

Pasteurized retail dairy enables genomic surveillance of H5N1 avian influenza virus in United States cattle

Andrew J. Lail¹, William C. Vuyk¹, Heather Machkovech¹, Nicholas R. Minor², Nura R. Hassan¹, Rhea Dalvie¹, Isla E. Emmen¹, Sydney Wolf¹, Annabelle Kalweit¹, Nancy Wilson¹, Christina M. Newman¹, Patrick Barros Tiburcio¹, Andrea Weiler², Thomas C. Friedrich^{2,3}, David H. O'Connor^{1,2*}

¹ School of Medicine and Public Health, University of Wisconsin, Madison, Wisconsin, United States of America

² Wisconsin National Primate Research Center, University of Wisconsin, Madison, Wisconsin, United States of America

³ School of Veterinary Medicine, University of Wisconsin, Madison, Wisconsin, United States of America

* Corresponding author

E-mail: dhoconno@wisc.edu

Abstract

Highly pathogenic avian influenza (HPAI) viruses with H5 hemagglutinin (HA) genes (clade 2.3.4.4b) are causing an ongoing panzootic in wild birds. Circulation of these viruses is associated with spillover infections in multiple species of mammals, including a large, unprecedented outbreak in American dairy cattle. With limited testing of dairy herds, there is an unmet need for genomic surveillance. Infected cattle can shed high amounts of HPAI H5N1 viruses in milk, allowing detection in pasteurized retail dairy samples. Over a 2-month sampling period in one Midwestern city, we obtained dairy products processed in 20 different states. Here we demonstrate that a tiled-amplicon sequencing approach produced over 90% genome coverage at greater than 20x depth from 5 of 13 viral RNA positive samples. A combination of RT-qPCR testing and sequencing from retail dairy products can be a useful component of a One Health framework for responding to the avian influenza outbreak in cattle.

Introduction

The One Health approach, as defined by the One Health High Level Expert Panel (OHHLEP), “recognizes the health of humans, domestic and wild animals, plants, and the wider environment (including ecosystems) are closely linked and interdependent” (1). Recent epidemic and pandemic zoonoses have underscored the relevance and utility of One Health approaches to the study of emerging infectious disease (2,3). The OHHLEP emphasizes the importance of communication, coordination, collaboration, and capacity building (the 4 Cs). Consistent with these guidelines, we developed an approach designed to build capacity for H5N1 genomic surveillance using retail milk as a pooled substrate, somewhat similar to wastewater, that is readily available for sampling, even when testing at individual farms or dairy processors is challenging.

In 2021, the HPAI clade 2.3.4.4b HA was introduced to North America as part of a multi-species panzootic affecting both birds and mammals, which has resulted in millions of wild bird and poultry deaths (4). An ongoing outbreak of HPAI (H5N1) B3.13 genotype in dairy cattle was confirmed on March 25, 2024 by the National Veterinary Services Laboratories (NVSL) following reports of decreased milk yield from multiple states (5). Phylogenetic analysis indicates this outbreak likely originated from a single introduction of an avian H5N1 virus into cattle, with subsequent spread facilitated by inter-farm movements of cattle, and potentially of people (6). Exposure to cattle has resulted in multiple human cases of HPAI infection in dairy workers and dozens of feline HPAI infections (5,7,8). Human infections have also been observed in poultry workers after presumed spillover of HPAI into poultry from cattle, underscoring the zoonotic potential of this outbreak (9).

It is made available under a [CC-BY 4.0 International license](#) .

H5N1 infection in dairy cattle can occur within the mammary glands (5), resulting in viral titers in milk of at least 10^7 TCID₅₀ per mL (10). In commercial milk production, milk from cows on a farm is collected into a refrigerated temporary storage tank (bulk tank), then transported by a milk-hauling truck to a processing plant. Depending on the scale of the plant, milk from hundreds to thousands of animals may be combined. This milk then undergoes clarification, standardization, pasteurization (typically High Temperature Short Time pasteurization at 72°C for 15 seconds), homogenization, and packaging. Despite the dilution effect of pooling and potential degradation during pasteurization, H5N1 vRNA remains detectable in retail products. This is perhaps due to the known RNA stabilizing properties of milk (11,12).

The presence of H5N1 vRNA in commercial pasteurized milk presents an opportunity for genomic surveillance. Several groups have used quantitative reverse transcriptase polymerase chain reaction (RT-qPCR) assays to test for H5N1 vRNA in commercial milk, observing geographic correlation to the known outbreak in cattle at the time (13,14). Buying milk from grocery stores as a sampling method leverages the national distribution network of dairy products to provide a cost-effective, broad-scale surveillance strategy. Because pasteurization neutralizes infectious virus, H5N1 genomic surveillance using pasteurized dairy products can be performed in any BSL1 facility (15).

Although states like Colorado and California have implemented regular bulk-tank testing of herds regardless of clinical signs (16,17), most affected states do not have ongoing surveillance of farms in the absence of symptoms, although this may soon change as USDA implements new policy (18). The lack of on-farm testing creates a surveillance gap, hampering efforts to contain spread through early detection. Retail milk testing allows surveillance in regions without on-farm testing, but studies so far have focused more on testing for vRNA presence than on sequencing due to a lack of existing approaches for sequencing influenza from retail milk.

Here we aimed to determine whether we could detect and sequence vRNA in retail dairy products from states with known infections. We also sampled widely within Wisconsin (the home state of the researchers on this study and the second-largest dairy producing state in the US), where, as of December 2024, there has not been a documented outbreak. Over a two-month period, we screened 66 commercial dairy products processed in 20 states using RT-qPCR. Of these, 13 samples from 5 states were positive. To sequence positive samples we developed a tiled-amplicon approach to amplify fragmented influenza genome products, resulting in sequence for 5 of 13 positive samples at over 90% genome coverage. This tiled-amplicon approach was designed to work with existing workflows and reagents from our work in SARS-CoV-2 sequencing, enabling rapid response to a

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

new pathogen of interest. Consistent with OHHLEP's 4 Cs, this strategy can provide a means to build capacity for HPAI genomic surveillance even when there is a lack of regional testing of individual farms or dairy processors. Furthermore, because HPAI infectivity is inactivated by pasteurization, our approach does not require access to high-containment laboratories in which to process raw milk. This study demonstrates that retail dairy products are a readily accessible sample type for HPAI genomic surveillance in cattle, and provides an example of how the OHHLEP's One Health approach recommendations can be applied in the rapid response to an emerging zoonotic pathogen.

