Method comparison of targeted influenza A virus typing and whole-genome sequencing from respiratory specimens of companion animals

Patrick K. Mitchell,* Brittany D. Cronk,* Ian E. H. Voorhees, Derek Rothenheber, Renee R. Anderson, Timothy H. Chan, Brian R. Wasik, Edward J. Dubovi, Colin R. Parrish, Laura B. Goodman[1]

Abstract. Epidemics of H3N8 and H3N2 influenza A viruses (IAVs) in dogs, along with recognition of spillover infections from IAV strains typically found in humans or other animals, have emphasized the importance of efficient laboratory testing. Given the lack of active IAV surveillance or immunization requirements for dogs, cats, or horses imported into the United States, serotype prediction and whole-genome sequencing of positive specimens detected at veterinary diagnostic laboratories are also needed. The conserved sequences at the ends of the viral genome segments facilitate universal amplification of all segments of viral genomes directly from respiratory specimens. Although several methods for genomic analysis have been reported, no optimization focusing on companion animal strains has been described, to our knowledge. We compared 2 sets of published universal amplification primers using 26 IAV-positive specimens from dogs, horses, and a cat. Libraries prepared from the resulting amplicons were sequenced using Illumina chemistry, and reference-based assemblies were generated from the data produced by both methods. Although both methods produced high-quality data, coverage profiles and base calling differed between the 2 methods. The sequence data were also used to identify the subtype of the IAV strains sequenced and then compared to standard PCR assays for neuraminidase types N2 and N8.

Key words: canine; equine; feline; influenza A virus; influenza matrix; M-RT-PCR; multi-segment PCR; neuraminidase; reverse-transcription PCR; single-nucleotide polymorphism; typing; whole-genome sequencing.

Introduction

Influenza A virus (IAV) testing and subtyping is an important component of the veterinary diagnostic laboratory repertoire. In addition to its natural reservoirs among different species of wild birds, IAV has demonstrated stable circulation in domestic poultry, humans, swine, mink, and horses, as well as causing epizootics in marine mammals.[7,29,33] Small companion animals (primarily dogs, cats, and ferrets) as well as horses are also susceptible to IAV infection. Human IAVs can infect dogs and cats, although likely inefficiently. Influenza A(H1N1)pdm09 virus has been documented to have infected several household pet cats and caused substantial mortality in a domestic cat colony.[8] A 2016 influenza A(H7N2) virus outbreak among shelter cats in New York City led to 2 human infections and prompted a major public health response.[2,24] Two recent transfers of IAV into dogs have produced sustained canine transmission. In ~1999, an equine influenza A(H3N8) virus entered the dog population, likely among racing greyhounds in Florida, and later became established in shelter dog populations.[6,21–23] The transfer of

an avian influenza A(H3N2)-like virus into the canine population likely occurred in China or Korea in ~2005. After circulating in East Asia, that virus was introduced into the United States in early 2015, causing an outbreak in several midwestern and southern states.[21,30,31] Multiple subsequent reintroductions of H3N2 strains from Asia to the United States and Canada have occurred since 2015.[31,34]

Given the propensity for zoonotic transmission of IAVs, veterinary diagnostic laboratories have a key role in surveillance of IAV distribution in animals, and in understanding the

Department of Population Medicine and Diagnostic Sciences (Mitchell, Cronk, Rothenheber, Anderson, Chan, Dubovi, Goodman), and Baker Institute for Animal Health and Department of Microbiology and Immunology (Voorhees, Wasik, Parrish), College of Veterinary Medicine, Cornell University, Ithaca, NY.

*These authors contributed equally.

[1]Corresponding author: Laura B. Goodman, Cornell University College of Veterinary Medicine, 235 Hungerford Hill Road, Ithaca, NY 14853.
laura.goodman@cornell.edu

Table 1. Primers for influenza A virus multi-segment reverse-transcription PCR (M-RT-PCR) for whole genome amplification.

M-RT-PCR primer	Full primer sequence	Influenza conserved region sequence
Method A[15]		
HFAdapter	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AGCAAAAGCAGG**	MBTuni-12
HF	TGTATAAGAGACAG**AGCAAAAGCAGG**	MBTuni-12
HRAdapter	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**AGTAGAAACAAGG**	MBTuni-13
HR	TGTATAAGAGACAG**AGTAGAAACAAGG**	MBTuni-13
Method B[18, 36]		
2322F	GTTACGCGCC**AGCAAAAGCAGG**	MBTuni-12
2323F	GTTACGCGCC**AGC**<u>**G**</u>**AAAGCAGG**	MBTuni-12 SNV
2324R	GTTACGCGCC**AGTAGAAACAAGG**	MBTuni-13

SNV = single-nucleotide variation. Conserved sequences are in bold. Underline indicates SNV in the MBTuni-12 SNV primer.

likelihood of human exposure. One strategy is detection of the matrix gene via reverse-transcription real-time PCR (RT-rtPCR), as well as targeted molecular subtyping of the hemagglutinin (HA) and/or neuraminidase (NA) gene segments.[11] However, specific detection of one or more those segments is prone to failure because of strain variation, and whole-genome sequencing (WGS), as it becomes more widely used by veterinary diagnostic laboratories, presents a more robust and flexible alternative to PCR for typing IAV submissions. Compared to basic subtyping, WGS also provides more detailed information on the composition and likely properties of any new virus, and thereby improves our understanding of IAV transmission within and between species.

WGS of IAV typically employs multi-segment RT-PCR (M-RT-PCR). This technique takes advantage of the fact that the 8 RNA segments in IAV share highly conserved regions at their ends: 12 nucleotides at the 3'-terminus (MBTuni-12), and 13 nucleotides at the 5'-terminus (MBTuni-13). M-RT-PCR therefore utilizes a set of primers that can amplify RNA segments of any IAV lineage by targeting these conserved regions at the ends of the genomic segments.[36] Several variations of this method, in terms of primer sequences and thermocycling profiles, have been proposed for use with commercial library preparation kits.

