

Exploring the Spread and Evolution of Avian Influenza through Bioinformatic Analyses

Jasmine Ma - 5/21/2025 - Period 3 - Dr. Yilmaz Syslab



Background

- Zoonotic spillover is the transmission of pathogens from animals to humans
 - Spillover events have increased due to direct contact with infected objects
- Examples of spillover events:
 - SARS-CoV-2 pandemic: likely emerged via wildlife trade (Garry, 2022)
 - Avian influenza: primarily impacting birds but capable of infecting humans (European Food Safety Authority [EFSA] et al., 2024)
- Phylogenetic trees can visualize genetic relationships between strains
- Wastewater surveillance is a new method of detecting presence of viruses in the environment, especially in a rapidly evolving state



Problem

- Limitations of current testing: Clinical/farm
 - **Costly, inefficient, inaccurate:** fails to capture variability in the environment, requires individual testing. (Schreiber, 2024)
- Current methods have limited implementation (**only monitoring, not long-term evolutionary tracking**) due to lack of widespread sewage-level surveillance programs
- **Many farms DO NOT test** due to financial concerns (Schreiber, 2024)
- With unique/evolving datasets, phylogenetic tree rooting can be challenging, especially without temporal data (Kinene et al., 2016)



Previous Work

- In 2023, I interned at Scripps Research through the SRTI program to process wastewater in San Diego
- Focused on building phylogenetic trees using COVID-19 wastewater samples
- Learned how to debug rooting methods, compile datasets, and build small trees using IQTree.
- Deliverables:
 - **Compiled/visualized 2023 COVID variants**
 - **Identified common issues with tree rooting specific to those variants**



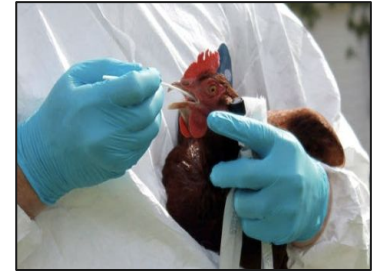
What's New?

- **New virus focus:** Exploring avian influenza, applying more advanced phylogenetic models to build more complex trees
- **Integration of more complex datasets** (1,600+ samples from all over the world, ranging from wastewater to environmental)
- **Solving a new tree-rooting issue:** Avian influenza is more divergent due to high variability in genetic sequences
- **Perform large-scale variant analysis:** Using new tool, Freyja, to identify emerging strains and track zoonotic risks in specific samples

Other Solutions

1) Traditional Surveillance

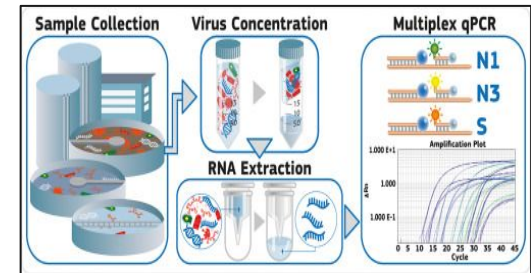
- a) Sampling birds the field (Hoye et al., 2010)
- b) **Issue:** only captures data from specific populations, delayed detection
- c) Very accurate!



(Benson, 2015)

2) COVID-specific wastewater techniques

- a) qPCR: (Quantitative Polymerase Chain Reaction) by measuring viral genetic material (Navarro et al., 2021)
- b) **Issue:** works well for well-known pathogens, but struggles for low-prevalence variants



(Navarro et al., 2021)



Why IS Ours BETTER?

1. Integrates temporal data: Unlike traditional methods that only focus on viral load, we **incorporate temporal changes to contextualize the tree**
2. Using lineage abundance data, we can **estimate when a specific strain first emerged** to pinpoint when mutations occur
3. Environmental/wastewater data is more representative because it **can capture contributions from multiple species in nature** (ex. A farm, reservoir leading to crossover/recombination)



Novelty

- Dwindling wastewater projects post-pandemic (Ladyzhets, 2023)
- **Avian influenza lacks comprehensive wastewater analysis** in the US
- Integrating **temporal, time-calibrated models** to track evolutionary timelines (Donoghue & Yang, 2016)
- **First attempt to quantify variant prevalence in avian influenza**, identifying emerging strains early
- **Integrates environmental data** such as water bodies, wildlife, agriculture through diverse data