Methods

Sample collection

Sixty-six cartons of pasteurized retail dairy products (whole and low-fat milk, whipping cream, coffee creamer, yogurt, kefir, and buttermilk) were purchased from 11 different stores in the Madison, WI, USA metro area from April 24 to June 17, 2024. Commercial dairy products are labeled with a plant code that includes the Federal Information Processing Standards code for the state where the dairy processing plant is located, combined with a unique number for that specific plant. This code can be used to trace the location of processing using the FDA's Interstate Milk Shippers list. The processing location does not necessarily correspond to the herd location, but it is the most accurate proxy for the herd location available at the point of sale. While we primarily collected fluid milk, we also bought half and half, cream, kefir, and yogurt if fluid milk was not available from a given plant code. We selected dairy products for RT-qPCR and sequencing based on processing plant code. Initially, samples were collected aiming for the broadest geographic sampling possible, with particular interest in states that had known active outbreaks. After successfully isolating and detecting vRNA, we aimed to resample from the same processing plants whose milk had detectable H5N1 HPAI vRNA in our initial testing. We also intensified our geographic focus on Wisconsin, testing as many processors as possible to enable early warning in this leading dairy state.

We took 50 mL aliquots of product from the original container, recording the plant code, expiration date, and type of dairy product. Products that were too viscous to pipette were diluted 1:1 in nuclease-free water. All products were kept at 4°C throughout this process.

RNA extraction

We have previously reported our protocol to extract viral RNA from milk on protocols.io (19).

We followed the MagMAX Wastewater Ultra Nucleic Acid Isolation Kit (Applied Biosystems) protocol on a Kingfisher Apex instrument (ThermoFisher Scientific) using 400 μ L of neat or dilute dairy product as the input sample. Each sample was isolated in replicates of 6. Samples were eluted in 50 μ L of elution buffer. We also isolated from 400 μ L of water, which we treated as a sample in all downstream processing steps (Control A).

After isolation, one replicate of each extracted RNA sample was directly used for RT-qPCR, and the other 5 replicates were pooled for cleaning and concentrating. The pooled samples were incubated at 37°C for 30 minutes with 20 μ L of 10X Turbo DNase buffer and 2 μ L of TURBO DNase (ThermoFisher Scientific). Then, samples containing vRNA were cleaned and concentrated using the RNA Clean and Concentrator-5 kit (Zymo Research) following the manufacturer's protocol, without the optional DNase I treatment. We also included a water control in this step (Control B). Since we used a large input volume, we increased the final elution volume from 15 to 30 μ L. Eluted RNA was then used in the reverse transcriptase reaction for sequencing.

Detection of HPAI using RT-qPCR

We used RT-qPCR to detect HPAI vRNA. Initially, we used a previously published assay targeting the influenza M gene segment (20). Any samples that tested positive (defined as a Ct less than 39) by the IAV M-gene assay were re-tested using the HPAI-specific assay described below.

For more specific HPAI detection, primer/probe sets were designed using Integrated DNA Technologies' PrimerQuest™ Tool (<https://www.idtdna.com/pages/tools/primerquest>) to target a consensus sequence of 6 randomly selected HA sequences from the HPAI outbreak in cattle (SRR28834851, SRR28834852, SRR28834853, SRR28834854, SRR28834855, SRR28834884) that we sourced from the Andersen Lab's avian-influenza github repository (<https://github.com/andersen-lab/avian-influenza>). The primer sequences were 5'-GGGAAGCTAT-GCGACCTAAAT-3' (forward) and 5'-CATTCGGCACTCTGATGAA-3' (reverse). The probe sequence was 5'-ACATTGGGTTTCCGAGGAGCCATC-3'. The primers and probes were aligned *in silico* to ensure they would bind to all reported 2.3.4.4b HA currently circulating in birds and to ensure they were specific to 2.3.4.4b HA (Supplemental data 1).

Our RT-qPCR assays used TaqMan Fast Virus 1-Step Master Mix (Thermo Fisher Scientific) with primers at a final concentration of 600 nM and probe at a final concentration of 100 nM. The RT-qPCR was conducted using the LightCycler® 480 System (Roche Diagnostics) with cycling conditions of 50°C for 5 minutes, 95°C

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

for 20 seconds, followed by 50 cycles of 95°C for 15 seconds and 60°C for one minute. Samples were tested in duplicate and considered positive when the average Ct was less than 39. For the HPAI-specific assay, absolute concentration within the reaction was determined by interpolation onto a standard curve with known concentrations of synthetic H5 dsDNA manufactured and quantified by IDT. The standard curve ranged from 1.93×10^6 to 1.93 copies per reaction.

Reverse Transcription

We performed RT-PCR, amplification, and library prep as described in our publicly available protocol (<https://dx.doi.org/10.17504/protocols.io.kqdg322kpv25/v1>). We used reagents from the QIAseq DIRECT SARS-CoV-2 Kit A and QIAseq DIRECT SARS-CoV-2 Enhancer (Qiagen). For each sample, we assembled a reaction master mix on ice with RP Primer (1 μ L), Multimodal RT Buffer (4 μ L), RNase Inhibitor (1 μ L), nuclease-free water (8 μ L), and EZ Reverse Transcriptase (1 μ L). We added 5 μ L of cleaned RNA sample to this reaction. The thermal cycling conditions were 25°C for 10 minutes, 42°C for 10 minutes, 85°C for 5 minutes, followed by a hold at 4°C. At this point we also added a no-template control sample consisting of nuclease-free water (Control C). After reverse transcription, samples were taken directly to amplicon PCR.

