Practical: Avian Influenza detection

GenEpi-BioTrain – Virtual training 4: Avian influenza NGS data analysis for outbreak investigation from a One Health perspective

Trainers: Marta Maria Ciucani & Yuan Liang

Background

Influenza, a highly contagious respiratory illness, continues to pose significant public health challenges worldwide. Each year, numerous countries face outbreaks during the annual influenza season, which necessitates a multidisciplinary approach involving epidemiologists, virologists, healthcare professionals, and public health authorities.

In addition to wild and domestic birds, many mammals have been affected by high pathogenicity avian influenza virus (HPAIV) in recent seasons.
In 2021 and 2022, five mammals tested positive for clade2.3.4.4b H5Nx HPAIVs in Denmark.
The first mammalian case was an adult male harbor seal (*Phoca vitulina*), found dead in September 2021 on a beach at Southwest Funen (Figure1). The seal was in a state of progressed decay and was assessed to have been dead for up to 1 week at sea. At necropsy, the seal was found emaciated. The heart and respiratory organs appeared unaffected, whereas the abdominal organs were too decayed for thorough examination. The organs were unsuitable for histology. The organs were also tested for the presence of pathogenic bacteria and morbillivirus as previously described,1in which none tested positive.

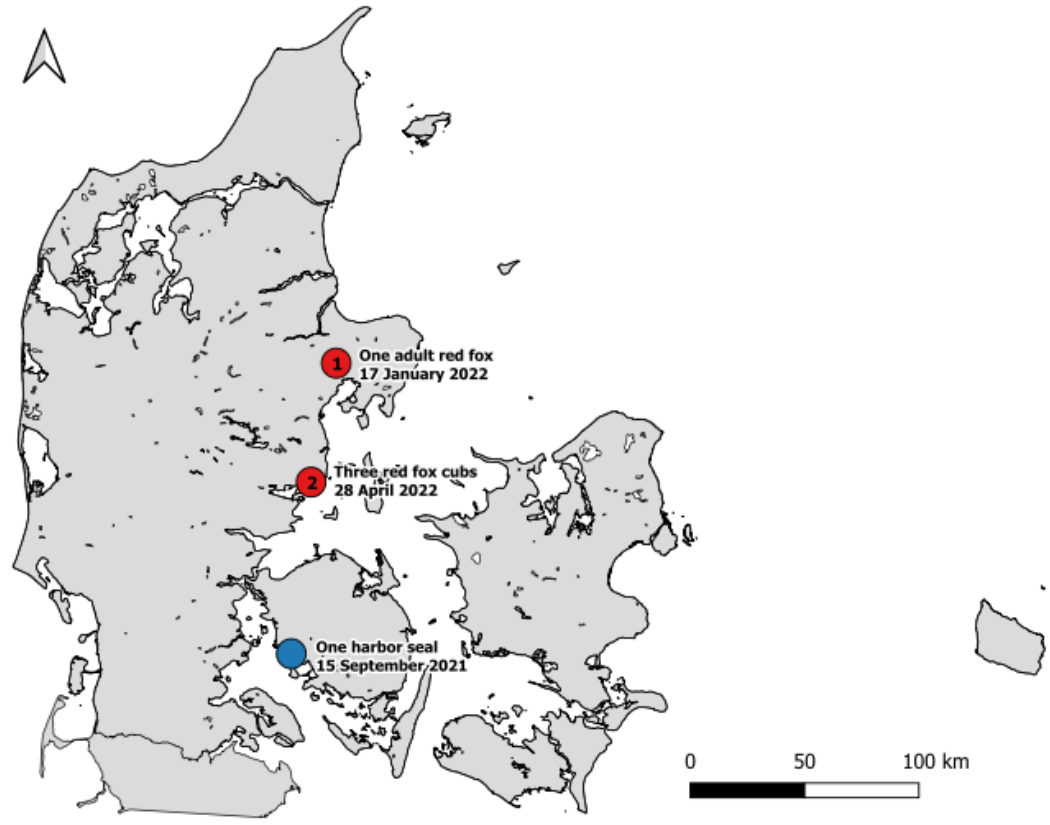

FIGURE 1 Map of Denmark showing where the HPAIV-positive harbor seal (blue dot), adult fox (red dot marked with "1"), and three red fox cubs (red dot marked with "2") were detected. The date of observation is also indicated next to the location site. The harbor seal was found at a beach on Southwestern Funen. The adult fox and fox cubs were all detected on the eastern part of Jutland.

In 2022, HPAIV H5N1 was detected in one adult red fox and three red fox cubs (Vulpes vulpes) by rRT-PCR and sequencing. Brain and/or lung tissues from the four foxes were collected for virological and histopathological analyses. The adult fox was found dead on January 17, 2022, close to a fox pit at Djursland in Eastern Jutland, where HPAIV H5N1 was detected in the lung (Table1). On April28, 2022, three fox cubs were found dead in another location, 77 km southwest of Djursland, next to a fox pit in Odder municipality. All three fox cubs were observed alive 3 days prior.

