Session II: Genomic Epidemiology

Daniel Evans, MS MB(ASCP)^{CM} CIC REHS CPH
Genomic Epidemiologist, Sequencing and Bioinformatics Unit
Minnesota Department of Health

About Me

Genomic Epidemiologist,
Minnesota Department of Health (full-time)

Environmental Science Officer,
United States Army Reserve (part-time)

My Teams and Zoom habitats:

- APHL NGS Subcommittee
- CSTE AMD Subcommittee (Workforce Development)
- Applied Genomic Epidemiology Community of Practice (AMD Platform)
- Pathogen Genomics Centers of Excellence
- AMD Regional Training and Workforce Development
- DOD Wastewater Surveillance Program

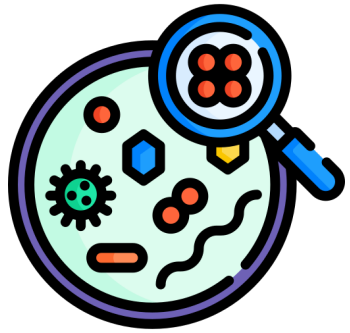
Bottom Line Up Front

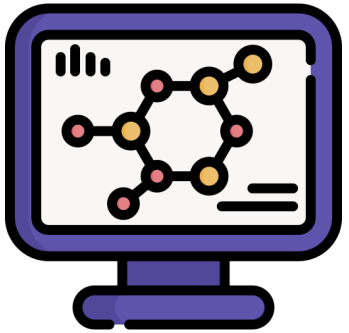
"Genomic epidemiology" broadly describes many tasks related to disease surveillance.

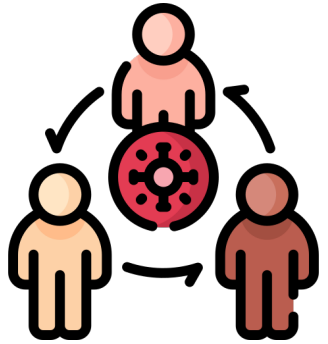
The practice of genomic epidemiology depends on state, pathogen, disease surveillance goals, and "side of the fence."

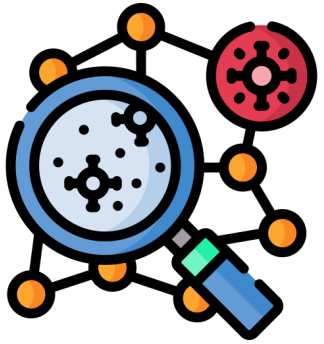
Priorities for pathogen genomic investigations will follow several general themes.

Genomic epidemiologists must develop and maintain cross-disciplinary skill sets.







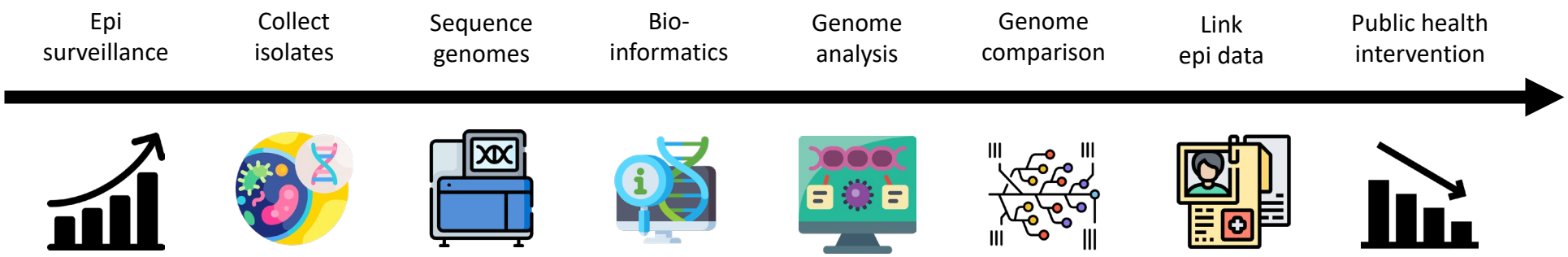


Flaticon: Freepik

The pipeline of genomic surveillance

Place the following terms on the pipeline figure:

