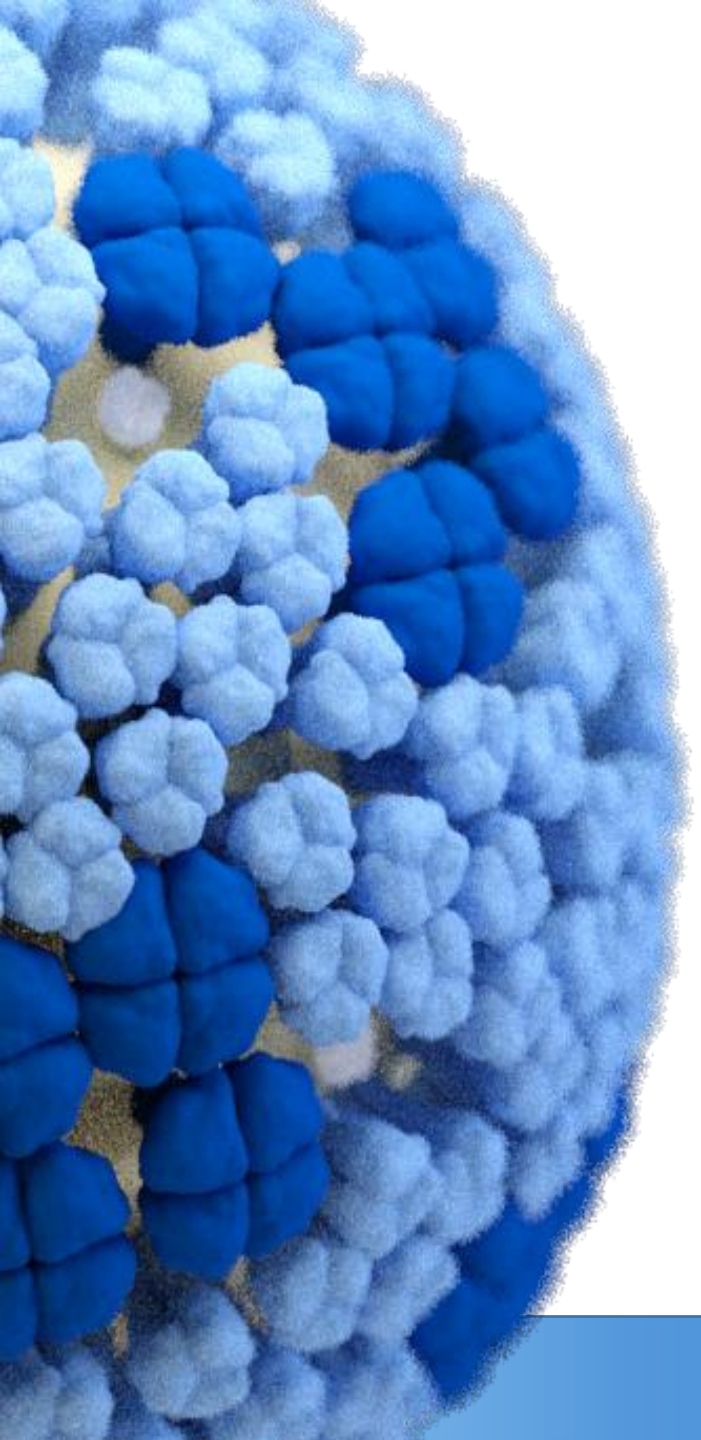


Genomic Epidemiology

Norman Hassell

Centers for Disease Control and Prevention

Disclaimer

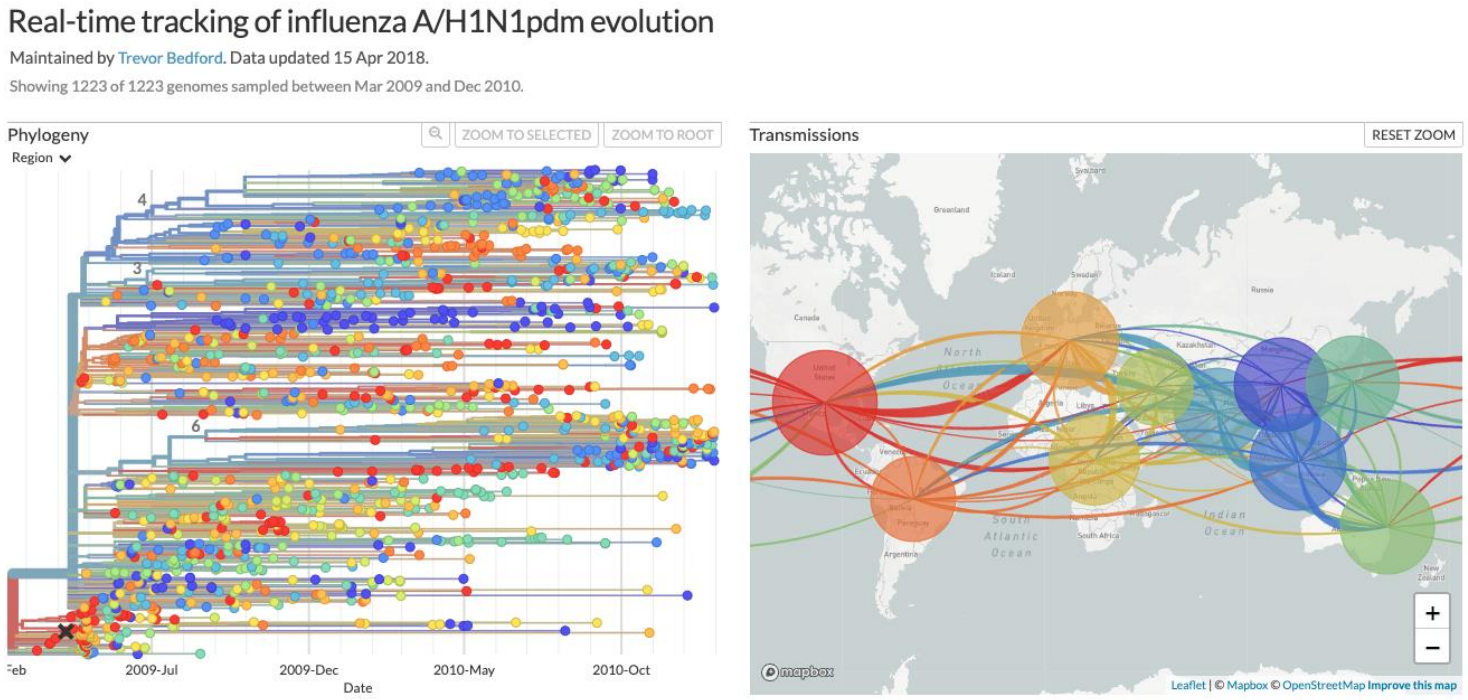
- The findings and conclusions in this presentation are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

- Use of trade names and commercial sources is for identification only and does not imply endorsement by the U.S. Department of Health and Human Services.