We compared 2 previously described M-RT-PCR methods for influenza WGS. The first, method A, was proposed in an Illumina technical note (https://www.illumina.com/content/dam/illumina-marketing/documents/products/appnotes/miseq-nextera-xt-influenza-application-note-770-2015-053.pdf) and modifies the MBTuni primers by adding Illumina Nextera transposon sequences to the end conserved region in order to improve coverage at the ends of segments.[15] The second, method B, adds different tails to the conserved region of the MBTuni primers and uses both versions of the variable MBTuni-12 position 4.[18,36] We used sequencing data generated following amplification by each of these methods for subtyping and reference-based assembly. The resulting assemblies were also utilized to evaluate published PCR primers for N2 and N8.[11]

Materials and methods

Two sets of samples were selected for inclusion in our study. Set 1 included 25 IAV-positive samples (21 canine, 3 equine, 1 feline) from routine diagnostic specimens submitted to the Cornell University Animal Health Diagnostic Center (Ithaca, NY) during 2017–2018. These were chosen to represent the typical host distribution of clinical submissions and to cover a variety of subtyping results. Set 2 consisted of 10 equine samples submitted during 2018–2019. Samples from both sets were originally determined positive for IAV by rtPCR, using the Centers for Disease Control and Prevention (CDC) universal IAV assay[26] on the OpenArray platform.[10] Virus isolation and propagation in embryonated chicken eggs was attempted for 7 of the respiratory swab supernatants from set 1 using standard methods.[4]

Influenza whole-genome amplification was performed on both sample sets using 2 M-RT-PCR methods that differ by primer sequences used and thermal cycling profiles. Method A uses primer concentrations of 300 nM for F1 and R1 and 200 nM for F2 and R2; method B uses 2 forward primers of 100 nM each and 1 reverse primer of 200 nM (Table 1). For both methods, M-RT-PCR was performed using the Superscript III one-step RT-PCR kit with Platinum Taq (Invitrogen) on the GeneAmp 7900 thermocycler (Applied Biosystems). The thermal cycling conditions for method A started with an RT step at 42°C for 60 min and 94°C for 2 min, followed by 2 sets of amplification cycles. The first set contained 5 cycles of 94°C for 30 s, 45°C for 30 s, and 68°C for 3 min. The second set contained 31 cycles of 94°C for 30 s, 57°C for 30 s, and 68°C for 3 min. The thermal cycling conditions for method B began with an RT step at 55°C for 30 s, followed by 42°C for 50 min and 94°C for 2 min. Following the RT step were 5 cycles of 94°C for 15 s, 42°C for 30 s, and 68°C for 3 min 20 s. The second set of cycling was 35 cycles of 94°C for 15 s, 57°C for 30 s, and 68°C for 3 min 20 s, with a final extension of 68°C for 5 min. PCR products were visualized using gel electrophoresis and purified with AMPure XP beads (Beckman Coulter) at a ratio of 0.4×.
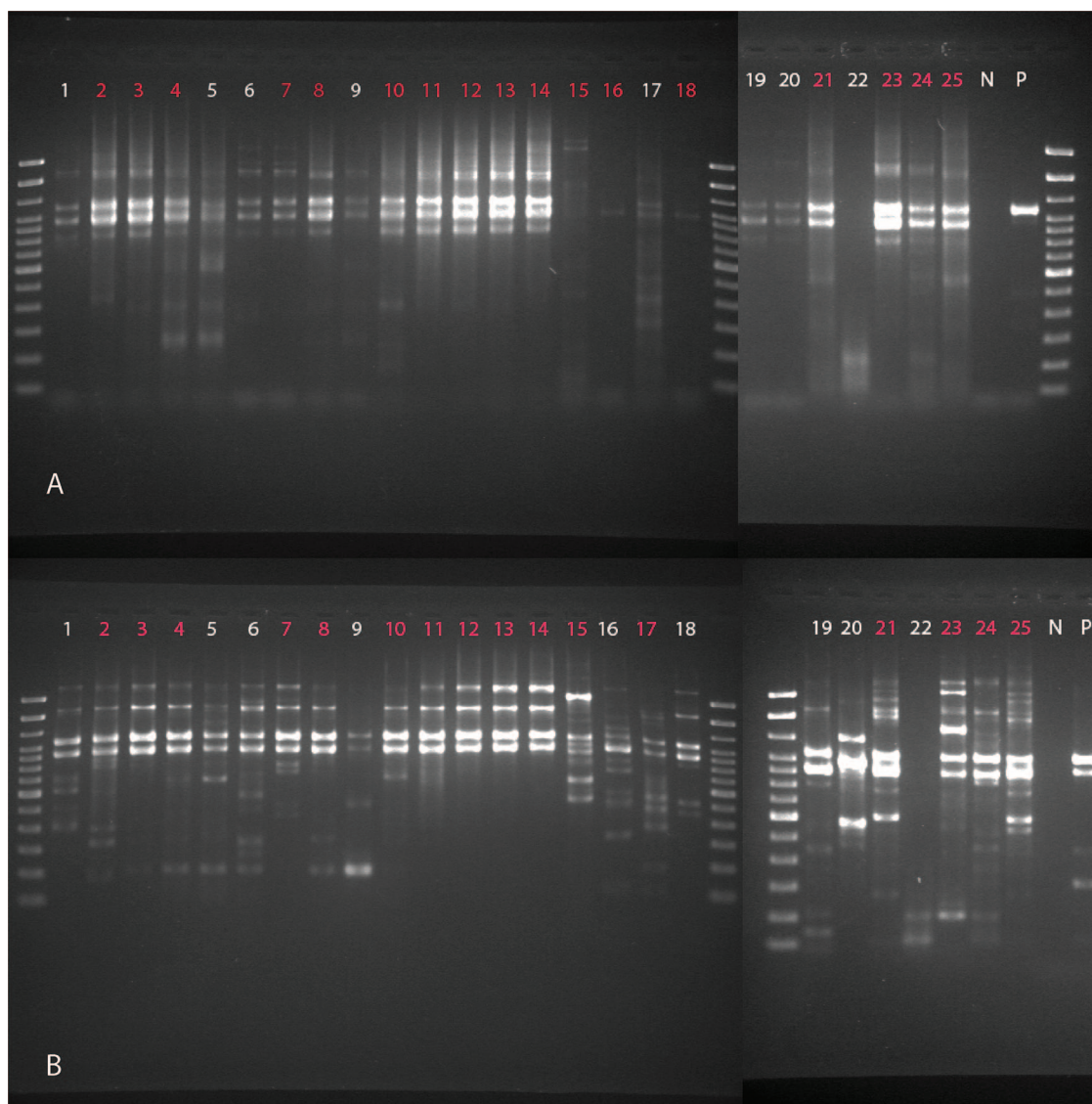
Figure 1. Gel image of M-RT-PCR product with the samples run with **A.** method A and **B.** method B conditions. Samples highlighted in red were chosen for sequencing.

