Current and predicted meat consumption by type

Oxford NANOPORE Technologies

APPLICATIONS

Rapid decentralized nanopore sequencing of full-length influenza A genomes

Samriti Midha[1], Richard Fetherston[1], Munir Iqbal[2], Sissel Juul[1], Phillip James[1]

[1]Oxford Nanopore Technologies plc, Oxford, UK, [2]The Pirbright Institute, Woking, UK

Abstract

Influenza A viruses, such as the H5N1 subtype currently circulating in birds in the United States and globally, can infect a wide range of animal groups important to the farming industry, including domestic poultry and swine. In addition, the currently circulating avian influenza A has been shown to be able to sporadically infect humans. For proper management of the disease, it is important to identify and subtype the pathogen correctly. Subtyping influenza A infections often revolves around classifying the haemagglutinin (H) and neuraminidase (N) genes. Additionally, molecular epidemiological patterns can be ascertained by analyzing individual variants in whole-genome sequencing data. Influenza A is a negative-sense single-stranded RNA virus, and its genome is split into 8 segments. Each segment has conserved ends, and primers have been designed which can reverse transcribe and PCR-amplify the individual segments in their entirety in a single reaction. Here we show how these amplified segments can be barcoded and sequenced in multiplex on a portable Oxford Nanopore sequencing device, to produce full-length intact sequences of individual influenza segments in single reads. We have created a NextFlow-based bioinformatic pipeline, which generates a consensus sequence and a subtype from nanopore data, allowing rapid decentralized sequencing of influenza A genomes. This is of particular importance during potential outbreaks of novel influenza A subtypes, especially within and between commercially important animal populations.

1. Influenza A can impact both animal and human populations

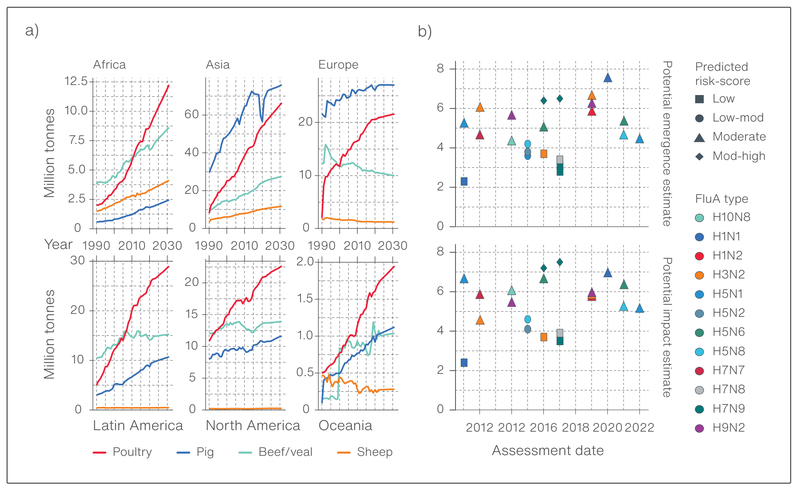

Fig. 1 a) current and predicted meat consumption b) IRAT emergence and impact scores

The OECD-FAO agricultural outlook 2022 predicts a global increase in meat consumption of 14 % by the year 2030, and that the two dominant types of meat consumed across different geographical regions continue to be poultry and pig meat (Fig. 1a)[1]. A number of diseases have repeatedly impacted meat markets, such as African swine fever, highly pathogenic avian influenza, and foot and mouth disease. In addition to the financial burden placed upon the farming industry, influenza A is a global health threat with the ability to infect various organisms including, but not limited to, humans, chickens, ducks, horses, dogs, pigs, and cats. It can also transfer between animals, and this ability for zoonotic transfer further enhances the risk to public health. The CDC produces an influenza risk- assessment tool using a variety of metrics, including genomic surveillance, to assess the potential risk posed by influenza A viruses not currently circulating in people (Fig. 1b)[2].

2. Influenza A H5N1 cases in avian populations in the US

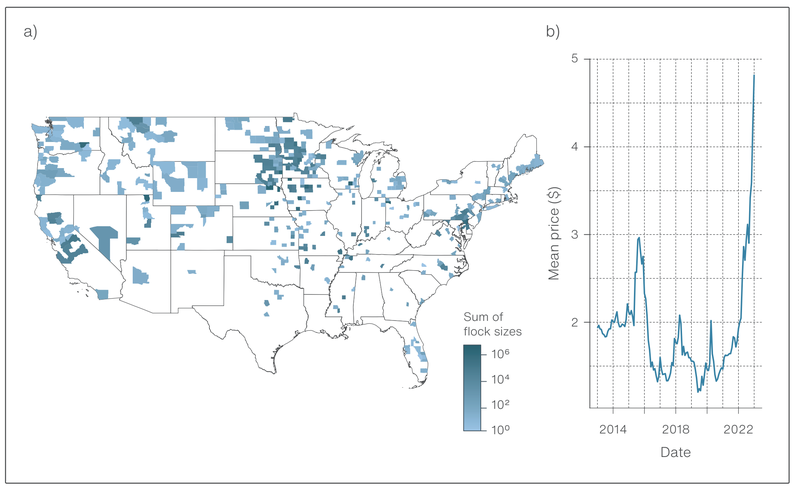

Fig. 2 a) H5N1 avian influenza cases across the US b) mean price of 12 grade A large eggs