Amplicon Generation

We generated a primer scheme for tiled amplicon sequencing using PrimalScheme (21). As our input to PrimalScheme, we used 239 sequences cattle sequences (S1 Table) from the Anderson lab's Github repository (<https://github.com/andersen-lab/avian-influenza>), which in turn sources sequences from the USDA's collection of sequences related to the cattle outbreak (BioProject accession number PRJNA1102327 in the NCBI BioProject database, <https://www.ncbi.nlm.nih.gov/bioproject/>). Duplicate sequences within each segment were removed to avoid biasing. The amplicon size was 250 bp. Each segment was input separately. The algorithm did not identify primer sequences at the 5' and 3' termini of each segment, so we manually designed additional primers to capture the ends of each segment. At the time primer sets were generated there were no sequences available for the complete noncoding regions (NCRs) in any gene segments from H5N1 viruses causing the cattle outbreak, so we inferred likely NCR sequences based on publicly available sequences of clade 2.3.4.4b H5N1 viruses isolated from North American birds and mammals. We designed 5' and 3' terminal primers for each segment using approaches we have described previously (22), which have recently been adapted by Jaeger et al. (23). The primer pools were solubilized in water to a stock concentration of 100 μ M total, then diluted to a

working concentration of 10 uM total.

Amplicon Sequencing PCR

We set up two polymerase chain reactions for each sample (one for each primer pool) using reagents from QIAseq DIRECT SARS-CoV-2 Kit A and QIAseq DIRECT SARS-CoV-2 Enhancer according to Table 1.

Table 1. PCR set up for tiled amplicon strategy.

Reagent	Amount	Final Concentration
cDNA from sample	8 uL	
Primer pool, 10 μ M	4 uL	1.6 uM
UPCR buffer, 5X	5 uL	1X
QN Taq Polymerase	1 uL	
Nuclease-free water	7 uL	

The cycling conditions were 98°C for 2 minutes, then 4 cycles of 98°C for 20 seconds and 63°C for 5 minutes, followed by 29 cycles of 98°C for 20 seconds and 63°C for 3 minutes with a final hold at 4°C. A Qubit fluorometer (ThermoFisher) was used to quantify the double-stranded DNA after the reaction.

Library Preparation and Sequencing

We prepared our library according to the protocol for the Oxford Nanopore Technology Native Barcoding Kit V14 (24). Nuclease-free water was included in the library preparation as a negative control for this step (Control D).

The library was sequenced on a minION mk1C sequencer using an R10 flow cell.

Sequencing analysis methods

To perform Oxford Nanopore base-calling, primer-trimming, and variant-calling, we developed a bespoke Nextflow pipeline available at <https://github.com/dholab/Lail-et-al-2024-analysis-pipeline> (25). The pipeline handles parallelizing tasks, running reproducible software containers via Docker, and providing a configurable command-line interface to the underlying tools it implements. These tools include the Oxford Nanopore Dorado base-caller v0.7.1, CutAdapt v4.8 for primer and adapter trimming, minimap2 v2.28 for read alignment, iVar v1.4.2 for variant-calling, and snpEff v5.2 for variant-effect annotation. Utilities in SeqKit2 v2.8.2, rusage v2.0.0, vsearch v2.28.1, and samtools v1.5 are used throughout the pipeline, alongside other tools (26–33). A full list of the pipeline’s dependencies can be found in the GitHub repository’s Python project configuration file,

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

pyproject.toml, available at <https://github.com/dholab/Lail-et-al-2024-analysis-pipeline/blob/main/pyproject.toml>. In short, the core pipeline steps for each sample are: base-calling with Dorado, trimming reads by quality and selecting those reads that span complete amplicons, mapping to a reference genome, and consensus-sequence- and variant-calling.

Reads for this manuscript were mapped to A/Bovine/texas/24-029328-01/2024(H5N1), a sequence isolated from an infected cow's milk early in the outbreak and reported by Burroughs et al (8). We added synthetic overhangs to our reference sequence to enable our primers to map to the reference. This is necessary since our end primers extend beyond the sequence available in the reference sequence. We used FLAN to annotate open reading frames within our custom reference to allow for mutations to be called at the amino acid level (34). For our analysis parameters, we used the containerless profile with a minimum read length cutoff of 150 and a maximum read length cutoff of 500, along with a minimum read quality cutoff of 10. We allowed up to 2 primer mismatches between a given primer and the reference sequence. Variants and consensus sequences were called using a minimum required depth of coverage of 20 reads and a minimum variant frequency of 0.1.

Sequences were uploaded to Nextclade ((35), <https://clades.nextstrain.org>) for analysis against other publically available sequences. HA gene segments were uploaded and analyzed against the H5Nx clade 2.3.4.4 tree (community/moncla-lab/iav-h5/ha/2.3.4.4). We used the dataset specific to the cattle outbreak (avian-flu/h5n1-cattle-outbreak/genome) to analyze the sequence of all 8 genome segments concatenated together. This dataset contains sequences from individual cattle as well as samples from other domestic and peri-domestic animals. Additionally, we used the nextstrain H5N1 cattle outbreak tree to understand how common the mutations we detected were. We filtered by genotype at a specific amino acid and recorded how many sequences contained the mutation out of the total number of sequences in the tree. To expand this analysis to avian influenza sequences broadly, not just those related to the cattle outbreak, we used the same strategy to count mutations present in the segment-specific H5N1 trees from the past two years (eg. for PB2, avian-flu/h5n1/pb2/2y).

We used the Augur pipeline's traits function to predict the processing state of our samples (36). This function is designed to provide an estimation of missing metadata based on metadata contained within the rest of a tree's nodes. To run the tool, we downloaded the H5N1 cattle outbreak nextclade tree with our sequences included and used that as input to the traits function.

Data Sharing and Availability

We created a github repository (github.com/dholab/dairy-hpai-monitoring) where RT-qPCR testing results

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

were shared with the scientific community in near-real-time. Results are reported at the state level only, balancing the need for some geographic information about viral spread with the need to maintain privacy for individual producers. Retail dairy purchase dates, expiration dates, Ct values, processor state, and SRA accession numbers are included as metadata associated with each sample.

Reads were uploaded to NCBI's Sequence Read Archive under the Bioproject accession PRJNA1121320. Accession numbers are available in S4 Table.

Results

RT-qPCR detects H5N1 vRNA in pasteurized dairy products

During a two-month surveillance period from April 24 to June 17, 2024, we tested 66 pasteurized commercial dairy products processed in 20 different states across the United States. At the end of the sampling period, the following states had outbreaks confirmed by NVSL: Texas, Kansas, Michigan, New Mexico, Ohio, Idaho, South Dakota, North Carolina, Colorado, Minnesota, Wyoming, and Iowa. Thirteen positive samples were identified in products processed from five states: Colorado, Indiana, Kentucky, Michigan, and Texas (S3 Table). It is important to note that the proportion of positive samples is not unbiased; some producers' products were tested repeatedly upon identifying H5+ genetic material. Three of the positives that we identified were sequenced as part of our method optimization and performed poorly. We did not attempt to resequence them with the final sequencing protocol. Full testing results are available in S3 Table.