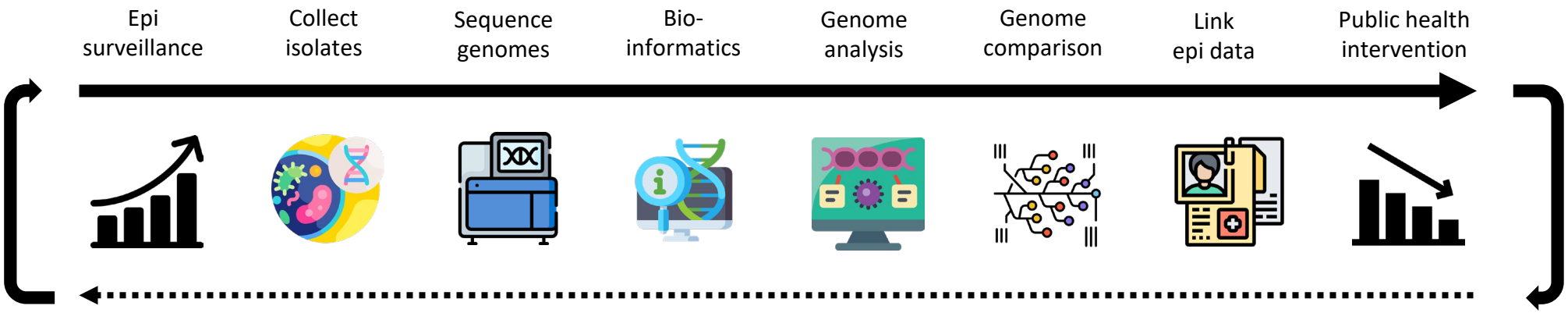
Laboratory
Bioinformatics
Epidemiology
Genomic Epidemiology



Flaticon: Freepik, Smashicons, Flat Icons, Winnievinzence, Eucalyp

The pipeline of genomic surveillance

"Genomic epidemiology" and "genomic epidemiologists" could potentially include / be included in **ANY of these tasks**.



Flaticon: Freepik, Smashicons, Flat Icons, Winnievinzence, Eucalyp

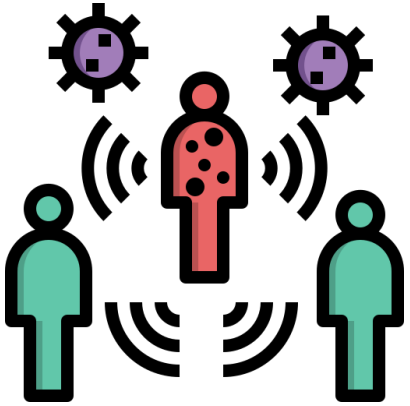
What is genomic epidemiology?

GENOMICS

The study of the structure, function, and evolution of genomes – the complete genetic "codes" of organisms

EPIDEMIOLOGY

The study of the incidence, causes, spread, and control of diseases that affect human health

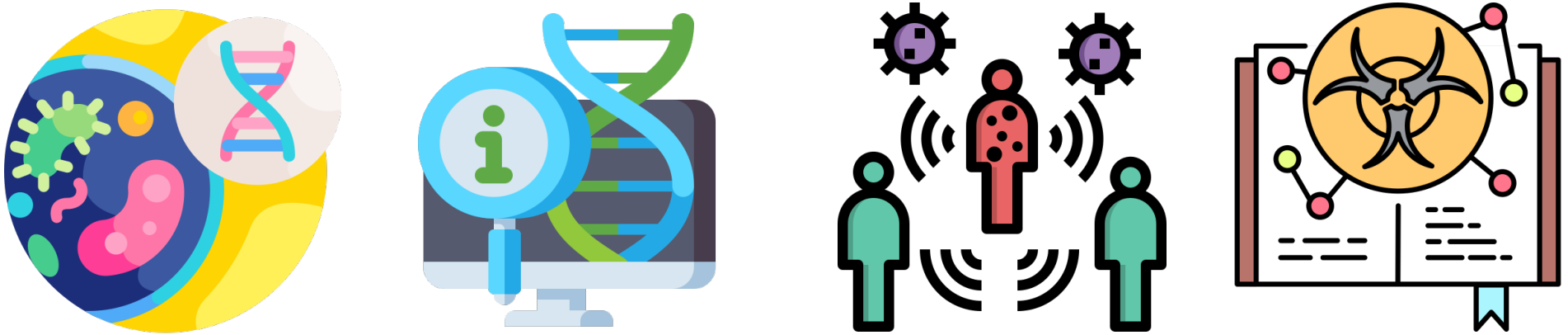


Flaticon: Freepik, Backwoods

What is genomic epidemiology?