Between late 2021 and early 2022 the predominant highly pathogenic avian influenza (HPAI) that was infecting poultry world wide was H5N1. In January 2022 the first HPAI H5N1 clade, 2.3.4.4b, was reported in the United States. Since then, it has been detected in 6,467 wild birds and affected over 58 million farmed birds across 47 states (Fig. 2a)[3] at the time of writing. In April 2022 the first detection in a human sample was reported. However, the distinction between detection and infection was unclear. The impact of the currently circulating HPAI on the poultry industry has been felt globally and a sharp rise in the cost of poultry-related products, such as eggs, has been seen by consumers (Fig. 2b)[4].

3. Sequencing influenza A with a one-pot rtPCR to produce full-length sequences

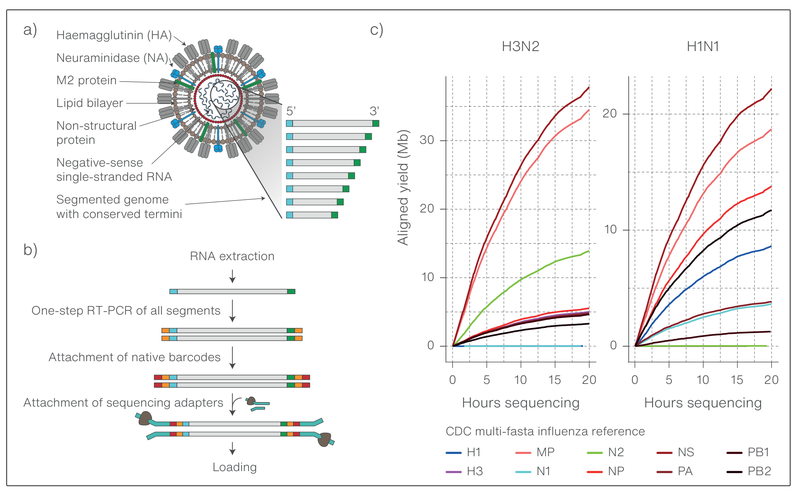

Fig. 3 Influenza A a) genome structure b) library prep c) aligned bases over sequencing time

The genome of influenza A is composed of 8 negative-sense single-stranded RNA molecules. These segments are often numbered by decreasing size. Segments 4 and 6 code for the surface glycoproteins hemagglutinin and neuraminidase, which are used for subtyping (Fig. 3a). The ends of each segment have highly conserved sequences which can be used as targets for reverse transcription and amplification of all segments in their entirety, using only 3 specific primers. Using a one-pot RT and PCR approach previously described[5], influenza A RNA segments can be amplified prior to ligation of sample-specific barcodes using Oxford Nanopore's native barcoding kit (Fig. 3b). Sequences from two samples, derived from cultured H3N2 A/Wisconsin/15/2009 and H1N1 r A/Swine/Iowa/15/30 and run as part of a 24-sample multiplex, were aligned to the CDC multi-fasta reference, and the number of sequences per unit sequencing time aligning to each segment was calculated (Fig. 3c).

