Master of Public Health
Applied Practice Experience

IDENTIFYING A ROLE FOR GENOMIC EPIDEMIOLOGY IN POST-PANDEMIC INFLUENZA SURVEILANCE IN KANSAS

by

Cole J. King

MPH Candidate

submitted in partial fulfillment of the requirements for the degree

MASTER OF PUBLIC HEALTH

Graduate Committee:
Bethany Plakke, PhD
Sherry Fleming, PhD
Ellyn Mulcahy, PhD, MPH

Applied Practice Experience Site:
Kansas Department of Health and Environment,
Bureau of Epidemiology and Public Health Informatics

Applied Practical Experience Preceptor:
John Anderson, MSc., PhD

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

Table of Contents

Chapter 1 - Summary of Portfolio Products

I completed my Applied Practice Experience (APE) at the Bureau of Epidemiology and Public Health Informatics at the Kansas Department of Health and Environment (KDHE) office in Topeka, Kansas. At KDHE, I completed a three-phase project to identify opportunities and challenges for implementing a genomic surveillance program for influenza in Kansas. Genomic epidemiology is an emerging field in public health, in large part due to the vast amount of genomic data made available during the SARS-CoV-2 pandemic. Now that infrastructure exists to generate genomic surveillance data, an open question remains about how best to use these resources to monitor other infectious diseases now that the peak of the pandemic has subsided. Therefore, the main goal of my APE was to review the current literature on influenza and SARS-CoV-2 to understand their genomic epidemiology, analyze the current limited genomic data for influenza in Kansas as a pilot case, and conduct system-level analysis of the U.S. influenza surveillance network to identify a role for genomic epidemiology at the state level.

The first phase of the project was a literature review on the genomic epidemiology of influenza viruses, namely regarding how the novel genomic epidemiology tools from the SARS-CoV-2 pandemic can be used to monitor influenza. From this literature review, I wrote two short, plain-language articles describing the genomic epidemiology of both SARS-CoV-2 and influenza (products 1 and 2) for inclusion in training materials for epidemiologists unfamiliar with genomic epidemiology. The article focuses on the nomenclature and designations of the two viruses and includes information on how genomic epidemiologists surveil them. The literature review highlighted transmission between animals and humans and informed the direction of my phylogenetic analysis of influenza A sequences from Kansas to include humans and non-human animals.

Following the literature review, the second phase of the project analyzed influenza A genomes collected in Kansas by the Kansas Health and Environment Laboratories (KHEL) and genomes obtained from the National Center for Biotechnology Information (NCBI) database. From these data, I made phylogenetic trees to visualize the evolutionary relationships between the viruses sampled. These trees were annotated with metadata, including host species, genome subtype, and date of collection for the hemagglutinin (HA) and neuraminidase (NA) genes (products 3 and 4). Notably, some samples from different hosts clustered phylogenetically and temporally on the hemagglutinin tree for the H1 subtype, suggesting possible transmission events between humans and non-human animals.

In the final phase of the project, I conducted systems-level analysis of the U.S. influenza surveillance system. First, I completed a process map to identify key players and systems involved in the complicated and often uncoordinated U.S. influenza surveillance system (product 5). Secondly, I created a causal loop diagram to identify the dynamics of the system and to identify challenges brought about by the disjointed system, particularly how they may lead to unrepresentative data collection and, therefore, bias decision making (product 6). Lastly, I utilized system archetypes to identify dynamics that influence public perceptions and how those perceptions feed back to influence the availability of resources for influenza surveillance (product 7). Following the analysis, I compiled my findings into an oral presentation which I delivered to epidemiologists at KDHE at the conclusion of the APE (product 8).

Table 1.1 Summary of Portfolio Products

Portfolio Product		Description
1	An Overview of SARS-CoV-2 Genomic Epidemiology	Plain-language article explaining the genomic epidemiology of SARS-CoV-2 and the practical use of genomic epidemiology in its surveillance
2	An Overview of Influenza Virus Genomic Epidemiology	Plain-language article explaining the genomic epidemiology of influenza and the practical use of genomic epidemiology in its surveillance
3	Hemagglutinin (HA) Phylogenetic Tree	Annotated phylogenetic tree of influenza hemagglutinin sequences combining state and federal data
4	Neuraminidase (NA) Phylogenetic tree	Annotated phylogenetic tree of influenza neuraminidase sequences combining state and federal data
5	Process Diagram of U.S. Influenza Surveillance	Diagram mapping the U.S. surveillance structure underlying influenza vaccine formulation and the directional transfer of data within it
6	Causal Loop Diagram of U.S. Influenza Surveillance	Diagram describing interdependencies and vulnerabilities in influenza surveillance in the United States using a common systems-thinking tool
7	System Archetypes Underlying the Success/Failure of Influenza Prevention	Diagram describing the interplay between public dynamics and influenza prevention in the United States
8	Identifying a Role for Genomic Epidemiology in Post-Pandemic Influenza Surveillance (Oral Presentation Slides)	Slides from oral presentation describing the necessity and potential for genomic epidemiology in future influenza surveillance

Table 1.2 Portfolio Products and Competency Addressed

Portfolio Product		Number and Competency Addressed	
1	An Overview of SARS-CoV-2 Genomic Epidemiology (Original Article)	18,19	Select communication strategies for different audiences and sectors. Communicate audience-appropriate public health content, both in writing and through oral presentation.
2	An Overview of Influenza Virus Genomic Epidemiology (Original Article)	18,19	Select communication strategies for different audiences and sectors. Communicate audience-appropriate public health content, both in writing and through oral presentation.
3	Hemagglutinin (HA) Phylogenetic Tree	3,4	Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming, and software, as appropriate. Interpret results of data analysis for public health research, policy, or practice.
4	Neuraminidase (NA) Phylogenetic Tree	3,4	Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming, and software, as appropriate. Interpret results of data analysis for public health research, policy, or practice.
5	Process Diagram of U.S. Influenza Surveillance	22	Apply systems thinking tools to a public health issue.
6	Causal Loop Diagram of U.S. Influenza Surveillance	22	Apply systems thinking tools to a public health issue.
7	System Archetypes Underlying Success/Failure of Influenza Prevention	22	Apply systems thinking tools to a public health issue.
8	Identifying a Role for Genomic Epidemiology in Post-Pandemic Influenza Surveillance (Slides)	4,18,19	Interpret results of data analysis for public health research, policy, or practice. Select communication strategies for different audience and sectors. Communicate audience-appropriate public health content, both in writing and through oral presentation.

Chapter 2 - Attainment of MPH Foundational Competencies

Table 2.1 Summary of MPH Foundational Competencies

Number and Competency		Description
3	Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming, and software, as appropriate.	Influenza sequencing data were aligned and analyzed quantitatively using a well-known algorithm. The resulting phylogenetic trees were then annotated and aesthetically mapped with available qualitative metadata.
4	Interpret results of data analysis for public health research, policy, or practice.	Phylogenetic analyses of influenza sequences were placed into a One Health and historical context to advocate for the use of genomic epidemiology in future influenza surveillance.
18	Select communication strategies for different audiences and sectors.	Two articles and an oral presentation were developed for audiences of differing familiarities with genomic epidemiology. The articles were written for an educated public, and the oral presentation was developed for an audience of epidemiologists.
19	Communicate audience-appropriate public health content, both in writing and through oral presentation.	The styles of the articles and presentations meaningfully differed, with the articles highlighting foundational concepts in accessible language. The oral presentation, on the other hand, highlighted theoretical concepts in technical and precise scientific language.
22	Apply systems thinking tools to a public health issue.	The U.S. influenza surveillance system was analyzed with a process map, a causal loop diagram, and system archetypes. This allowed for long-range cause-and-effect relationships to be integrated and visually displayed.

3. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming, and software, as appropriate.

Hemagglutinin (HA) and neuraminidase (NA) sequences of samples collected from Kansas were obtained from multiple sources and aligned using multiple sequence alignment tools. Once aligned, a matrix of the genetic distance between samples was made using the Jukes-Cantor model in the computer program R. The R programming language was also used to import and clean metadata associated with the samples. Once sequences and metadata were prepared, phylogenetic trees of the sequences were constructed and annotated with patient- or sample-derived metadata. This allowed for visualization of phylogenetic data and integration of qualitative metadata like host species and date of collection into the analysis.

4. Interpret results of data analysis for public health research, policy, or practice.

The findings of the phylogenetic analysis speculatively suggested that human and non-human animal viruses are related and may be subject to back-and-forth transmission, a finding consistent with literature on the subject. Though speculative, I used these data to advocate for an integrated One Health approach to influenza surveillance that utilizes representative data from multiple sources. Since the human-animal sequence clusters were predominantly in the hemagglutinin subtype H1 gene, these data were also placed in the historical context of the likely emergence of the pandemic 1918 H1N1 influenza virus in rural Kansas.

18. Select communication strategies for different audiences and sectors.

Original media directed at different audiences were developed over the course of the APE. First, two original, plain-language articles were written for an educated lay audience describing the core tenets of their genomic epidemiology and relevance to public health. These articles will be used in future training materials for genomic epidemiology at KDHE. Second, an

oral presentation (and accompanying slide deck) was made regarding the potential for genomic epidemiology in next-generation influenza surveillance. This presentation was directed toward an audience of epidemiologists with field-specific knowledge and context.

19. Communicate audience-appropriate public health content, both in writing and through oral presentation.

In the original media developed over the course of the APE, careful consideration was made regarding which content was appropriate for each of the two audiences. The first pieces of media (the two original articles) communicated, through writing, basic and foundational information about the practice and implementation of genomic epidemiology in accessible terms. The majority of this information was collated from public facing government webpages from public health authorities and supplemented with additional information from academic sources. This is in contrast with the oral presentation developed with an audience of epidemiologists in mind. The information in the oral presentation was primarily derived from primary investigation and academic sources, and the language used in the presentation is thus more technical and precise.

22. Apply systems thinking tools to a public health issue.

Qualitative analysis of the U.S. influenza surveillance system was conducted using common systems thinking tools: process maps, causal loop diagrams, and systems archetypes. The process map allowed for the development of a foundational understanding of the U.S. influenza surveillance system, particularly the flow of information and data between entities and the multiple agencies and jurisdictions involved. The causal loop analysis took this one step further and allowed for analysis of not only the flow of information, but also the consequences of interagency competition and the role of dynamic public processes in shaping outcomes. In the

same vein, system archetype analysis allowed for conceptualization of stochastic social processes and how they play into the resource landscape available to combat infectious diseases.

Table 2.2 MPH Foundational Competencies Course Mapping

22 Public Health Foundational Competencies Course Mapping	MPH 701	MPH 720	MPH 754	MPH 802	MPH 818
Evidence-based Approaches to Public Health					
1. Apply epidemiological methods to the breadth of settings and situations in public health practice	x		x		
2. Select quantitative and qualitative data collection methods appropriate for a given public health context	x	x	x		
3. Analyze quantitative and qualitative data using biostatistics, informatics, computer-based programming and software, as appropriate	x	x	x		
4. Interpret results of data analysis for public health research, policy or practice	x		x		
Public Health and Health Care Systems					
5. Compare the organization, structure and function of health care, public health and regulatory systems across national and international settings		x			
6. Discuss the means by which structural bias, social inequities and racism undermine health and create challenges to achieving health equity at organizational, community and societal levels					x
Planning and Management to Promote Health					
7. Assess population needs, assets and capacities that affect communities' health		x		x	
8. Apply awareness of cultural values and practices to the design or implementation of public health policies or programs					x
9. Design a population-based policy, program, project or intervention			x		
10. Explain basic principles and tools of budget and resource management		x	x		
11. Select methods to evaluate public health programs	x	x	x		
Policy in Public Health					
12. Discuss multiple dimensions of the policy-making process, including the roles of ethics and evidence		x	x	x	
13. Propose strategies to identify stakeholders and build coalitions and partnerships for influencing public health outcomes		x		x	x
14. Advocate for political, social or economic policies and programs that will improve health in diverse populations		x			x
15. Evaluate policies for their impact on public health and health equity		x		x	
Leadership					
16. Apply principles of leadership, governance and management, which include creating a vision, empowering others, fostering collaboration and guiding decision making		x			x
17. Apply negotiation and mediation skills to address organizational or community challenges		x			
Communication					

22 Public Health Foundational Competencies Course Mapping	MPH 701	MPH 720	MPH 754	MPH 802	MPH 818
18. Select communication strategies for different audiences and sectors	DMP 815				
19. Communicate audience-appropriate public health content, both in writing and through oral presentation	DMP 815				
20. Describe the importance of cultural competence in communicating public health content		x			x
Interprofessional Practice					
21. Perform effectively on interprofessional teams		x			x
Systems Thinking					
22. Apply systems thinking tools to a public health issue			x	x	

Chapter 3 - Attainment of MPH Emphasis Area Competencies

My ILE experience was the completion of a thesis (submitted separately from this report) entitled "Maternal Choline Supplementation Modulates Cognition and Induces Anti-Inflammatory Signaling in the Prefrontal Cortex of Adolescent Rats Exposed to Maternal Immune Activation." This study used an animal model of maternal immune activation (MIA) by a viral mimetic to simulate maternal infection and found that exposure to MIA was sufficient to induce behavioral and inflammatory changes in the prefrontal cortices of adolescent rats. This effect was modulated by a maternal diet supplemented with the essential micronutrient choline. This study achieved all IDZ emphasis area competencies as documented below and contributes to scholarship in the fields of public health and developmental neuroscience.