Since our initial determination of positive and negative were based only on Ct values, we re-tested freshly isolated vRNA from all of our previously positive samples in a single RT-qPCR assay to quantify vRNA concentrations. The viral RNA copy numbers ranged from 83-31,500 copies per mL of dairy product, with a median of 683 copies per mL (Table 2).

Table 2: Geographic origin, qPCR results, and sequencing coverage for samples and controls sequenced in this study.

Sample Name	State	Purchase Date	Expiration Date	Copies per mL dairy product	% Sequence coverage at 20X depth	Reads mapped
Carton 05	CO	4/24/2024	6/20/2024	778	90.0	331,732
Carton 06	CO	4/24/2024	5/19/2024	375	84.5	236,964
Carton 21	TX	5/2/2024	6/4/2024	1233	89.2	272,222
Carton 24	MI	5/2/2024	7/11/2024	31500	95.7	1,950,789
Carton 33	CO	5/20/2024	9/12/2024	^a	65.8	71,809
Carton 36	CO	5/20/2024	7/8/2024	83	44.2	36,290

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/) .

Carton 46	CO	6/5/2024	8/16/2024	291	92.7	753,937
Carton 48	MI	6/5/2024	8/22/2024	683	74.3	21,685
Carton 63	MI	6/18/2024	8/29/2024	651	92.7	760,342
Carton 65	CO	6/18/2024	8/23/2024	3613	92.6	1,016,270
Control A					0	171
Control B					0	126
Control C					0	104
Control D					0	119

^aCarton 33 was initially tested by RT-qPCR with the M-gene assay. In that assay, the Ct was 38. When it was re-tested with the HPAI-specific assay, no amplification was observed. We still attempted to sequence it because of the M-gene assay result.

Tiled-amplicon sequencing covers all H5N1 gene segments

We developed a tiled amplicon approach to analyze H5N1 HPAI vRNA in pasteurized commercial dairy products. The method yielded an overall average sequencing depth of 8,176x across all samples, with individual samples varying from 325x to 29,256x. This correlated with the input vRNA amount, with less input vRNA yielding lower depth in sequencing. The average depth across all samples was lowest for the NA segment, with a depth of 199x. The average depth was highest for the PB1 gene segment with a depth of 419x. Regions within the same gene segment may be sequenced to different depths because amplification efficiency may differ by amplicon. We used a 20x depth cutoff at a given position to determine the percent of the segment that was covered by our sequencing (Table 3).

Table 3: Percent coverage at 20x depth by segment for each sample.

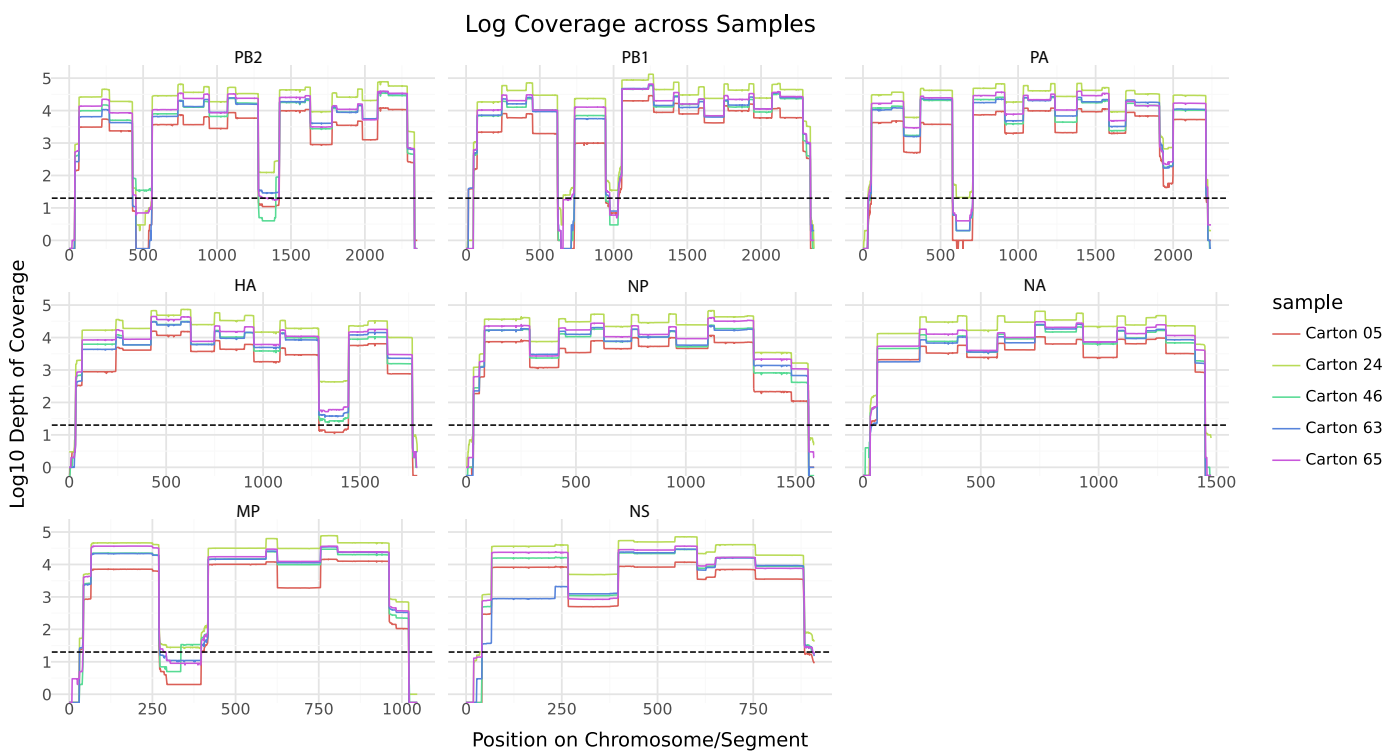
Segment	Carton 05	Carton 06	Carton 21	Carton 24	Carton 33	Carton 36	Carton 46	Carton 48	Carton 63	Carton 65
PB2	90%	88%	88%	95%	76%	33%	88%	77%	91%	91%
PB1	86%	86%	86%	93%	42%	35%	93%	72%	92%	91%
PA	90%	87%	90%	98%	71%	44%	91%	82%	91%	92%
HA	88%	88%	88%	97%	81%	63%	97%	69%	97%	97%
NP	96%	67%	96%	96%	52%	52%	96%	77%	96%	96%
NA	96%	94%	94%	96%	62%	62%	96%	68%	96%	96%
M	82%	80%	80%	95%	75%	20%	87%	74%	83%	83%
NS	93%	78%	93%	96%	78%	46%	95%	71%	95%	95%
Overall	90%	85%	89%	96%	66%	44%	93%	74%	93%	93%
Copies/mL	778	375	1233	31500	^a	83	291	683	651	3613