4. Epi2me labs nanopore-specific and self-contained influenza workflow

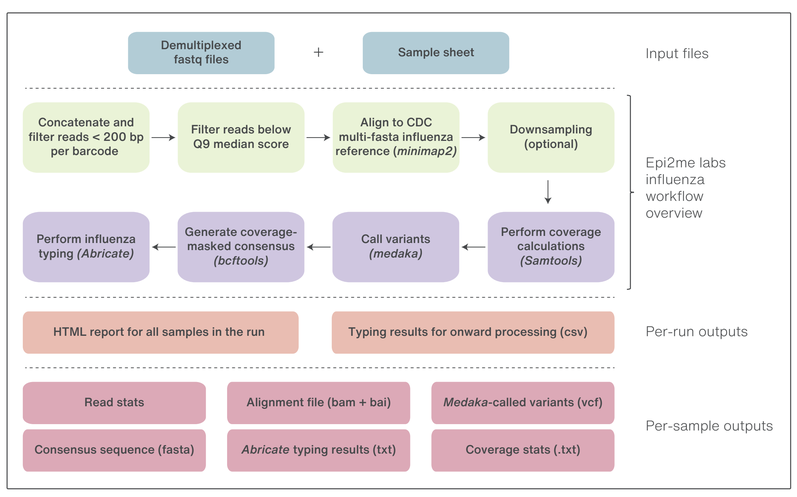

Fig. 4 Schematic of the Epi2me labs influenza workflow

Epi2me labs provides self-contained, nanopore-specific, bioinformatic workflows to analyse a wide variety of research questions. The wf-flu workflow types and subtypes influenza A or B from multi- or single-plexed influenza whole-genome-sequencing data. Briefly, the workflow concatenates fastq files and removes reads below 200 bp, and the remaining reads are then aligned to the CDC multi-fasta influenza reference, and coverage statistics are calculated using samtools. An optional down-sampling step can be implemented to speed up subsequent steps whereby reads are filtered to fall within +/- 10% of the reference segment length. Variants are called with medaka and a coverage-masked consensus is created using bcftools. Finally, typing is performed with Abricate using the INSaFLU database[6]. An html report is generated for each sample, with information on type and subtype based on the H and N genes, bam files, vcf files of variant calls, and a consensus fasta file (Fig. 4).

5. Nanopore sequencing of influenza samples generates complete consensus genomes

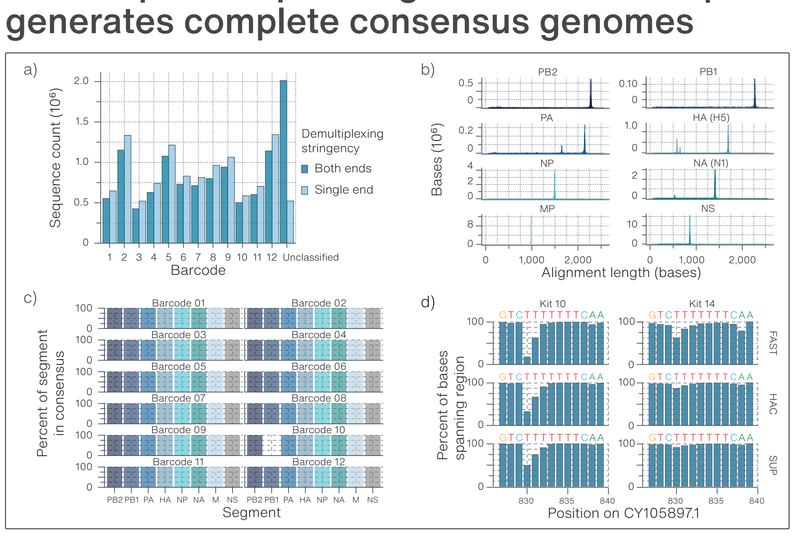

Fig. 5 a) Barcode balancing b) sequence length c) complete genomes d) homopolymer calling

A number of influenza RNA samples representing different avian and swine influeza strains were barcoded and sequenced concurrently on a single flow cell. We demultiplexed barcodes using guppy under two different stringencies: detecting a native barcode on either end of a molecule or detecting a matching barcode on both ends (Fig. 5a). After demultiplexing reads we aligned them to the CDC multi-fasta influenza A reference. The resulting alignments show full segment-length reads (Fig. 5b). After running the data through the *EPI2ME Labs wf-flu* workflow, we obtained full consensus genomes for 11 of the 12 samples. We saw dropout of an internal region of the PB1 segment for barcode 10. This phenomenon is often seen in cultured virions (Fig. 5c). The most recent nanopore sequencing kits and base-calling models show a notable improvement in coverage across the longest homopolymer in the reference sequence, which lies in the segment encoding for the matrix protein (Fig. 5d).