1. Evaluate modes of disease causation of infectious agents.

Maternal infection has been described as a risk factor for psychiatric and neurodevelopmental sequelae in the offspring. Though not infectious in the traditional sense, a growing body of literature supports a link between exposure to infectious disease and neurodevelopmental disorders through activation of the maternal immune response. The immunological responses to infectious diseases have direct consequences for brain development and, if timed at critical developmental windows, can skew neurodevelopment and facilitate the onset of neurodevelopmental disorders. To analyze this infectious etiology of neurodevelopmental disease, an animal model of MIA by a viral mimetic that evokes antiviral innate immunity was used. Offspring exposed to the simulated infection were behaviorally and

cognitively tested in adolescence, and then inflammatory outcomes were assessed in their prefrontal cortices.

2. Investigate the host immune response to infection.

The immune response to maternal infection was assessed in offspring exposed to MIA. Indeed, it is understood that the host (maternal) response to infection is the driving force of behavioral and inflammatory changes in humans and animals exposed to MIA. Consistent with this, persistent changes in inflammatory signaling were found in the prefrontal cortex in adolescent MIA-exposed offspring. These persistent responses to maternal infection are consistent with adolescents with autism spectrum disorders (ASD), schizophrenia, and attention deficit hyperactivity disorder (ADHD).

3. Examine the influence of environmental and ecological forces on infectious diseases.

A high-choline maternal diet was tested in the study as an external force that affects the offspring immune response to MIA. Choline is a critical micronutrient for neurodevelopment that is present in several foods like eggs, meats, and leafy greens. However, most people, especially pregnant people, do not receive the recommended levels of dietary choline. Despite this, choline is absent from many prenatal vitamins. Therefore, to establish evidence for choline as a neuroprotective factor, maternal choline was supplemented to a subset of pregnant rats exposed to MIA. It was found that offspring of choline-supplemented MIA mothers were less behaviorally impacted by MIA and had increased anti-inflammatory signaling in their prefrontal cortices. This adds to the current body of literature supporting choline's neuroprotective effects and suggests that exposure to a nutrient-rich maternal diet may be protective against the inflammatory and neurodevelopmental impacts of infectious diseases encountered in the womb.

4. Analyze disease risk factors and select appropriate surveillance.

This study analyzed risk factors for neurodevelopmental disorders, namely exposures to infectious diseases and a non-choline-supplemented maternal diet. It was identified that maternal choline supplementation is protective against the behavioral and inflammatory effects of MIA evoked by viral diseases. A need for further longitudinal human studies of choline's neuroprotective impact against maternal COVID-19 was articulated in the proposal for this study since large birth cohorts exposed to MIA by COVID-19 are now available. Animal studies like the one here are also important for the identification of biomarkers consistent with neurodevelopmental disorders. For example, deficiencies in protein kinase B (Akt) expression were identified in the prefrontal cortices of animals exposed to MIA. These deficiencies were reversed by maternal choline supplementation. This finding in the brain is consistent with decreases in Akt expression found peripherally in humans with schizophrenia. Therefore, it is possible that this may be a biomarker that can be used in future studies to predict the development of schizophrenia in people exposed to MIA.

5. Investigate the role of vectors, toxic plants, and other toxins in infectious diseases.

One advantage of MIA studies using the viral mimetic used in this study is that the evoked maternal immune response is generalizable to a variety of viral pathogens. Indeed, vector borne viruses like Zika virus have been identified as MIA inducers that can act through the maternal immune response and by crossing the placental barrier. Therefore, the MIA paradigm in this study is generalizable to other vector borne diseases as well since the resultant consequences of MIA are largely similar. Therefore, a need for differing surveillance paradigms based on the distribution of vector populations and environmental toxins was identified. In addition, the literature review for this study identified a need for more "two-hit" animal models that examine

the cumulative effects of MIA and environmental toxin exposure on the risk of developing neurodevelopmental disorders.

Table 3.1 Summary of MPH Emphasis Area Competencies

MPH Emphasis Area: Infectious Diseases and Zoonoses (IDZ)		
Number and Competency		**Description**
1	Evaluate modes of disease causation of infectious agents.	The infectious origins of neurodevelopmental disorders were analyzed in a rodent model, and differences in prefrontal cortex immune signaling were observed between offspring exposed to a viral mimetic in the womb and those who were not.
2	Investigate the host immune response to infection.	The core tenet of the maternal immune activation (MIA) hypothesis is that the maternal immune response to a pathogen, more so than the pathogen itself, is responsible for perturbed neurodevelopment. Therefore, the persistent inflammatory responses in the brains of offspring exposed to MIA were analyzed long after the inflammatory insult to investigate the persistent host response to infection *in utero*.
3	Examine the influence of environmental and ecological forces on infectious diseases.	The essential micronutrient choline was supplemented to a subset of pregnant rats exposed to MIA to analyze the effects of maternal dietary choline intake on neurodevelopmental resilience. Evidence was found that maternal choline blunts the inflammatory impacts of MIA exposure and may therefore be protective in immunity through its metabolism by innate immune cells.
4	Analyze disease risk factors and select appropriate surveillance.	The main purpose of this study was to analyze inflammatory risk factors for neurodevelopmental disorders and the protective effects of choline against inflammatory insult. Based on the evidence supporting the protective nature of maternal choline supplementation, it was recommended that future human studies with COVID-19-exposed birth cohorts examine the effect of COVID-19-evoked MIA and maternal choline on future neurodevelopmental disease risk.
5	Investigate the role of vectors, toxic plants, and other toxins in infectious diseases.	The MIA paradigm in this study is generalizable to those evoked by vector borne viral diseases. Future human studies should consider the distribution of vector populations and environmental toxins in MIA susceptibility. Animal studies can use “two-hit” models to analyze the contributions of environmental factors in conjunction with MIA, as well.

Appendix A - Portfolio Products

A.1 An Overview of SARS-CoV-2 Genomic Epidemiology (Original Article)

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is an enveloped RNA virus of the family *Coronaviridae* that causes COVID-19. It was first identified in Wuhan, China in 2019 after causing outbreaks of respiratory disease linked to a seafood wholesale market. SARS-CoV-2 infects a wide range of host cells via different mechanisms, but its primary mode of cell entry is through the angiotensin-converting enzyme 2 (ACE2) receptor with its spike (S) protein. SARS-CoV-2 has a large and complicated genome encoding 29 proteins, so it evolves very quickly due to mutations introduced by its error-prone RNA-dependent RNA polymerase, which is required for genome replication.[1] Because of this, genomic epidemiology has been at the forefront of defining new SARS-CoV-2 variants and informing the public health response to the rapidly changing virus.

SARS-CoV-2 Lineages and Variants

There are three main naming conventions for SARS-CoV-2: Pangolin, NextStrain, and WHO. All three are valuable and communicate information describing different characteristics of SARS-CoV-2 subgroups.

Pangolin: Pangolin, or Pango, lineage naming is the most commonly used classification system among scientists. Under this naming scheme, groups of related SARS-CoV-2 sequences are divided into **lineages**. For a lineage to be designated, there must be multiple individual, high-quality sequences showing significant changes in the virus' genome, either in the number of mutations or altered function of genes. A lineage cannot be comprised of the exact same sequence many times; it must show evidence of continued transmission and evolution, which manifests in (often neutral) mutations. When a new lineage is created, it is assigned a name with an alphabetical prefix and a numerical suffix. The alphabetical prefixes can be any letter except I, O, and X. X is reserved to denote **recombination** events, where two viruses exchange genetic material to create a new, third lineage. When X is used, it is placed *before* the prefix of the lineage it is describing. After the letter, a number is designated based on the order of lineages branching off of the parent. For example, if a new lineage is the second branch off of lineage A, it will be given the number 2. This will be added to the parent lineage's name to specify its relationship to other known lineages (e.g., the second lineage from lineage A would be A.2). As additional descendants are discovered and added, numbers are added to the chain, separated by periods, to show the ancestral lineages the new one descended from (e.g., the third lineage from A.2 above would be A.2.3). If a lineage would be the fourth number in the chain, the lineage is given the next letter prefix available, which will act as an alias for the long chain (e.g., A.2.3.4.1 may instead be F.1, where F represents "A.2.3.4"). When the alphabetical prefixes are exhausted, the prefixes start over, but with two letters (e.g., AA, AB, AC).[2] Although Pango lineages may be confusing, they are currently the most specific and robust way to describe subtypes of SARS-CoV-2 viruses.

Nextstrain: NextStrain naming is a bit more straightforward than Pangolin naming. In the Nextstrain scheme, virus subtypes are called **clades**, which represent large groups of related viruses. For a NextStrain clade to be designated, the viral subtype must either be designated as a variant of concern or interest by the World Health Organization, or it must reach high regional or global prevalence *and* contain at least one biologically meaningful mutation. NextStrain clade names consist of a two-digit prefix and an alphabetical suffix. The two digits represent the year that the clade was created (e.g., 23 corresponds to 2023), and the alphabetic suffixes are simply given in order of designation in that particular year.[3]

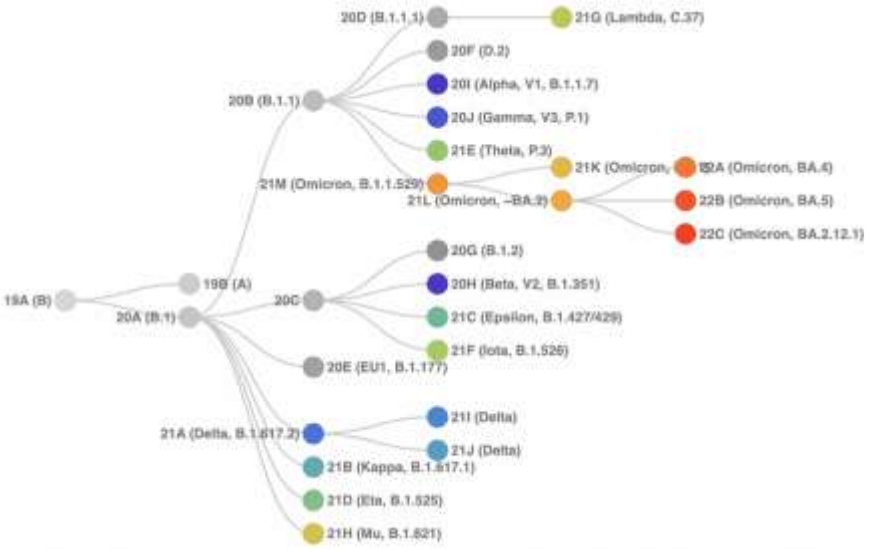


Fig. 1: A schematic showing NextStrain, WHO, and Pangolin names for major clades of SARS-CoV-2, respectively (Credit NextStrain, https://nextstrain.org/blog/2022-04-29-SARS-CoV-2-clade-naming-2022).

WHO: WHO, or World Health Organization, naming is the most commonly used in media or communication with the lay public. The WHO assigns major variants of SARS-CoV-2 Greek letters in alphabetical order, with the exceptions of nu and xi, which were skipped when the most recent variant, omicron, was named.[4] What constitutes a major variant is debated, but a new variant designation is generally reserved for lineages with major changes of great public health significance.

Classifying Lineages/Variants of Public Health Importance

In the United States, an interagency governmental panel classifies SARS-CoV-2 lineages and variants of public health importance into four groups. In increasing order of severity, important variants are given the following designations:[5]

- **Variant Being Monitored (VBM):** VBMs are variants that have genetic changes that suggest that they may become a public health threat. VBMs can also be previous variants from the more severe categories that are no longer prevalent.

- **Variant of Interest (VOI):** A VOI is a variant that is resistant to immunity generated from vaccines or prior infection. Variants that may cause more severe disease or those that evade detection may also be classified as VOIs.