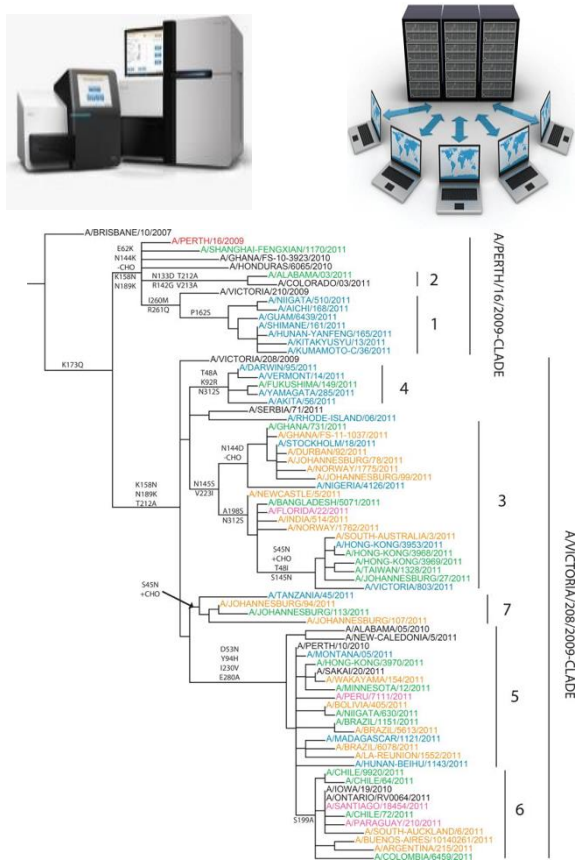
- References to non-CDC sites on the Internet do not constitute or imply endorsement of these organizations or their programs by CDC or the U.S. Department of Health and Human Services. CDC is not responsible for the content of pages found at these sites.

Reiterating the Points of Genomic Epidemiology

- Genomic Epidemiology (aka Molecular Epidemiology):
  - The use of genetic sequence data to study the distribution, transmission, and evolution of diseases in populations.



Goals of Global Influenza Genomic Surveillance: understand circulating viral diversity, where and to what extent



Klimov et al, *Vaccine*. 2012 Oct 5;30(45):6461-71.

- **<u>Create quality sequence of a virus and share data publicly</u>**

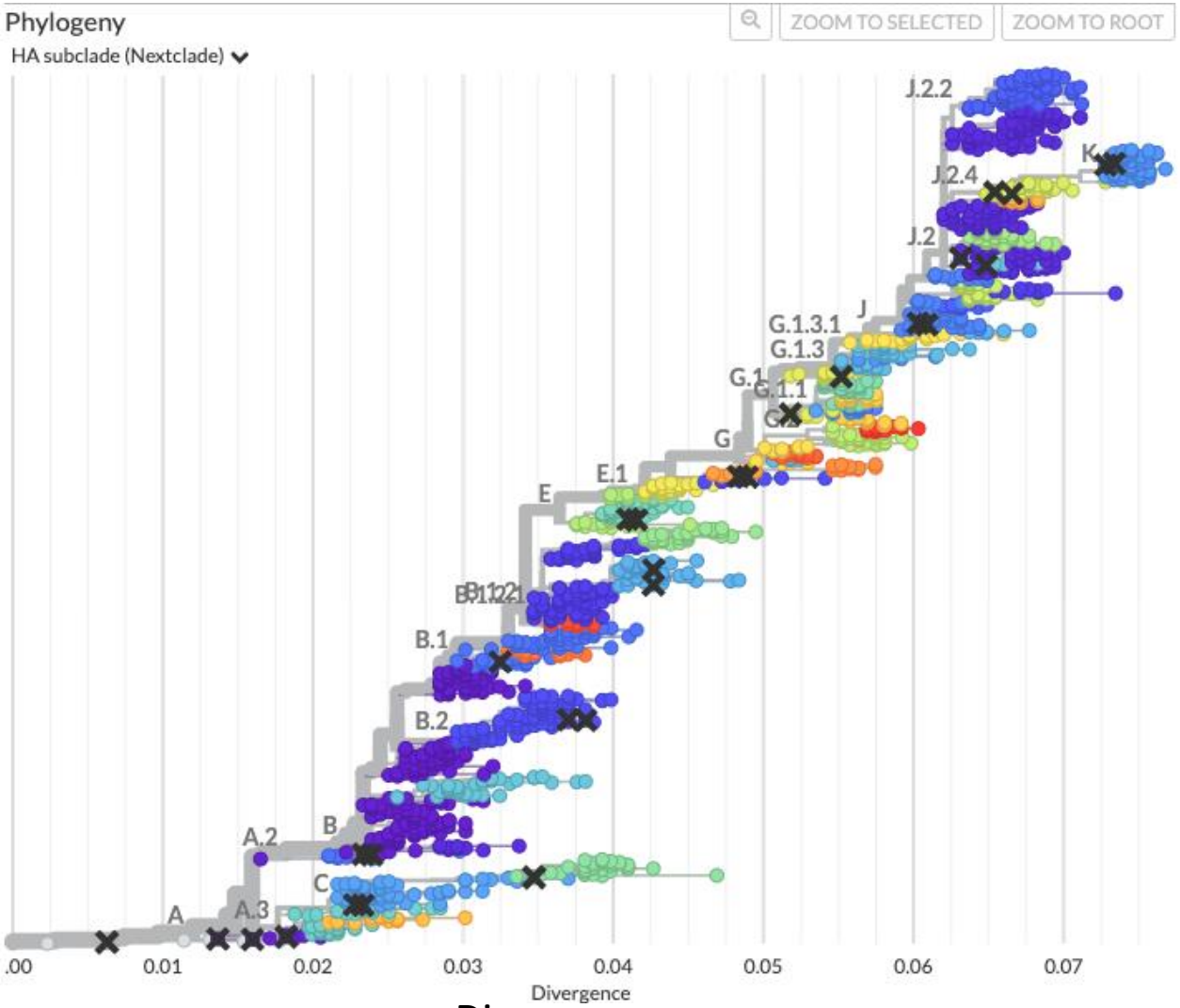
Molecular Epidemiology

- Multiple sequence alignment with relevant sequences of viruses/reference virus for your question
- Create phylogenetic tree(s) – Maximum-likelihood and Time-scaled
- Data integration
  - Correlate genetic change with metadata such as phenotypic, proteomic and epidemiological data
- Analysis is retrospective and a moving window by season – timeliness of data critical for situational awareness