Impact

- 1) **Earlier detection** of emerging variants of avian influenza
- 2) Guiding public health interventions: if phylogenetics can **predict where/when strains emerge first**, interventions can be **more targeted**
- 3) Wildlife/environmental management: if water bodies show high lineage prevalence, we can **guide environmental policies** to reduce contamination
- 4) **Lower cost** compared to traditional surveillance, **more scalable**



Global Initiative on Sharing All Influenza Data (GISAID) Dataset) (Shu, 2017)

- Comprehensive collection of genomic sequences and metadata
- Contains info on virus subtype, host species, location, date of collection
- Helps study viral evolution and cross-species transmission
- Frequently updated





Method

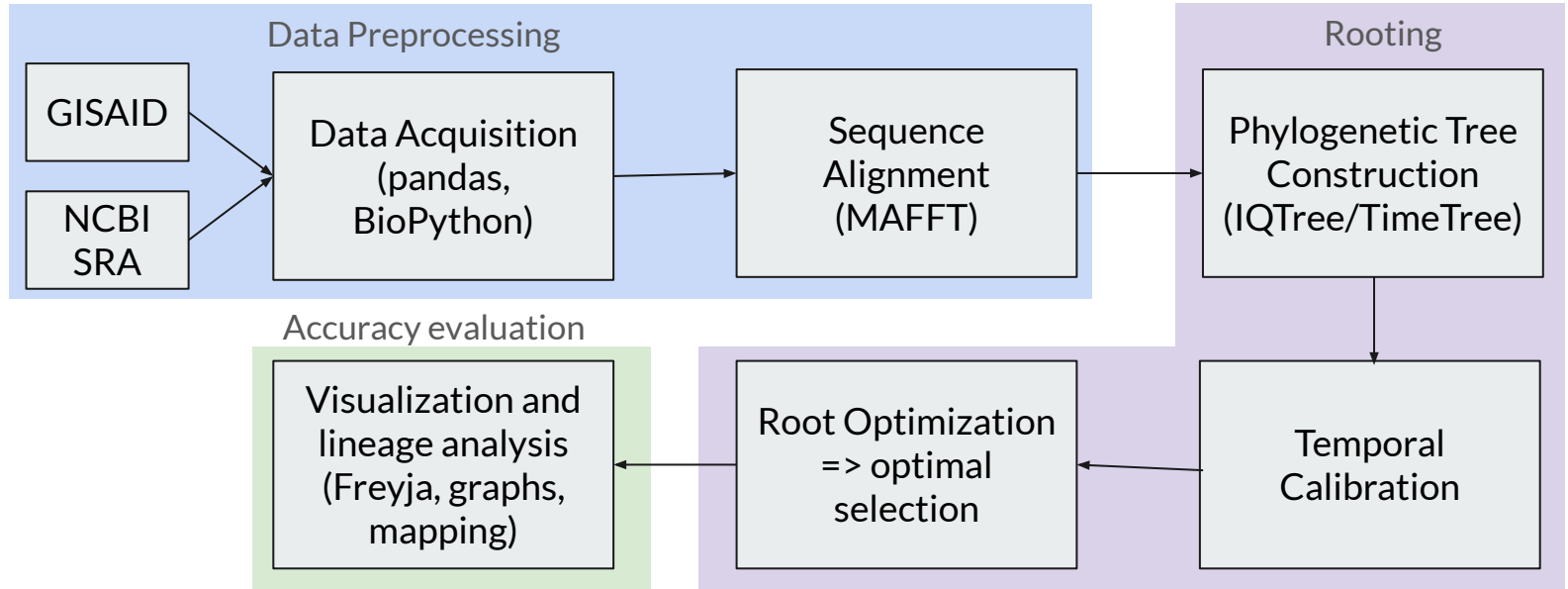
- 1) **Data acquisition:** Obtain 1689 influenza sequences from GISAID and their respective temporal information
- 2) **Sequence alignment:** Align all flu sequences using MAFFT (Multiple Alignment using Fast Fourier Transform)
- 3) **Preprocess temporal metadata** for each sequence to include sample collection dates
- 4) **Construct a base phylogenetic tree** using IQTree



Method (cont.)

