Aubry Landsom

Phylogenetics Seminar

Viral Phylodynamic Applications to Vaccine Prediction – Annotated Bibliography

Carnegie, L., J Raghwani, G Fournié, & Hill, S. C. (2023). Phylodynamic approaches to studying avian influenza virus. *Avian Pathology*, *52*(5), 289–308. https://doi.org/10.1080/03079457.2023.2236568

This research focuses on using phylodynamic methods that can be used to study avian influenza. These methods can be helpful in studying how the virus spreads between populations over geographical distances. The paper also works to identify virus dispersal factors as well as factors that influence issues with phylodynamic interpretation. The data analyzes the different factors of viral dispersal between wild and domestic birds as well as the spread within domestic bird populations. These data included single species transmission rates and transmission between different species. The paper found that there are many ways to improve the current viral tracking methods. One such way is with a surveillance system such as Nextstrain. Another important change could be using bird travel history in phylogenetic analysis. Controlling the spread of avian influenza can largely impact the health of humans and other mammals. Understanding the spaciotemporal spread of this virus is incredibly helpful in this goal.

Dolan, P. T., Whitfield, Z. J., & Andino, R. (2018). Mechanisms and Concepts in RNA Virus Population Dynamics and Evolution. *Annual Review of Virology*, *5*(1), 69–92. https://doi.org/10.1146/annurev-virology-101416-041718

RNA viruses have the highest mutation rate, can infect many different hosts, and can exist in a variety of ecosystems. Scientists have determined that there is one mutation per replication on average for RNA viruses. Recombination also occurs in RNA viruses. It occurs through a process called copy choice where the polymerase jumps from one transcript to another when transcribing the genome. Homologous recombination creates genetic variation while nonhomologous recombination often results in defective interfering particles which act as decoys. Reassortment in viruses requires two viral genomes to infect the same cell an create virions with a mix of segments. The robustness of a virus is separated into the categories intrinsic and extrinsic. Intrinsic robustness is based on the genome and its impact on factors such as codon bias and protein structure. Extrinsic robustness is based on the factors that the host provides, specifically the cellular chaperone system. The vast diversity in RNA viruses with large population sizes leads to a quasispecies or many genotypes that surround a master sequence. Several viral species have been known to show cooperativity which describes a situation when a new phenotype results from two genotypes. Population sizes can grow through en bloc transmission where multiple virions are transmitted with multivesicular bodies or bacteria.

Viruses can evolve across time and space through host-to-host, intraspecies, and cell-to-cell transmission.

***Joseph K. Agor & Osman Y. Özaltın (2018) Models for predicting the evolution of influenza to inform vaccine strain selection, Human Vaccines & Immunotherapeutics, 14:3, 678-683, DOI: 10.1080/21645515.2017.1423152**

This research paper introduces the basics of why the influenza vaccine is necessary as well as the specific of how many strains of influenza A and B are normally included in the vaccine. Currently, surveillance of emerging viruses and animal models are used to predict future strains of the virus but is difficult to predict antigenic drift as well as the fate of the strains that do emerge in the population. These are important issues to consider because vaccine effectiveness is based on the accuracy of the strain predications and these predications must occur far before the strain emerges in the population due to the amount of time that it takes to create a vaccine and circulate it. To look at the connection between antigenic differences and mutations, researchers use evolution data. This is where phylogenetic trees, such nextflu, come into play. nextflu analyzes circulation patterns to provide projections for future years. Allele dynamic plots can also be used to infer a phylogenetic tree and the ancestral character state to look for characters with a selective advantage. Researchers also use antigenic maps that represent serology data based off hemagglutinin inhibition assays (antigens) and reference antisera (antibodies). Unfortunately, ferrets, which are used as model systems due to the respiratory systems, have different immune systems than humans and don't factor in how a human's immune history can impact their response to new strains. These models can help predict alleles that are most likely to become predominant or estimate fitness of a virus. Fitness is examined specifically to the mutation locations. Antigenic analysis shows that only certain amino acid mutations have antigenic impacts. Vaccine immunity and efficacy is also possible to determine through a computational framework called DAMIP. Effectiveness of vaccines is also determined by vaccination history and are less effective for those who were not recently vaccinated.