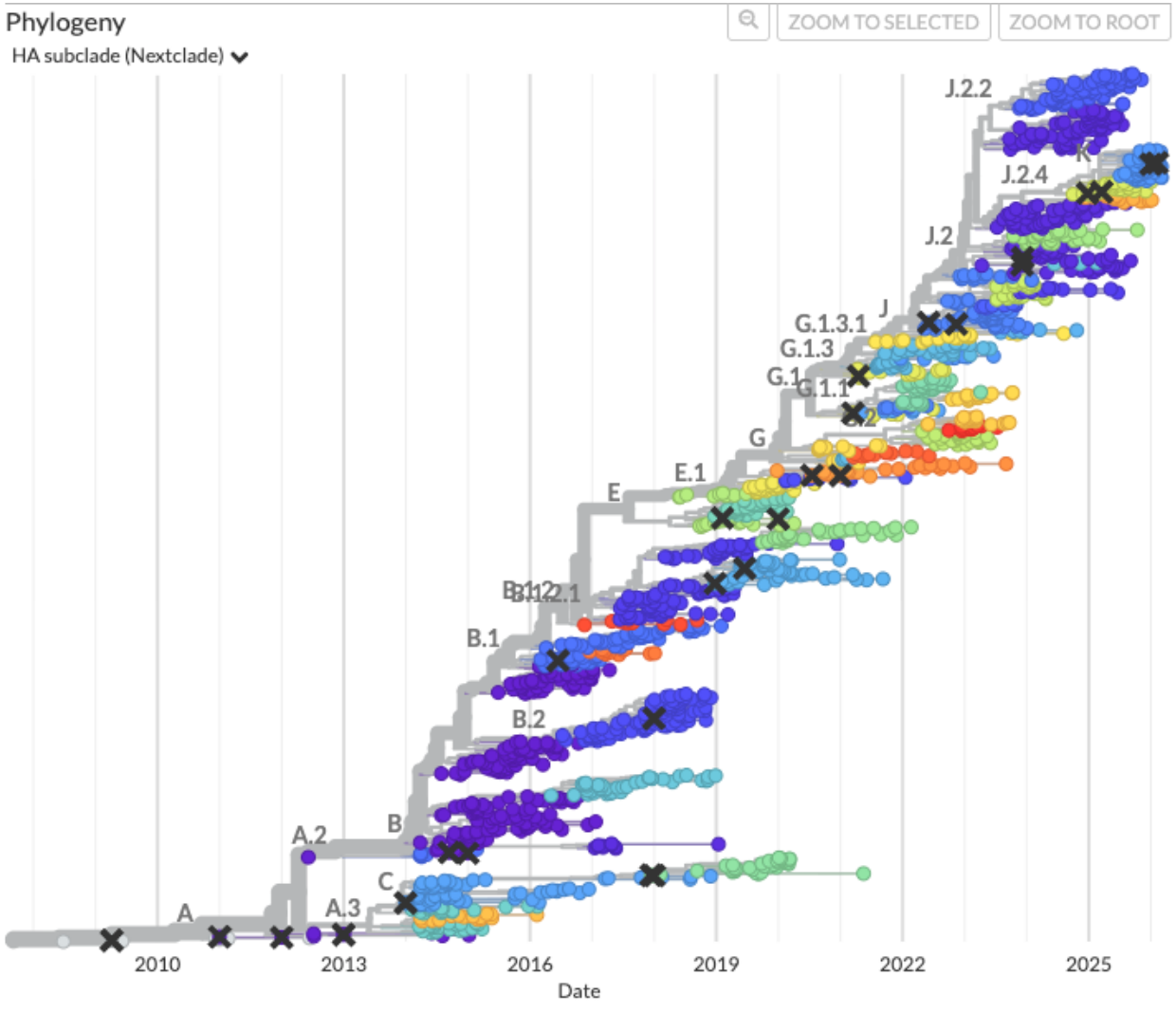
HA Genetic Groups/Clades/Subclades

- Classification mechanism to compare **between** WHO CC labs
  - In general, genetic groups are a **clade** of the phylogenetic tree which shares specific amino acid changes
  - Viruses in genetic groups share the similar amino acid changes – and should generally have similar antigenic profiles
- Timeline
  - pre-2012 – general naming like "A/Brisbane/10/2007"
  - 2012 – Alphanumerics designations annotated the strain name
    - Examples: 6B, 6B.1, 6B.1A for H1s; 3C.2, 3C.2a, 3C.2a.1 for H3s and V, V1 and V1.A
  - 2020 – Subclades –
    - not necessarily distinct phenotypes but meant to assist in discussing genetic diversity and frequency dynamics of co-circulating viral variants without distinct properties
    - Subclades described for both the HA and NA proteins
      - Examples: G.1.1.2 (H3N2), C.1.7.2 (H1pdm09), C.5.6 (B Vic)
      - influenza-clade-nomenclature · GitHub

Nextstrain



Divergence



Time-scaled

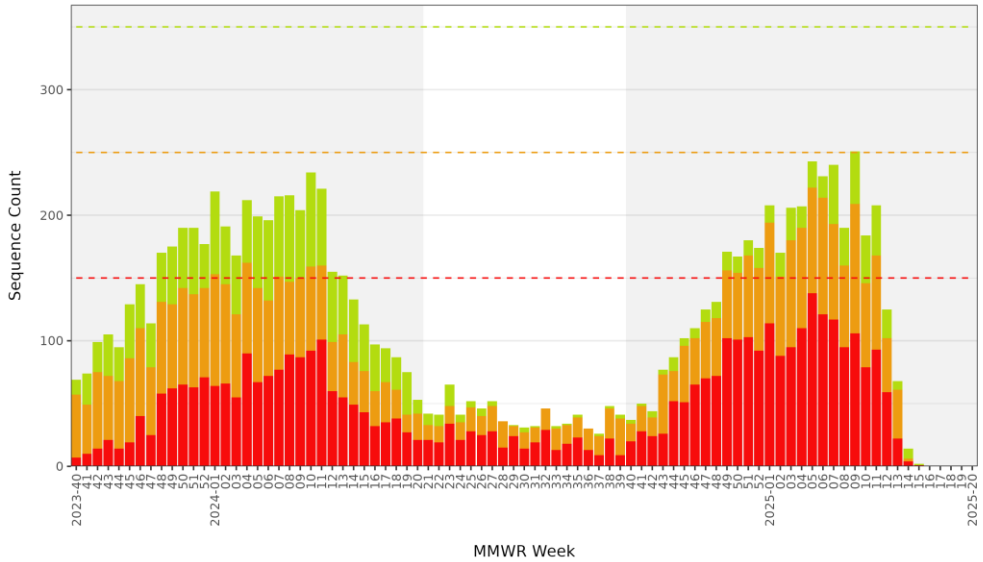
https://nextstrain.org/seasonal-flu/h3n2/ha/12y