^aCarton 33 was initially tested by RT-qPCR with the M-gene assay. In that assay, the Ct was 38. When it was re-tested with the HPAI-specific assay, no amplification was observed. We still attempted to sequence it because of the M-gene assay result.

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Coverage per segment ranged from 20% to 98%. We considered any segments over 90% coverage in subsequent analyses. We chose this cutoff to allow for some amplicon failure while still retaining enough of the sequence to allow robust comparisons with other sequences. With this threshold we obtained passable sequences for all the segments of cartons 21, 24, 46, 63, and 65 (Fig 1). Carton 21 was very close to the quality cutoff at 89%, along with carton 6 at 85%. Several individual segments in carton 21 were at greater than 90% coverage, but only one segment from carton 6 (NA) reached this threshold. No individual segments from cartons 33, 36, or 48 had coverage at over 90%.

Fig 1: Genomic coverage of samples at 20X depth. Coverage at each base is plotted per-segment, with each color line representing a different carton. The vertical axis is given in log base 10, with a dashed line at $\log_{10}(20)$ to indicate our depth cutoff.



Sequences cluster within reported outbreak sequences

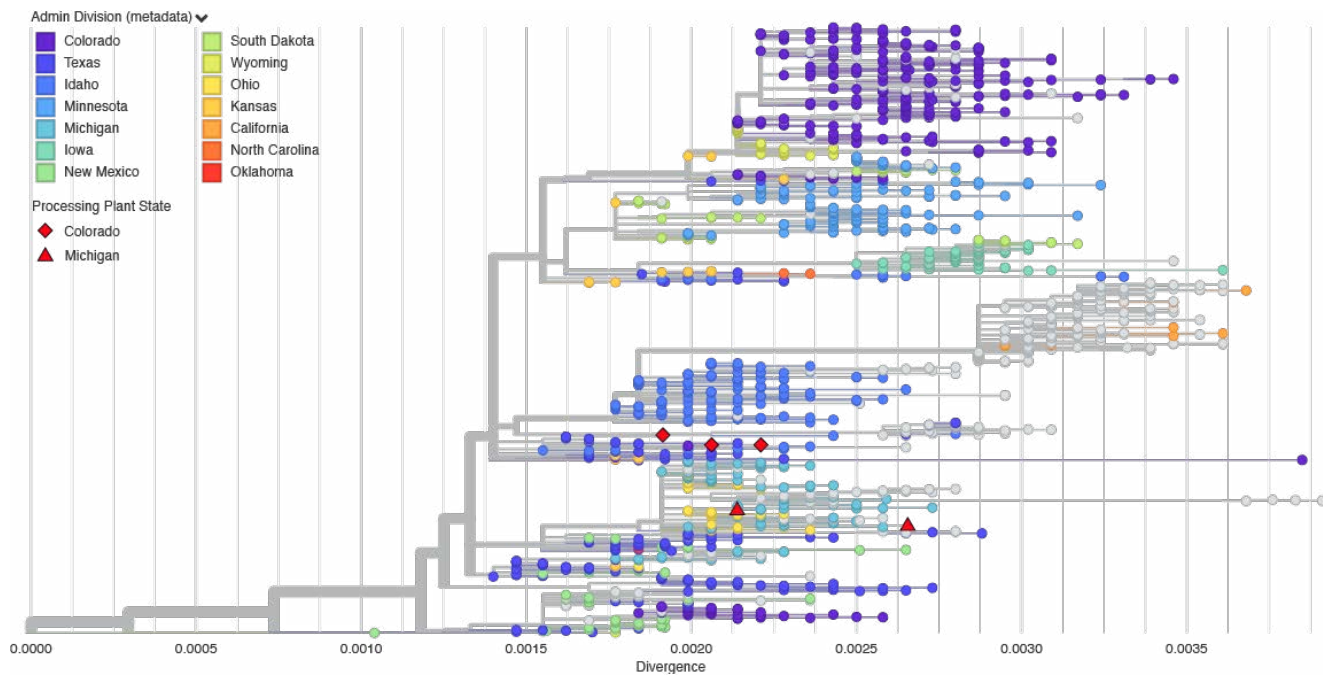
Initially, we confirmed that our H5 consensus sequences were clade 2.3.4.4b HA by clustering with the H5Nx clade 2.3.4.4 dataset. All of our HA sequences were contained within the clade 2.3.4.4b lineage (S1 Fig).

We also wanted to confirm that our consensus sequences were contained within the diversity of other sequences from the HPAI outbreak in cattle. All of our samples clustered within the known diversity of the current North American H5N1 outbreak (Fig 2). We anticipated that our sequences would cluster closely with sequenc-

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

es from their processing plant state. For example, we expected viruses in milk processed in Michigan to cluster with outbreak sequences from Michigan. We masked our state-level metadata and used Augur's "traits" function to infer the original processing state (36). It provided the correct state for three out of five samples.

Fig 2: Phylogenetic clustering in Nextclade. The consensus sequences from each dairy product are clustered against the inferred ancestral sequence using the <https://nextstrain.org/avian-flu/h5n1-cattle-outbreak/genome> dataset in Nextclade.