Łuksza, M., Lässig, M. A predictive fitness model for influenza. *Nature* 507, 57–61 (2014). https://doi.org/10.1038/nature13087

This paper discusses the important role of the epitope region of the hemagglutinin segment of the influenza genome. Population frequency predictions are possible to determine once researchers identify fitness based on genotype and host environment. Models are used to predict evolution based on fitness. Looking at the different HA sequences leads to the finding that there have been several point mutations at epitope and non-epitope positions. These sequences are put into a phylogenetics tree where each clade lasts up to about 5 years before it becomes a fixation or a loss. The different clades can be distinguished based off their different

combinations of mutations. The model calculates observed frequency by summing strain frequencies. The fitness model uses epitope and non-epitope genotypes and observes fitness of strains against individuals who have partial immunity due to previous exposure to similar strains. As expected, similarity of the strains impacts the cross-immunity amplitude. They also find that epitope mutations are under positive selection, while non-epitope mutations face negative selection. The fitness is then determined based on the expected impact of the fitness costs from non-epitope mutations combined with reduced cost as epitope mutations occur. Other fitness models were compared to establish an estimated accuracy of the model. Points of origination are used to look at connections between fitness and geographical location. The model selected vaccine strains that were always smaller in amino acid distance that the actual vaccine strains. The model is also useful in discovering how influenza evolution is impacted by vaccination. Another quantity that is helpful is the cumulative fitness flux which reveal the total amount of adaption in each clade. It was found that high-fitness clades lead to high-fitness clades through beneficial mutations and lead to loss through detrimental mutations.

***Ma, Y., Liu, K., Yin, Y., Qin, J., Zhou, Y.-H., Yang, J., Li, S., Poon, L. L. M., & Zhang, C. (2020). The Phylodynamics of Seasonal Influenza A/H1N1pdm Virus in China Between 2009 and 2019. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/fmicb.2020.00735**

The different origins of the strains as well as the time frame that they infected humans is described with the swine-origin strain being the focus of this research. Due to the inaccuracy of the RNA polymerase used in viruses, the mutation rate is extremely high. Previous characterizations of strains have not occurred recently and left a hole in the data. There is also a recent increase in cases, which required characterization of recent strains to understand and analyze. For this reason, this study works to fill in the gap using the complete genome as well as hemagglutinin and neuraminidase sequences to map the genetic variation over in is the A/H1N1 pdm virus in China from 2009-2019. To understand selection for specific traits the non-synonymous to synonymous substation rates were utilized. MEME and BEB methods were used to infer positively selected sites. MCC trees show the evolution of influenza using eight segments of the genome. The results showed vary selection pressure on the different segments with HA, NA, and NS (prevents inhibitory response) genes experiencing the most pressure. Within the HA and NA glycoproteins, there have been mutations that cause fixed changes in the amino acid content of specific sites on each protein. Unsurprisingly, HA is essential for binding to the host cell and NA is necessary for virion release. The researchers also mapped out spatial dynamics on the transmission patterns of the virus. The variations in the dN/dS ratio over time are suggested as being caused be times when the virus was entering a new host population (increase, 2009-2010), maintaining fitness (decrease, 2011-2017) or increased selection from vaccination or previous infection (increase, 2018-2019). A comparison of several different strains

of influenza shows some similarities. For example, all four had high levels of activity during 2018-2019.