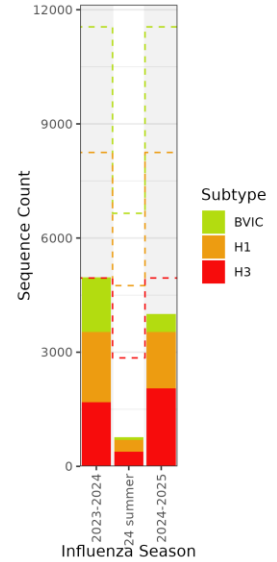
Goals of Global Influenza Genomic Surveillance: understand circulating viral diversity, where and to what extent

- Determine type of analysis and confidence levels to set sampling strategies
  - E.g., Random sampling of specimens from ILI surveillance

US Sample Goals vs Actuals

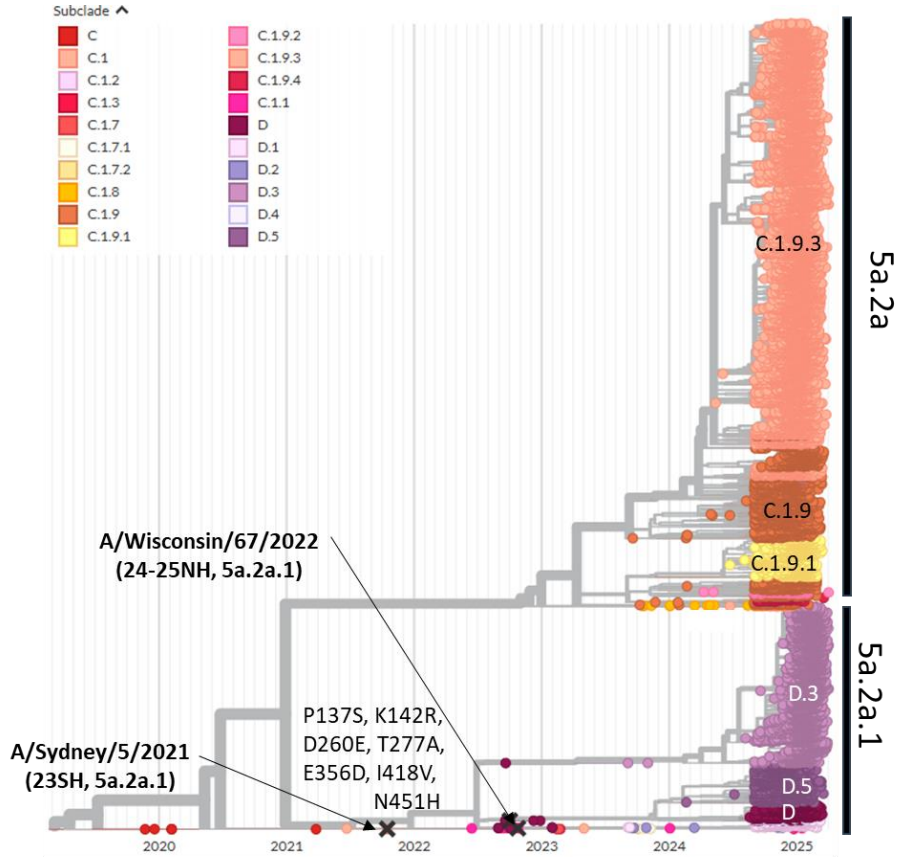
B/Vic	= 100/wk	= 3300/season
H1N1	= 100/wk	= 3300/season
H3N2	= 150/wk	= 4950/season
Total	= 350/wk	= 11,550/season





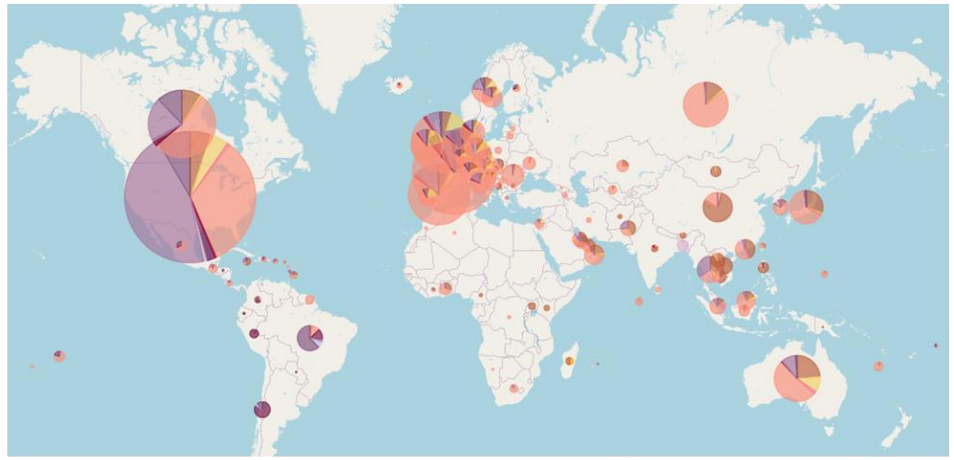
Understand the genetic diversity in epidemic

- Predominant clade/subclade, HA sequence by geography
  - Confidence dependent on sample size