Of the 25 set 1 samples amplified, 16 (12 canine, 3 equine, 1 feline) were chosen for WGS based on the quality of gel images and concentrations post-cleanup (Fig. 1). Purified PCR products were prepared for sequencing (Nextera XT DNA library preparation kit; Illumina) using an AMPure XP bead ratio of 1.8× for IAV libraries during the final cleanup step. The bead-based normalization step was omitted and replaced with concentration-based normalization using Qubit fluorometric quantification with the Qubit dsDNA HS assay kit (Thermo Fisher Scientific). Sequencing libraries were prepared from purified PCR products for all set 2 samples using a different protocol (Nextera DNA Flex library preparation kit; Illumina). Quality of the DNA fragmentation and library preparation was assessed (set 1: AATI fragment analyzer, set 2: DNA ScreenTape analysis; Agilent). Pooled libraries were then sequenced (set 1: MiSeq sequencer, set 2: iSeq sequencer; Illumina) using 2 × 150 bp chemistry. Read quality was assessed (FastQC v.0.11.8, https://www.bioin-formatics.babraham.ac.uk/projects/fastqc/).

NA typing by PCR was attempted on all samples from set 1 using assays for N2 and N8 described previously.[11] The primers were also searched against the assemblies from both sets using blastn[5] with a word size of 7 in order to assess how well they matched the sequence data and to investigate typing failures. Coverage and identity thresholds were set to 95% and relaxed until alignments were found.

HA and NA subtypes were also determined by using Mash v.2.1[19,20] to screen reads against custom databases of segment 4 and segment 6 sequences. To build this database, all non-redundant segment 4 and segment 6 sequences were

downloaded from NCBI Influenza Virus Resource (https://www.ncbi.nlm.nih.gov/genomes/FLU/Database/nph-select.cgi?go=database)[1] and filtered to remove sequences containing <99% canonical nucleotides and sequences not within 100 bp of the median length for that segment. The remaining sequences were then clustered at 95% identity using CD-HIT v.4.7.[9,17] Subtype annotations for each cluster were checked by identifying clusters in which multiple subtypes were present and by manually inspecting phylogenetic trees built using RAxML v.8.2.9[28] from alignments of representative sequences generated using MAFFT v.7.427.[13] When the subtype of a cluster representative appeared to be misidentified on the basis of phylogenetic position and/or annotations for other cluster members, a new representative was chosen that matched the expected subtype. The final sketches are available as part of FluStAR (https://github.com/pkmitchell/FluStAR).

Reference-based assemblies were then generated for each sample using Snippy v.4.3.6 (https://github.com/tseemann/snippy) with default variant calling parameters. Differences in base calls between the 2 methods were investigated by identifying sites that differed between each pair of assemblies. Prior to assembly, reads files were error-corrected using Lighter v.1.1.2[27] and trimmed with Trimmomatic v.0.36[3] to remove Nextera adapters and poor-quality bases with the following parameters: (Leading: 2, Trailing: 2, Minlen: 25, SlidingWindow: 4:5). Reference sequences were chosen based on subtype and host species. The IAV A/canine/Illinois/41915/2015(H3N2) genome was used for canine H3N2 samples, A/equine/Concepcion/RO7C/2018(H3N8) for equine H3N8 samples, and A/cheetah/California/D0912239/2009(H1N1) for feline H1N1 samples. If the subtype could not be determined from the sequence data, Kraken2[35] was used to assess whether the read set consisted primarily of non-IAV sequence data, which would indicate that viral amplification failed. One assembly for each sample was submitted to NCBI (Suppl. Table 1).

Per-base read pileup depths were determined for each assembly using SAMtools v.1.9.[16] Overall coverage depth was calculated as the average depth per base for each amplification method. In order to assess the coverage profiles of the 2 amplification methods, the coverage for each segment relative to the whole genome was first calculated as the average read depth for that segment divided by the average read depth across the entire assembly for each sequence. Wilcoxon signed-rank tests comparing these coverage values between the 2 methods were then computed for each of the 8 genomic segments. The p-values were adjusted for multiple comparisons using the Benjamini–Hochberg method. To assess within-segment variations in coverage, the coverage for each site relative to the average coverage for the segment was calculated and plotted for each assembly. All statistical analyses were performed using R v.3.4.2.[25] Coverage analyses were conducted separately for set 1 and set 2; for set 1,

analyses were restricted to canine H3N2 samples given the low number of samples of other subtypes.

Results

Library preparation and sequencing quality control

Among set 1 libraries prepared using method A, 14 of 16 had a Fragment Analyzer peak of 144–146 bp, and 1 had a peak at 184, suggesting the presence of empty adapters (https://support.illumina.com/bulletins/2019/10/what-are-adapter-dimers-.html?langsel=/us/). Relative concentrations for these peaks ranged 4–27.9%. The remaining method A sample did not have a Fragment Analyzer peak, suggesting that it did not contain empty adapters. No libraries prepared using method B produced these peaks. Libraries from method B also yielded a very even fragment size distribution, with a single peak at 100% concentration for all samples. In contrast, 11 of 16 libraries from method A had multiple peaks in addition to the previously mentioned empty adapter peaks (Fig. 2, Table 2).

Empty adapter peaks were not observed for any set 2 samples, and 9 of 10 libraries from both amplification methods produced a single peak of 464–524 bp. The remaining method A sample had a primary peak (83.3% relative concentration) at 489 bp and a secondary peak (16.7% relative concentration) at 751 bp. The remaining method B sample had a primary peak (99.1% relative concentration) at 508 bp and a secondary peak (0.9% relative concentration) at 2,054 bp.

Method A produced more sequencing reads per sample than method B. However, reads from method B were longer and had higher quality scores. These differences were considerably less pronounced in set 2 samples compared to set 1 samples (Fig. 3).

