

**Title:**

**Virome Sequencing Identifies H5N1 Avian Influenza in Wastewater from Nine Cities.**

**Authors:**

Michael J. Tisza<sup>1,2</sup> †, Blake M. Hanson<sup>4</sup> †, Justin R. Clark<sup>2,3</sup>, Li Wang<sup>1,2</sup>, Katelyn Payne<sup>1,2</sup>, Matthew C. Ross<sup>1,2</sup>, Kristina D. Mena<sup>4,5,6</sup>, Anna Gitter<sup>4,5,6</sup>, Sara J. Javornik-Cregeen<sup>1,2</sup>, Juwan Cormier<sup>1,2</sup>, Vasanthi Avadhanula<sup>2</sup>, Austen Terwilliger<sup>2,3</sup>, John Balliew<sup>7</sup>, Fuqing Wu<sup>4,5,6</sup>, Janelle Rios<sup>4,5</sup>, Jennifer Deegan<sup>4,5</sup>, Pedro A. Piedra<sup>2</sup>, Joseph F. Petrosino<sup>\*1,2</sup>, Eric Boerwinkle<sup>4,5</sup> \*, and Anthony W. Maresso<sup>2,3\*</sup>

**Affiliations:**

<sup>1</sup> The Alkek Center for Metagenomics and Microbiome Research, Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, 77030, USA

<sup>2</sup> Department of Molecular Virology and Microbiology, Baylor College of Medicine, Houston, TX, 77030, USA

<sup>3</sup> TAILOR Labs, Baylor College of Medicine, Houston, TX, 77030, USA

<sup>4</sup> Department of Epidemiology, School of Public Health, The University of Texas Health Science Center at Houston, TX, 77030, USA

<sup>5</sup> Texas Epidemic Public Health Institute (TEPHI), Houston, TX, USA

<sup>6</sup> Department of Environmental and Occupational Health Sciences, The University of Texas Health Science Center at Houston, TX, 77030, USA

<sup>7</sup> El Paso Water Utility, El Paso, TX, USA

† These authors contributed equally

\* Corresponding authors

**Abstract:**

Avian influenza (serotype H5N1) is a highly pathogenic virus that emerged in domestic waterfowl in 1996. Over the past decade, zoonotic transmission to mammals, including humans, has been reported. Although human to human transmission is rare, infection has been fatal in nearly half of patients who have contracted the virus in past outbreaks. The increasing presence of the virus in domesticated animals raises substantial concerns that viral adaptation to immunologically naïve humans may result in the next flu pandemic. Wastewater-based epidemiology (WBE) to track viruses was historically used to track polio and has recently been implemented for SARS-CoV2 monitoring during the COVID-19 pandemic. Here, using an agnostic, hybrid-capture sequencing approach, we report the detection of H5N1 in wastewater in nine Texas cities, with a total catchment area population in the millions, over a two-month period from March 4<sup>th</sup> to April 25<sup>th</sup>, 2024. Sequencing reads uniquely aligning to H5N1 covered all eight genome segments, with best alignments to clade 2.3.4.4b. Notably, 19 of 23 monitored sites

had at least one detection event, and the H5N1 serotype became dominant over seasonal influenza over time. A variant analysis suggests avian or bovine origin but other potential sources, especially humans, could not be excluded. We report the value of wastewater sequencing to track avian influenza.

## **Main Text and Results:**

Highly pathogenic avian influenza viruses are extremely virulent members of the H5 and H7 subtypes of Influenza A, cause systemic disease in avian species, and incur significant cost to agricultural production because of mass culling of infected animals.<sup>1-3</sup> Adaptations in the hemagglutinin (HA) surface protein that influence cleavage by the furin-like proteases is thought to be a key driver of high virulence.<sup>1,4,5</sup> Since 1955, most outbreaks have been caused by three lineages, with the so-called A/goose/Guangdong/1996 H5N1 lineage widespread in poultry and responsible for human infections.<sup>4</sup> The 2.3.4.4b clade evolved within this lineage and has spread throughout the world, including North America via wild birds in December 2021.<sup>2,4,6-10</sup> On March 25<sup>th</sup> 2024, H5N1 2.3.4.4b was detected in Texas dairy cattle herds concomitantly with herds in Michigan and Kansas.<sup>11</sup> The first human case of 2024 was detected shortly thereafter in Texas (March 28<sup>th</sup>, 2024)<sup>12</sup> in an individual with exposure to symptomatic cattle. Detections have now been observed in 36 dairy herds across nine states.<sup>13</sup> Another group recently reported influenza with an H5 hemagglutinin gene present in March and April of 2024 in high levels in two wastewater sheds known to accumulate industrial discharges of cow milk.<sup>15</sup> Genomic comparisons of the human and cattle cases indicate a close genetic relatedness with a single amino acid substitution of E627K within the polymerase basic protein 2 (PB2) segment.<sup>14</sup>