Few major amino acid substitutions were detected

We next identified major (consensus-defining) and minor (sub-consensus) mutations present in each sequence we generated, relative to the A/Bovine/texas/24-029328-01/2024(H5N1) reference. We detected a total of 32 major mutations across all our samples, of which 8 were non-synonymous (Table 4). We did not detect non-synonymous mutations in HA in any of these sequences. A major mutation we observed was PB2:A255V, which was present in all five samples. In this case, the alanine in the reference is actually relatively rare, with the majority of cattle outbreak sequences (99.4%) and H5N1 avian influenza (80%) containing PB2:255V (Table 4). We identified PB1:V171M in two of our sequences, a relatively rare mutation through the rest of the cattle outbreak. PB1:V171M is much more common throughout avian influenza sequences worldwide (72%). In PA, we found PA:Y24H, PA:A36T, and PA:I596V, all of which are present in less than 0.5% of the rest of the cattle sequences. Of these, PA:A36T has been noted as a potentially virulence-enhancing mutation in mice (37).

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

NA:S71N is present in over 30% of cattle outbreak sequences. This mutation is associated with modulating host-range (6).

We used a threshold of 10% frequency to identify any sub-consensus mutations present in our sequences. There was only one sub-consensus mutation of note. We observed HA:P324L (H3 numbering) in carton 24 at a 36% frequency, which is a mutation implicated in increased virulence (6).

Table 4: Amino acid variants at the consensus level in each sample.

Gene	AA Position	Reference	Carton 5	Carton 24	Carton 46	Carton 63	Carton 65	% of nextstrain H5N1 cattle outbreak sequences with mutation
PB2	255	A	V	V	V	V	V	99.38%
PB1	171	V	V	V	M	V	M	0.19%
PA	24	Y	Y	Y	H	Y	H	0.43%
PA	36	A	A	A	A	T	A	0.12%
PA	596	I	I	I	I	V	I	0.00%
NS	79	M	M	M	M	I	M	0.25%
MP	28	I	T	I	I	I	I	5.49%
NA	71	S	S	N	S	N	S	30.29%

These variants are relative to A/Bovine/texas/24-029328-01/2024(H5N1).

Discussion

The HPAI H5N1 2.3.4.4b virus currently spreading in dairy cows is a prime example of pathogen exchange between wild animals, domestic animals, and humans, as the One Health concept reminds us (1). The outbreak is concerning due to its potential for mammalian adaptation and spillover while circulating in a domestic mammal with close proximity to humans. Furthermore, there is still a need for expanded genomic surveillance. Each state sets its own testing policies, so regulations for testing dairy cattle and/or milk vary. Although only 14 states have reported dairy herds infected with HPAI H5N1 viruses (as of November 2024), those states are distributed throughout the US; only California and Colorado are conducting regular surveillance testing for HPAI H5N1 viruses in milk, although there have been some recent moves towards nationwide testing of bulk tanks (18).

Retail dairy sampling is an orthogonal approach to current sporadic bulk tank testing. Previous studies established that HPAI H5N1 vRNA can be detected in retail dairy products using RT-qPCR. Here we show that such retail dairy products can also be used to generate near-complete HPAI H5N1 viral genome sequences suitable for phylogenetic and evolutionary analyses. In line with the OHHLEP's 4 Cs (communication, coordination, collaboration, and capacity building), we have shared all of our protocols and results publicly throughout this study such that researchers at any properly equipped BSL-1 facility could obtain retail dairy samples from a

local grocery store and contribute to HPAI genomic surveillance.

Limitations of sampling retail milk for HPAI H5N1

Only a small fraction of nationwide dairy production is allocated to fluid milk. According to the USDA Economic Research Service, "most of the milk supply is used to produce manufactured dairy products" like cheese and butter (38). For example, in Wisconsin 90% of fluid milk is used for cheese production (39). This reduces the effectiveness of retail milk sampling for surveillance purposes, because some farms and dairy processors produce little or no product that could be sampled as retail fluid milk. We have attempted to detect and sequence H5N1 vRNA in other dairy products, but all of those have potential pitfalls. Many manufactured dairy products contain multiple dairy ingredients, like cream, milk powder, or whey. These ingredients may be sourced from multiple geographically distant processing plants, obfuscating the original geographic origin of the dairy material containing HPAI H5N1 RNA. We operate under the assumption that retail milk is likely processed at a plant that is close to its origin farm, since it is not economically viable to haul milk over long distances (40).

Additionally, consensus sequences obtained from retail dairy may represent viruses from a single cow or, more likely, multiple animals. This limits the utility of phylogenetic analyses, which assume that each sequence represents a single infected individual. Various approaches have been used to deconvolute inputs from multiple infections in other pooled samples, but given the short length of our reads and the relatively low genetic diversity within the current HPAI H5N1 cattle outbreak, these are not helpful for deconvoluting our dataset (41–43).

In our phylogenetic analysis, HPAI sequences from retail dairy samples were not always part of the same lineage of HPAI sequences as infected cattle from the same state. There are several potential explanations for this. Due to the fairly recent common ancestor of the outbreak, the existing variation within the HPAI outbreak in cattle is constrained, limiting our ability to confidently assign lineages. In addition, while some portion of the HPAI cattle sequences do cluster by state, many sequences do not. Moreover, there may be movement of virus among states, increasing within state virus diversity (reducing between state diversity) and reducing our ability to accurately identify state of origin for a given sample. Movement of virus between states was more likely to occur before interstate testing requirements were introduced (44). Because our samples were all purchased within 2 months of those requirements being introduced, it is possible that interstate transmission makes our samples more difficult to place in this analysis.

Open science for emerging threats

Our team's ability to quickly adapt to sequencing HPAI from the emerging cattle outbreak was due in part to our ability to pivot existing tools, protocols, and collaborations established for SARS-CoV-2 genomic monitoring projects. We repurposed the reagent kits we had on hand for SARS-CoV-2 sample processing and sequencing. This allowed us to process new samples almost immediately with protocols that were already established. Additionally, since retail dairy sampling can be done in BSL-1 facilities, this work could be democratized to laboratories that do not have containment facilities required to handle pathogenic HPAI H5N1.

Public data sharing networks strengthened by the COVID-19 pandemic facilitated our use of initial outbreak sequences posted by colleagues. Access to these sequences was necessary for the rapid generation of a whole-genome tiled-amplicon primer scheme with PrimalScheme, which allows for automatic generation of a tiled-amplicon primer scheme of arbitrary length based on a reference fasta sequence (21). Shared data analysis platforms like Nextstrain, and the data submitters that power them (Supplemental Table 5) were also key to understanding the sequencing results.

Transitioning from SARS-CoV-2 sequence surveillance to HPAI sequence surveillance speaks broadly to the importance of maintaining inter-pandemic expertise, the use and development of flexible multi-pathogen lab workflows, and networks of open data and protocol sharing, as has been encouraged by the OHHLEP and the CDC SPHERES Consortium (1,45).