GENOMIC EPIDEMIOLOGY

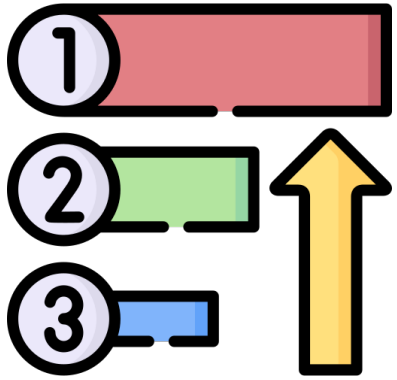
The practice of identifying, studying, tracking, and controlling the spread of disease
by integrating the study of genomes into public health surveillance.



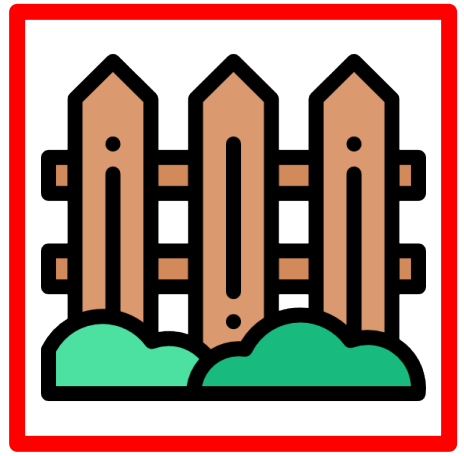
Flaticon: Freepik, Backwoods

Genomic epi looks different in different jurisdictions

Every health jurisdiction has **unique public health needs and challenges** that affect genomic epidemiology.