- 5) **Calibrate the tree** using TimeTree, integrating temporal data
- 6) **Apply rooting techniques** to determine most accurate relationships
- 7) Analyze sequences using GeoPandas/Matplotlib to **map mutation patterns**
- 8) Use Freyja to decompose sequence data into **lineage abundances**, **visualizing variant/mutation frequencies**
- 9) **Create plots** to map lineages and emerging strains

Method: Systems Architecture





Overall Results

- Freyja covariant and mixed lineage analyses show a high prevalence of mutated lineages within avian influenza environmental samples
- Analysis shows signs of mixed lineages show risks of zoonotic spillover and genetic reassortment.
- Evolving lineages were detected in environmental samples of avian influenza, showing Freyja's adaptability to avian influenza viral strains
- Phylogenetic analyses of the influenza sequences suggest an evolutionary relationship
- Geospatial analyses using environmental data showed effects of migration on virus spread

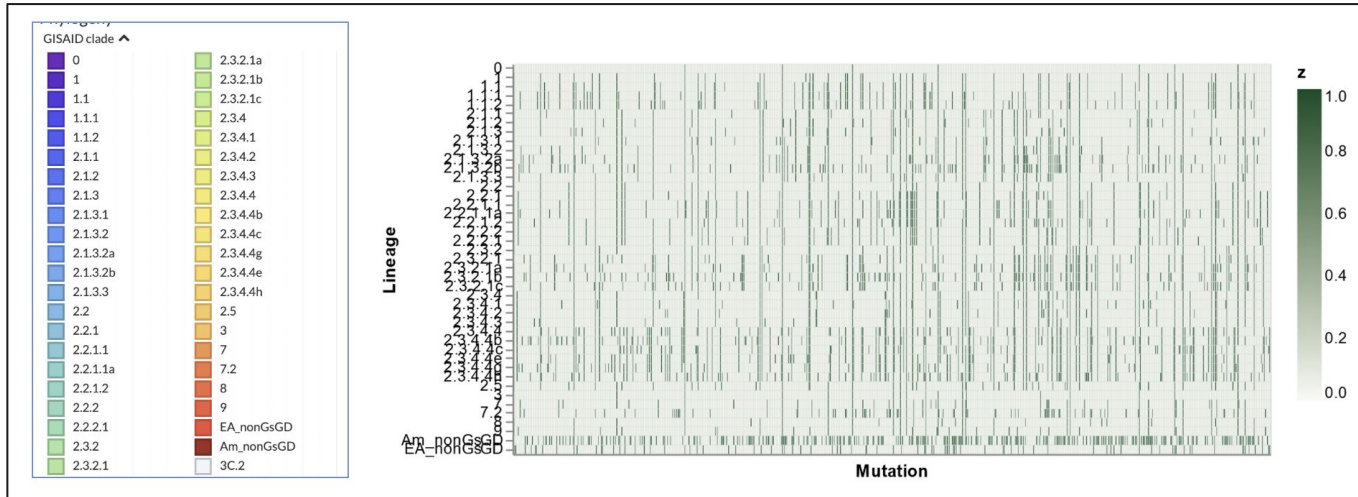
Results (cont.)

Learned how to align
sequences for avian
influenza analysis



Results (cont.)