Given the diverse geographic origins of milk available at our local grocery stores, it is possible that additional sampling hubs could sample the majority of retail fluid milk processors nationwide. There are several states with high milk production that our group has been unable to sample rigorously, like California, Texas, Idaho, and New York. This concept is already being demonstrated in Canada, where a network has been formed to sample retail milk longitudinally to detect HPAI in dairy cattle (46).

Finally, it is worth considering what other types of environmental sampling might be valuable for monitoring future zoonotic risks posed by livestock. Would consistent environmental sampling of wastewater or air at pig farms be helpful for detecting new variants of avian influenza viruses? What other grocery store or restaurant products could be used for pathogen monitoring? The unexpected detection of HPAI H5N1 in retail milk and the utility of this data for viral sequencing highlights how unorthodox approaches to genomic epidemiology can complement conventional "shoe-leather" epidemiology based on testing individual animals.

References

1. Panel (OHHLEP) OHHLE, Adisasmito WB, Almuhairi S, Behraves CB, Bilivogui P, Bukachi SA, et al. One Health: A new definition for a sustainable and healthy future. *PLOS Pathog*. 2022 Jun 23;18(6):e1010537.
2. Rahman MT, Sobur MA, Islam MS, Ievy S, Hossain MJ, El Zowalaty ME, et al. Zoonotic Diseases: Etiology, Impact, and Control. *Microorganisms*. 2020 Sep;8(9):1405.
3. Morens DM, Folkers GK, Fauci AS. The challenge of emerging and re-emerging infectious diseases. *Nature*. 2004 Jul;430(6996):242–9.
4. 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 May 1;14(9):1372.
5. Hu X, Saxena A, Magstadt DR, Gauger PC, Burrough E, Zhang J, et al. Highly Pathogenic Avian Influenza A (H5N1) clade 2.3.4.4b Virus detected in dairy cattle [Internet]. *bioRxiv*; 2024 [cited 2024 Jul 3]. p. 2024.04.16.588916. Available from: <https://www.biorxiv.org/content/10.1101/2024.04.16.588916v1>
6. Nguyen TQ, Hutter C, Markin A, Thomas M, Lantz K, Killian ML, et al. Emergence and interstate spread of highly pathogenic avian influenza A(H5N1) in dairy cattle [Internet]. *bioRxiv*; 2024 [cited 2024 Jun 6]. p. 2024.05.01.591751. Available from: <https://www.biorxiv.org/content/10.1101/2024.05.01.591751v1>
7. Garg S, Reed C, Davis CT, Uyeki TM, Behraves CB, Kniss K, et al. Outbreak of Highly Pathogenic Avian Influenza A(H5N1) Viruses in U.S. Dairy Cattle and Detection of Two Human Cases - United States, 2024. *MMWR Morb Mortal Wkly Rep*. 2024 May 1;73(21):501–5.
8. Burrough ER, Magstadt DR, Petersen B, Timmermans SJ, Gauger PC, Zhang J, 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 Apr 29;30(7).
9. Drehoff CC. Cluster of Influenza A(H5) Cases Associated with Poultry Exposure at Two Facilities — Colorado, July 2024. *MMWR Morb Mortal Wkly Rep* [Internet]. 2024 [cited 2024 Oct 7];73. Available from: <https://www.cdc.gov/mmwr/volumes/73/wr/mm7334a1.htm>
10. Guan Lizheng, Eisfeld Amie J., Pattinson David, Gu Chunyang, Biswas Asim, Maemura Tadashi, et al. Cow's Milk Containing Avian Influenza A(H5N1) Virus — Heat Inactivation and Infectivity in Mice. *N Engl J Med* [Internet]. [cited 2024 Jun 5];0(0). Available from: <https://www.nejm.org/doi/full/10.1056/NEJM2405495>
11. Izumi H, Kosaka N, Shimizu T, Sekine K, Ochiya T, Takase M. Bovine milk contains microRNA and mes-

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

- senger RNA that are stable under degradative conditions. *J Dairy Sci.* 2012 Sep 1;95(9):4831–41.
12. Izumi H, Tsuda M, Sato Y, Kosaka N, Ochiya T, Iwamoto H, et al. Bovine milk exosomes contain microRNA and mRNA and are taken up by human macrophages. *J Dairy Sci.* 2015 May 1;98(5):2920–33.
 13. Spackman E, Jones DR, McCoig AM, Colonius TJ, Goraichuk I, Suarez DL. Characterization of highly pathogenic avian influenza virus in retail dairy products in the US. 2024 May 22 [cited 2024 Jun 7]; Available from: <https://www.medrxiv.org/content/10.1101/2024.05.21.24307706v1>
 14. Luring AS, Martin ET, Eisenberg M, Fitzsimmons WJ, Salzman E, Leis AM, et al. Surveillance of H5 HPAI in Michigan using retail milk [Internet]. *bioRxiv*; 2024 [cited 2024 Aug 2]. p. 2024.07.04.602115. Available from: <https://www.biorxiv.org/content/10.1101/2024.07.04.602115v1>
 15. Spackman E, Anderson N, Walker S, Suarez DL, Jones DR, McCoig A, et al. Inactivation of Highly Pathogenic Avian Influenza Virus with High-temperature Short Time Continuous Flow Pasteurization and Virus Detection in Bulk Milk Tanks. *J Food Prot.* 2024 Oct 1;87(10):100349.
 16. Colorado State Veterinarian Now Requiring HPAI Testing of Commercial Dairy Cow Operations | Department of Agriculture [Internet]. [cited 2024 Oct 9]. Available from: <https://ag.colorado.gov/press-release/colorado-state-veterinarian-now-requiring-hpai-testing-of-commercial-dairy-cow>
 17. Healthy Dairy Cattle HPAI Testing Protocols for Monitored Herd, Pre-Movement, and Non-Monitored Herd Surveillance FOR PRODUCERS [Internet]. [cited 2024 Oct 18]. Available from: https://www.cdffa.ca.gov/AHFSS/Animal_Health/docs/surveillance_testing_of_dairy_cattle_protocol-for_producers.pdf
 18. USDA Builds on Actions to Protect Livestock and Public Health from H5N1 Avian Influenza | Animal and Plant Health Inspection Service [Internet]. [cited 2024 Nov 19]. Available from: <https://www.aphis.usda.gov/news/agency-announcements/usda-builds-actions-protect-livestock-public-health-h5n1-avian-influenza>
 19. Lail A, Vuyk W, Emmen I, O'Connor D. RNA extraction from milk for HPAI surveillance. 2024 May 10 [cited 2024 Jul 25]; Available from: <https://www.protocols.io/view/rna-extraction-from-milk-for-hpai-surveillance-dczp2x5n>
 20. Moncla LH, Ross TM, Dinis JM, Weinfurter JT, Mortimer TD, Schultz-Darken N, et al. A Novel Nonhuman Primate Model for Influenza Transmission. *PLoS ONE.* 2013 Nov 14;8(11):e78750.
 21. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc.* 2017 Jun;12(6):1261–76.
 22. Braun KM, Haddock III LA, Crooks CM, Barry GL, Lalli J, Neumann G, et al. Avian H7N9 influenza virus-