Neher, R. A., Bedford, T., Daniels, R. S., Russell, C. A., & Shraiman, B. I. (2016). Prediction, dynamics, and visualization of antigenic phenotypes of seasonal influenza viruses. *Proceedings of the National Academy of Sciences*, *113*(12), E1701–E1709. https://doi.org/10.1073/pnas.1525578113

Hemagglutinin inhibition assays work by finding the minimal antiserum concentrations needed to stop crosslinking red blood cells by a specific amount of virus. These assays use ferret antiserum which is effective against the homologous virus and some heterologous viruses when at higher concentrations. The HI differences can be utilized to infer the antigenic evolution of individual amino acid substitutions. These findings reveal that HA phylogenetic trees are accurate models for antigenic properties of influenza viruses. Using this data can be helpful in prediction of dominant clades. There are two models that work to predict HI titers based on the sequence differences. The first model is the tree model which looks at the connection between the branches on a phylogenetic tree while the second, a substitution model uses the number of amino acid substitutions to make these same inferences. HI measurements are also based on the reactivity of antisera and viruses. As expected, as genetic variance increases, titers tend to decrease. The prediction values for the tree and substation models were recorded for a variety of strains for both titers and viruses. The next section of the paper works to find the total past antigenic change. It found that rates of change vary dramatically with the different types of virus lineages. For A(H3N2) half of the evolution is estimated to be many substitutions that caused a small effect. The HI titers also work to approximate the distance on a tree. As stated in many other papers, most antigenic evolution occurs at certain sites, specifically the seven positions near the HA receptor binding site. They also found that large antigenic changes tend to persevere longer than small changes, but there are exceptions that may be due to fitness costs from changes.

Richard A. Neher, Trevor Bedford, nextflu: real-time tracking of seasonal influenza virus evolution in humans, *Bioinformatics*, Volume 31, Issue 21, November 2015, Pages 3546–3548, https://doi.org/10.1093/bioinformatics/btv381

As the influenza virus is constantly evolving, so must the vaccine made to combat this virus. To ensure that the vaccine is most effective for the strains that are present in the population, there must be a system for recording all news strains and analyzing the evolution of the virus in real-time. The nextflu program hopes to do just that by acting as a computational method of comparing relationships between the virus strains and creating a visual of the genetics and epidemiological features of the virus. nextflu is currently tracking all four strains: A/H3N2, A/H1N1pdm, B/Victoria and B/Yamagata. The system works by analyzing virus sequence data

and displaying the results visually. It requires the name of the strain. The date it was sampled, and passage history. Filtering of the data works to find a representative sequence for a local outbreak or incomplete information. Based on this data, a phylogenetic tree is made using FastTree and then refined. The system infers internal nodes using a marginal maximum likelihood method as well as estimates frequency trajectories of mutations, genotypes, and clades. The user can then select viruses based on dates or color the tree based on epitope mutations, receptor binding mutations, local branching index, HA genotype or geographic region. The display also offers a frequency plot that shows the trajectory of selected clades. Finally, the programs is also able to plot trajectories of individual mutations, two mutations combined, or predefined clades.

Steinbrück L, Klingen TR, McHardy AC2014.Computational Prediction of Vaccine Strains for Human Influenza A (H3N2) Viruses. J Virol88:.https://doi.org/10.1128/jvi.01861-14

The antigenic drift of the influenza virus depends mainly on epitope variation of hemagglutinin and neuraminidase, but this study looks only at hemagglutinin data. Different strains are used to create a vaccine every year. Each vaccine contains four strains total for both the Northern Hemisphere and the Southern Hemisphere. Due to the increased data on present and past viruses, phylogenetic analysis and antigenic cartography have become the primary methods for vaccine prediction. Another useful technique is allele dynamics plots which looks at the different alleles of a gene over time and find allele that have the largest probability of experiencing directional selection, but the weakness of this method is that specific alleles will score best every time regardless of the comparison of current predominant strains. A recent method uses nonnegative least-squares optimization to infer antigenic trees and is applied in this study. The methods attempted to looked for alleles most likely to be on the rise to propose for a vaccine for the next year. If no new allele showed predominance, the vaccine was left unchanged. The allele that represented the predominant virus in the next season was considered positive while all other alleles were considered negatives. This categorization system was used to list predictions as true positives, true negatives, false positives, and false negatives. A false positive would result in the creation of a mismatching vaccine while a false negative would lead to a vaccine that is not updated. The results showed that the predictions had a 78% accuracy compared to the WHO (World Health Organization) which had predictions at 66% accurate.