Since May of 2022, the Texas Epidemic Public Health Institute (TEPHI) has been using hybrid-capture sequencing to test weekly wastewater samples throughout Texas<sup>16</sup>, detecting over 400 human and animal viruses to date, several of which (e.g. SARS-CoV-2, Influenza, and Mpox) correlate to clinical case data.<sup>17</sup> Seasonal influenza serotypes H3N2 and H1N1 are routinely detected in TEPHI wastewater samples, and levels have corresponded to clinical case loads from May 2022 through the beginning of March 2024. Until that point, serotype H5N1 was never detected (0 out of 1,337 wastewater samples). However, in samples from March 4<sup>th</sup> through April 25<sup>th</sup> (most recent available data), H5N1 is detected in 9 of 10 cities, 19 (of 23) sites, and in 46 of 163 samples (Fig. 1B, C, Fig. S1). The agnostic nature of this methodology means that these signals were observed without any change to routine protocols. The abundance of H5N1 sequences has not correlated with influenza-related hospitalizations, which have continued to decline during the spring (Fig. 1B).

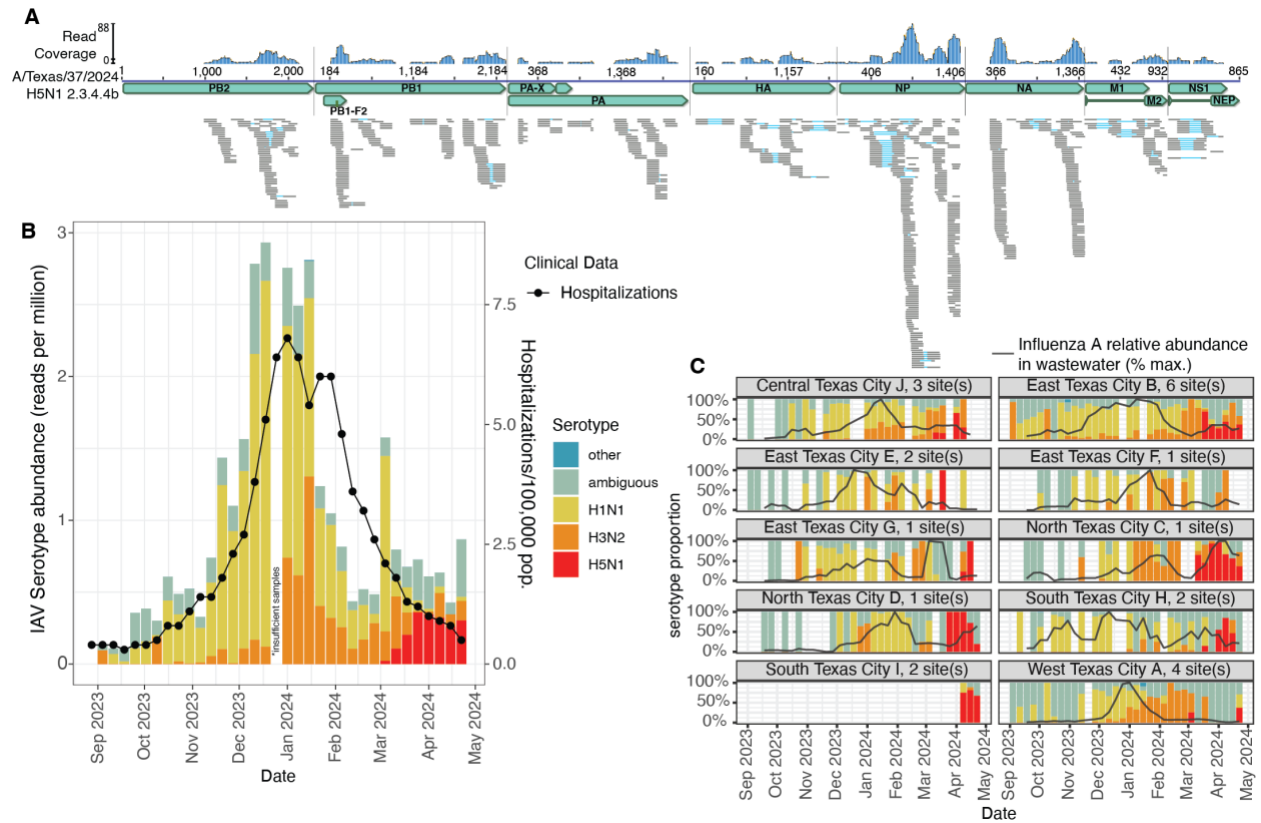
Sequences were assigned to H5N1 by a competitive alignment approach to an extensive influenza genome database, requiring read pairs to align closely to H5N1 references and better to H5N1 than other serotypes. Putative H5N1 sequences were then manually validated and documented by three independent genomics researchers associated with TEPHI (see methods). Within the set of validated H5N1 sequencing reads, there are alignments to all eight genome segments including the PB2 (on segment 1), HA/hemagglutinin (on segment 4) and NA/neuraminidase (on segment 6) genes (Fig. 1A). All sequencing reads best match H5N1