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

- es are evolutionarily constrained by stochastic processes during replication and transmission in mammals. *Virus Evol.* 2023 Jan 1;9(1):vead004.
23. Jaeger, Anna. GitHub. [cited 2024 Nov 20]. Reverse Transcription and PCR. Available from: https://github.com/moncla-lab/h5-sequencing-protocol-dev/blob/main/RT_and_PCR_protocol.md
 24. Oxford Nanopore Technologies [Internet]. 2022 [cited 2024 Oct 23]. Ligation sequencing amplicons - Native Barcoding Kit 96 V14 (SQK-NBD114.96) (NBA_9170_v114_revN_15Sep2022). Available from: <https://nanoporetech.com/document/ligation-sequencing-amplicons-native-barcoding-v14-sqk-nbd114-96>
 25. Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. *Nat Biotechnol.* 2017 Apr;35(4):316–9.
 26. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal.* 2011 May 2;17(1):10–2.
 27. Li H. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics.* 2018 Sep 15;34(18):3094–100.
 28. Castellano S, Cestari F, Faglioni G, Tenedini E, Marino M, Artuso L, et al. iVar, an Interpretation-Oriented Tool to Manage the Update and Revision of Variant Annotation and Classification. *Genes.* 2021 Mar 8;12(3):384.
 29. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin).* 2012 Apr 1;6(2):80–92.
 30. Shen W, Sipos B, Zhao L. SeqKit2: A Swiss army knife for sequence and alignment processing. *iMeta.* 2024;3(3):e191.
 31. Hall MB. Rasusa: Randomly subsample sequencing reads to a specified coverage. *J Open Source Softw.* 2022 Jan 29;7(69):3941.
 32. Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. *PeerJ.* 2016 Oct 18;4:e2584.
 33. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. *GigaScience.* 2021 Feb 1;10(2):giab008.
 34. Bao Y, Bolotov P, Dernovoy D, Kiryutin B, Tatusova T. FLAN: a web server for influenza virus genome annotation. *Nucleic Acids Res.* 2007 Jul;35(Web Server issue):W280-284.
 35. Aksamentov I, Roemer C, Hodcroft EB, Neher RA. Nextclade: clade assignment, mutation calling and qual-

It is made available under a [CC-BY 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

ity control for viral genomes. *J Open Source Softw.* 2021 Nov 30;6(67):3773.

36. Huddleston J, Hadfield J, Sibley TR, Lee J, Fay K, Ilcisin M, et al. Augur: a bioinformatics toolkit for phylogenetic analyses of human pathogens. *J Open Source Softw.* 2021 Jan 7;6(57):2906.
37. Zhu W, Zhu Y, Qin K, Yu Z, Gao R, Yu H, et al. Mutations in Polymerase Genes Enhanced the Virulence of 2009 Pandemic H1N1 Influenza Virus in Mice. *PLoS ONE.* 2012 Mar 15;7(3):e33383.
38. USDA ERS - Background [Internet]. [cited 2024 Oct 4]. Available from: <https://www.ers.usda.gov/topics/animal-products/dairy/background/>
39. Dairy Farmers of Wisconsin [Internet]. WEDC. [cited 2024 Oct 23]. Available from: <https://wedc.org/success-in-wisconsin/success-stories/dairy-farmers-of-wisconsin/>
40. Freije C. MILK HAULING CHARGES IN THE UPPER MIDWEST MARKETING AREA MAY 2007.
41. Karthikeyan S, Levy JI, De Hoff P, Humphrey G, Birmingham A, Jepsen K, et al. Wastewater sequencing reveals early cryptic SARS-CoV-2 variant transmission. *Nature.* 2022 Sep;609(7925):101–8.
42. Jahn K, Dreifuss D, Topolsky I, Kull A, Ganesanandamoorthy P, Fernandez-Cassi X, et al. Early detection and surveillance of SARS-CoV-2 genomic variants in wastewater using COJAC. *Nat Microbiol.* 2022 Aug;7(8):1151–60.
43. Baaijens JA, Zulli A, Ott IM, Nika I, van der Lugt MJ, Petrone ME, et al. Lineage abundance estimation for SARS-CoV-2 in wastewater using transcriptome quantification techniques. *Genome Biol.* 2022 Nov 8;23(1):236.
44. Tomlinson SMA. Federal Order Requiring Testing for and Reporting of Highly Pathogenic Avian Influenza (HPAI) in Livestock.
45. CDC. Advanced Molecular Detection (AMD). 2024 [cited 2024 Oct 25]. SPHERES. Available from: <https://www.cdc.gov/advanced-molecular-detection/php/spheres/index.html>
46. Wallace HL, Wight J, Dowding B, Baz M, Flamand L, Hobman T, et al. Longitudinal Influenza A Virus Screening of Retail Milk from Canadian Provinces (Rolling Updates) [Internet]. medRxiv; 2024 [cited 2024 Jun 5]. p. 2024.05.28.24308052. Available from: <https://www.medrxiv.org/content/10.1101/2024.05.28.24308052v2>
47. Recommended composition of influenza virus vaccines for use in the 2024-2025 northern hemisphere influenza season [Internet]. [cited 2024 Jul 8]. Available from: <https://www.who.int/publications/m/item/recommended-composition-of-influenza-virus-vaccines-for-use-in-the-2024-2025-northern-hemisphere-influenza-season>