5a.2a and **5a.2a.1** co-circulated globally with regional trends

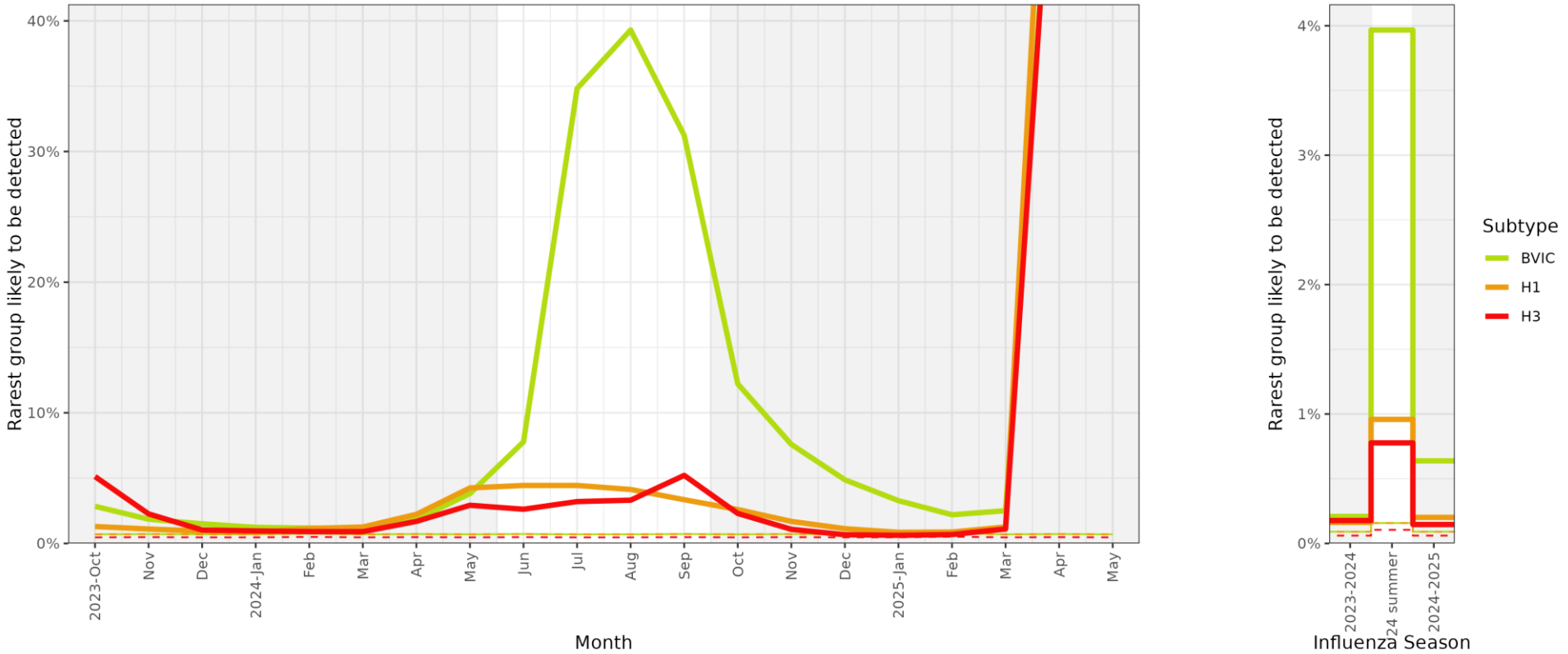
- 5a.2a: Predominant in Asia, Africa, Oceania and Europe, co-circulated in the Americas
  - C.1.9.3 is the largest subclade
- 5a.2a.1: Predominant in the Americas
  - D.3 is the predominant subclade



Approximate Subclade Proportion – Collected since 9/1/2024 Data as of 4/28/2025

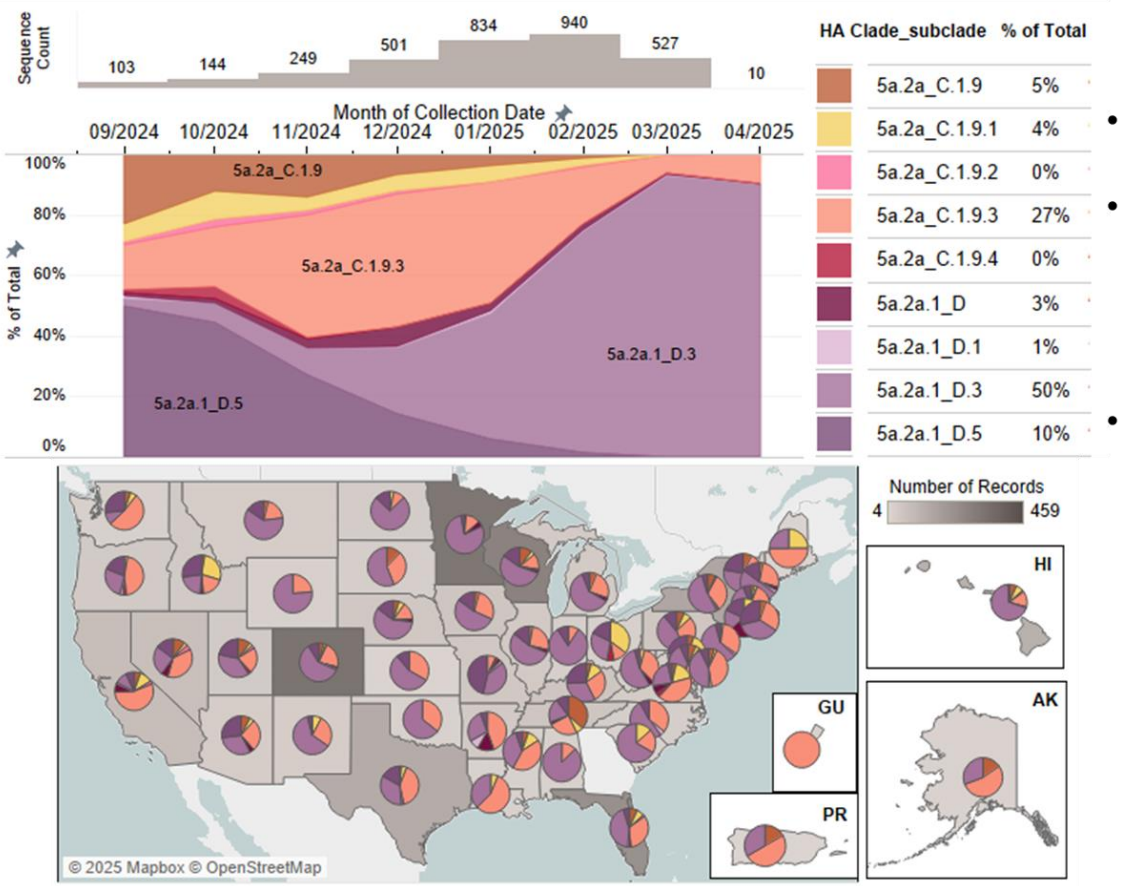
Understand the genetic diversity in epidemic

- What is the minimum proportion we are likely (i.e., confidence = 95%) to detect with a given sample size?
  - Goal to identify emerging genetic variants early