genomes from birds and mammals (including the human case) collected since 2023 and are assigned to the 2.3.4.4b clade. A variant and SNP analysis found mutations consistent with either avian or cattle origin; mainly, the presence of a glutamic acid, instead of a lysine, in position 627 of the PB2 gene supports a non-human source (Supplemental Table 1).<sup>18</sup>

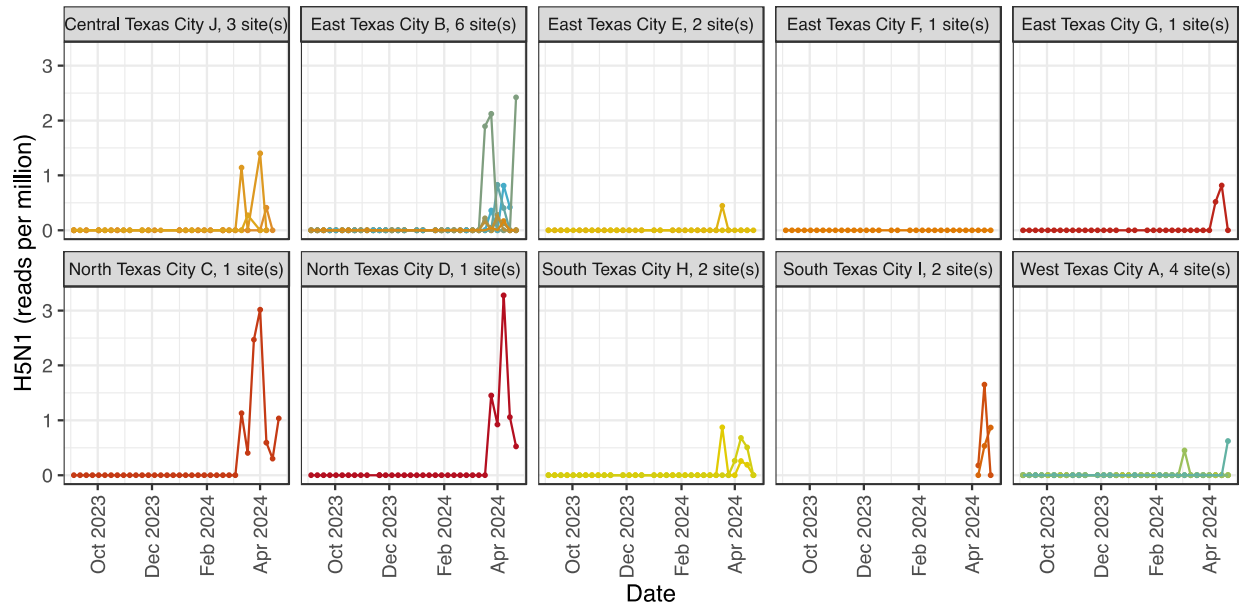
In conclusion, we report the widespread detection of Influenza A H5N1 virus in wastewater from nine U.S. cities during the spring of 2024. Although the exact cause of the signal is currently unknown, lack of clinical burden along with genomic information suggests avian or bovine origin. Given the now widespread presence of the virus in dairy cows, the concerning findings that unpasteurized milk may contain live virus, and that these two recent factors will increase the number of viral interactions with our species, wastewater monitoring should be readily considered as a sentinel surveillance tool that augments and accelerates our detection of evolutionary adaptations of significant concern.