Viral subtyping and isolation

Of the 16 set 1 samples selected for sequencing, 7 were typed as N2 by PCR, 1 was typed as N8, and 8 were inconclusive. Using the sequence-based method, 12 were identified as H3N2, 2 as H3N8, and 1 as H1N1. Of the 10 set 2 samples, 9 were identified as H3N8 using the sequence-based method. One sample from set 1 and one sample from set 2 could not be subtyped. Sequence-based subtyping results were identical between method A and method B, and all samples for which a NA subtype was identified by PCR had matching results by sequencing. All H3N2 samples were from canine hosts, both H3N8s were equine, and the H1N1 was from a cat, but represented a human H1N1 pandemic virus (Table 3). Sequencing reads from both amplification methods for the samples that could not be subtyped were found to be primarily non-IAV, and these samples were excluded from further analysis.
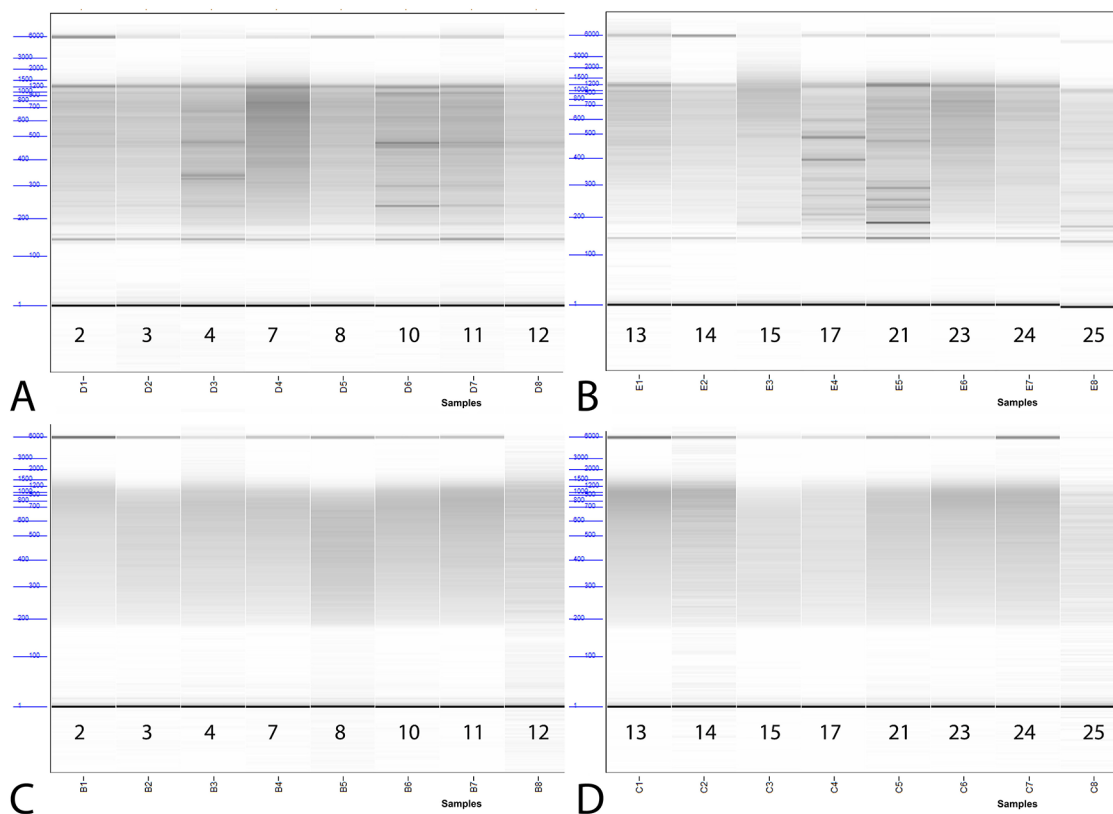
Figure 2. Fragment analyzer results of libraries prepared for sequencing using method A (**A, B**) and method B (**C, D**) conditions.

Table 2. Fragment analyzer peak size data for each set 1 library run.

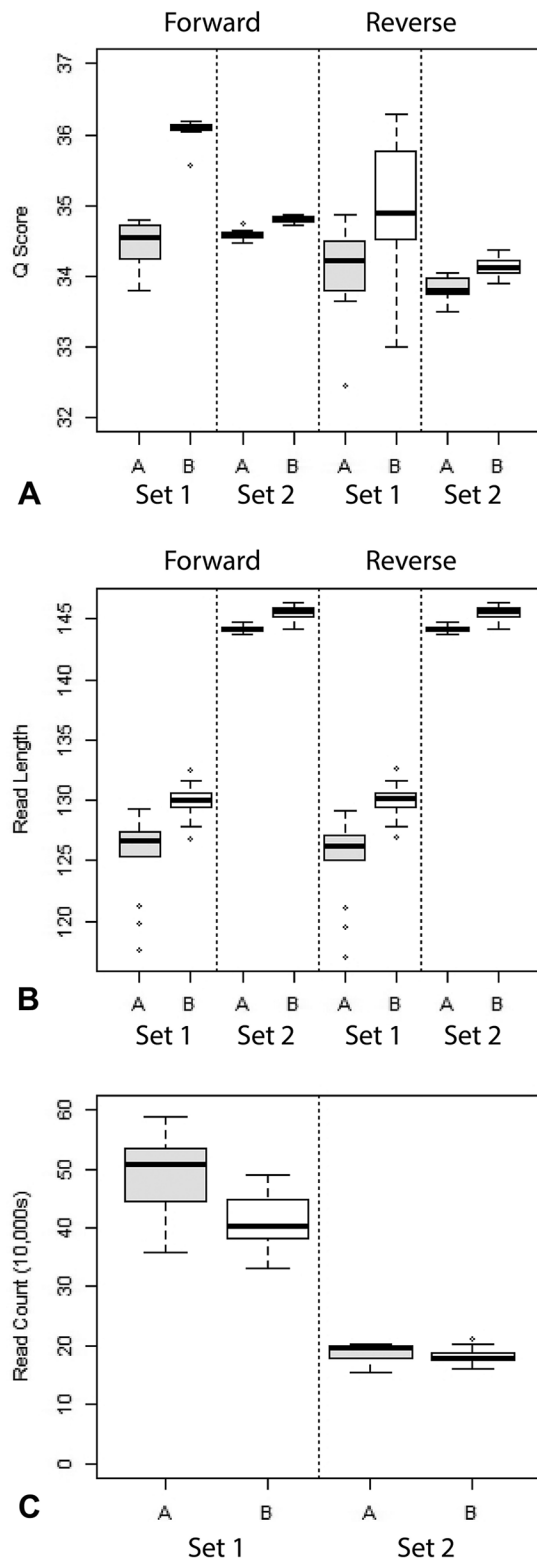|		Method A			Method B
		Empty peak adapter			
Gel	Strain	Size (bp)	Relative concentration (%)	Library peak sizes (bp)	Library peak sizes (bp)
2	A/Canis lupus familiaris/USA/188191/2017	145	9.7	301; 515; 1,218	869
3	A/Canis lupus familiaris/USA/188203/2017	146	14.5	315; 1,218	701
4	A/Canis lupus familiaris/USA/188226/2017	146	7.2	239; 341; 672; 1,201	740
7	A/Canis lupus familiaris/USA/188297/2017	144	24.9	784; 1,200	767
8	A/Canis lupus familiaris/USA/188305/2017	146	13.5	706; 1,218	510
10	A/Canis lupus familiaris/USA/218592/2017	145	8.6	240; 471; 1,200	492
11	A/Canis lupus familiaris/USA/000915/2018	146	8	241; 474; 1,218	832
12	A/Canis lupus familiaris/USA/006974/2018	146	16.8	474; 1,218	712
13	A/Canis lupus familiaris/USA/007780/2018	144	4	1,200	912
14	A/Canis lupus familiaris/USA/007781/2018	144	9.8	1,200	888
15	A/Equus caballus/USA/038358/2018	184	13.6	1,036	404
17	A/Felis catus/USA/047732/2018	146	27.9	396; 488; 1,141	881
21	A/Equus caballus/USA/149632/2018	144	8.9	185; 474; 1,188	800
23	A/Canis lupus familiaris/USA/150620/2018	145	8.6	1,177	864
24	A/Canis lupus familiaris/USA/152386/2018	144	24	1,153	789
25	A/Equus caballus/USA/154390/2018	NA	0	442; 978	655