- **Variant of Concern (VOC):** VOCs may be variants that escape detection and evade immunity and also cause more severe disease. VOCs are usually highly transmittable.

- **Variant of High Consequence (VOHC):** The VOHC distinction is the most severe. VOHCs are those variants that encompass the VOC designation, but are also resistant to medical countermeasures like vaccination, therapeutics, and diagnostic technologies.

References

1. Wu Can-rong, Yin W chao, Jiang Y, Xu HE. Structure genomics of SARS-CoV-2 and its Omicron variant: drug design templates for COVID-19. *Acta Pharmacol Sin*. 2022;43(12):3021-3033. doi:10.1038/s41401-021-00851-w

2. Statement of Nomenclature Rules – Pango Network. Accessed November 21, 2023. https://www.pango.network/the-pango-nomenclature-system/statement-of-nomenclature-rules/

3. SARS-CoV-2 clade naming strategy for 2022. Accessed November 21, 2023. https://nextstrain.org//blog/2022-04-29-SARS-CoV-2-clade-naming-2022

4. WHO announces simple, easy-to-say labels for SARS-CoV-2 Variants of Interest and Concern. Accessed November 21, 2023. https://www.who.int/news/item/31-05-2021-who-announces-simple-easy-to-say-labels-for-sars-cov-2-variants-of-interest-and-concern

5. CDC. Coronavirus Disease 2019 (COVID-19). Centers for Disease Control and Prevention. Published February 11, 2020. Accessed November 21, 2023. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-classifications.html

A.2 An Overview of Influenza Virus Genomic Epidemiology (Original Article)

Introduction

Influenza viruses (family *Orthomyxoviridae*) are lipid-enveloped, single-stranded RNA viruses that infect a wide range of hosts. Their RNA is negative-sense, meaning that they rely on an error-prone RNA-dependent RNA polymerase to generate readable mRNA that can be translated into viral proteins by the host cell. Influenza viruses primarily infect the respiratory epithelium of the upper respiratory tract in humans. The virus binds to cells using its hemagglutinin (H) protein, which allows it to enter, and cellular escape is mediated by the neuraminidase (N) protein. These antigens are strongly targeted by vaccines and host immune responses, so they are constantly under evolutionary pressure to change. Therefore, genomic surveillance is critical to identify the molecular underpinnings of this evolution so that outbreaks can be controlled, vaccines can be developed, and future pandemics can be predicted and avoided.

Naming Influenza Viruses

Influenza viruses are named by their antigenic type (A, B, C, or D), where they were isolated, species, sample identifier, isolation year, and virus subtype.[1,2] This nomenclature allows genomic epidemiologists to pack in as much spatial and temporal information as is reasonably possible so that changes in influenza genomes can be tracked across space and time. Note that no species identifier is included in human infections.

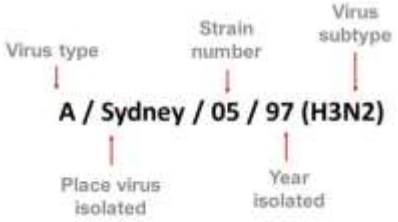


Fig. 1: An example from the CDC showing the naming of influenza viruses. Since this virus was isolated from a human host, there is no species identifier for this sample. The "strain number" is another name for the sample identifier; it is unique to each sample. This allows epidemiologists to link the sequence of that virus' genome to the case. This virus is an influenza A virus (H3N2 subtype) isolated from a human in Sydney, Australia in 1997 (Credit CDC, https://www.cdc.gov/flu/about/viruses/types.htm).

Types of Influenza Viruses

Influenza A: Influenza A viruses infect humans and many non-human animals and can cause seasonal influenza epidemics and, occasionally, pandemics.[3] Influenza A viruses are of highest concern to epidemiologists due to their ability to circulate in animal species undetected.[4] These viruses are known for their ability to "jump" from animals to humans and vice versa, meaning that surveillance of animal strains is important to predict the next strains that will affect humans.[5–7]

Influenza A viruses are difficult to group based on epidemiological characteristics, like the species they infect or where they circulate. Therefore, subtypes are named based on the properties of their surface proteins, hemagglutinin (H) and neuraminidase (N). To date, 18 variants of hemagglutinin and 11 variants of neuraminidase have been described.[4] The most commonly circulating influenza A viruses in humans are subtypes H3N2 and H1N1. Both of these subtypes have caused human pandemics, in 1968-69 and 2009-10, respectively.

Influenza B: Influenza B, unlike influenza A, infects humans almost exclusively.[2] Like influenza A, though, it circulates seasonally and contributes to seasonal influenza outbreaks. Influenza B evolves more slowly than influenza A and lacks a non-human reservoir, making surveillance generally easier. Rather than subtypes, influenza B is divided into two major *lineages*: B/Victoria and B/Yamagata. In naming, this lineage name takes the place of subtype. In recent years, B/Victoria has dominated circulation, while B/Yamagata viruses have been relatively rarer.[1]

Influenza C: Influenza C generally causes mild infection and is not frequently detected. Therefore, it is not commonly considered to be of public health significance. Recent studies, however, have identified high seroprevalence of influenza C viruses in children and young adults, suggesting that it may circulate widely in children.[8] However, this high seroprevalence is not associated with adverse clinical outcomes.

Influenza D: Influenza D viruses primarily infect cattle and are not believed to cause illness in humans. Studies have identified anti-influenza D antibodies in humans, particularly in agricultural workers, suggesting that the virus can spill over into humans.[9] However, despite these exposures, the virus is likely not virulent enough to cause disease in humans.

Clades and Subclades

Within the subtypes of the major influenza virus classes, viruses can be further categorized based on their genetic characteristics. While the antigens (H and N proteins) between viruses of the same subtype are largely the same, small differences in gene sequences allow genomic epidemiologists to study the evolution of viruses through time. These minor changes may rarely allow the virus to infect the host more optimally, although most of them are deleterious. The beneficial changes will be selected for by natural selection, meaning that they will replicate faster and predominate over other variants. Sometimes, these mutations confer such an advantage that that particular variant will drive others to extinction.

To describe genomic within subgroups/lineages, genomic epidemiologists create phylogenetic trees based on sequence data. Every sample is given one branch on the tree. Computer algorithms then sort the branches based on differences in sequence and link them together based on which ones are likely to have descended from a hypothetical common ancestor. This process continues until all sequences have been linked to a common ancestor through one or many "generations." A group of sequences linked through a hypothetical common ancestor is called a **clade**, and a smaller group of linked sequences within a clade is, predictably, termed a **subclade**.[1] When combined with spatial and temporal information, the transmission dynamics of specific clades and subclades can be pieced together, making phylogenetic analysis a powerful tool for

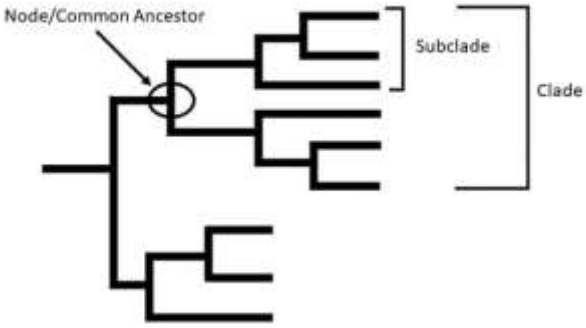


Fig. 2: A simple phylogenetic tree depicting clades and subclades. Note that all sequences in a clade or subclade diverge from a node representing a *hypothetical* common ancestor. Note that these common ancestors are not directly observed, but rather inferred based on the relationships between the observed sequences (Credit CDC, https://www.cdc.gov/flu/about/viruses/types.htm).

References

1. CDC. Types of Influenza Viruses. Centers for Disease Control and Prevention. Published March 30, 2023. Accessed November 20, 2023. https://www.cdc.gov/flu/about/viruses/types.htm

2. Influenza (Seasonal). Accessed November 20, 2023. https://www.who.int/news-room/fact-sheets/detail/influenza-(seasonal)

3. Lam TTY, Zhu H, Guan Y, Holmes EC. Genomic Analysis of the Emergence, Evolution, and Spread of Human Respiratory RNA Viruses. *Annu Rev Genomics Hum Genet*. 2016;17(1):193-218. doi:10.1146/annurev-genom-083115-022628

4. Peteranderl C, Herold S, Schmoldt C. Human Influenza Virus Infections. *Semin Respir Crit Care Med*. 2016;37(04):487-500. doi:10.1055/s-0036-1584801

5. Lam TT, Pybus OG. Genomic surveillance of avian-origin influenza A viruses causing human disease. *Genome Med*. 2018;10(1):50. doi:10.1186/s13073-018-0560-3

6. Nelson MI, Stratton J, Killian ML, Janas-Martindale A, Vincent AL. Continual Reintroduction of Human Pandemic H1N1 Influenza A Viruses into Swine in the United States, 2009 to 2014. *J Virol*. 2015;89(12):6218-6226. doi:10.1128/JVI.00459-15

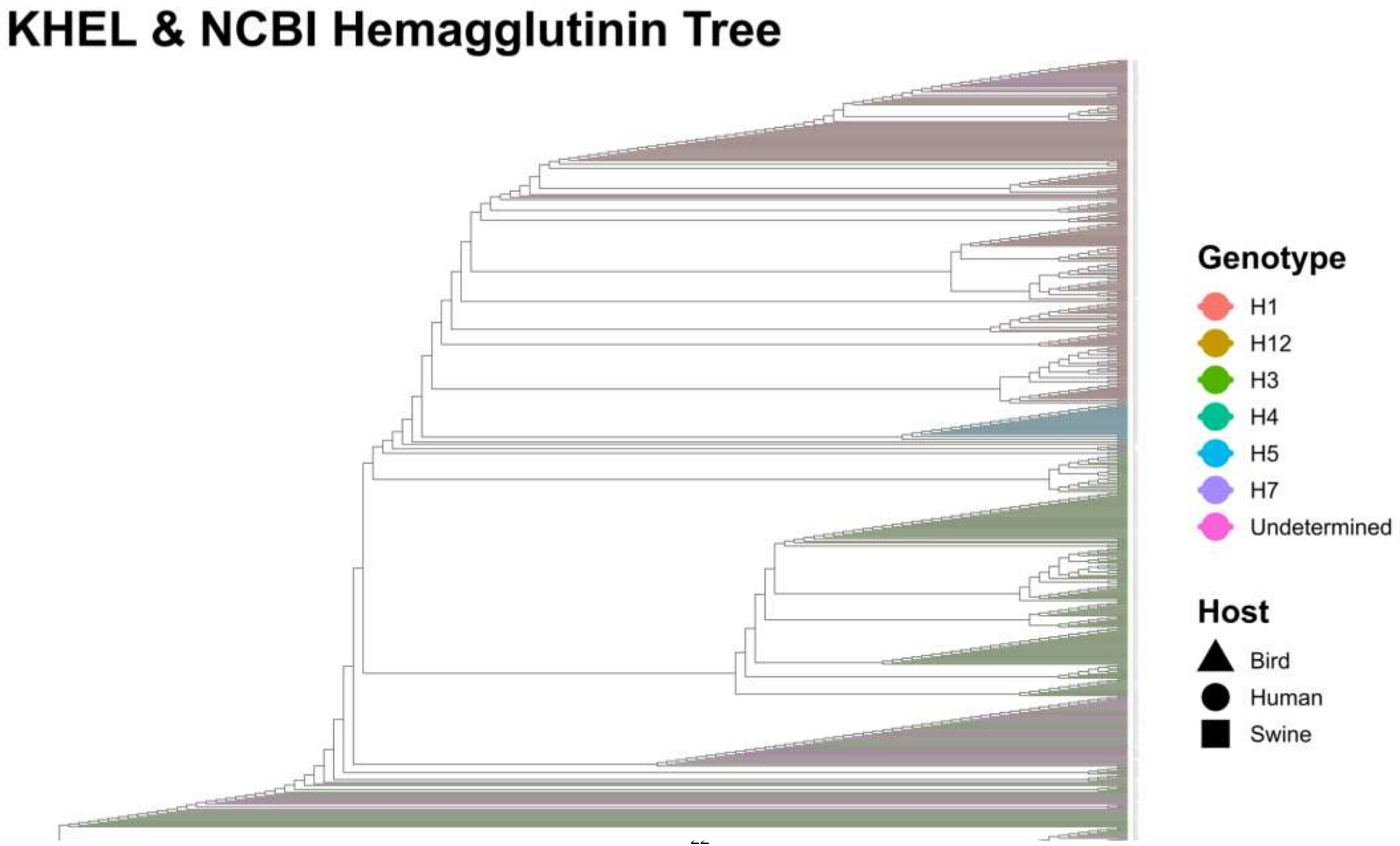
7. Rambo-Martin BL, Keller MW, Wilson MM, et al. Influenza A Virus Field Surveillance at a Swine-Human Interface. Lowen AC, ed. *mSphere*. 2020;5(1):e00822-19. doi:10.1128/mSphere.00822-19