**Figure 1: Influenza A H5N1 serotype is detectable in wastewater via virome sequencing. (A)**

Dereplicated reads from wastewater virome sequencing assigned to H5N1 from all samples with validated H5N1 signal (see methods) aligned to a recently collected H5N1 2.3.4.4b reference sequence (GISAID: EPI\_ISL\_19027114). (B) Temporal chart (September 1<sup>st</sup>, 2023 – April 25<sup>th</sup>, 2024) measuring Influenza A serotype and abundance averaged across all 23 Texas wastewater treatment sites (colored bars) and hospitalizations due to influenza across Texas as reported by Texas Department of Health and Human services (black line). Note that hospitalizations also include those caused by Influenza B, which is not measured in wastewater in this chart. (C) Bars represent proportion of influenza A serotypes in wastewater samples from each city. Grey line shows the 3-week moving average of influenza A relative abundance in wastewater, normalized to each city’s seasonal maximum. Note that sampling of South Texas City I sites began April 8<sup>th</sup>, 2024.



**Supplementary Figure 1:** Site-by-site (n=23) detection and quantification of H5N1 sequencing reads from September 1<sup>st</sup>, 2023 to April 25<sup>th</sup>, 2024.



**Supplemental Table 1: Variant Analysis of Pooled H5N1 Reads Detected in Wastewater.**

The table presents the results of a variant analysis of H5N1 detected in wastewater. The ‘Nucleotide Position’ refers to the concatenated position in the GISAID EPI\_ISL\_19027114 genome. ‘Polymorphism Type’ describes the type of genetic variation, with SNP standing for Single Nucleotide Polymorphism. ‘Change’ indicates the alteration in the nucleotide sequence. ‘Variant Frequency’ is the proportion of reads that contain the variant. ‘Unique Detections’ represents the number of times the reference or variant alleles were detected in the read data on different days and/or at different sites. ‘Coverage’ is the number of reads that include the site of the variant. ‘Amino Acid Change’ and ‘CDS Codon Number’ describe the resulting change in the protein sequence and the position of the changed codon in the coding sequence, respectively. ‘Codon Change’ indicates the alteration in protein sequence. ‘Protein Effect’ describes the impact of the variant on the protein, such as a substitution or a frameshift. ‘Variant P-Value’ is an approximate value representing the probability of a sequencing error resulting in observing bases with at least the given sum of qualities. The lower the p-value, the more likely the variation at the given position represents an allele. Abbreviations: CDS, Coding Sequence; SNP, Single Nucleotide Polymorphism; PB2, Polymerase Basic Protein 2; PB1, Polymerase Basic Protein 1; PA, Polymerase Acidic Protein; HA, Hemagglutinin; NP, Nucleoprotein; NA, Neuraminidase; M1, Matrix Protein 1; NS1, Non-structural Protein 1; PA-X, Polymerase Acidic Fusion Protein.