Understand the genetic diversity in epidemic

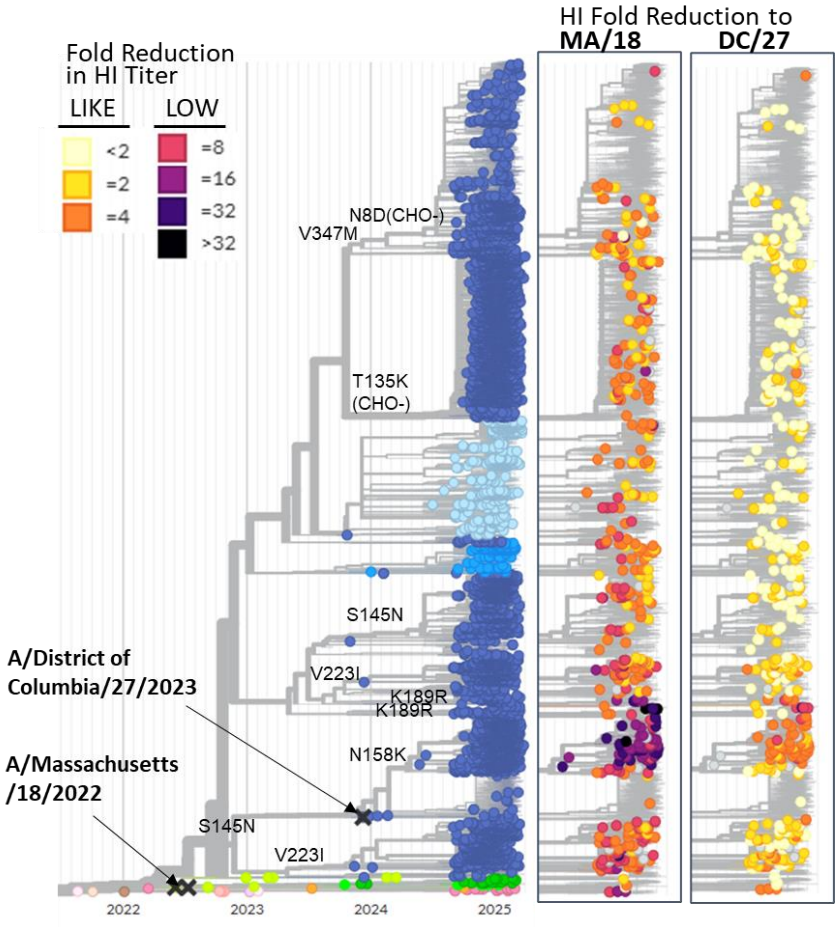
- Change in clade/subclade, HA sequence over time



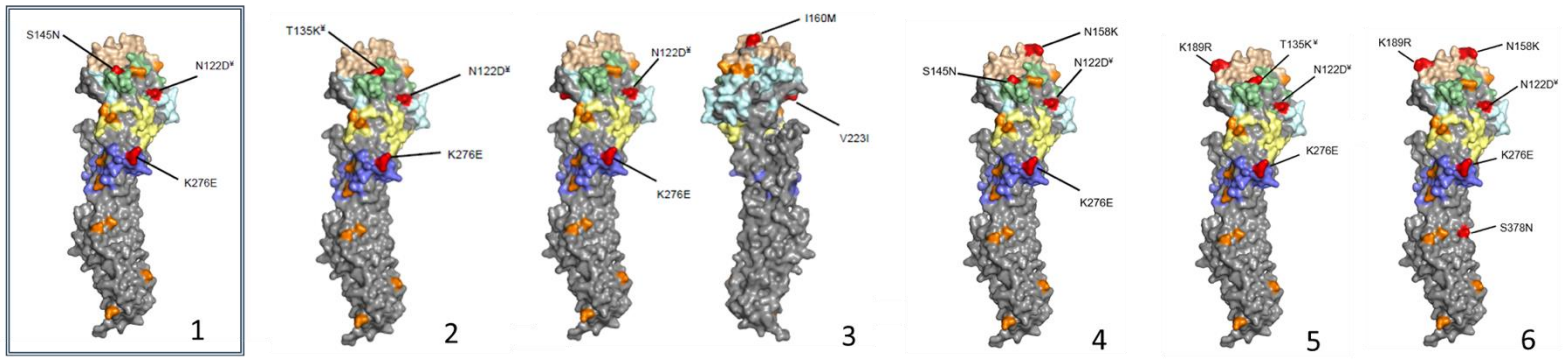
Collected in U.S. since 9/1/2024

- 5a.2a and 5a.2a.1 HA co-circulated
- 5a.2a: Predominant in early season
  - C.1.9.3 was the largest subclade
  - C.1.9.3 predominant in a few states
- 5a.2a.1: Predominant during peak and late season
  - D.3 increased in mid-season and became predominant
  - D.3 is predominant in most states
  - The increase of D.3 proportion is observed in most states, even in C.1.9.3-predominant states

Genotype to Phenotype



- Create phenotypic data representing the circulating genetic diversity
  - Prioritize viruses by their HA sequence and global observance
  - Create testing goal per HA sequence
- Determine molecular determinants of antigenic change through genetic and antigenic data integration
  - E.g., Amino acid positions N158K and K189R associated with decreased antibody recognition using post-infection ferret antisera
  - Identify genetic and antigenic lineages for additional analysis using human sera

