Exploring seasonal influenza evolution in the lab and in nature

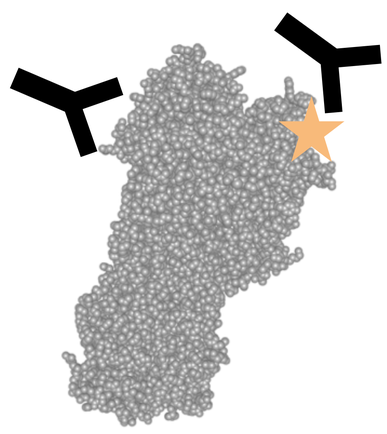
Allison Black, Sarah Hilton, John Huddleston, and Khrystyna North

PROBLEM

Every few years the flu vaccine needs to be updated because the virus has evolved to escape human immunity. As computational and laboratory biologists interested in influenza, we want to understand how flu evolves to better predict vaccines.

Antibodies in our body bind to the influenza protein HA and prevent the virus from infecting cells. However, mutations at specific sites in HA can allow the virus to "escape" the antibodies and seed an infection.

New experimental approaches allow us to measure the ability of every possible mutation in HA to escape antibodies giving us a map of immune selection on the gene. The full power of these large, experimental datasets is only really achieved when the data is contextualized with the three-dimensional protein structure and the naturally occurring amino-acid frequencies



MOTIVATION

Years of work and multiple PhDs have been invested in designing, refining, and implementing the experimental workflow to generate immune selection data. However, the strength of the experiment is limited by the ability of scientists to easily interpret and contextualize the results.

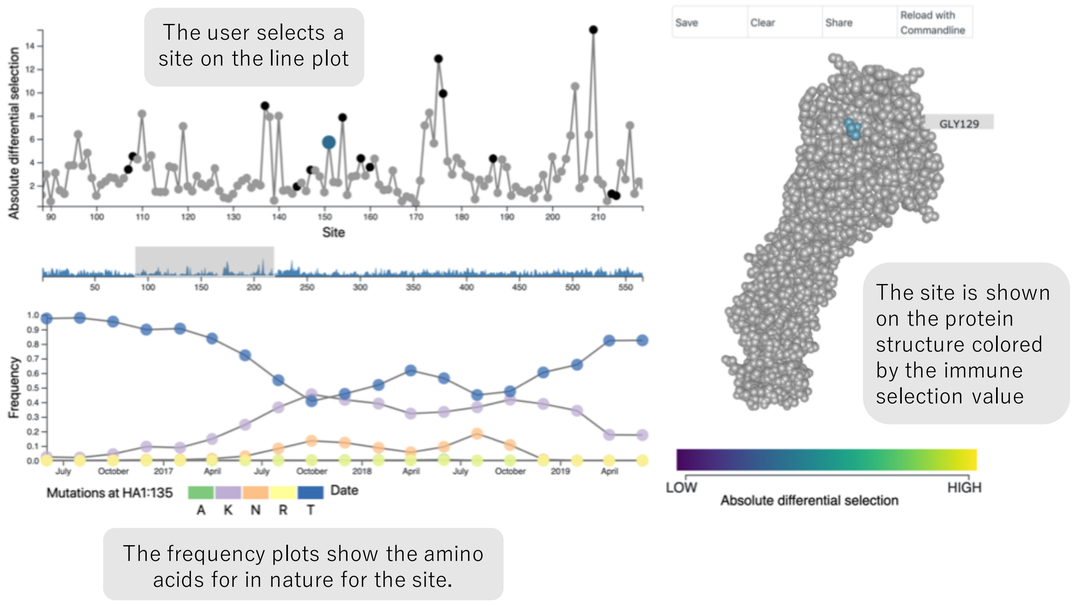
Currently, the computational workflow to analyze the data involves iteratively generating static graphs for all three streams of data: immune selection, protein structure, and natural frequencies. The task of integrating these three sources falls on the user and requires extensive domain knowledge about the protein and the plotting software.

The power of these datasets should not be limited by computational expertise. We want to make a platform that can be used by the scientists who generate the data, their collaborators, and any other interested scientists to explore the data and generate hypotheses.

APPROACH + RESULTS

full display

The purpose of our visualization is to link three pieces of information, laboratory immune selection data, the 3-D protein structure and natural amino-acid frequencies, by their common feature, "site".

