Delving Below the Species Level To Characterize the Ecological Diversity in the Global Virome: An Exploration of Avian Influenza
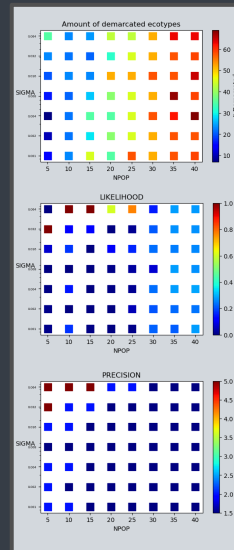
Schuyler Sloman , Rachel Hsu, Danny Krizanc, Fred Cohan

Department of Mathematics & Computer Science and the Department of Biology

Introduction:

Our research is a contribution to the Global Virome Project, which focuses on finding new viruses that have important biological differences from already known species[2]. Our project is a continuation of the work started in the Cohan lab West Nile Virus Project[3].

We have aimed to discover identify putative ecotypes – or new ecologically similar groups – for different Influenza viruses and create a phylogenetic tree. Specifically, we focused on finding evidence of ecological divergence among H3N8 ecotypes in Alaska and whether they differ in host bird which they are sampled from. To do so we searched through the NCBI database for genome sequences that fit our parameters. The choice to examine Minto Flats, Alaska was made because of the sheer amount of data that has been collected from the region.

Preparation:

Due to the lack of standardization among the data, we cleaned the GenBank files from the NCBI database and discarded all the incomplete data. We were able to locate 103 samples of H3N8 among ducks within a 30 mile radius of each other. Using MEGAX we aligned and concatenated each segment of strains, and we were able to begin modeling the divergence of H3N8 ecotypes.

Although some other states and regions had equivalent number of strains, finding over 103 strains within a 30-mile radius was the highest concentration we could find on NCBI.

Upon choosing the data we were forced to remove some data primarily due to two factors. (1) The fact that some researchers failed to upload precise coordinates (2) and the fact that certain segment sequences were missing or unusually short.

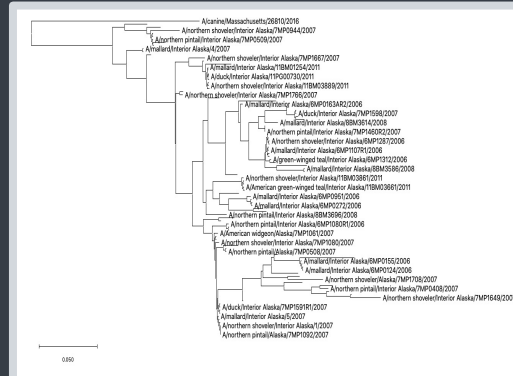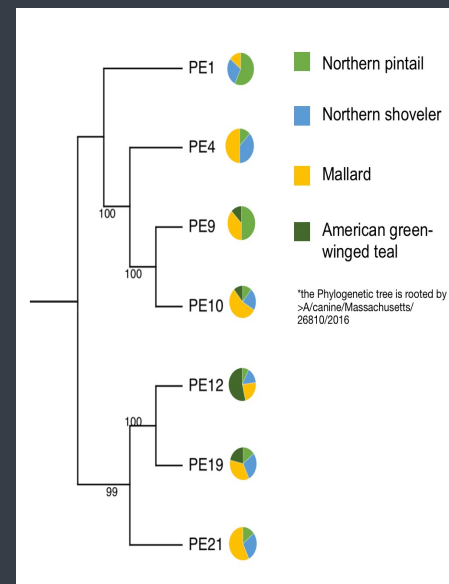Mallard   G-W Teals   N. Shoveler   N. Pintails

Methods:

Our analysis relied on Ecotype Simulation 2, written and developed by Jason M. Wood, Eric D. Becraft, Daniel Krizanc, Frederick M. Cohan, David M. Ward[1].

First, we generated a phylogeny tree from the cleaned NCBI data using FastTree. We fed in the corresponding Fasta files into the 32-bit Ecotype Simulation Program written for Windows.

Ecotype simulation, a Monte-Carlo approach to modeling the evolutionary dynamics of a microbial population based on the Stable Ecotype Model of microbial speciation, estimates the demarcation of a species and predicts the number of Putative ecotypes, in this case n=21.

In terms of statistical analysis, we first eliminated Ecotypes 2,3,5,6,7,8,11,13,14,15,16,17,18,20 because the host count was too low (<7). To determine if remaining Ecotypes (n$\geq$7) were significant based on host bird association, we used R to run a G-test.

Findings:

Our results suggest that there is a significant relationship between host birds and putative ecotypes. If the null hypothesis is that there is no association between species of host and H3N8 virus ecotype, then we would reject the null hypothesis for the G-test because the the G value = 31.271 (p=0.02678).

This suggests that are multiple ecologically distinct species of H3N8 in wild water birds in the interior Alaska region.

Northern pintail
Northern shoveler
Mallard
American green-winged teal

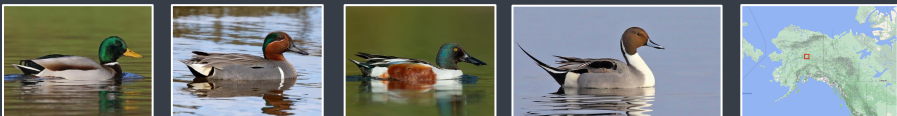*the Phylogenetic tree is rooted by >A/canine/Massachusetts/26810/2016

Conclusion:

Moving forward, the goal of this project is to expand the Ecotype simulation modelling to analyze other viruses such as H4N6 and H1N1. Eventually, we hope to model putative ecotypes for many different viruses concentrated within an area.

We also look to examine the influenza viruses with human as host species as well model closely related viruses with different ecological features to better predict when divergence begins and new viruses emerge.

References:

1. Jason M. Wood, Eric D. Becraft, Daniel Krizanc, Frederick M. Cohan, David M. Ward bioRxiv 2020.02.10.940734; doi: https://doi.org/10.1101/2020.02.10.940734

2. Carroll, D., P. Daszak, N. D. Wolfe, G. F. Gao, C. M. Morel, S. Morzaria, A. Pablos-Mendez, O. Tomori, and J. A. K. Mazet. 2018. The Global Virome Project. Science 359:872-874.

3. Tong Kong, Kelly Mei, Ammie Wang, Danny Krizanc, Frederick M. Cohan bioRxiv 2019.12.12.874214; doi: https://doi.org/10.1101/2019.12.12.874214

4. B. Perry, Johannes Sikorski, Danny Krizanc, Andrew Warner, David M. Ward, Alejandro P. Rooney, Evelyne Brambilla, Nora Connor, Rodney M. Ratcliff, Eviatar Nevo, Frederick M. Cohan Proceedings of the National Academy of Sciences Feb 2008, 105 (7) 2504-2509; DOI: 10.1073/pnas.0712205105