Figure 3. Comparison of **A.** read quality scores, **B.** read count, and **C.** read length between methods. Values for method A are shown in gray and values for method B in white.

All 12 H3N2 samples had perfect matches to the N2 reverse primer and had 1 single-nucleotide variation (SNV)

relative to the N2 forward primer, with a G in position 10 rather than an A. The 11 H3N8 samples differed from the forward primer sequence at 3 or 4 of 25 sites, with 1 sample from set 1 differing from the remaining 10 samples by 1 SNV. The reverse primer had lower homology with the 11 H3N8 sequences as a result of structural changes, and 2 sites in the primer binding region varied among H3N8 samples (Table 4). The sequence identified at the N2 and N8 primer binding regions did not differ between the 2 M-RT-PCR amplification methods for any pair of sequences.

IAV was successfully isolated from one of the samples that was completely sequenced and published in NCBI (A/ Canis lupus familiaris/USA/006974/2018; Suppl. Table 1). Isolation was attempted unsuccessfully for samples A/ Canis lupus familiaris/USA/218592/2017 and A/Felis catus/USA/047732/2018. The other 4 samples for which isolation was attempted (1 of which was successful) were not selected for sequencing.

Coverage

Overall coverage depth in H3N2 samples from set 1 was higher for method B than for method A, although this was not consistent across all segments. Method A produced higher coverage for segments 4 and 6, the HA and NA segments, and there was no difference in coverage between the 2 methods for segments 7 and 8. The remaining segments had higher coverage using method B. In set 1, method B produced significantly higher average relative coverage for segments 1 and 2, whereas method A produced higher coverage for segments 4, 6, 7, and 8 after normalizing to overall coverage depth (Fig. 4, Table 5). Method B also had higher overall coverage depth in set 2, but the average relative coverage did not differ between amplification methods for any segment (Table 5). In set 1, which had high overall coverage, unaligned and low coverage sites were rare, with most samples having 0 such sites by either method. However, in set 2, which had lower overall coverage, the median number of low coverage and unaligned sites for method 1 samples was 888, whereas for method 2 samples, the median was 0. The feline H1N1 sample in set 1, and 2 of the equine H3N8 samples in set 2, produced low coverage over most of segment 2 by both amplification methods. These were excluded from NCBI upload (Suppl. Table 1).

In H3N2 samples from set 1 and H3N8 samples from set 2, method A generally produced higher coverage at the ends of segments, with a drop in coverage in the middle section of the longer segments, whereas coverage for method B tended to be more consistent through the segment with drop-offs at the beginning and end. In set 1 samples, method A also produced a mid-segment jump in coverage for segment 3, starting at position 1,482. This effect was not seen using method B. A similar, although less pronounced, jump occurred at position 1,076 of segment 4. This jump was present in both methods but was of greater magnitude for method A (Fig. 5, Suppl. Fig. 1). Both jumps begin at

Table 3. PCR and sequence-based subtyping for set 1 samples.

Strain	PCR	Sequence-based*
A/Canis lupus familiaris/USA/188191/2017	Inconclusive	H3N2
A/Canis lupus familiaris/USA/188203/2017	Inconclusive	H3N2
A/Canis lupus familiaris/USA/188226/2017	Inconclusive	H3N2
A/Canis lupus familiaris/USA/188297/2017	Inconclusive	H3N2
A/Canis lupus familiaris/USA/188305/2017	Inconclusive	H3N2
A/Canis lupus familiaris/USA/218592/2017	N2	H3N2
A/Canis lupus familiaris/USA/000915/2018	N2	H3N2
A/Canis lupus familiaris/USA/006974/2018	N2	H3N2
A/Canis lupus familiaris/USA/007780/2018	N2	H3N2
A/Canis lupus familiaris/USA/007781/2018	N2	H3N2
A/Canis lupus familiaris/USA/150620/2018	N2	H3N2
A/Canis lupus familiaris/USA/152386/2018	N2	H3N2
A/Felis catus/USA/047732/2018	Inconclusive	H1N1
A/Equus caballus/USA/149632/2018	Inconclusive	H3N8
A/Equus caballus/USA/154390/2018	N8	H3N8
A/Equus caballus/USA/038358/2018†	Inconclusive	Inconclusive

* Sequence-based typing results were identical between methods.
† Sample excluded from further analysis.

Table 4. N8 reverse primer comparison.