Nucleotide Position	CDS	Polymorphism Type	Nucleotide Change	Variant Frequency	Unique Detections (Reference Allele)	Unique Detections (Variant Allele)	Coverage	Amino Acid Change	CDS Codon Number	Codon Change	Protein Effect	Variant P-Value
1100	PB2	SNP (transition)	A->G	100.00%	0	2	8	E->G	362	GAG->GGG	Substitution	2.5E-30
1894	PB2	SNP (transition)	A->G	100.00%	0	3	12	K->E	627	AAA->GAA	Substitution	6.3E-44
1906	PB2	SNP (transversion)	A->C	100.00%	0	3	12	M->L	631	AUG->CUG	Substitution	4.0E-45
1955	PB2	SNP (transition)	U->C	26.70%	2	1	15	I->T	647	AUA->ACA	Substitution	2.2E-12
1960	PB2	Substitution	AUA->GAG	26.70%	2	1	15	I->E	649	AUA->GAG	Substitution	3.4E-11
1964	PB2	Substitution	GAG->AUU	21.4%>26.7%	2	1	14->15	RG->NC	650	AGA,GGC->AAU,UGC	Substitution	2.9E-9
1977	PB2	SNP (transition)	G->A	100.00%	0	2	11		654	CCG->CCA	None	2.0E-41
2091	PB2	SNP (transition)	A->G	42.90%	1	1	7		692	AGA->AGG	None	2.8E-10
2829	PB1	SNP (transition)	C->U	40.00%	1	1	5		167	CUC->CUU	None	4.0E-7
2963	PB1	SNP (transition)	U->C	30.00%	2	1	10	L->S	212	UUA->UCA	Substitution	9.5E-10
3502	PB1	SNP (transition)	G->A	100.00%	0	1	3	V->I	392	GUA->AUA	Substitution	7.9E-12
3612	PB1	SNP (transition)	G->A	100.00%	0	1	3		428	GAG->CAA	None	7.9E-12
4186	PB1	SNP (transversion)	U->G	25.00%	2	1	12	Y->D	620	UAU->GAU	Substitution	8.7E-7
4191	PB1	SNP (transition)	A->G	100.00%	0	2	12		621	CAA->CAG	None	4.0E-45
4269	PB1	Deletion	-A	75.00%	1	1	8		647		Frame Shift	1.8E-15
4881	PA	SNP (transversion)	C->A	100.00%	0	1	5		79	AUC->AUA	None	3.2E-19
4881	PA-X	SNP (transversion)	C->A	100.00%	0	1	5		79	AUC->AUA	None	3.2E-19
4882	PA	Substitution	GA->UG	100.00%	0	1	5		80		Truncation	3.2E-19
4882	PA-X	Substitution	GA->UG	100.00%	0	1	5		80		Truncation	3.2E-19
5068	PA	SNP (transition)	G->A	100.00%	0	1	4	E->K	142	GAA->AAA	Substitution	1.6E-15
5068	PAX	SNP (transition)	G->A	100.00%	0	1	4	E->K	142	GAA->AAA	Substitution	1.6E-15
5439	PA	SNP (transition)	A->G	100.00%	0	1	3		265	CCA->CCG	None	7.9E-12
5473	PA	SNP (transition)	C->U	100.00%	0	1	3	P->S	277	CCU->UCU	Substitution	7.9E-12
6134	PA	SNP (transition)	A->G	60.00%	1	1	10	K->R	497	AGG->AGG	Substitution	1.3E-17
6226	PA	Substitution	AGUGACCCGAG->GCCUCGGGUA	60.00%	2	1	10	TDPK->ASGQ	528	ACU,GAC,CCG,AGG->GCC,UCG,GGU,CAG	Substitution	2.1E-19
6238	PA	SNP (transition)	C->U	42.90%	2	1	7		532	CUA->UUA	None	2.8E-10
7585	HA	SNP (transition)	C->U	62.50%	1	2	8		243	GAC->GAU	None	1.8E-17
7587	HA	SNP (transversion)	U->G	50.00%	2	1	8	F->C	244	UUC->UGC	Substitution	4.1E-13
7723	HA	SNP (transition)	C->U	100.00%	0	1	7		289	CAC->CAU	None	7.9E-24
7750	HA	SNP (transition)	A->G	30.00%	1	1	10		298	GUA->GUG	None	9.5E-10
8161	HA	SNP (transversion)	U->A	30.00%	2	1	10	D->E	435	GAU->GAA	Substitution	9.5E-10
8164	HA	Substitution	CUG->UCA	23.1%>30.0%	3	1	10->13	VW->VQ	436	GUC,UGG->GUU,CAG	Substitution	2.3E-9
8168	HA	Substitution	ACC->CAU	23.10%	1	1	13	T->H	438	ACC->CAU	Substitution	2.3E-9
8170	HA	SNP (transversion)	C->A	76.90%	1	3	13		438	ACC->ACA	None	2.9E-33
8175	HA	Substitution	AUGCGUA->UGUCCAG	21.4%>37.5%	3	1	8->14	NAE->MSR	440	AUU,GCU,GAA->AUG,UCC,AGA	Substitution	2.9E-9
9029	NP	SNP (transition)	U->C	100.00%	0	4	17		135	CAU->CAC	None	1.3E-43
9219	NP	SNP (transition)	C->U	40.00%	2	1	5		199		Truncation	4.0E-7
9272	NP	SNP (transition)	A->G	100.00%	0	1	23		216	AGA->AGG	None	7.9E-66
9357	NP	SNP (transversion)	A->U	13.00%	4	1	23	S->C	245	AGC->UGC	Substitution	1.4E-8
9443	NP	SNP (transition)	G->A	85.10%	3	2	74		273	AAG->AAA	None	1.3E-208
9551	NP	Deletion	-C	13.00%	5	1	46		309		Frame Shift	2.3E-12
9632	NP	SNP (transition)	G->A	100.00%	0	4	14		336	GCG->GCA	None	1.6E-52
11157	NA	Substitution	UC->CA	80.00%	1	1	5	S->H	340	UCU->CAU	Substitution	7.9E-15
11160	NA	Substitution	AAUG->UAGA	80.00%	1	1	5		341		Truncation	1.3E-13
11170	NA	Substitution	AU->UG	66.70%	1	1	5	Y->L	344	UAU->UUG	Substitution	2.4E-14
11288	NA	SNP (transition)	A->G	100.00%	0	4	10		383	ACA->ACG	None	7.9E-37
11547	NA	SNP (transversion)	U->G	100.00%	0	3	23		470		Extension	7.9E-86
11553	SNP (transversion)	G->U	100.00%	0	3	23					6.2E-81	
11556	SNP (transversion)	C->G	91.30%	1	2	23					6.0E-76	
11559	SNP (transversion)	A->C	91.30%	1	2	23					6.0E-76	
11565	Substitution	UAGAUUAUGA->ACUUGUCAAU	77.8%>93.8%	1	3	8->16					4.5E-25	
11896	M1	SNP (transversion)	A->U	42.90%	1	2	7	I->L	107	AUA->UUA	Substitution	2.8E-10
12274	M1	Substitution	CUCCUUGA->UCUAAACC	26.7%>35.3%	3	1	15->17	LLE->SNR	233	CUC,CUU,GAA->UCU,AAC,CGA	Substitution	9.5E-16
12283	M1	Substitution	AA->GG	41.20%	3	1	17	N->G	236	AUU->GUU	Substitution	2.4E-22
12286	M1	Substitution	UUUGCAG->CGAAAC	41.2%>58.3%	3	1	12->17	LQ->RN	237	UUG,CAG->CGA,AAC	Substitution	2.4E-22
12677	NS1	SNP (transition)	C->U	100.00%	1	1	8	L->F	33	CUU->UUU	Substitution	1.0E-28