8. Sederdahl BK, Williams JV. Epidemiology and Clinical Characteristics of Influenza C Virus. *Viruses*. 2020;12(1):89. doi:10.3390/v12010089

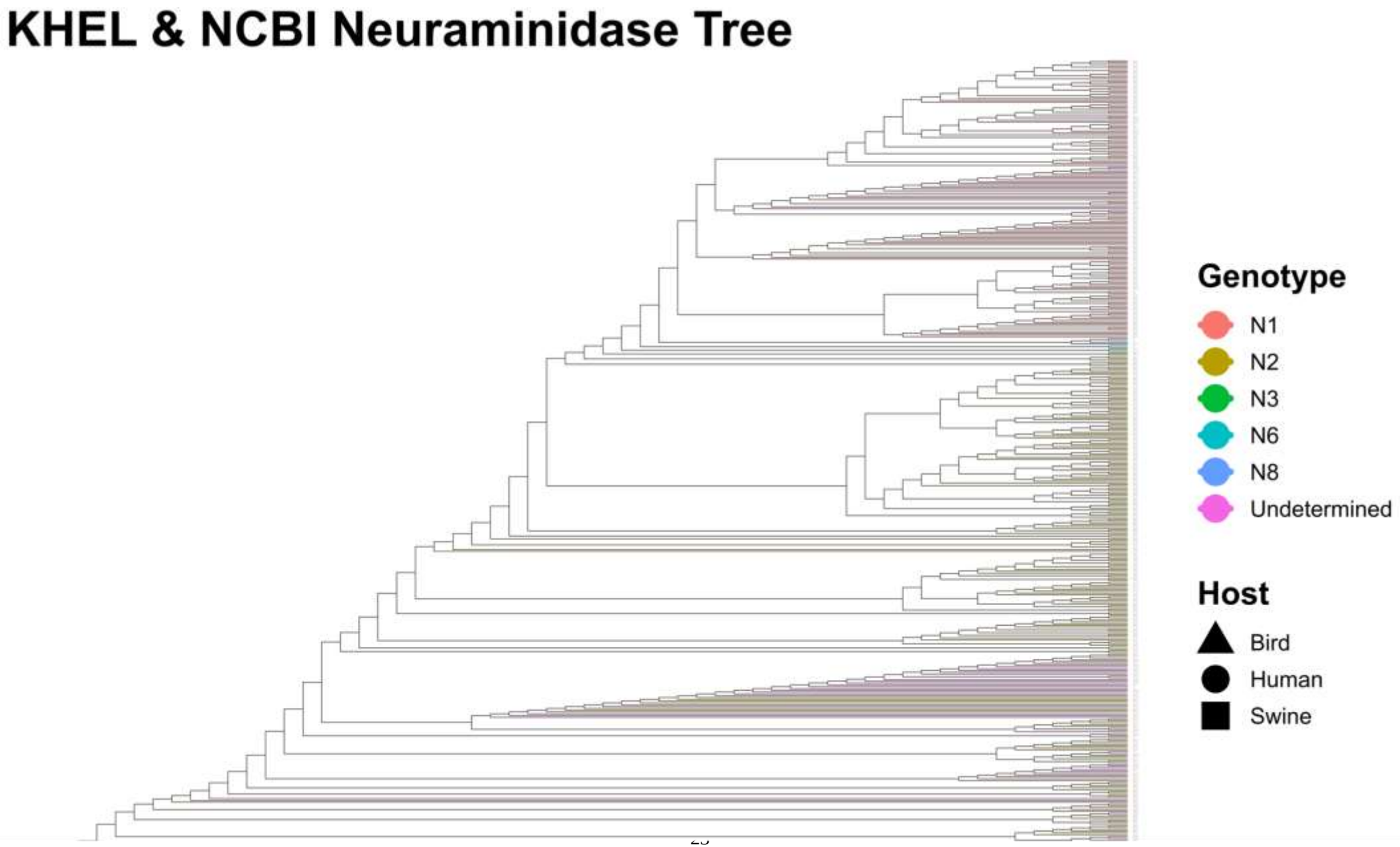
9. Liu R, Sheng Z, Huang C, Wang D, Li F. Influenza D Virus. *Curr Opin Virol*. 2020;44:154-161. doi:10.1016/j.coviro.2020.08.004

10. Dudas AB and G. *An Applied Genomic Epidemiological Handbook.* Accessed November 20, 2023. https://alliblk.github.io/genepi-book/

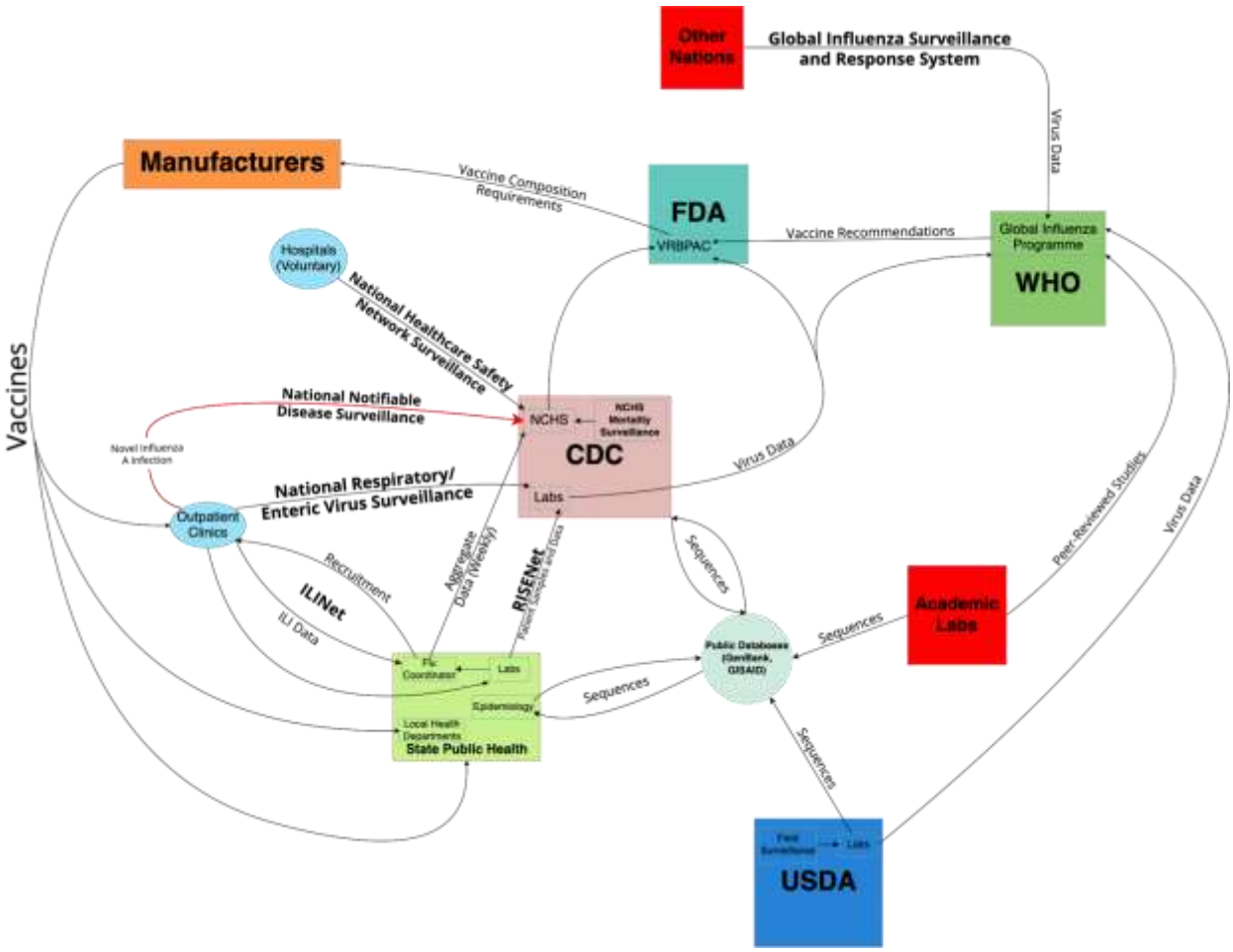
A.3 Hemagglutinin (HA) Phylogenetic Tree



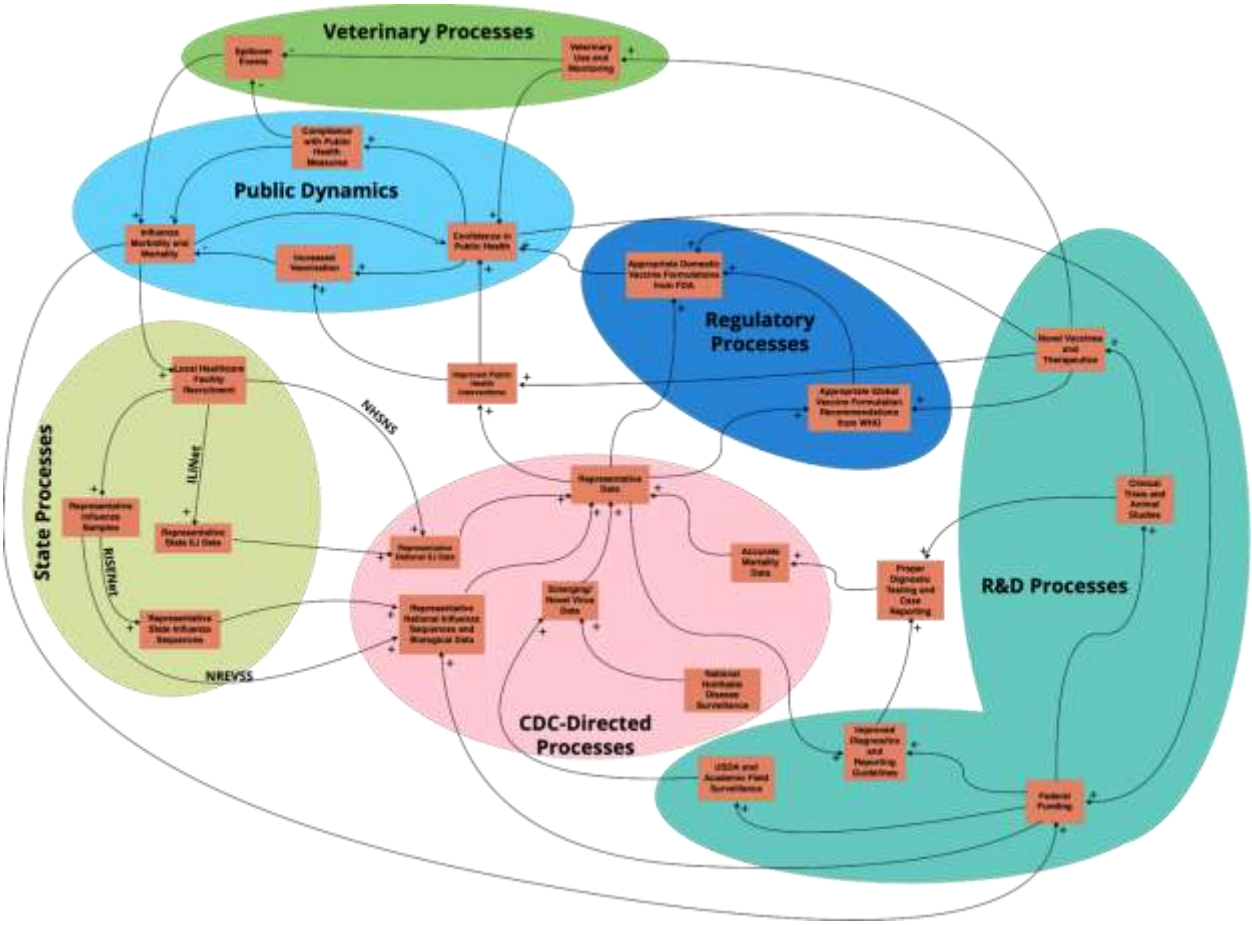
A.4 Neuraminidase (NA) Phylogenetic Tree



