal of Avian Influenza Reveals Exposed Bird Orders. *Nat. Commun.* **2024**, *15* (1), 1126. https://doi.org/10.1038/s41467-024-45462-1.

(50) Tian, H.; Zhou, S.; Dong, L.; Van Boeckel, T. P.; Cui, Y.; Newman, S. H.; Takekawa, J. Y.; Prosser, D. J.; Xiao, X.; Wu, Y.; Cazelles, B.; Huang, S.; Yang, R.; Grenfell, B. T.; Xu, B. Avian Influenza H5N1 Viral and Bird Migration Networks in Asia. *Proc. Natl. Acad. Sci.* **2015**, *112* (1), 172–177. https://doi.org/10.1073/pnas.1405216112.

(51) THE GLOBAL CONSORTIUM FOR H5N8 AND RELATED INFLUENZA VIRUSES. Role for Migratory Wild Birds in the Global Spread of Avian Influenza H5N8. *Science* **2016**, *354* (6309), 213–217. https://doi.org/10.1126/science.aaf8852.

(52) Lycett, S. J.; Pohlmann, A.; Staubach, C.; Caliendo, V.; Woolhouse, M.; Beer, M.; Kuiken, T.; Global Consortium for H5N8 and Related Influenza Viruses. Genesis and Spread of Multiple Reassortants during the 2016/2017 H5 Avian Influenza Epidemic in Eurasia. *Proc. Natl. Acad. Sci.* **2020**, *117* (34), 20814–20825. https://doi.org/10.1073/pnas.2001813117.

(53) Gass Jr, J. D.; Dusek, R. J.; Hall, J. S.; Hallgrimsson, G. T.; Halldórsson, H. P.; Vignisson, S. R.; Ragnarsdottir, S. B.; Jónsson, J. E.; Krauss, S.; Wong, S.-S.; Wan, X.-F.; Akter, S.; Sreevatsan, S.; Trovão, N. S.; Nutter, F. B.; Runstadler, J. A.; Hill, N. J. Global Dissemination of Influenza A Virus Is Driven by Wild Bird Migration through Arctic and Subarctic Zones. *Mol. Ecol.* **2023**, *32* (1), 198–213. https://doi.org/10.1111/mec.16738.

(54) Siembieda, J. L.; Johnson, C. K.; Cardona, C.; Anchell, N.; Dao, N.; Reisen, W.; Boyce, W. Influenza A Viruses in Wild Birds of the Pacific Flyway, 2005–2008. *Vector-Borne Zoonotic Dis.* **2010**, *10* (8), 793–800. https://doi.org/10.1089/vbz.2009.0095.