Steinbrück L, McHardy AC, Allele dynamics plots for the study of evolutionary dynamics in viral populations, *Nucleic Acids Research*, Volume 39, Issue 1, 1 January 2011, Page e4, https://doi.org/10.1093/nar/gkq909

This paper analyzes the use of allele dynamics plots for visualizing the evolution of a gene and analyzing alleles that are likely to face directional selection. This technique is used on two different influenza A strains: H3N2 and H1N1. AD plots are constructed using information from phylogenetic inference and combining it with ancestral character state reconstruction. At a population level, AD plots show alleles that might show selective advantage. Though dN/dS (non-synonymous-to-synonymous mutations) ratios have been used in past research, they may not be the most effective method for looking at selection within a population. A novel method called antigenic cartography which uses multidimensional scaling of assay data showed that antigenic evolution is more clustered than genetic evolution. AD plots work using a series of steps. First, they infer evolutionary relationships, ancestral character states, and evolutionary intermediates. Next, genetic changes are mapped to tree topology so the prevalence of different alleles at different times can be determined. Lastly, the AD plots reveal how fast the genetic variants or alleles propagate in the population. Alleles with selective advantage will increase in frequency faster than those without an advantage. The plot itself visually represents this data by charting frequency of the allele versus time to make it obvious which alleles increase in prevalence most rapidly over two consecutive influenza seasons. These plots can help identify the strains that are the more relevant targets for vaccine creation. In fact, the allele that scored high in the 1998/99 Northern season was used as a vaccine component in the 1999-2000 Northern season. Targeting these amino acids that are unlikely to change helps to create a vaccine that is usefully across a larger time. A limitation of this method is the ability for a specific allele to score best for every time period, with no support for or against if it is antigenically similar or different to the current vaccine strain. Therefore, antigenic information is also important for the consideration of the next vaccine.

Volz EM, Koelle K, Bedford T (2013) Viral Phylodynamics. PLOS Computational Biology 9(3): e1002947. https://doi.org/10.1371/journal.pcbi.1002947

This paper defines viral phylodynamics as the study of how epidemiological, immunological, and evolutionary processes impact viral phylogenies. More broadly phylodynamics in virology highlights the many factors that impact genealogy that are specific to viruses. Population size impacts the changes in branch size over time such that which exponential growth in the population branch length increase towards the present whereas a constant population size will result in shorter branch lengths closer to the present. Host populations affect taxa clustering due to geographic residence of populations. Selection influence viral phylogenies through immune escape process variance. Understanding these factors can help identify viral origins, predict viral spread, and dictate viral control efforts such as vaccination or antiretroviral drug therapies. The coalescent theory which shows a nonrecombining gene copy ancestry is often used to map of the phylogeny of a virus. Another factor is phylogeography which looks at how geographical location in related to genetics. Finally, simulations look at the sequence data to find geographic patterns. The paper then utilizes influenza and HIV as models

of comparison for viral phylodynamics. It explains influenza as a virus that primarily accumulates mutations in the exposed hemagglutinin and neuraminidase regions and therefore faces strong selective pressure in these regions. Influenza also has a rapid population turnover rate meaning that geographic spread often occurs very quickly. HIV, on the other hand, has-subtypes that are dominant in different parts of the globe. These subtypes vary in several ways including how they are transmitted and the therapies necessary to treat them. Transmission often occurs at the beginning of an infection period in most HIV cases. HIV evolution is the result of host-to-host mutations as well as within-host mutations. Finally, the paper suggests that future research look at DNA viruses in more depth.