6. Defective interfering RNAs and phylogenetic analysis

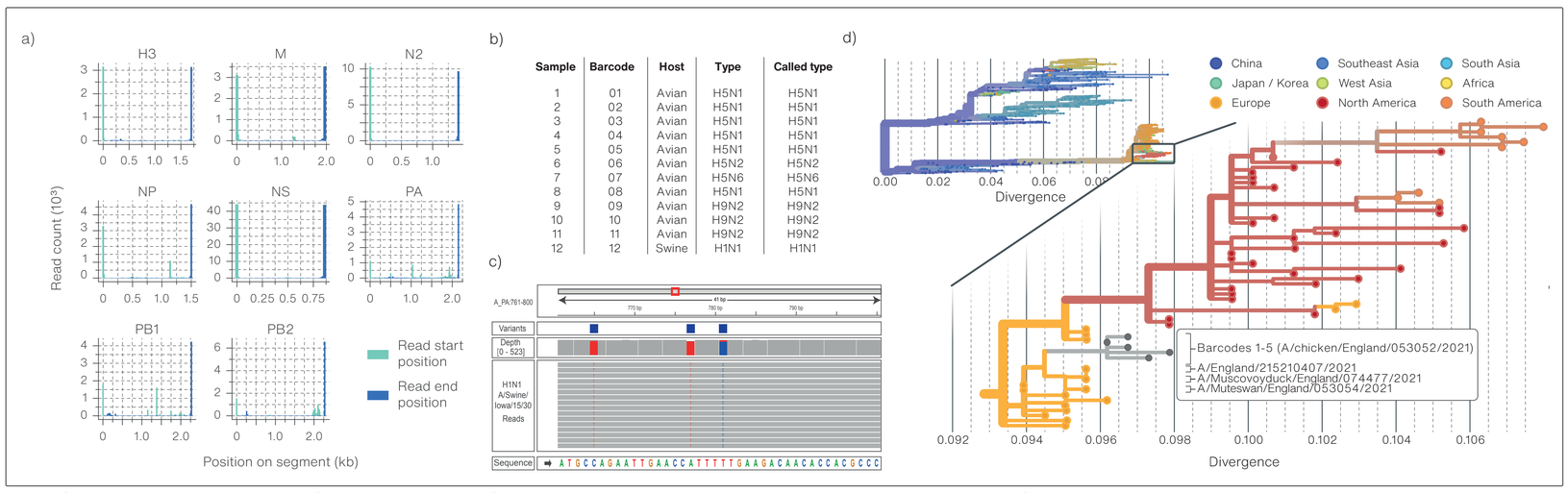

Fig. 6 a) detecting potential DI RNAs b) sample subtyping c) Epi2me labs Kit14 variant calling in and around a homopolymer d) Phylogenetic tree of GISAID H5N1 sequences and barcodes 1-5

As nanopore sequencing enables the sequencing of full amplicons produced by the RT and PCR method, early terminations or breakpoints can also be deduced by looking at the start and end positions of alignments (Fig. 6a). Defective interfering RNAs (DI RNAs) have long been seen in cultured influenza A preparations such as those used here, where large internal regions of one or more segments are removed. It is thought that these compromised segments are a result of incorrect replication but are still packaged in a virion. There is evidence to suggest that DI RNAs exist in wild-type populations[7], but their roles remain unclear. Full-length next-generation amplicon sequencing may provide a useful research tool to study these molecules. The 12 RNA samples representing a number of different avian influenza and swine influenza strains were sequenced and analysed by the EPI2ME Labs wf-flu workflow and the correct subtype was assigned to each based on the consensus haemagglutinin and neuraminidase sequences (Fig. 6b). Correct variant calling is paramount including in and around homopolymer regions present in the flu genome (figure 6c). Using the Nextstrain quick-start build for avian influenza, we constructed a phylogenetic tree with the H5N1 HA sequences available on GISAID* and our consensus HA sequences for barcodes 1 – 5. The tree splits the sequences into H5 clades 2.3.2.1a-c, 2.3.4.4 and 2.3.4.3 (Fig. 6d). The consensus HA segments from barcodes 1-5 sit within the correct 2.3.4.4 clade next to other samples in the GISAID database from the same geographical region and time point. Detailed pathogen surveillance, in both animal and human populations, has the potential to aid in rapid and focused implementations of biosecurity precautions and public health measures.

*We gratefully acknowledge the authors and submitting laboratories, details of which can be found https://github.com/nanoporetech/lc23-poster-avian-flu

Conclusion

Here we show that a simple one-pot reverse transcription and PCR amplification, using only three primers, can generate full-length influenza A molecules which can be barcoded and multiplexed on a single flow cell. The bioinformatic workflow is able to type and subtype flu A as well as generate consensus sequences after variant calling. The ability to interrogate the genetic information of currently circulating influenza A viruses in animal populations can aid in a number of processes, such as epidemiological analysis and risk management. Furthermore, certain animal populations such as swine may be vaccinated against influenza subtypes, and understanding which virus subtypes are currently circulating can inform vaccine choice. Tracking large-scale outbreaks of influenza, such as that seen globally during the current HPAI outbreak, may allow for the rapid implementation of control measures at different geographic scales.

Nanopore influenza workflow

Download poster

References

1. https://stats.oecd.org/
2. https://www.cdc.gov/flu/pandemic-resources/monitoring/irat-virus-summaries.htm
3. https://www.cdc.gov/flu/avianflu/modules/commercial-backyard-flocks.csv
4. https://data.bls.gov/timeseries/APU0000708111
5. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human influenza a viruses. Zhou et al. *Journal of Virology* **83** (19) 2009
6. INSaFLU: an automated open web-based bioinformatics suite "from-reads" for influenza whole-genome-sequencing-based surveillance. Borges et al. *Genome Medicine* **10** (1) 2018
7. Influenza A virus defective viral genomes are inefficiently packaged into virions relative to wild-type genomic RNAs. Alnaji et al. *mBio* **12** (6) 2021