## Methods:

### Sample collection and processing

All collection, processing, and sequencing methods have been described in an earlier publication<sup>17</sup> and were followed precisely. To restate, between 100–500 mL of raw wastewater influent was collected into 500 ml leak-proof pre-labeled sample bottles at 23 wastewater treatment plants in 10 Texas cities, the identities of these cities are masked on the request of local public health officials and water utilities. From South Texas City H, sewage from three treatment plants is combined into a single sample before shipment to the sequencing laboratory; all other sites had samples processed individually. Treatment plants were coded upon the request of public health officials. The surface of the sample bottles was decontaminated with 10% bleach and moved to a “clean” zone, where the samples were sealed into biohazard bags in shipping boxes with absorbent pads and ice packs for overnight shipping to the Alkek Center for Metagenomics and Microbiome Research at Baylor College of Medicine, Houston, TX.

Wastewater samples were barcoded upon arrival and stored at 4 °C until processing. First, 50 mL of wastewater was decanted and centrifuged at 3374 x g for 10 min, separating the solid and liquid fractions. The supernatant was then vacuum filtered using an ion-based cellulose filter paper and the virus-containing cellulose filter was placed into a bead-beating tube with lysis buffer. The tube was run on a homogenizer for 1 min at 5 m/s, rested for 1 min, then run on the homogenizer for 1 more minute. Following bead beating the samples were centrifuged at 14–17×1000 RPM for 2 min. DNA and RNA were extracted using the Qiagen QIAamp VIRAL RNA Mini Kit.

### *Library preparation, probe-based virome capture and sequencing*

RNA extracts were converted to cDNA using Protoscript II First Strand cDNA Synthesis Kit (New England Biolabs Inc.), NEBNext Ultra II Non-Directional RNA Second Strand Module (New England Biolabs Inc.), and Random Primer 6 (New England Biolabs Inc.). A total of 25 ng of the cDNA and DNA mix was used for library construction using Twist Library Preparation EF 2.0 Kit and Twist Universal Adaptor System (Twist Biosciences). The libraries were pooled, a maximum of 18 samples per pool, at equal mass to a total 1,500 ng per pool. The Twist Comprehensive Viral Research Panel (Twist Biosciences) was used to hybridize the probes at 70 °C for 16 h. The post-capture pool was further PCR amplified for 12 cycles and final libraries were sequenced on Illumina NovaSeq 6000 SP flow cell, to generate 2×150 bp paired-end reads. Following sequencing, raw data files in binary base call (BCL) format were converted into FASTQs and demultiplexed based on the dual-index barcodes using the Illumina ‘bcl2fastq’ software.

### *Influenza A read detection and serotype assignment*

Reads were processed and analyzed using the publicly available *iav\_serotype* (v0.1.1) software package ([https://github.com/mtisza1/influenza\\_a\\_serotype](https://github.com/mtisza1/influenza_a_serotype)). For context, *iav\_serotype* utilizes a database of 207,488 publicly available, complete Influenza A segments from all known serotypes (Influenza\_A\_segment\_sequences database v1.1, available on [Zenodo](https://zenodo.org/record/10000000)). In detail, paired-end reads from wastewater virome sequencing are quality filtered using *fastp* (v0.23.4)<sup>19</sup> with default settings. Reads are aligned to Influenza\_A\_segment\_sequences database v1.1 using *minimap2*<sup>20</sup> (v2.28-r1209) and flags “-cx sr --secondary=yes” to allow secondary alignments. Average nucleotide identity (ANI) and alignment fraction (AF) of each alignment are calculated. Then read alignments are assigned to a particular serotype if the best alignment is exclusive to one serotype and (ANI\*AF >= 0.9). Else, reads are assigned as “ambiguous”.