Flaticon: Freepik

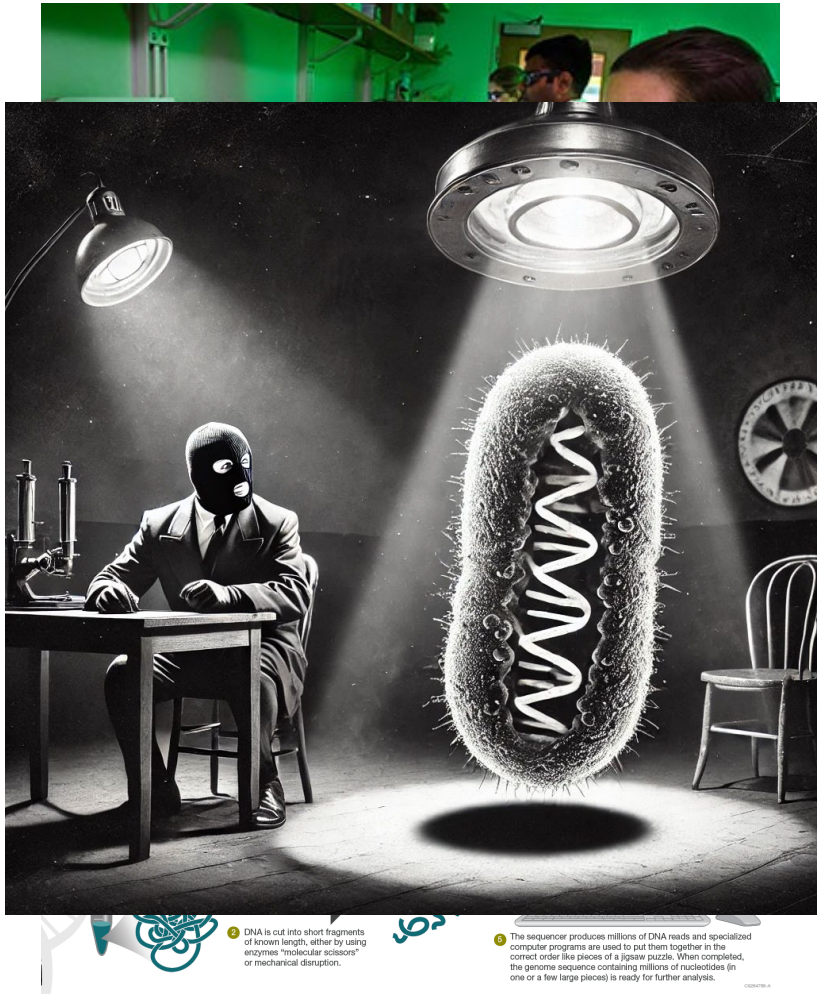
On the lab “side”…

“Can you start running weekly phylogenetic builds of SARS-CoV-2 genomes, and then share the results with the Epidemiology and Data Unit?”

“Here are a bunch of *C. difficile* genomes that we sequenced before you started here. Good luck!”

“You’re good at interrogating genomes. Can you help us analyze these weird metagenomics results?”

“We’re adding you to the lab chore schedule.”



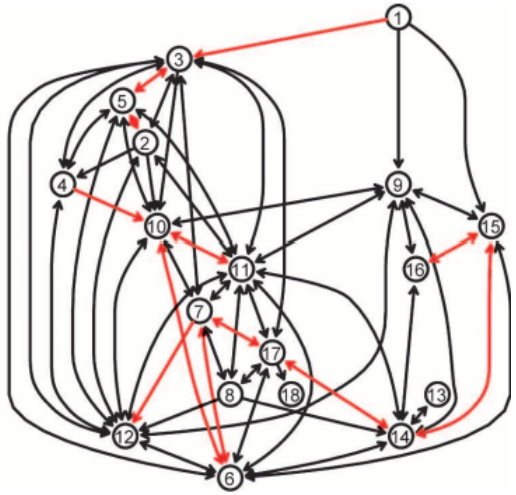
PulseNet, CDC, ChatGPT

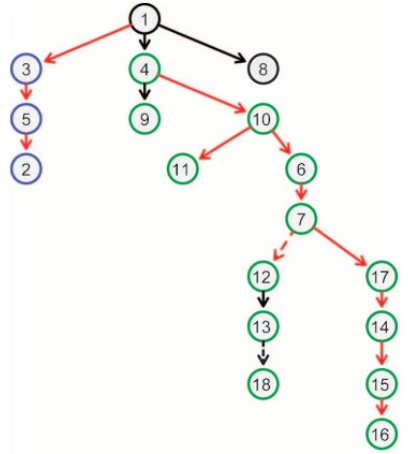
On the epi “side”…

“We’d like to pitch genome sequencing for HAIs in our hospitals and nursing homes. Can you give a presentation at our quarterly statewide call for infection preventionists?”

“There’s a tribal health department that wants to do COVID wastewater surveillance. Can you make some figures for them?”

“We want to automate data entry of COVID variants into our disease surveillance system. Can you help us?”