	Sequence
N8 reverse primer complement	5′-TTGGAGTAACGGGCCCTGATTCC-3′
Pattern 1 ($n=6$)*	5′-TTGG**A**GTCACAGG**G**CCTGACAACC-3′
Pattern 2 ($n=3$)	5′-TTGG**A**GTCACAGG**A**CCTGACAACC-3′
Pattern 3 ($n=1$)	5′-TTGG**G**GTCACAGG**G**CCTGACAACC-3′

Bold indicates observed single-nucleotide variations.
* An additional sample pair had an ambiguous base call at the second variable site.
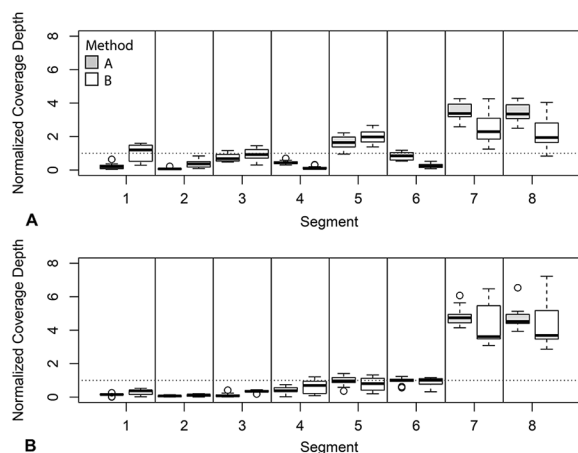

Figure 4. Boxplot of normalized coverage for each segment of influenza A virus using both M-RT-PCR methods for **A.** H3N2 samples from set 1 and **B.** H3N8 samples from set 2. The dotted line is at 1, representing the mean coverage across the whole genome. Values for method A are shown in gray and method B in white.

regions with high similarity to the MBTuni-12 sequence, suggesting that the forward primer is binding at this site leading to additional amplification of a portion of the gene. A coverage jump was also evident in segment 3 of set 2 samples amplified by method B beginning at position 1,218; this jump did not correspond to a site region that was similar to MBTuni-12 (Suppl. Fig. 2).

Base calls

Considering all forms of variation (SNVs, indels, and uncalled sites), there were 27 alignment positions that varied in >5 of the 12 set 1 H3N2 pairs. Of these, 16 are near the end (positions 1,445–1,460) of segment 6 (NA), which was uncalled in 5 sequences from method A and 1 from method B. An additional method A sequence called a deletion at position 1,446 but was identical to the reference and the rest of the samples over the rest of this span. This region overlaps with the longest single nucleotide repeat region in the reference genome and appears to correspond to a potential deletion or difficulties in accurately sequencing and mapping this region.

Another 3 positions correspond to the previously noted coverage jumps, suggesting ambiguity resulting from off-target primer binding rather than the template viral sequence.

Table 5. Wilcoxon signed-rank test comparison of normalized coverage by segment between methods.

	Set 1		Set 2	
Segment	Median shift*	Adjusted p-value†	Median shift*	Adjusted p-value†
1(PB2)	−0.85 (−1.16, −0.49)	0.002	−0.16 (−0.28, −0.02)	0.104
2(PB1)	−0.29 (−0.45, −0.16)	0.001	−0.05 (−0.08, 0.01)	0.129
3(PA)	−0.2 (−0.41, 0.05)	0.092	−0.25 (−0.31, −0.1)	0.063
4(HA)	0.32 (0.26, 0.38)	0.001	−0.27 (−0.45, −0.01)	0.104
5(NP)	−0.26 (−0.73, 0.04)	0.088	0.17 (−0.03, 0.4)	0.119
6(NA)	0.57 (0.36, 0.79)	0.001	0.11 (−0.07, 0.26)	0.129
7(M)	1.24 (0.37, 1.75)	0.012	0.87 (−0.1, 1.2)	0.119
8(NS)	1.52 (0.55, 1.96)	0.004	0.48 (−0.32, 1.06)	0.129

Numbers in parentheses are 95% confidence intervals.
* Positive values indicate higher values from method A compared to method B.
† p-values adjusted for 8 comparisons using the Benjamini–Hochberg method.

Position 1,482 of segment 3 was marked as ambiguous for all method A sequences, whereas it was called as an A for all method B sequences. The ambiguity in the method A sequences results from a split between A and G reads, with the Gs primarily being the first base in the read and likely corresponding to the last adapter base before the MBTuni-12 sequence in the primers. At position 1,489 of segment 3, there were 5 instances in which the base was ambiguous in the method A sequences and T in the method B sequences, 3 in which it was A in the method A sequences and T in the method B sequences, and 2 in which it was ambiguous in the method B sequences and A in the method A sequences. For the remaining 2 pairs, this position was ambiguous for both methods. This uncertainty was likely the result of homology with the MBTuni-12 sequence, which is identical to positions 1,483–1,494 of the reference sequence, aside from position 1,489, which is a T in the reference sequence and an A in MBTuni-12. Hence, some reads reflected the underlying viral sequence, whereas others corresponded to the primer sequence. Similarly, position 1,076 of segment 4 was called as a C in 7 sequences (1 in method A, 6 in method B), but was ambiguous in all others given conflict with the A at the beginning of the MBTuni-12 sequence.

The remaining 8 of these variable sites were position 4 of each segment, which varied in 10–12 of 12 pairs. These were also the only sites that varied among >5 of the 9 H3N8 pairs from set 2 H3N8. At these positions, method A consistently called an A, whereas most sequences from method B marked these positions as ambiguous. For 3 set 1 samples, however, method B instead called a G at this position in segment 4. The differences in base calling at this site result from the 2 variations of the MBTuni-12 sequence used by method B.