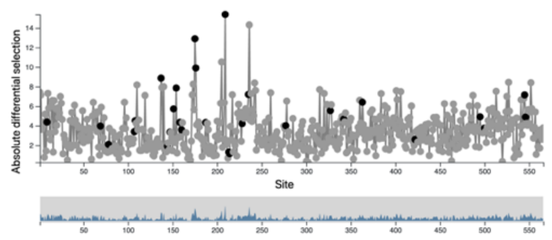


With one view, the user can answer three important questions:
(1) Is my site of interest important for immune escape?
(2) Where is my site of interest on the protein structure?
(3) Did my site of interest have any amino-acid substitutions in nature?

high-throughput immune selection assays

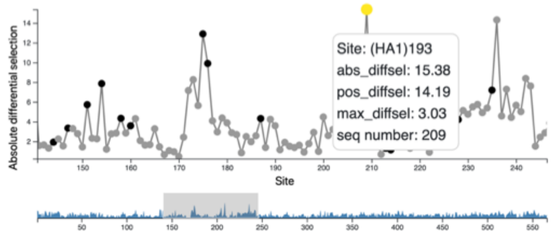
The goal for the visualization of the immune selection data is to allow the user to
- view the entire to gene to visually identify epitopes (peaks)
- interrogate the specific-sites
- identify which sites have natural frequency data

Black dots indicate sites with natural frequency data



The context plot shows the entire gene

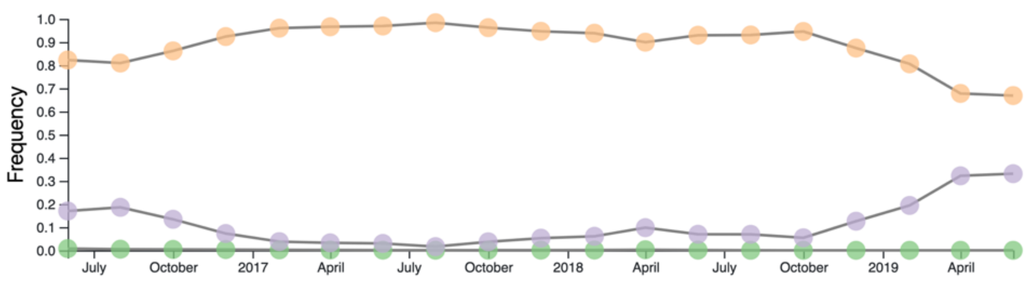
The tooltip returns specific values from the experiment



Brushing on the context plot zooms in on specific regions

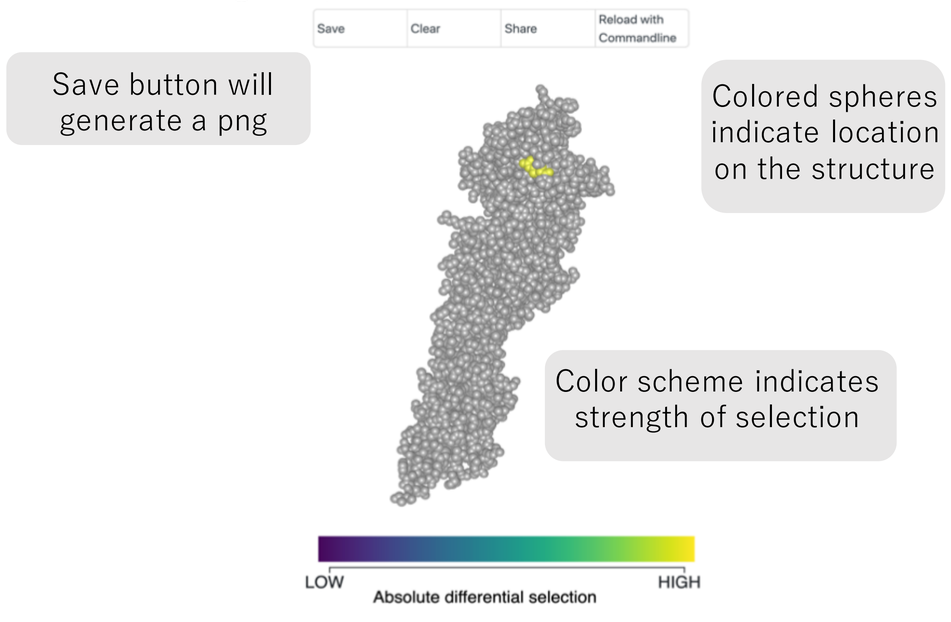
amino-acid frequencies in nature

The goal for the natural frequency visualization was for the user to be able to
- Identify the specific amino acids observed at their site of interest
- Determine the frequency of these amino acids over time



three-dimensional protein structure

The goal for the protein structure visualization was for the user to be able to see
- where on the protein structure their site of interest falls
- identify the strength of immune selection on their site of interest
- save a static image for presentations and papers



FUTURE DIRECTIONS

Currently, our prototype displays data for a single experiment on a single protein. However, there are multiple experiments performed on influenza HA and similar experiments performed on other viral proteins, such as HIV env and Zika E. Moving forward, we would like our tool to be flexible enough that others could upload their own data.

ACKNOWLEDGMENTS

Thank you to Drs. Adam Dingens and Juhye Lee for their data from and expertise on the experimental assay and to Halden Lin and Jane Hoffswell for their feedback and design review.