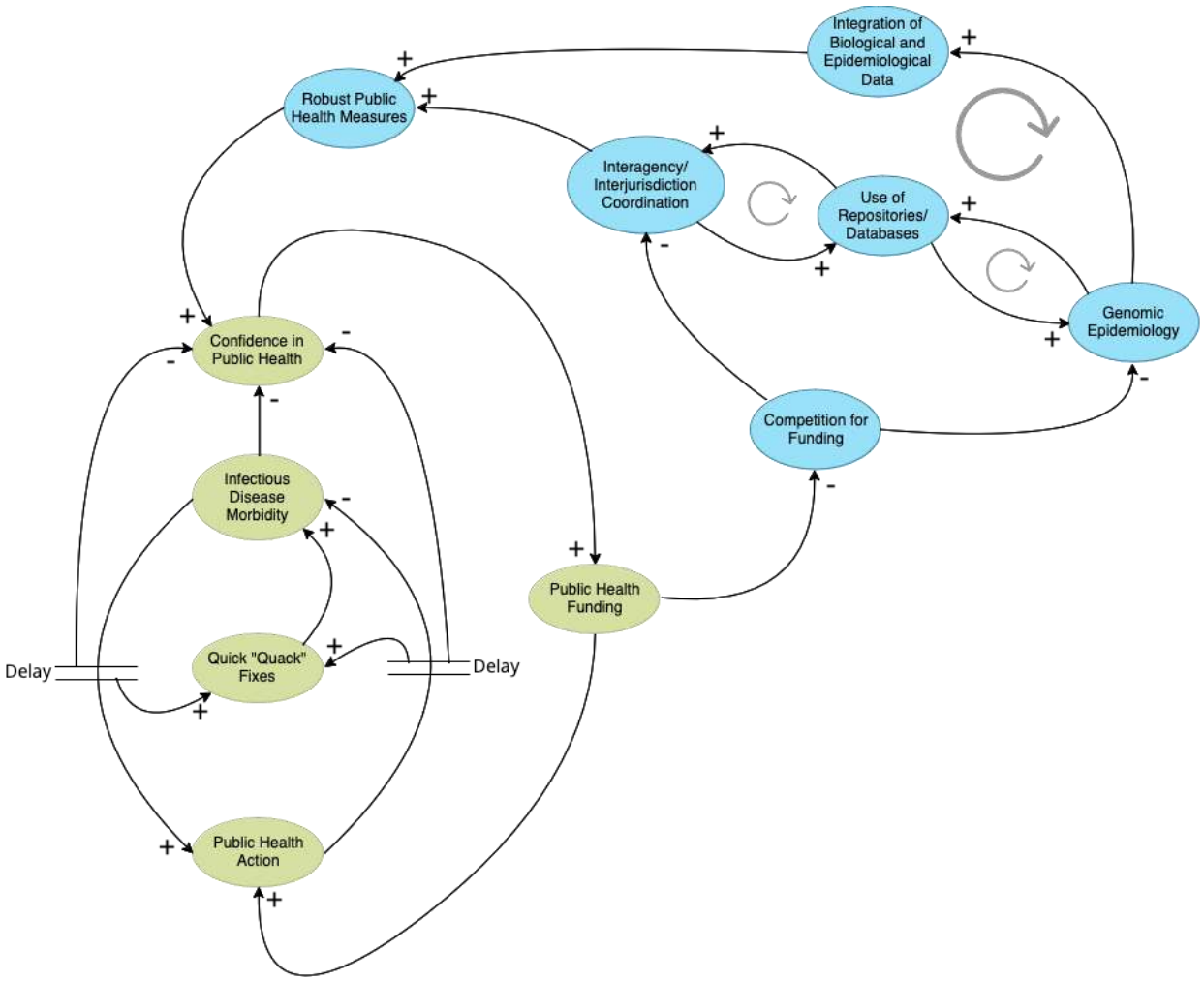
A.5 Process Diagram of U.S. Influenza Surveillance



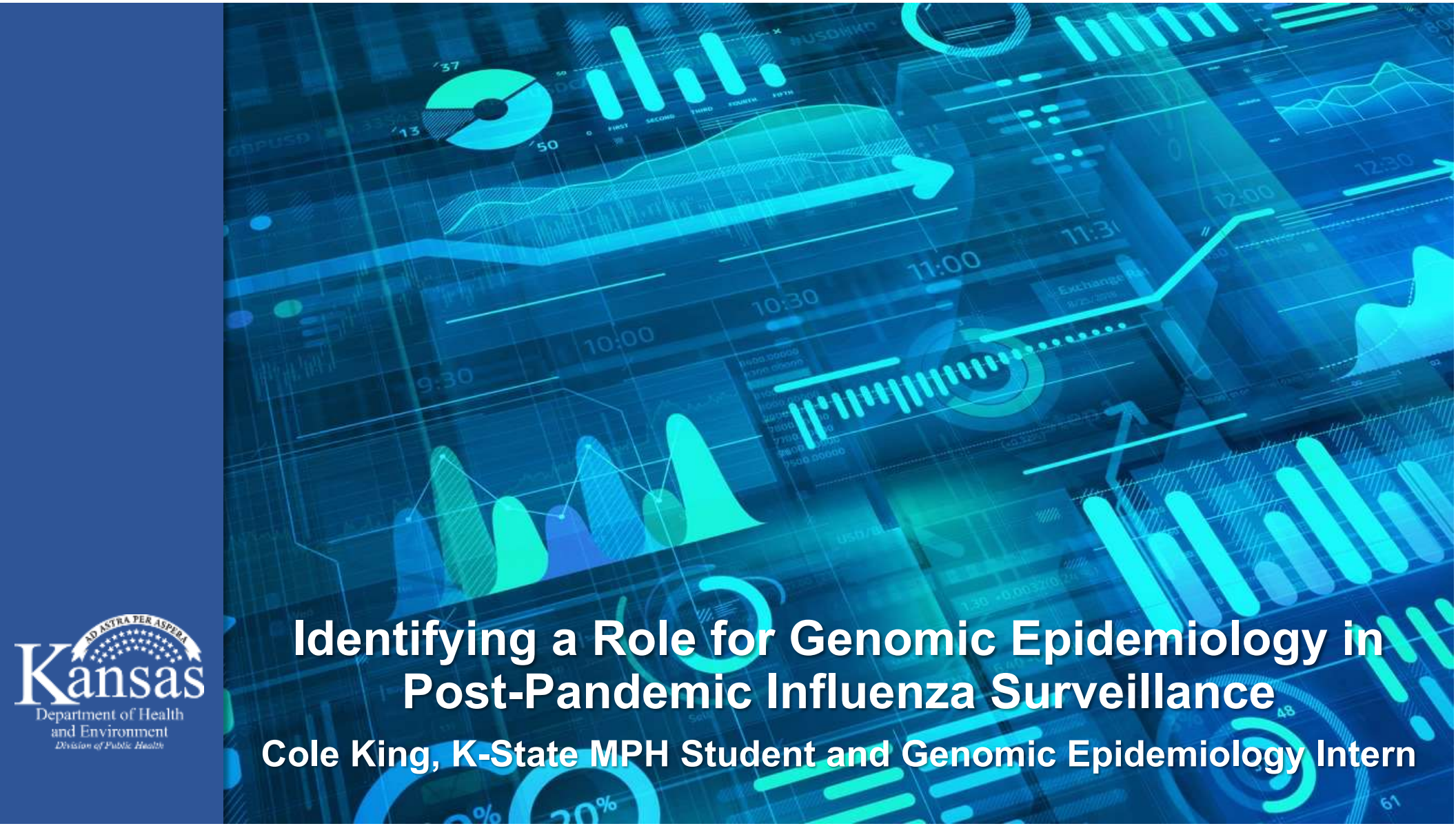
A.6 Causal Loop Diagram of U.S. Influenza Surveillance



A.7 System Archetypes Underlying Success/Failure of Influenza Prevention



A.8 Identifying a Role for Genomic Epidemiology in Post-Pandemic Influenza Surveillance (Slides)



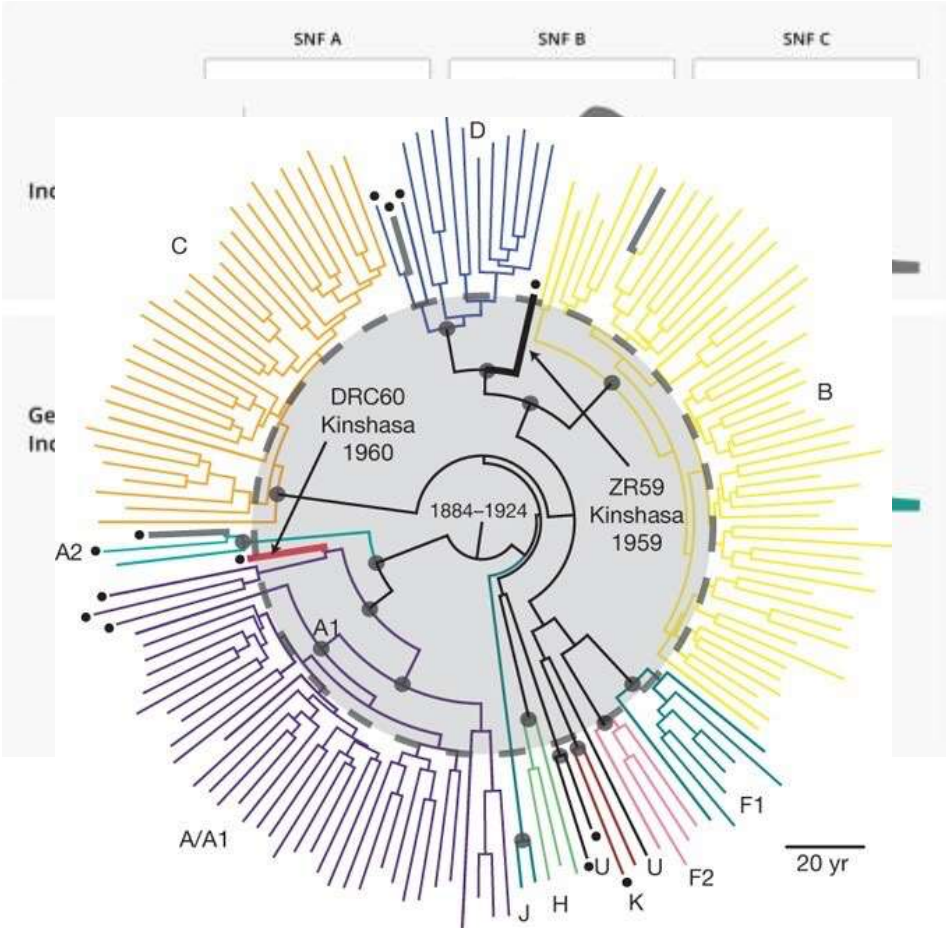
What is Genomic Epidemiology?

- The application of ***genomic analyses*** to understand and reduce chronic, infectious, environmental, and occupational diseases

- With quickly evolving and emerging pathogens, this information is *extremely* useful

What Can Genomic Epidemiology Do?

- Help ***distinguish*** related and unrelated cases
- ***Identify evolutionary changes*** contributing to new outbreaks
- Help ***reconstruct the history*** of an outbreak



Images from "An applied genomic epidemiological handbook" by Dudas, A. B. and G. Dudas, 2023, January 9 from https://alliblk.github.io/genepi-book/

Internship Focus

- Genomic epidemiology of SARS-CoV-2 and influenza
  - Apply lessons from SARS-CoV-2 to influenza
- Analyze the influenza surveillance system to identify a role for genomic epidemiology
- Analyze KS influenza sequences as a pilot case

An Overview of SARS-CoV-2 Genomic Epidemiology

Introduction

Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) is an enveloped RNA virus of the family *Coronaviridae* that causes COVID-19. It was first identified in Wuhan, China in 2019 after causing outbreaks of respiratory disease linked to a seafood wholesale market. SARS-CoV-2 infects a wide range of host cells via different mechanisms, but its primary mode of cell entry is through the angiotensin-converting enzyme 2 (ACE2) receptor with its spike (S) protein. SARS-CoV-2 has a large and complicated genome encoding 29 proteins, so it evolves very quickly due to mutations introduced by its error-prone RNA-dependent RNA polymerase, which is required for genome replication.[1] Because of this, genomic epidemiology has been at the forefront of defining new SARS-CoV-2 variants and informing the public health response to the rapidly changing virus.