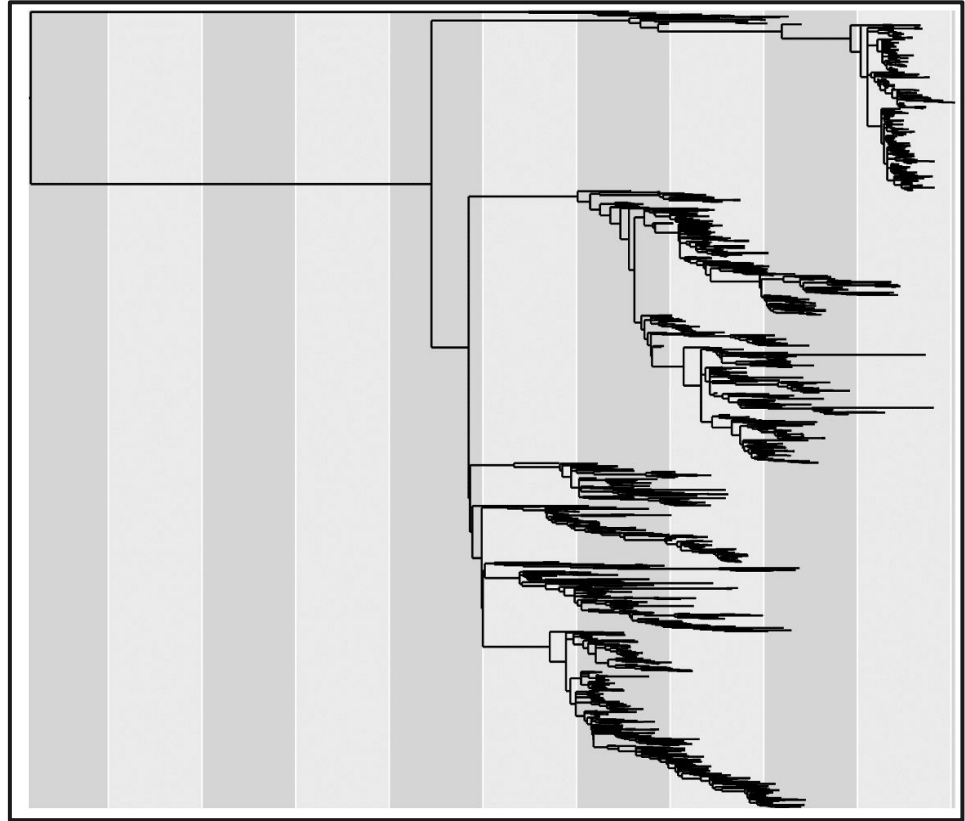
Using barcodes (gene reading) for wastewater is helpful because the presence of multiple genes can be more accurate for calculations





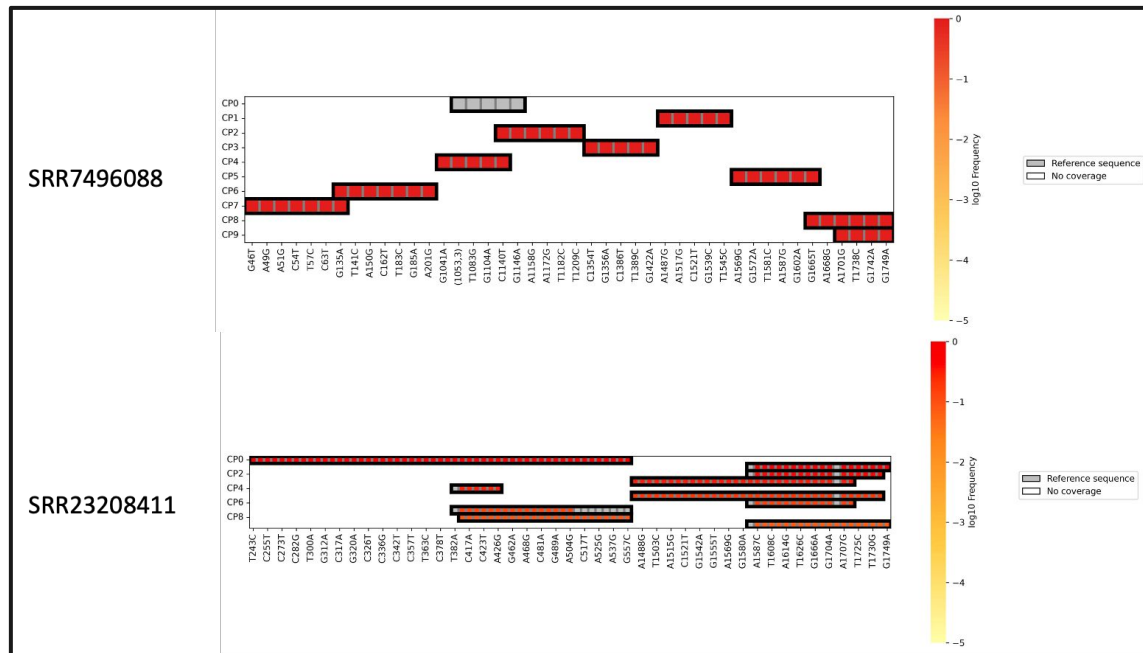
Results (cont.)