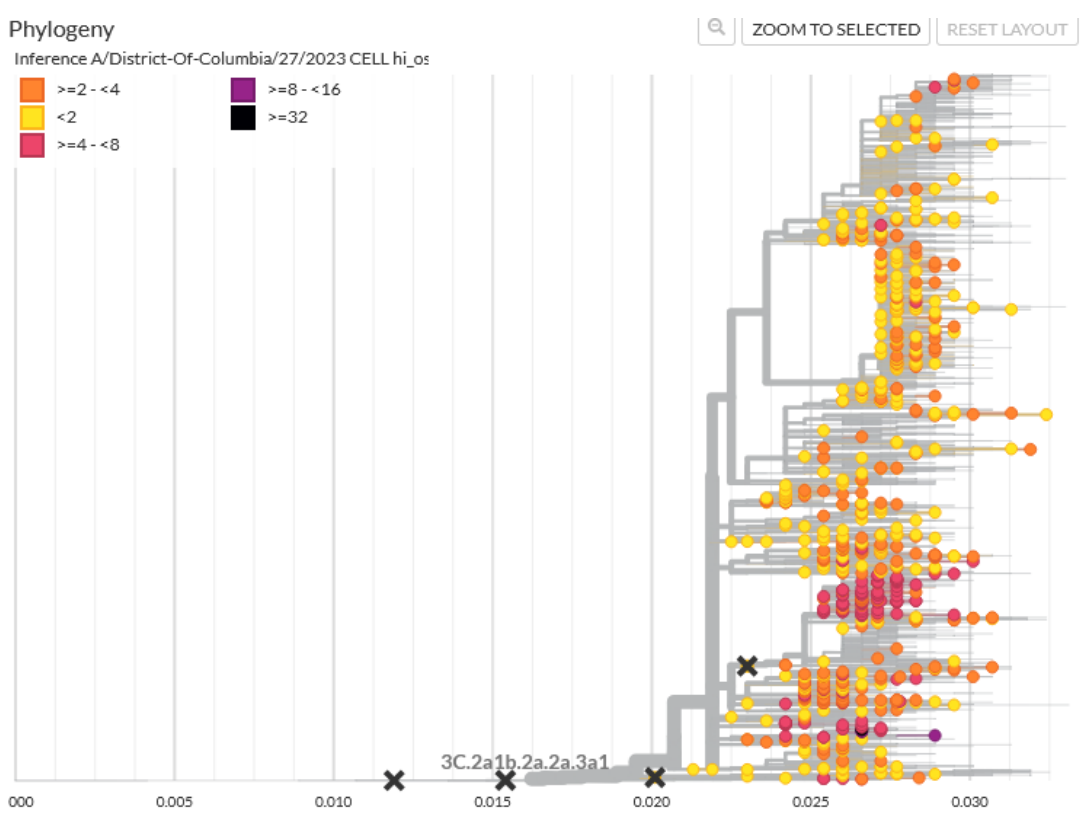


Emerging subclades in J.2 September 2024 through January 2025

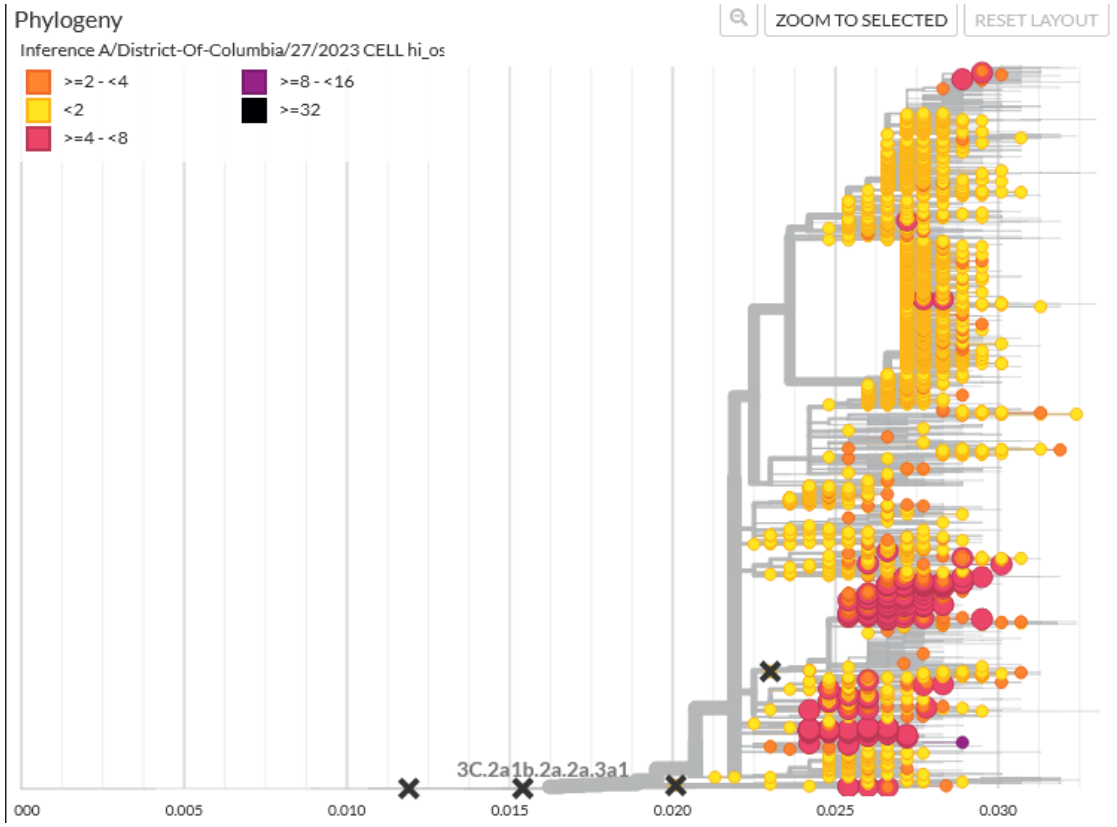
Genotype to Phenotype

Infer Antigenic Characterization for HA proteins which have met testing and result quality thresholds

- *more of the genetic variation seen within a season to have actional antigenic results in time for the vaccine selection meetings*



Test Results

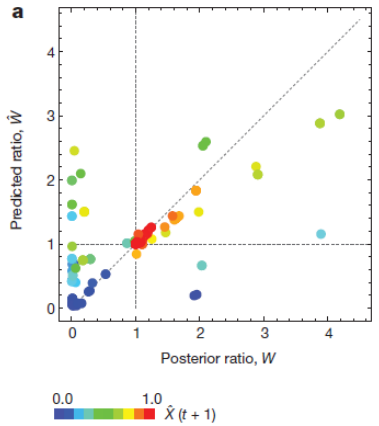


Test and Inferred Results

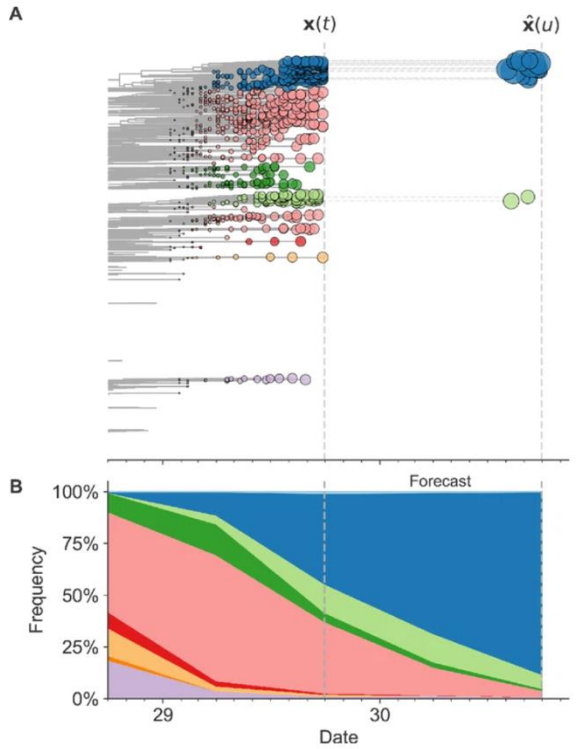
Predictive Models

Goals: predict the *clade* which will *continue* to circulate in the next 6 months/1 year

- NextFlu – Richard Neher and Trevor Bedford
- Previr - Marta Luksza and Michael Lassig
- Clade fitness models
  - Sequence data from GISAID
    - Clade frequency
    - Mutational load
    - Epitope sites
  - Antigenic data from HI or neutralization assays
  - Corresponding meta-data
  - Correlated epidemiological data



Luksza M, Lässig M. *Nature*. 2014 Mar 6;507(7490):57-61



Huddleston J, et al. *eLife 9:e60067*

Neher RA, Bedford T. *Bioinformatics*. 2015 Nov 1;31(21):3546-8.