### *Manual Inspection of Putative H5N1 Sequences*

After automated (*iav\_serotype*) detection of H5N1 in wastewater samples was noted by MJT, sequencing reads were shared for independent validation by co-authors BMH and JRC. After sequence inspection confirmed the general findings, parameters for final validation were agreed upon. Using web BLAST<sup>21,22</sup> against all of NT (final inspection on May 4<sup>th</sup> and 5<sup>th</sup>), forward and reverse reads from each pair were inspected separately. PCR duplicates were assumed to share origin. Read pairs were marked as “H5N1 specific”, otherwise marked as “ambiguous”. “H5N1 specific” label was used only if neither forward or reverse read alignments had a higher bit score to non-H5N1 genomes and either of the following conditions were met: 1) forward and/or reverse reads aligned with 100% identity and alignment fraction to H5N1 genomes and alignments to other serotype were worse by 1 or more bases. 2) forward and/or reverse reads aligned with less than 100% identity to H5N1 and non-H5N1 alignments were worse by 2 or more bases. Of 49 sequencing libraries with automatically assigned reads (per *iav\_serotype*), 46 libraries were manually validated as containing H5N1 specific reads.

### *Variant Analysis of Pooled Reads*

The variant analysis was conducted on pooled reads, independent of the serotype assignment. Here, raw reads underwent quality trimming to a score of 20 and reads with fewer than 30 base pairs were discarded using *BBDuk* (v38.84)<sup>23</sup>. Duplicate reads were removed using *Dedupe* (v38.84). The quality-controlled reads were then pooled and aligned to the GISAID

EPI\_ISL\_19027114 lineage using the Geneious Assembler in Geneious Prime 2024.0.5, with medium sensitivity settings to detect variations of any size.

We excluded variant calls with a Strand Bias P-value less than 0.05 and those resulting from ambiguities in the reference sequence. Strand Bias P-value assesses whether observed strand bias is likely due to chance or systematic effects. Variants with significant strand bias were filtered out to maintain data quality.

### *Influenza Hospitalization Data*

Weekly statewide and regional hospitalization data was downloaded directly from the Texas Respiratory Illness Interactive Dashboard (<https://texas-respiratory-illness-dashboard-txdshsea.hub.arcgis.com/>) on May 4<sup>th</sup>, 2024.

### **Acknowledgements:**

The authors thank the health departments and water utilities of all TEPHI network cities and counties for their support in the contribution of wastewater samples to the project.

### **Funding:**

This work was supported by S.B. 1780, 87th Legislature, 2021 Reg. Sess. (Texas 2021) (E.B., A.W.M., and J.F.P.), NIH/NIAID (Grant number U19 AI44297) (A.W.M.), Baylor College of Medicine Melnick Seed (A.W.M) and Alkek Foundation Seed (J.F.P.), and Pandemic Threat Technology Center (P.A.P.).

### **Author Contributions:**

Conceptualization: M.J.T., B.M.H., M.C.R., S.J.J-C., P.A.P., A.T., A.W.M., J.F.P., E.B.  
Methodology: M.J.T., B.M.H., K.P., L.W., M.C.R., S.J.J-C., V.A., A.W.M., J.F.P., E.B., F.W.  
Investigation: M.J.T., B.M.H., J.R.C., A.W.M.  
Visualization: M.J.T., B.M.H., J.R.C., A.W.M.  
Funding acquisition: B.M.H., J.F.P., J.D., E.B., A.W.M., P.A.P.  
Project administration: J.D., J.F.P., A.W.M.  
Supervision: M.J.T., M.C.R., J.F.P., E.B., A.W.M..  
Writing—original draft: M.J.T., B.M.H., J.R.C., A.W.M..  
Writing—review and editing: M.J.T., B.M.H., J.R.C., K.P., L.W., J.D., J.C., P.A.P., M.C.R., P.A.P., A.G., A.W.M., J.F.P., J.R., E.B., F.W.

### **Competing Interests:**

None

### **Data Availability:**

Read data from virome sequencing of all wastewater samples will be publicly available and deposited in SRA under project accession PRJNA966185.

H5N1 reads by sample are deposited at <https://zenodo.org/doi/10.5281/zenodo.11175923>.