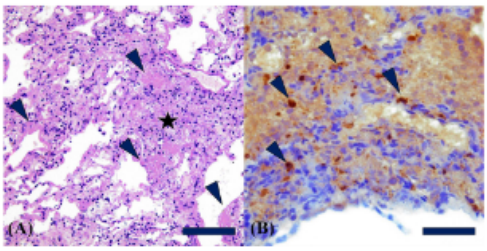

FIGURE 2 Lung tissue from two red foxes with pneumonia consistent with influenza infection. The lungs tested positive for HPAIV by rRT-PCR. (A) Adult fox with fibrinous to necrotizing lesions. Clots of fibrin (arrowhead) in alveoli and lining the terminal bronchi intermixed with small necrosis (star). Hematoxylin and eosin, bar 100 µm. (B) Fox cub with non-suppurative interstitial to fibrinous pneumonia with multiple cells staining positive for influenza virus (dark brown/arrow head) by immunohistochemistry using a mouse monoclonal antibody against influenza A virus NP, bar 50 µm.

HPAIV H5N1 was detected by rRT-PCR in brain tissue from one of the fox cubs and in the lung of the other two cubs. The rRT-PCR analysis was repeated with the same results.

Part 1: Determine mutations and genotypes.

Note: If you want to start from the raw data (fastq files) go to Part 2 of this document.

In this part of the exercise we are going to use the already generated consensus sequences of the avian influenza case.
The idea is to compare the sequences to other avian influenza consensus sequences to determine their mutations and genotypes.
First, you should download the data folder on your computer. This folder will contain the fasta files with the sequences from this study, other reference sequences and other material (see folder description on EVA).

a) Compare the sequences on AliView, MEGA or Unipro UGENE tools. Play a bit with the tool, perhaps by making an alignment and do a pairwise comparison.

b) Translate the DNA sequences to amino acids and focus on the PB2 segment.
Q1. Do you notice any amino acid mutation that could be interesting to investigate?
Q2. If yes, in which positions?
Q3. Which ones could be associated to mammalian adaptation?

Note: Remember to compare the mutations that you see with the ones in the review paper (Biological_markers_AIV.pdf)..

Q4. What are the subtypes and genotypes of the samples?
Note: here you can choose if you want to subtype/genotype based on a phylogeny or use blast. Follow the instructions provided during the lecture.

Make a phylogeny with whichever program you prefer and compare all the mammal sequences to the references.
Q5. Do all mammalian sequence cluster together?
Q6. Can you explain the pattern that you see?
Note: we will provide already made phylogenies in case you run out of time so that you can answer these questions.

Part 2: Generate Consensus sequences using IRMA. – Skip if you don't use IRMA

In this part of the exercise we are going to use the command line tool IRMA to perform consensus sequence reconstruction on Illumina sequencing data.

1. Locate the Illumina reads that you downloaded at the end of Session 1.
2. Activate the Irma environment.

```
conda activate irmaenv
```

3. And now we should think how we want to treat our data!
   For example:
   - What is the minimum depth and quality that we would like to have to assign a nucleotide?
   - Are we only interested in the consensus sequence?
   - Would we like to pay attention also to minority variants?
   - What frequency should a minority variant have to call an ambiguous site?

Let's have a look at the init.sh file and see what we could modify.

```
less /path_to_environment/irmaenv/bin/IRMA_RES/modules/FLU/init.sh
```

```
### REFERENCE ###
MIN_CONS_SUPPORT=50    # minimum allele coverage depth to call plurality consensus, otherwise calls "N".

### READ GATHERING ###
QUAL_THRESHOLD=30              # minimum read statistic
MIN_LEN=125                    # minimum read length

### FINISHING ASSEMBLY ###
INS_T=0.25              # minimum frquenncy threshold for insertion refinement
DEL_T=0.60              # minimum frequency threshold for deletion refinement
MIN_AMBIG=0.40          # minimum called SNV frequency for mixed base in amended consensus folder


### VARIANT CALLING ###
# HEURISTICS
AUTO_F=1                # auto-adjust frequency threshold [1,0]
MIN_FI=0.005            # minimum insertion variant frequency
MIN_FD=0.005            # minimum deletion variant frequency
MIN_F=0.008             # minimum frequency for single nucleotide variants
MIN_C=2                 # minimum count for variants
MIN_AQ=25               # minimum average variant quality, does not apply to deletions
MIN_TCC=100             # minimum non-ambiguous column coverage
MIN_CONF=0.80           # minimum confidence not machine error
```

Remember that IRMA has many configuration files and they can be used and modified to refine your assembly.

For example, it has both a FLU.sh and a FLU-avian.sh, instead of changing the init.sh file we can modify these configuration files!
In this case we can use nano to modify the FLU-avian.sh configuration file:

```
nano /path_to_environment/irmaenv2/bin/IRMA_RES/modules/FLU/config/FLU-avian.sh
```

Let's say that we are interested in samples that have at least 50x depth and we want to call for ambiguous sites only if the minority variant has at least 0.4 frequency.

Q7: Can you modify the config file accordingly?
Q8: What are the names of the parameters you should modify?

4. Run IRMA on your samples.
   - Enter the folder where you have your sequences:
   - Create 1 more folder that you can call run and enter this folder.

Example of the command to run the program using your files:

```
IRMA FLU-avian ../sample1_R1.fq.gz ../sample1_R2.fq.gz      sample1
```

Change the sample1_R1.fq.gz and sample1_R2.fq.gz with your sample's name.

5. Have a look at the output.

Q9: Do you notice any missing sites?
Q10.What does this tell you on the overall quality or coverage of the sample?