Current state of predictive modeling

Still a lot of work to do

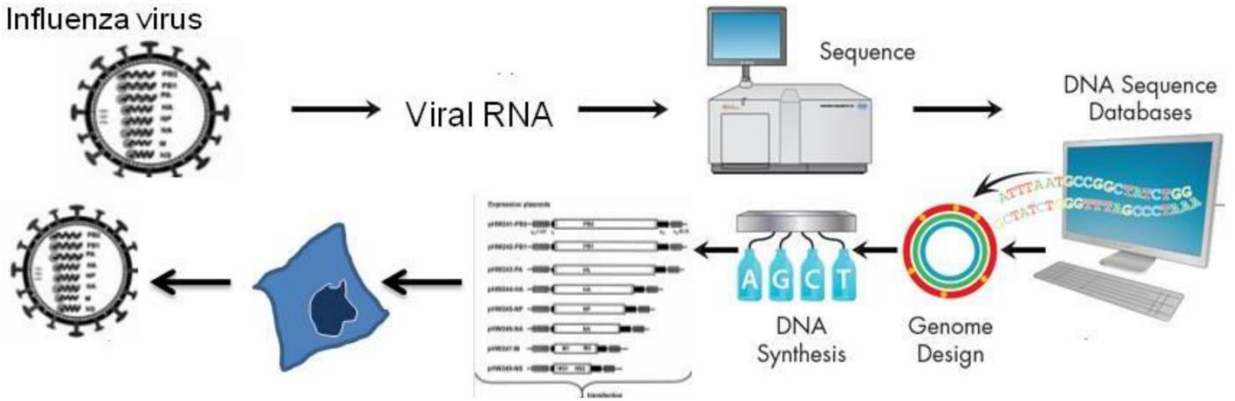
- Unable to predict which clade will predominate – but can have confidence that it will persist
- Unable to predict which new clades will emerge
- Unable to predict which A subtype or B lineage will predominate
- Virus phenotype changes
- Geographic and temporal differences in clade proportions
- Clade turnover rate differences

Influenza B and A(H1N1)pdm09 – need to assess the how the models based on A(H3N2) work for other subtypes

Future Landscape of Influenza Genomics

- Improved lab-epi data integration
  - Automated electronic lab reporting
  - Data lake approach
  - Cloud computing
- Synthetic genomics
  - Rapidly generated viruses from sequence data
- New vaccine strategies may require different data analysis for recommendations

Use of genetic data to design and synthesize candidate vaccine viruses



Courtesy of Dr. Dave Wentworth

Genomic Epi in Action: Subclade K CVV Generation Timeline

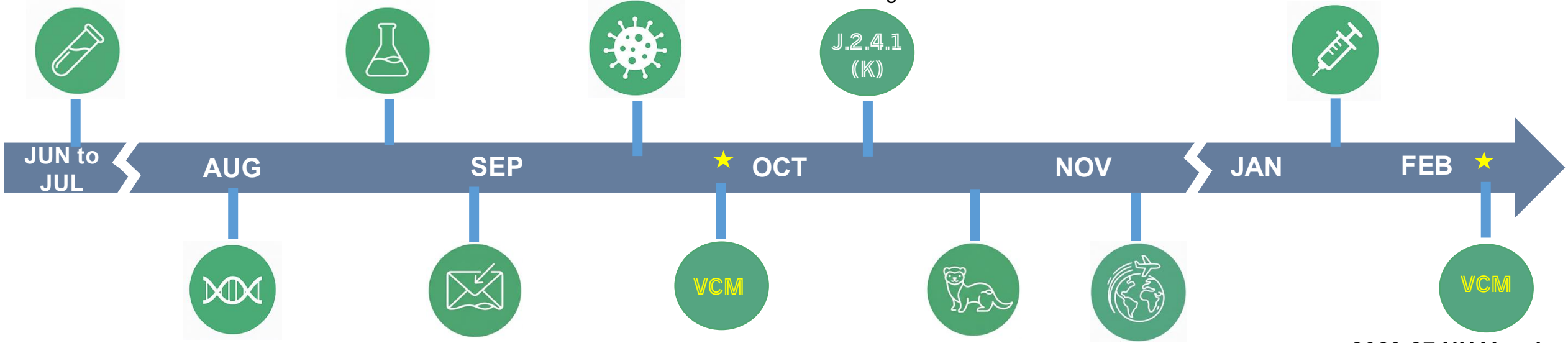
Specimen Collection
First K specimens collected through CDC National Influenza Surveillance and Traveler-Based Genomic Surveillance

CVV Initiation
Initial isolation attempts via provision of respiratory specimens National Influenza Surveillance

CVVs Generated
Egg and cell reference viruses generated to newly emerging subclade of J.2.4 viruses

Subclade Naming
Emerging subclade of J.2.4 viruses with specific amino acid substitutions were designated as subclade J.2.4.1 and then given alias K

Reassortants
High growth reassortants started to arrive at CDC to subclade K



Detection
Sequence analysis highlighted J.2.4 viruses with ten amino acid changes relative to vaccine and promptly performed antigenic characterization;
Evidence of antigenic drift

Collaborators Notified
CDC notified other GISRS collaborating centers of their findings for newly emerging subclade of J.2.4 viruses

2026 SH Vaccine Recommendation
Report: J.2.4 viruses well including those with ***notable additional HA substitutions*** S144N (a potential addition of an N-glycosylation site), N158D, I160K and Q173R, which have recently emerged

Reference ferret antisera
Ferret antisera generated to K reference viruses and tested

CVV Distribution
Reference viruses for subclade K distributed to collaborators and reassorting laboratories

2026-27 NH Vaccine Recommendation: Subclade K viruses recommended as A(H3N2) component

Pre-print Interim US 2025-26 Season SSRN

NOTE – timeline length is not to scale

Thank you!