Discussion

With WGS becoming more accessible to veterinary diagnostic laboratories, it is important to evaluate different methods and to assess their impact on downstream analysis, in order to establish well-characterized methods that can be shared more widely. Here, we compared 2 M-RT-PCR amplification methods. These methods differed in their thermocycling profiles, the oligonucleotide tails added to the conserved region of the amplification primers, and the presence (method B) or absence (method A) of a variable base at position 4 of the conserved region of the forward primer. Two sets of IAV samples were amplified using each of these methods, and the resulting PCR products were sequenced using 1 of 2 different library preparation kits and sequencing platforms.[15,18] Set 1 consisted primarily of canine H3N2 samples; set 2 consisted of equine H3N8 samples. In set 1, although both methods produced adequate data, they differed in 2 key respects. First, analysis of fragment size following amplification with method A consistently showed peaks corresponding to empty adapters, which likely led to the lower read quality and mapping depth observed for this method. This difference was not apparent in set 2, suggesting that the difference in library preparation protocol led to more consistent fragment sizes. Second, method A appeared more prone to off-target amplification in the H3N2 samples from set 1, particularly in segments 3 and 4, resulting in dramatic variation in coverage depth across those segments. Method A also produced lower relative coverage in the middle of longer segments, corresponding to the increased number of low coverage and unaligned bases in the method A samples from set 1. However, method A generally produced less of a drop in coverage at the ends of segments in both the H3N2 and H3N8 sets. This did not result in base calling differences in these regions but could be more influential if coverage depth is low. Both methods had ambiguous base calls in regions with homology to the MBTuni-12 sequence, although, in method A, these were typically found in regions of off-target binding whereas in method B they were located consistently at the fourth base of each segment. Because of the
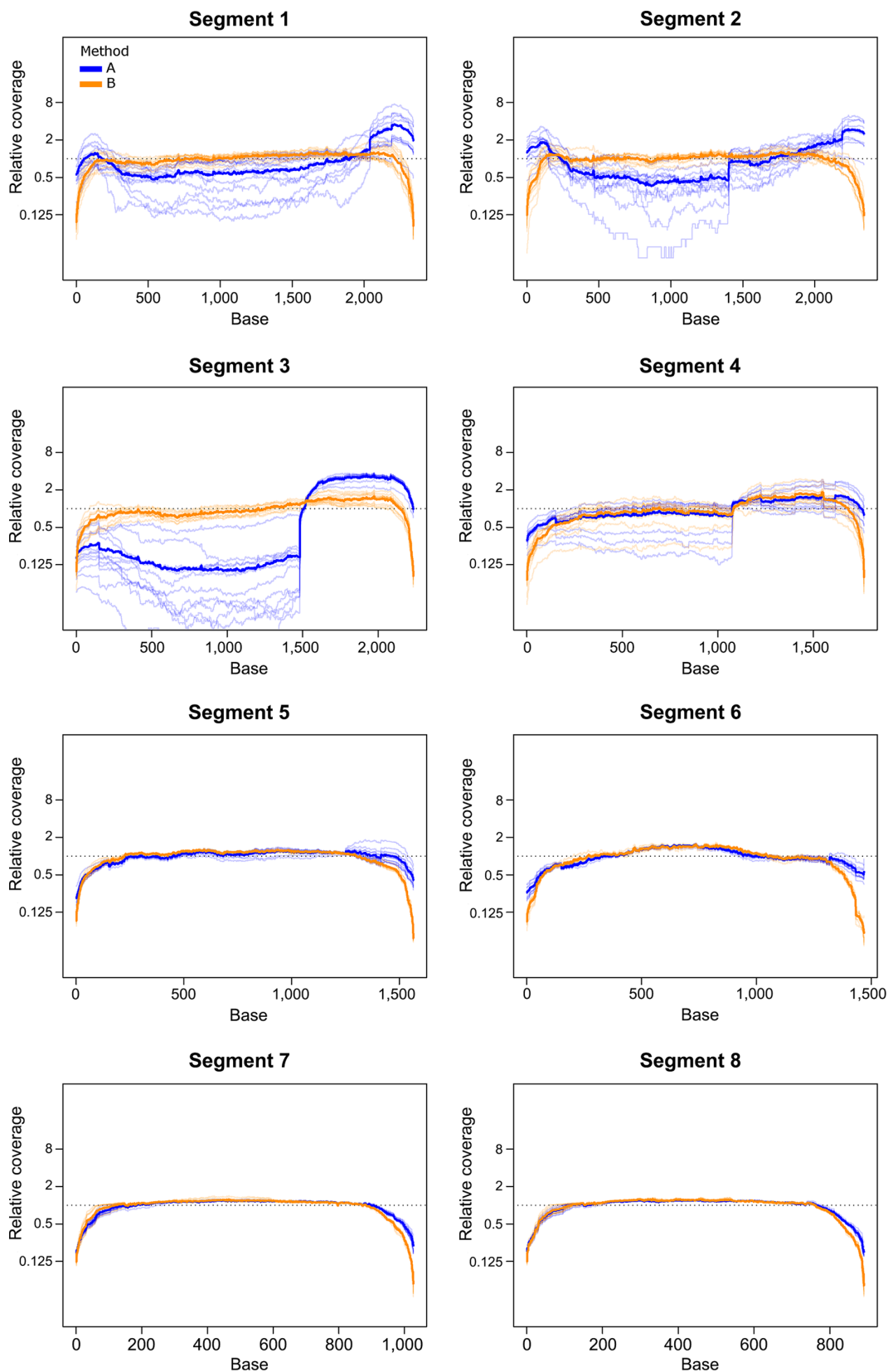
Figure 5. Relative coverage across each segment for set 1 H3N2 samples, with data for method A shown in blue and method B in orange. Each sample is represented by a light line, and the mean value for the method is shown as a dark line. The relative coverage for each base position is the number of reads at the base divided by the average coverage depth of the segment for a given sample.

higher quality sequencing output and more consistent ambiguity profile, method B may be preferable for routine testing and surveillance. Understanding these types of consistent features of different sequencing methods is critical for accurately interpreting comparisons with sequence data produced by other methods and developing an accurate picture of influenza viral dynamics.

Reliable subtyping is also a key component of influenza viral surveillance. We found that 5 of our 12 H3N2 samples had inconclusive PCR typing results, despite having strong matches to the N2 primers. Although these samples with inconclusive PCR results were ultimately found to match an expected subtype, PCR failures could also indicate the presence of an unexpected lineage. Sequencing allows these scenarios to be distinguished easily from one another and provides more detailed data to characterize the strains present. Furthermore, the time and costs associated with PCR subtyping rise as additional reactions are needed to identify a broader set of subtypes, whereas the cost of WGS remains constant. Given this increased resolution, reliability, and flexibility, M-RT-PCR and WGS for IAV characterization is more robust than PCR-based subtyping alone.