An Overview Influenza Virus Genomic Epidemiology

Introduction

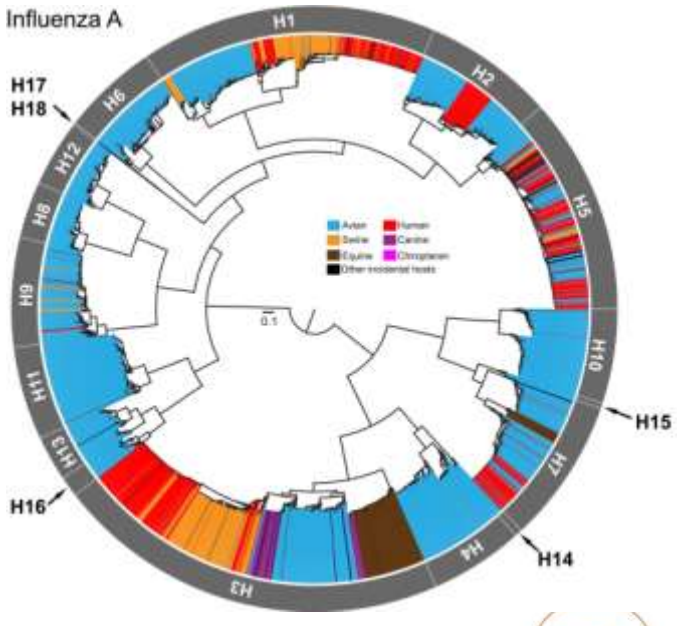
Influenza viruses (family *Orthomyxoviridae*) are lipid-enveloped, single-stranded RNA viruses that infect a wide range of hosts. Their RNA is negative-sense, meaning that they rely on an error-prone RNA-dependent RNA polymerase to generate readable mRNA that can be translated into viral proteins by the host cell. Influenza viruses primarily infect the respiratory epithelium of the upper respiratory tract in humans. The virus binds to cells using its hemagglutinin (H) protein, which allows it to enter, and cellular escape is mediated by the neuraminidase (N) protein. These antigens are strongly targeted by vaccines and host immune responses, so they are constantly under evolutionary pressure to change. Therefore, genomic surveillance is critical to identify the molecular underpinnings of this evolution so that outbreaks can be controlled, vaccines can be developed, and future pandemics can be predicted and avoided.

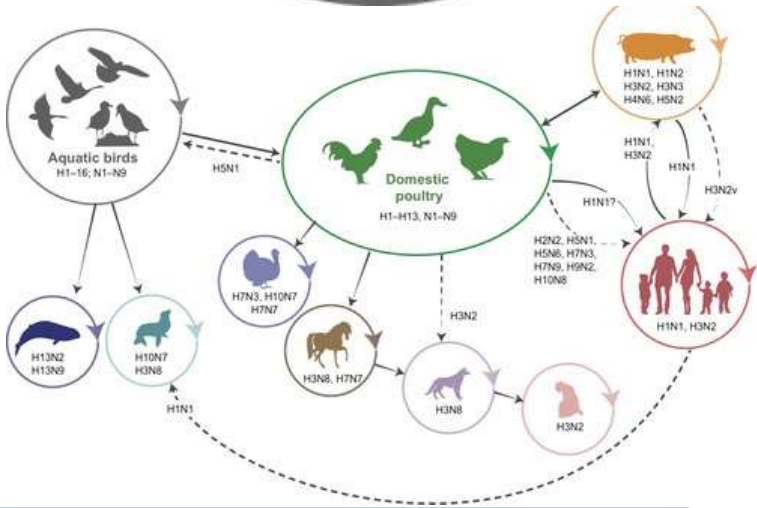
Why Influenza?

- Influenza evolves *extremely* rapidly
  - Unstable RNA genome prone to mutation
    - Natural selection and genetic shift/drift act upon these changes

- Influenza viruses are a constantly moving target
  - This makes it hard to:
    - Predict dominant strains and develop vaccines against them
    - Budget resources
    - Direct public health responses

- Influenza is also prevalent in many hosts
  - This adds *even more* opportunities to replicate, mutate, and recombine



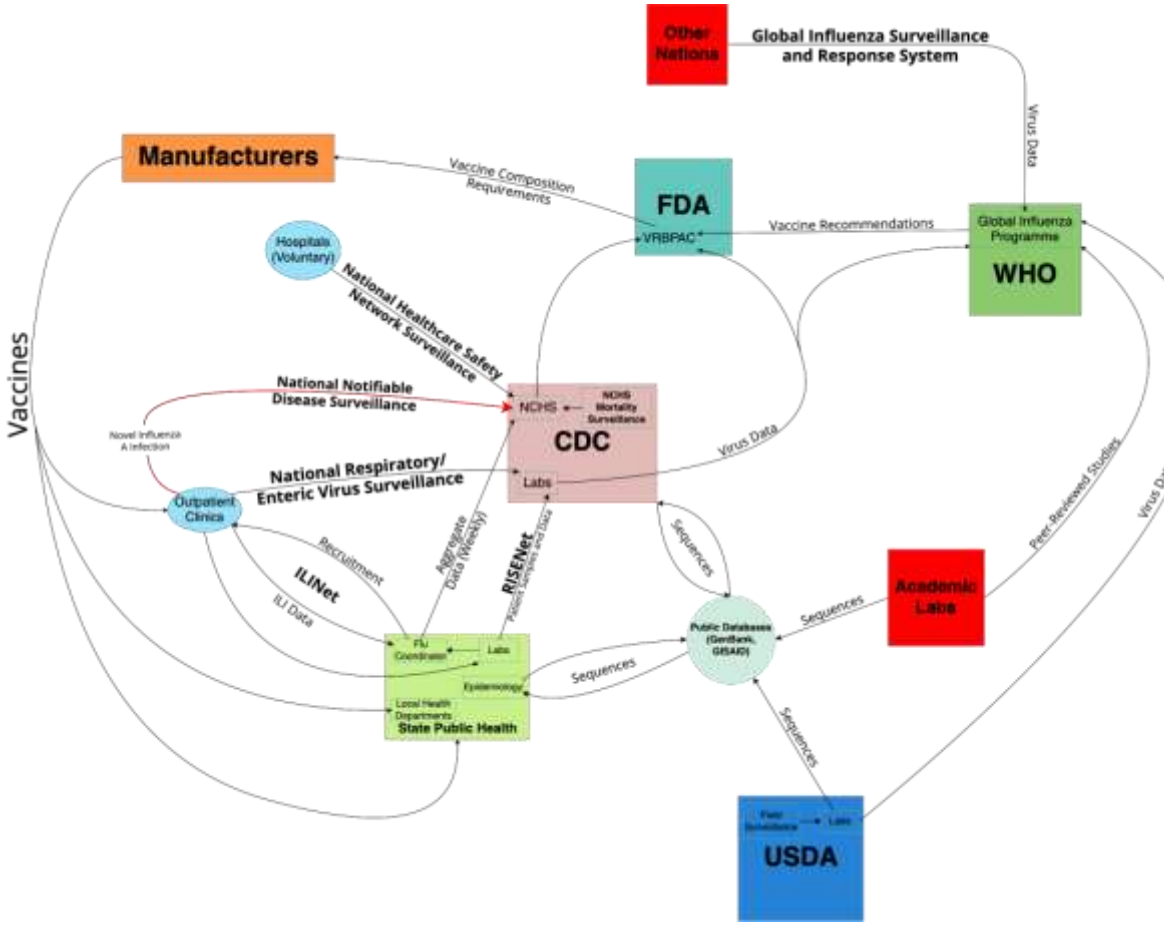


Images from "The ecology and adaptive evolution of influenza A interspecies transmission" by Joseph, U. et al, 2016, July 18, Copyright 2016 *Influenza and Other Respiratory Viruses*, from https://onlinelibrary.wiley.com/doi/10.1111/irv.12412

How Does U.S. Flu Surveillance Work?

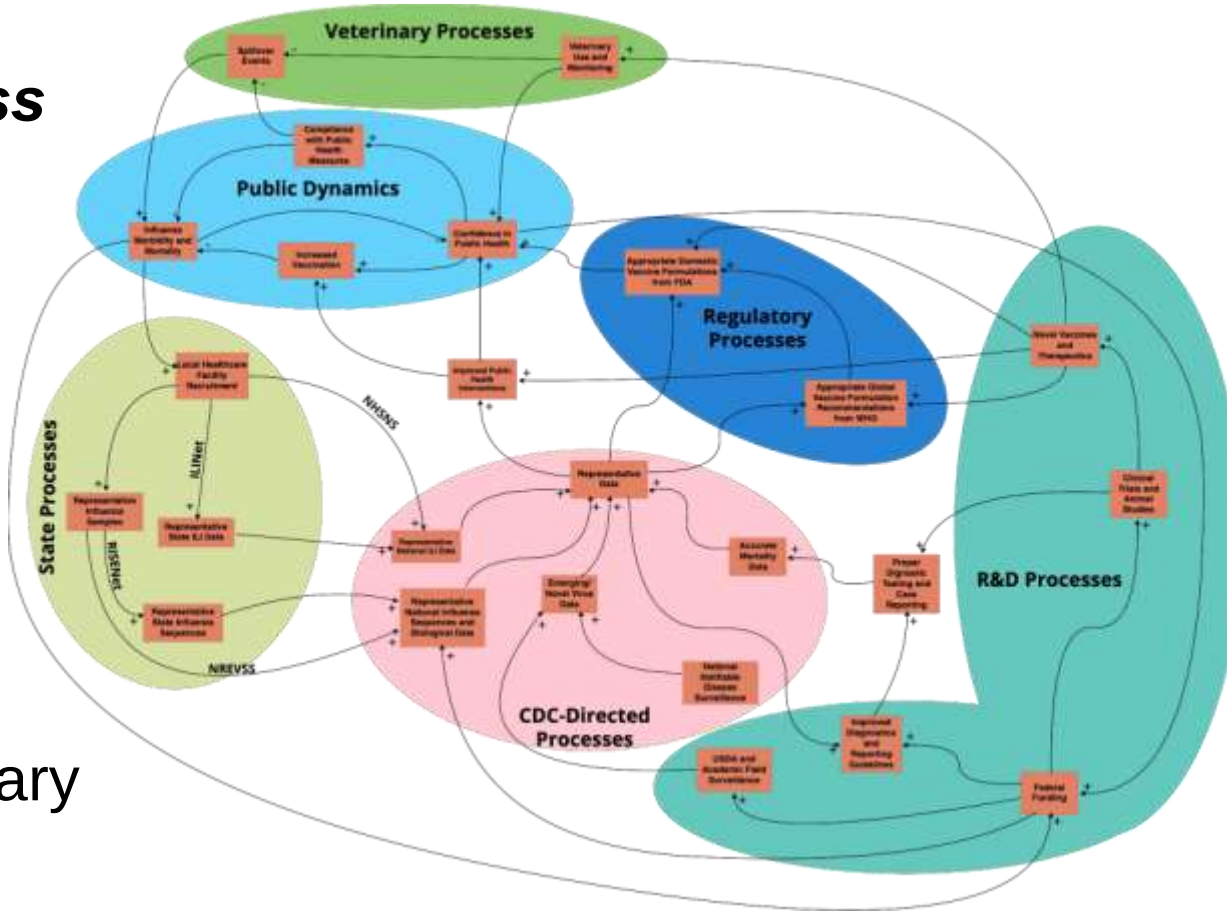
- Long story short, it's complicated…

- Highlights:
  - Lots of data ***bypasses state public health***
  - Human and veterinary surveillance ***not coordinated***
  - Vaccine decisions made on ***small data sets*** from population centers
    - Plus limited USDA surveillance