Created a phylogenetic tree using Avian Influenza data that showed GEOGRAPHIC and TIME-BASED evolutionary relationships among clustered samples



Results (cont.)

- Large-scale covariant analyses with Freyja. Ran large datasets to find precise mutation similarities and noticed patterns





Results (cont.)

- Previously observed confusion around data and source variability in prev. chart resolved through further research:
 - Sediment-collected samples contain more noise than root-collected samples like avian feces
 - No fundamental biological difference between sources if collection is accurate

Results (cont.)

- Mapping also revealed a large trend in GEOGRAPHIC spread that aligned with MIGRATION routes.



Tiles (C) Esri — Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye, Getmapping, Aerogrid, IGN, IGP, UP



Conclusion

- Phylogenetic analysis reveals virus evolution and mutation trends over time from environmental samples
- Geospatial mapping tracks spread and highlights emerging hotspots along migratory routes
- Freyja's mixed sample analysis WORKS for avian influenza and detects low-concentration variants, helping early surveillance
- Targeted sampling along migration routes can prevent transmission and enhance outbreak preparedness



Limitations

- **Efficiency:** Covariant data is hard to obtain because it requires a long processing time and must be analyzed one-by-one
- **Limited data:** Wastewater programs are being defunded by NIH budget cuts, and covariant datasets must be very detailed
- **Loss of genetic material:** When pathogens end up in a mixed environment like sewage pipes/water bodies, genetic info can be lost due to the diverse microbes present (Williams et al., 2024)



Next Steps

I hope to deploy a finished pipeline (Bash script) on GitHub that can automate this process, taking in a sequencing file and automatically generating the phylogenetic tree with:

- Rooting already fixed
- Easy-to-read labels
- Clearer way to run Freyja

Working around limitations w/ new software: potentially explore options outside of Freyja (difficult due to existing sequencing “barcodes”)



Vision for Future Improvement

- Could possibly expand sampling to analyze seasonal trends and mutation rates over time
- Use finer-scale mapping to detect micro-level outbreak zones
- Integrate with climate and migration data for predictive modeling
- Apply methods to other viruses (e.g., monkeypox, other coronaviruses)



References

- Donoghue, P. C., & Yang, Z. (2016). The evolution of methods for establishing evolutionary timescales. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 371(1699), 20160020. <https://doi.org/10.1098/rstb.2016.0020>
- Garry, R. F. (2022). The evidence remains clear: SARS-CoV-2 emerged via the wildlife trade. *Proceedings of the National Academy of Sciences of the United States of America*, 119(47), e2214427119. <https://doi.org/10.1073/pnas.2214427119>
- Hoye, B. J., Munster, V. J., Nishiura, H., Klaassen, M., & Fouchier, R. A. (2010). Surveillance of wild birds for avian influenza virus. *Emerging Infectious Diseases*, 16(12), 1827–1834. <https://doi.org/10.3201/eid1612.100589>
- Kinene, T., Wainaina, J., Maina, S., & Boykin, L. M. (2016). Rooting trees, methods for. *Encyclopedia of Evolutionary Biology*, 489–493. <https://doi.org/10.1016/B978-0-12-800049-6.00215-8>
- Ladyzhets, B. (2023, March 30). A valuable early-warning system for disease outbreaks could be shut down. *Scientific American*. Retrieved from <https://www.scientificamerican.com>



References

Navarro, A., Gómez, L., Sanseverino, I., Niegowska, M., Roka, E., Pedraccini, R., Vargha, M., & Lettieri, T. (2021). SARS-CoV-2 detection in wastewater using multiplex quantitative PCR. *The Science of the Total Environment*, 797, 148890. <https://doi.org/10.1016/j.scitotenv.2021.148890>

Schreiber, M. (2024, May 12). Lack of bird flu testing may be hiding true spread of virus on US farms. *The Guardian*. Retrieved from <https://www.theguardian.com>

Shu, Y., & McCauley, J. (2017). GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveillance: Bulletin Européen sur les Maladies Transmissibles = European Communicable Disease Bulletin*, 22(13), 30494. <https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494>



**Thank you for listening to my
presentation!**



Q and A?