As new sequencing technologies gain wider usage in influenza viral surveillance, the effect of different amplification procedures and sequencing methods on the resulting data should be considered. Although our formal analyses were limited to comparisons between amplification methods, the differences in results between set 1, which was prepared using Nextera XT, and set 2, which was prepared using the Nextera DNA Flex kit (Illumina), suggest that the Flex kit may produce more consistent outcomes regardless of amplification method. Looking to the future, newer technologies such as nanopore sequencing may become more widely used for WGS in veterinary diagnostic laboratories. Several adaptations of the M-RT-PCR IAV amplification method have been published as a component of nanopore sequencing library preparation methods.[12,14,32] Because of the long reads produced by nanopore sequencing, coverage variation across segments is likely to be less important than differences in the amplification between segments. However, further analyses will be needed to determine best practices.

Comprehensive genomic surveillance of veterinary IAV is necessary to characterize transmission of the virus in different host species and to understand the risk of species crossover events. Although our study focused primarily on canine H3N2 and equine H3N8 subtypes, the methods presented should be applicable to many hosts and viral lineages. However, issues such as inconclusive subtyping and within-segment coverage variation may differ in other host–virus systems. As WGS becomes more widely used in veterinary testing, the development of standardized methods and an understanding of the data produced by those methods are key to maximizing the animal and public health value of veterinary influenza viral surveillance.

ORCID iDs

Patrick K. Mitchell https://orcid.org/0000-0001-6848-0846
Brittany D. Cronk https://orcid.org/0000-0001-7239-4262
Brian R. Wasik https://orcid.org/0000-0001-5442-3883
Colin R. Parrish https://orcid.org/0000-0002-1836-6655
Laura B. Goodman https://orcid.org/0000-0002-8327-3092

Supplementary material

Supplementary material for this article is available online.

References

1. Bao Y, et al. The influenza virus resource at the National Center for Biotechnology Information. J Virol 2008;82:596–601.
2. Blanton L, et al. Update: influenza activity in the United States during the 2016–17 season and composition of the 2017–18 influenza vaccine. MMWR Morb Mortal Wkly Rep 2017;66:668–676.
3. Bolger AM, et al. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 2014;30:2114–2120.
4. Brauer R, Chen P. Influenza virus propagation in embryonated chicken eggs. J Vis Exp 2015;97:52421.
5. Camacho C, et al. BLAST+: architecture and applications. BMC Bioinformatics 2009;10:421.
6. Crawford PC, et al. Transmission of equine influenza virus to dogs. Science 2005;310:482–485.
7. Fereidouni S, et al. Influenza virus infection of marine mammals. Ecohealth 2016;13:161–170.
8. Fiorentini L, et al. Influenza A pandemic (H1N1) 2009 virus outbreak in a cat colony in Italy. Zoonoses Public Health 2011;58:573–581.
9. Fu L, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics 2012;28:3150–3152.
10. Goodman LB, et al. High-throughput detection of respiratory pathogens in animal specimens by nanoscale PCR. J Vis Exp 2016;117:e54781.
11. Huang Y, et al. Neuraminidase subtyping of avian influenza viruses with PrimerHunter-designed primers and quadruplicate primer pools. PLoS One 2013;8:e81842.

12. Imai K, et al. Whole genome sequencing of influenza A and B viruses with the MinION sequencer in the clinical setting: a pilot study. Front Microbiol 2018;9:2748.

13. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol Biol Evol 2013;30:772–780.

14. Keller MW, et al. Direct RNA sequencing of the coding complete influenza A virus genome. Sci Rep 2018;8:14408.

15. Lee HK, et al. Contamination-controlled high-throughput whole genome sequencing for influenza A viruses using the MiSeq sequencer. Sci Rep 2016;6:33318.

16. Li H, et al. The sequence alignment/map format and SAMtools. Bioinformatics 2009;25:2078–2079.

17. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 2006;22:1658–1659.

18. Mena I, et al. Origins of the 2009 H1N1 influenza pandemic in swine in Mexico. Elife 2016;5.

19. Ondov BD, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol 2016;17:132.

20. Ondov BD, et al. Mash Screen: high-throughput sequence containment estimation for genome discovery. Genome Biol 2019;20:232.

21. Parrish CR, et al. Influenza virus reservoirs and intermediate hosts: dogs, horses, and new possibilities for influenza virus exposure of humans. J Virol 2015;89:2990–2994.

22. Parrish CR, Voorhees IEH. H3N8 and H3N2 canine influenza viruses: understanding these new viruses in dogs. Vet Clin North Am Small Anim Pract 2019;49:643–649.

23. Pecoraro HL, et al. Epidemiology and ecology of H3N8 canine influenza viruses in US shelter dogs. J Vet Intern Med 2014;28:311–318.

24. Poirot E, et al. Detection of avian influenza A(H7N2) virus infection among animal shelter workers using a novel serological approach-New York City, 2016–2017. J Infect Dis 2019;219:1688–1696.

25. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, 2017.

26. Shu B, et al. Design and performance of the CDC real-time reverse transcriptase PCR swine flu panel for detection of 2009 A (H1N1) pandemic influenza virus. J Clin Microbiol 2011;49:2614–2619.

27. Song L, et al. Lighter: fast and memory-efficient sequencing error correction without counting. Genome Biol 2014; 15:509.

28. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 2014;30:1312–1313.

29. Taubenberger JK, Kash JC. Influenza virus evolution, host adaptation, and pandemic formation. Cell Host Microbe 2010;7:440–451.

30. Voorhees IEH, et al. Spread of canine influenza A(H3N2) virus, United States. Emerg Infect Dis 2017;23:1950–1957.

31. Voorhees IEH, et al. Multiple incursions and recurrent epidemic fade-out of H3N2 canine influenza A virus in the United States. J Virol 2018;92:e00323-18.

32. Wang J, et al. MinION nanopore sequencing of an influenza genome. Front Microbiol 2015;6:766.

33. Webster RG, et al. Evolution and ecology of influenza A viruses. Microbiol Rev 1992;56:152–179.

34. Weese JS, et al. Emergence and containment of canine influenza virus A(H3N2), Ontario, Canada, 2017–2018. Emerg Infect Dis 2019;25:1810–1816.

35. Wood DE, Salzberg SL. Kraken: ultrafast metagenomic sequence classification using exact alignments. Genome Biol 2014;15:R46.

36. Zhou B, et al. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human influenza A viruses. J Virol 2009;83:10309–10313.