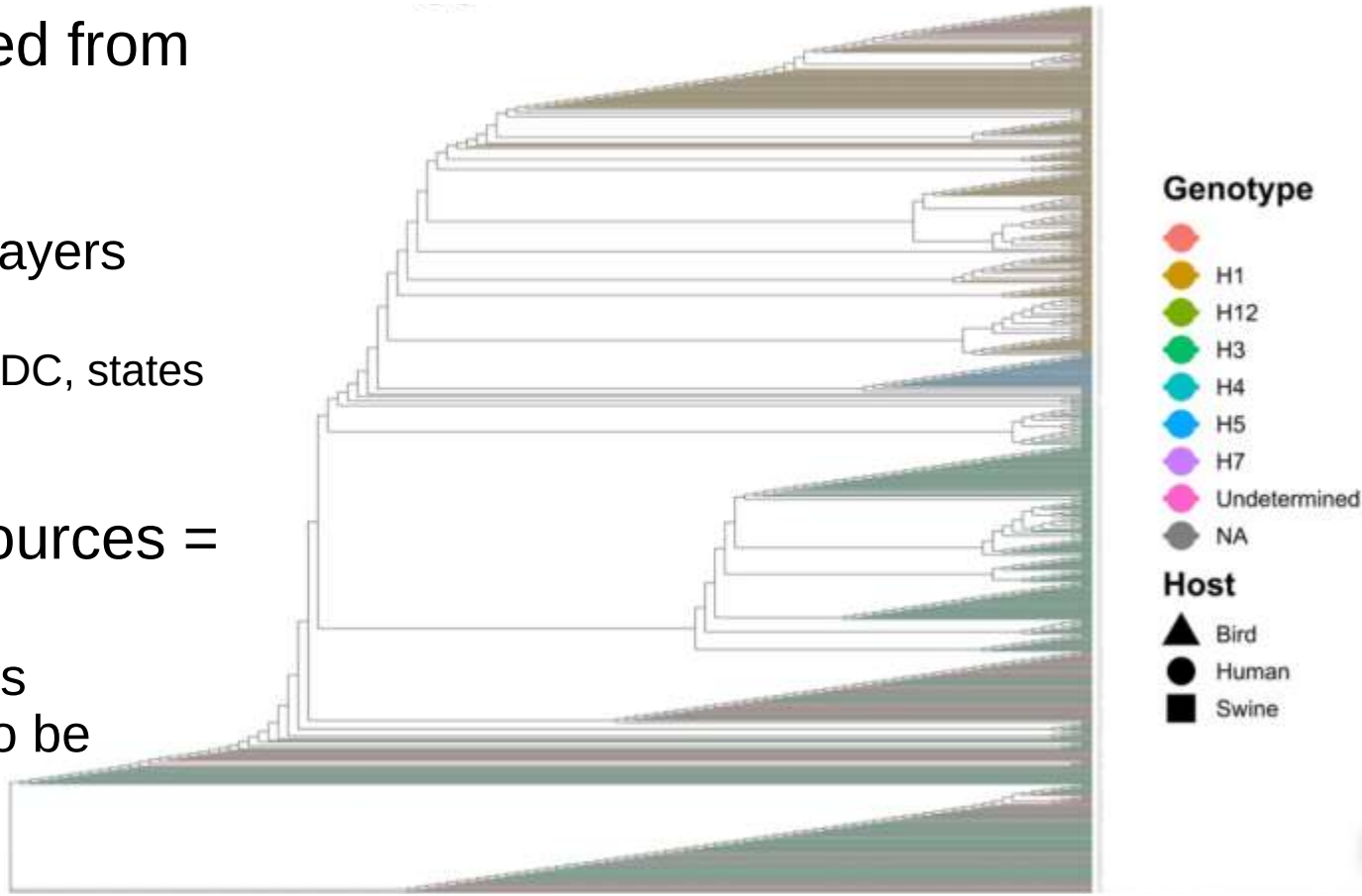
Weaknesses in U.S. Flu Surveillance

- Several data streams ***bypass state public health***
- Voluntary programs may produce ***unrepresentative data***
  - Underserved areas too burdened to participate
- ***Little communication*** between human and veterinary health



How Can Genomic Epidemiology Help?

- Sequences can be pulled from public databases
  - GISAID, GenBank, etc.
  - Several disconnected players upload to these
    - USDA, academic labs, CDC, states
- More data from more sources = more representative
  - More host species allows evolutionary dynamics to be observed



Genomic Epidemiology Can ***Bridge the Gap*** between Human and Veterinary Data

- Host adaptations manifest in *genetic changes*
  - Genetic shift and drift
  - Codon use bias
- Genetic changes underlie pathogen changes
  - Immune evasion
  - Virulence
  - Transmissibility
- Many of these changes may happen in animals *before* "spilling over" into humans

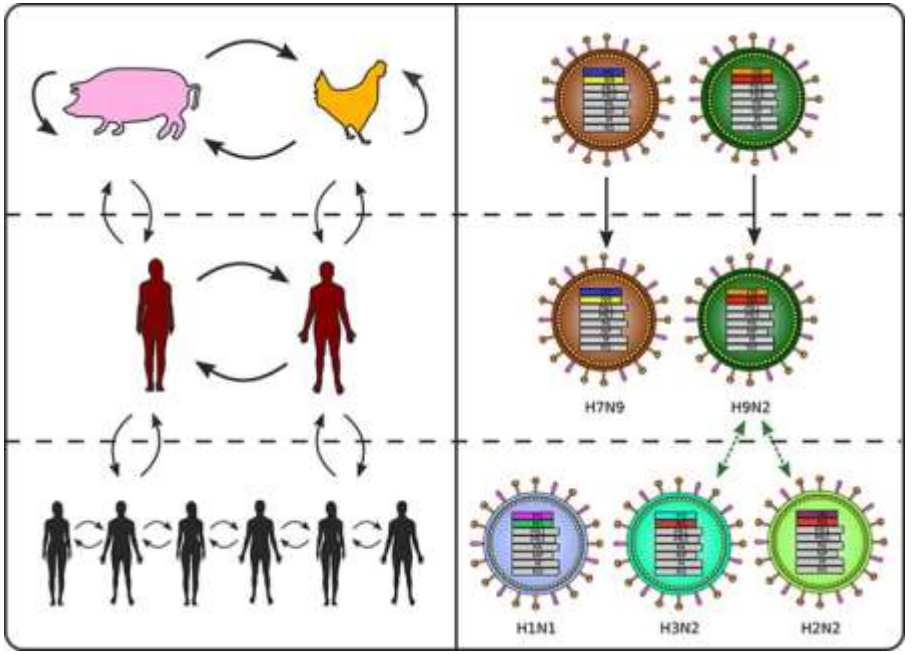
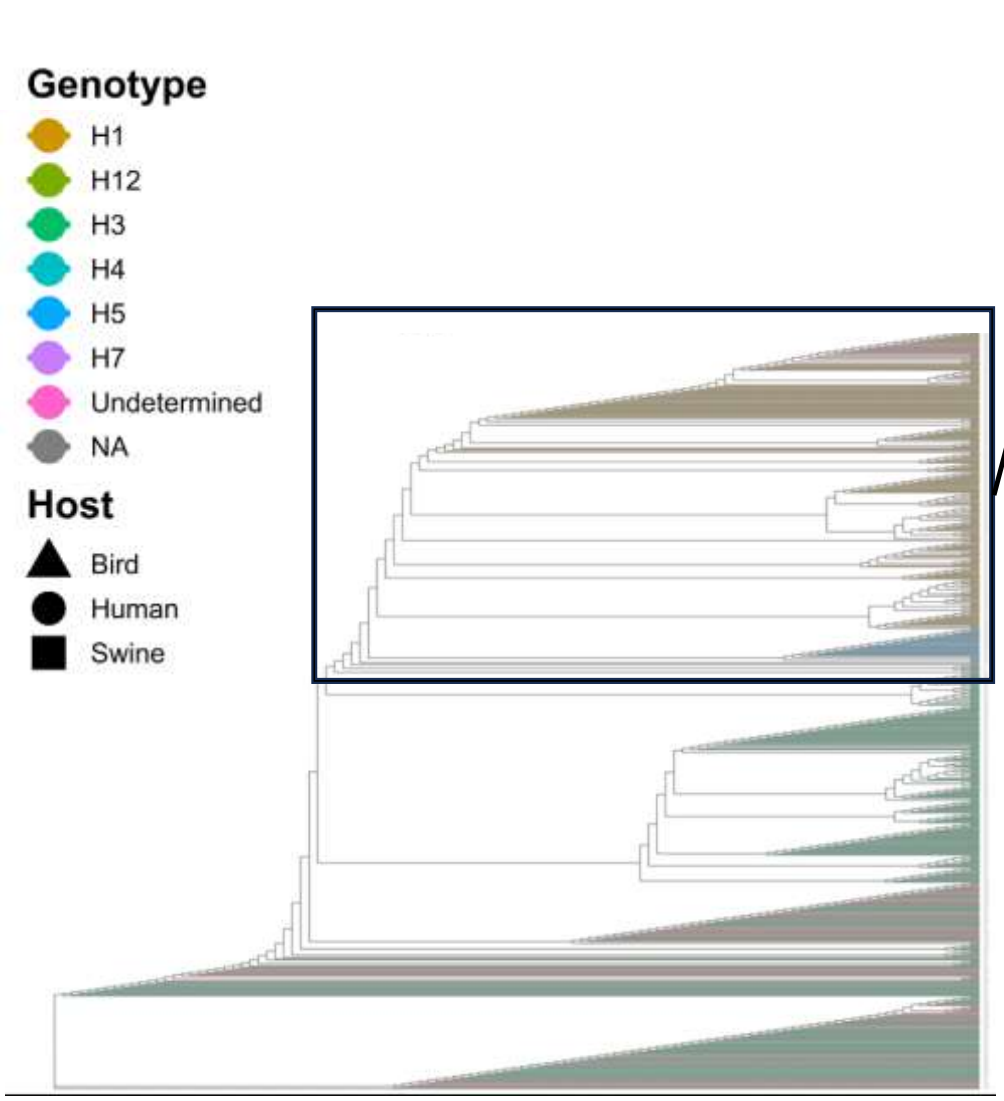
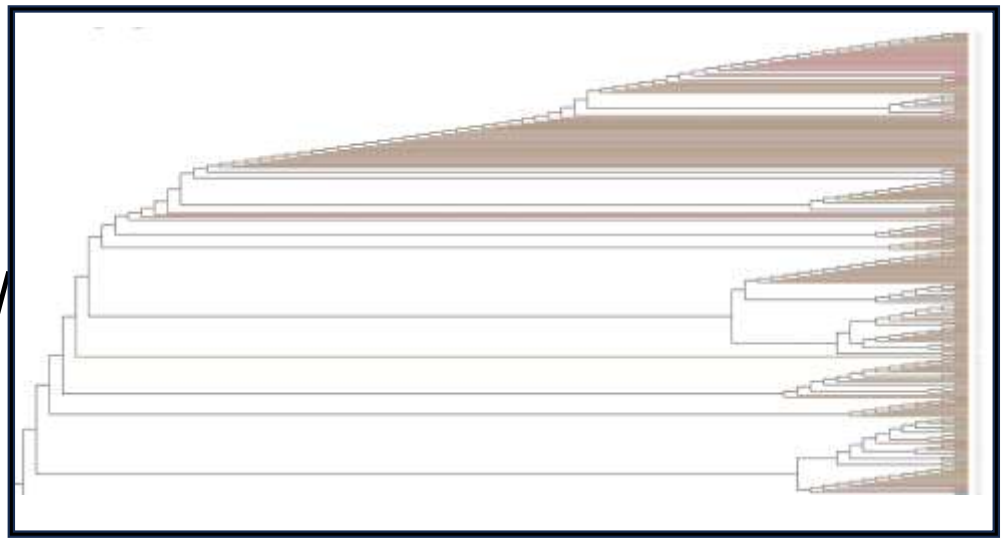
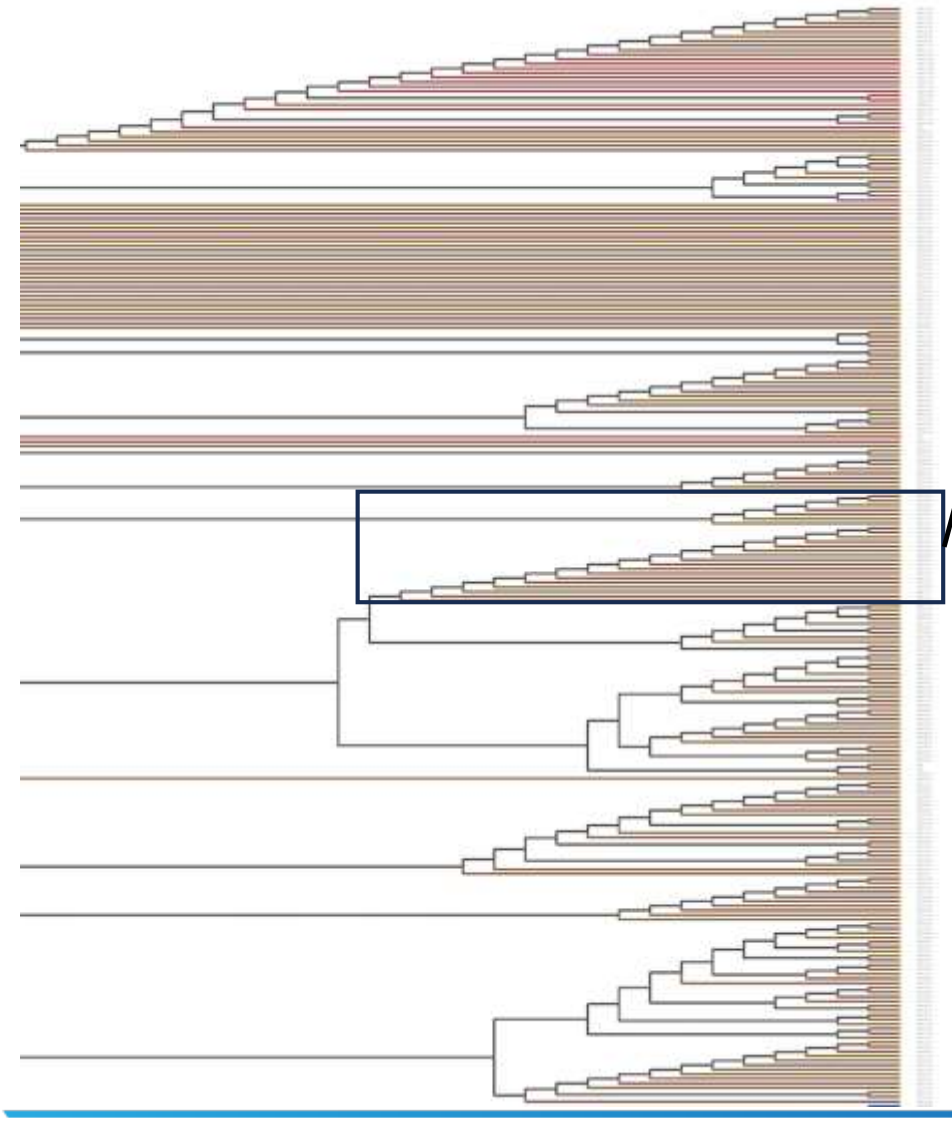