### **References:**



1. Graziosi G, Lupini C., Catelli E., Carnaccini S. Highly Pathogenic Avian Influenza (HPAI) H5 Clade 2.3.4.4b Virus Infection in Birds and Mammals Animals 2024.
2. Lee DH, Bertran K, Kwon JH, Swayne DE. Evolution, global spread, and pathogenicity of highly pathogenic avian influenza H5Nx clade 2.3.4.4. *J Vet Sci* 2017;18:269-80.
3. Rohm C, Horimoto T, Kawaoka Y, Suss J, Webster RG. Do hemagglutinin genes of highly pathogenic avian influenza viruses constitute unique phylogenetic lineages? *Virology* 1995;209:664-70.
4. Lee DH, Criado MF, Swayne DE. Pathobiological Origins and Evolutionary History of Highly Pathogenic Avian Influenza Viruses. *Cold Spring Harb Perspect Med* 2021;11.
5. David E. Swayne DLS, Leslie D. Sims. *Influenza. Diseases of Poultry*. 14th ed: Wiley; 2019.
6. Kandeil A, Patton C, Jones JC, et al. Rapid evolution of A(H5N1) influenza viruses after intercontinental spread to North America. *Nat Commun* 2023;14:3082.
7. Pantin-Jackwood MJ, Swayne DE. Pathobiology of Asian highly pathogenic avian influenza H5N1 virus infections in ducks. *Avian Dis* 2007;51:250-9.
8. Puryear W, Sawatzki K, Hill N, et al. Highly Pathogenic Avian Influenza A(H5N1) Virus Outbreak in New England Seals, United States. *Emerg Infect Dis* 2023;29:786-91.
9. Elsmo EJ, Wunschmann A, Beckmen KB, et al. Highly Pathogenic Avian Influenza A(H5N1) Virus Clade 2.3.4.4b Infections in Wild Terrestrial Mammals, United States, 2022. *Emerg Infect Dis* 2023;29:2451-60.
10. Caliendo V, Lewis NS, Pohlmann A, et al. Transatlantic spread of highly pathogenic avian influenza H5N1 by wild birds from Europe to North America in 2021. *Sci Rep* 2022;12:11729.
11. Burrough ER, Magstadt DR, Petersen B, et al. Highly Pathogenic Avian Influenza A(H5N1) Clade 2.3.4.4b Virus Infection in Domestic Dairy Cattle and Cats, United States, 2024. *Emerg Infect Dis* 2024;30.
12. Disease Outbreak News; Avian Influenza A (H5N1) – the United States of America. 2024. (Accessed May 7, 2024, at <http://www.who.int/emergencies/disease-outbreak-news/item/2024-DON512>.)
13. H5N1 Bird Flu: Current Situation Summary. 2024. (Accessed May 7th, 2024, at <https://www.cdc.gov/flu/avianflu/avian-flu-summary.htm>.)
14. Technical Report: Highly Pathogenic Avian Influenza A(H5N1) Viruses. 2024. (Accessed May 7th, 2024, at [https://www.cdc.gov/flu/avianflu/spotlights/2023-2024/h5n1-technical-report\\_april-2024.htm](https://www.cdc.gov/flu/avianflu/spotlights/2023-2024/h5n1-technical-report_april-2024.htm).)
15. Marlene K. Wolfe DD, Bridgette Shelden, Elana M. G. Chan, Vikram Chan-Herur, Stephen Hilton, Abigail Harvey Paulos, Alessandro Zulli, Bradley J. White, Alexandria B. Boehm. Detection of hemagglutinin H5 influenza A virus sequence in municipal wastewater solids at wastewater treatment plants with increases in influenza A in spring, 2024. *medRxiv* 2024.
16. Clark JR, Terwilliger A, Avadhanula V, et al. Wastewater pandemic preparedness: Toward an end-to-end pathogen monitoring program. *Front Public Health* 2023;11:1137881.
17. Tisza M, Javornik Cregeen S, Avadhanula V, et al. Wastewater sequencing reveals community and variant dynamics of the collective human virome. *Nat Commun* 2023;14:6878.
18. Zhou JY, Shen HG, Chen HX, et al. Characterization of a highly pathogenic H5N1 influenza virus derived from bar-headed geese in China. *J Gen Virol* 2006;87:1823-33.

19. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884-i90.
20. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 2018;34:3094-100.
21. Camacho C, Coulouris G, Avagyan V, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
22. Johnson M, Zaretskaya I, Raytselis Y, Merezhuk Y, McGinnis S, Madden TL. NCBI BLAST: a better web interface. *Nucleic Acids Res* 2008;36:W5-9.
23. Bushnell B. BBMap: A Fast, Accurate, Splice-Aware Aligner. Lawrence Berkeley National Laboratory 2014.