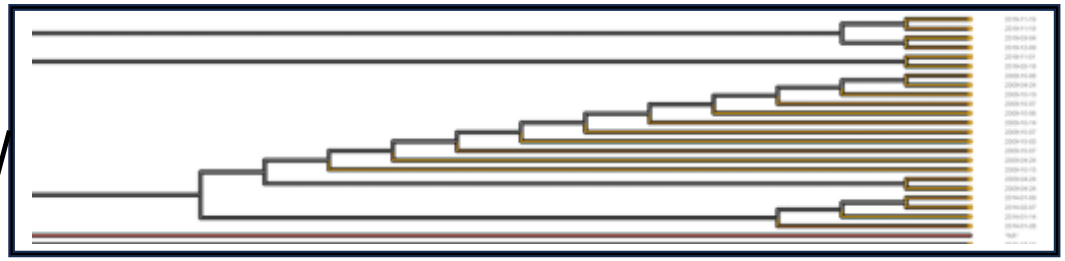
Image from "Science Forum: Viral factors in influenza pandemic risk assessment" by Lipsitch, M. et al, 2016, November 11, Copyright 2016 *eLife Life Sciences Magazine*, from https://elifesciences.org/articles/18491





- Let's look at some data from those public databases for the Hemagglutinin gene
  - Samples from humans, swine, and birds from several sources over 30 years in Kansas





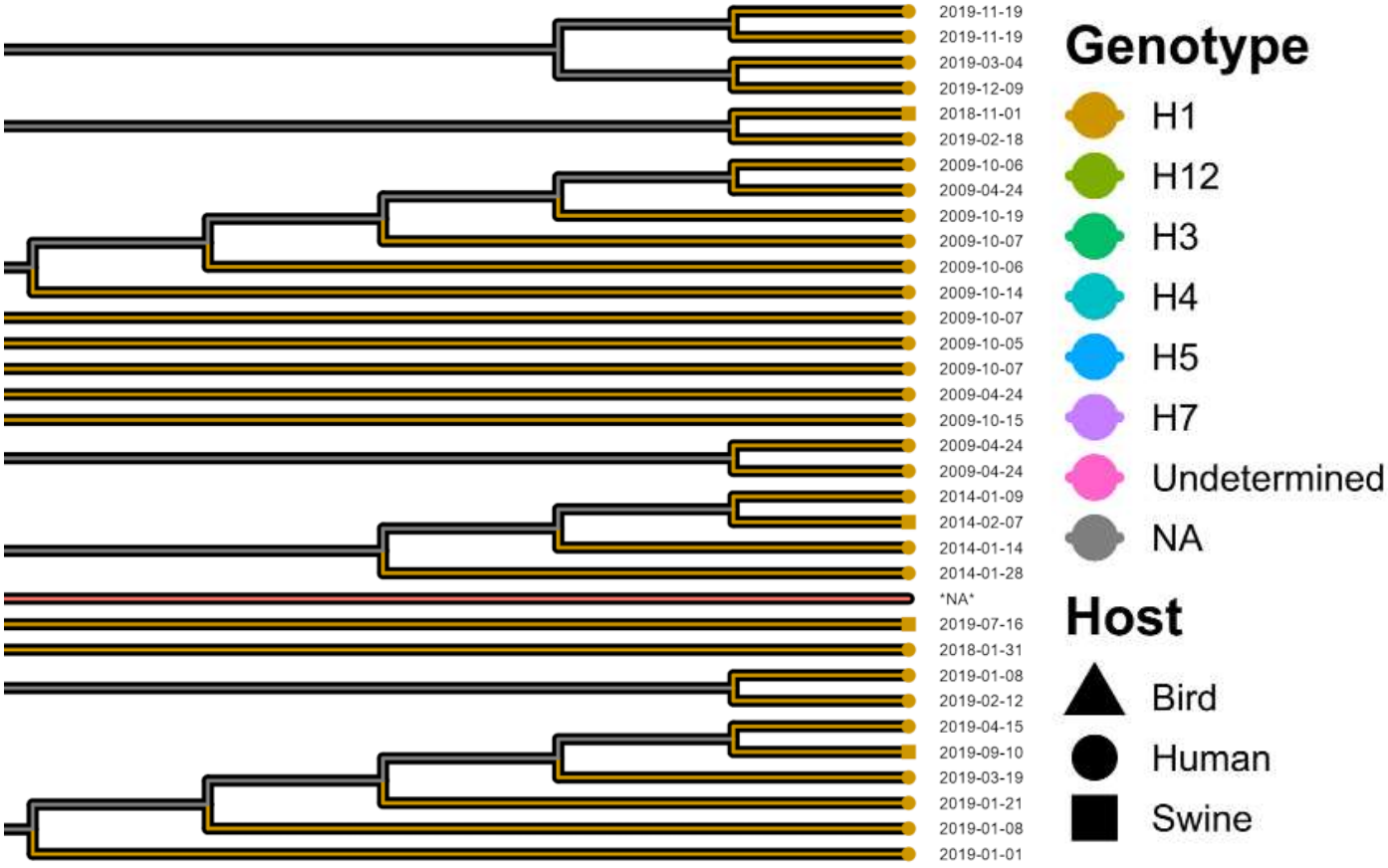
- Zooming in even more in the H1 clade
  - H1 viruses are those that cause “swine flu”

Genotype
- H1
- H12
- H3
- H4
- H5
- H7
- Undetermined
- NA

Host
- Bird
- Human
- Swine

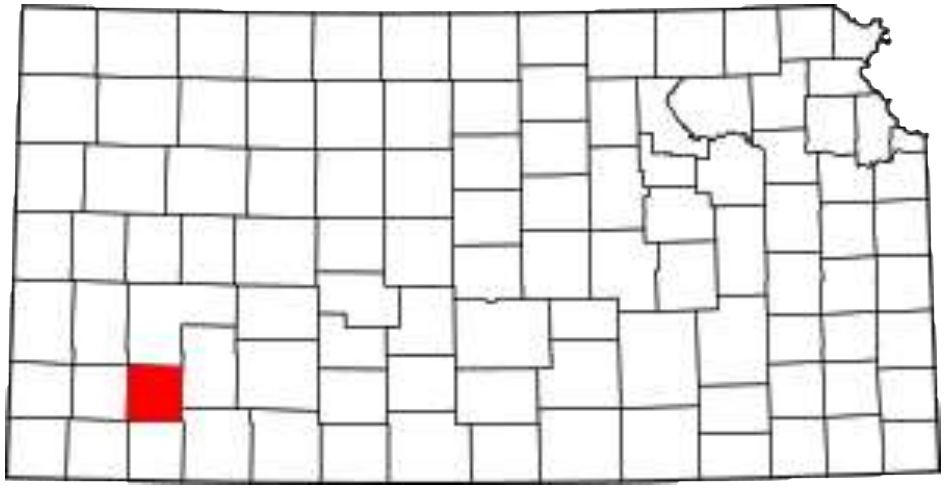
Proof of Concept from Public/KHEL Data

- What do you see?
  - Swine sequences clustering with human sequences?
- Not convincing?
  - Let's put up the date of collection for these sequences.



A One Health Approach is ***Critical***

- Animal reservoirs/populations ***cannot*** be ignored
  - The 1918 **H1**N1 pandemic likely started in Haskell County
    - Spillover from swine?



- Humans and swine pass influenza viruses back and forth
  - We infect them more than they infect us!

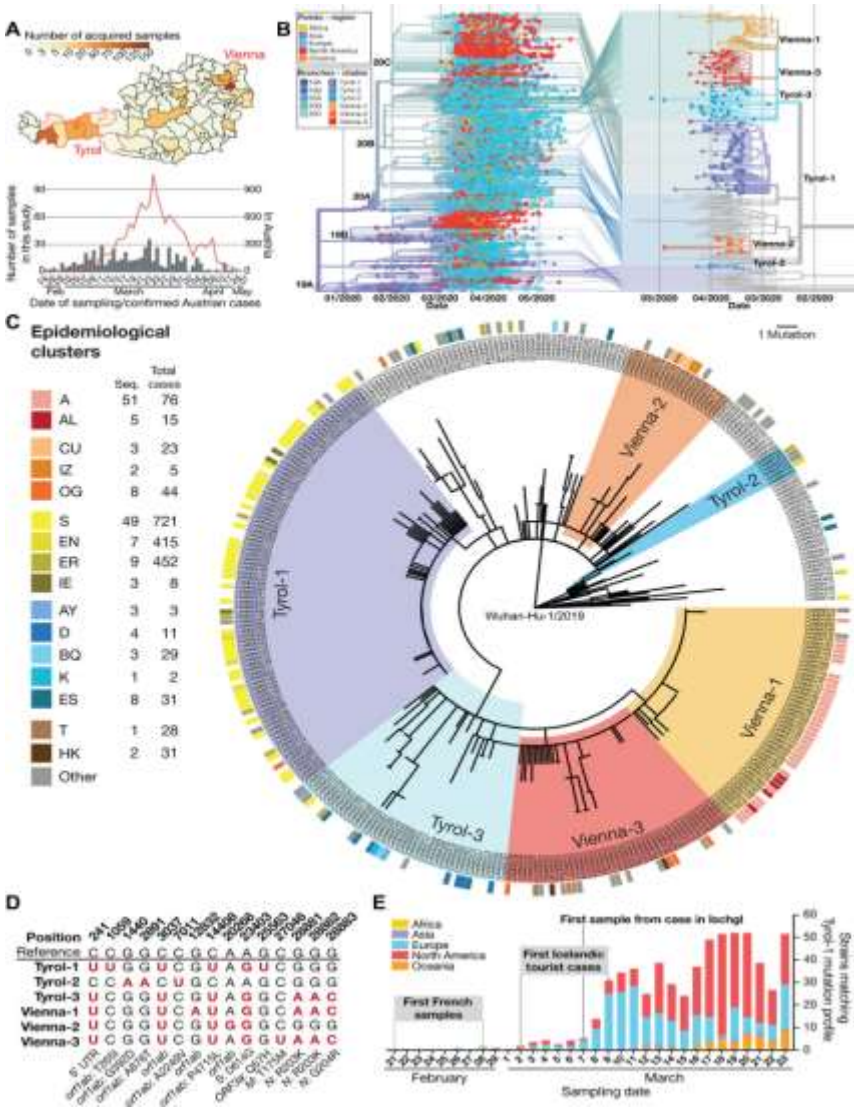
- Genomic surveillance can provide insights into possible spillover events, provided there is enough sampling and investigation.

Where Does Genomic Epidemiology Fit In?

- ***Integrates*** data from disparate sources
  - ***Enhances representativeness*** of data
    - Representative data = better decision-making

- ***Complements*** ongoing influenza surveillance
  - More info per sample = more data

- Eventually, community-level surveillance via wastewater may ***improve spatial resolution*** of data
  - Real-time monitoring *before* clinical signs



Summary

- Genomic epidemiology is an ***emerging tool*** for public health
  - First major use case in SARS-CoV-2 pandemic
    - Where do we use these tools now, though?

- Influenza has similar genomic epidemiology to SARS-CoV-2
  - Genomic material
  - Molecular evolution
  - Wide host variety

- Despite wide host range, influenza surveillance is ***disjointed***
  - Little connection between human and veterinary systems
  - Genomic epidemiology can help fill this gap