Snitkin et al, Science Translational Medicine 2012

“Classes” of genomic epidemiology

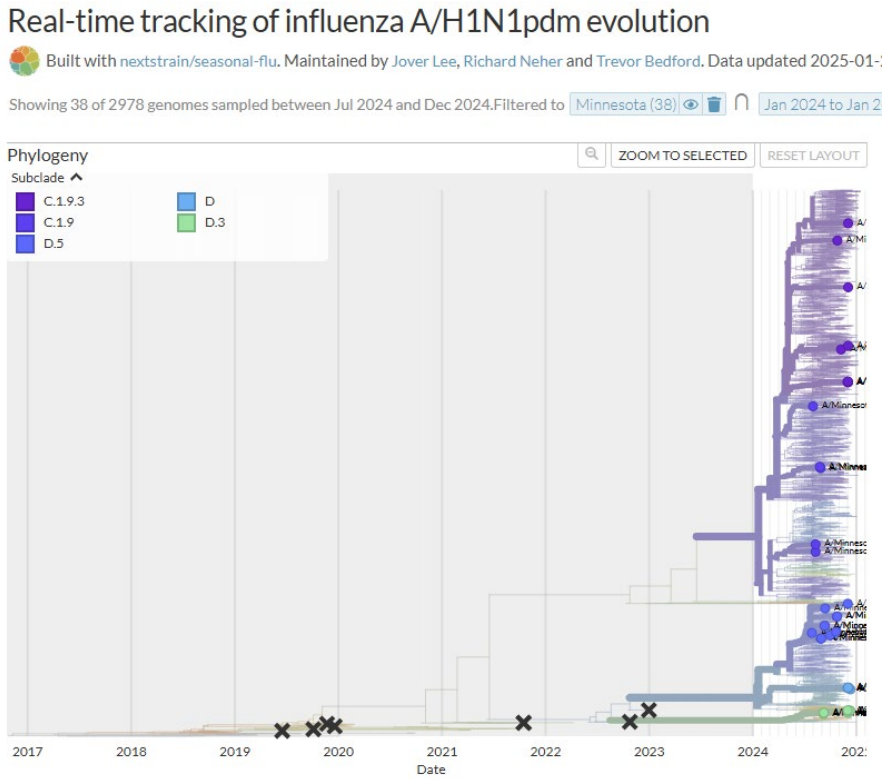
Genomic Epi Class	Example Questions	Example Lines of Effort

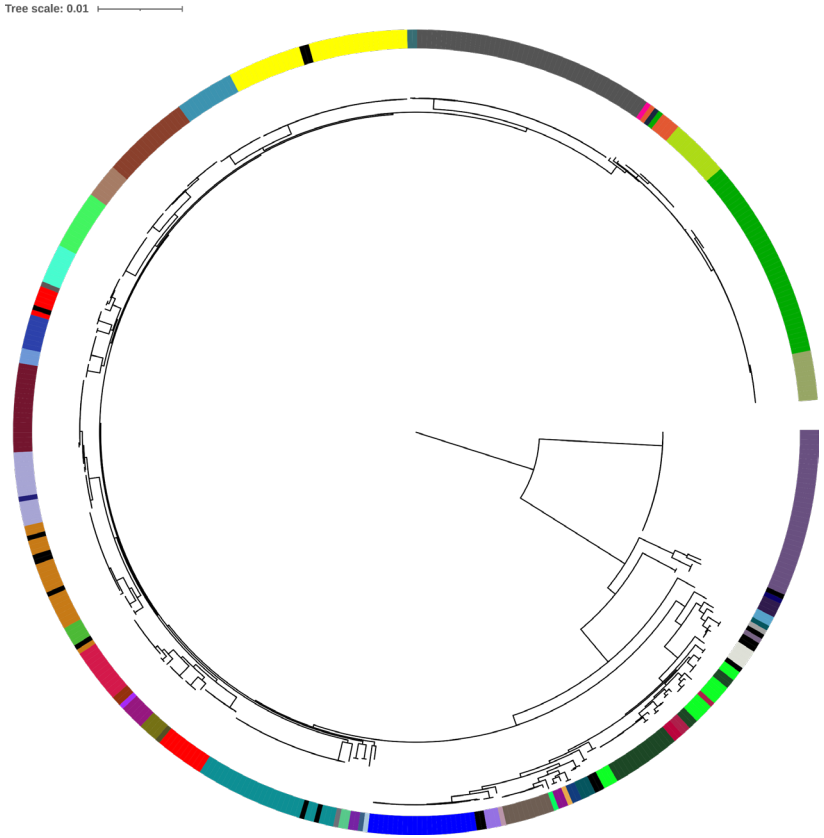
“Classes” of genomic epidemiology

Genomic Epi Class	Example Questions	Example Lines of Effort
Descriptive	What [pathogen subtypes / variants] are we seeing in this population?	Genetic typing of pathogens Variant dashboards
Detective	Is [variant] in our population? Are there undetected outbreaks?	Subsampled or active genomic surveillance of a specific pathogen
Investigative	Is this an outbreak? Why is this outbreak happening?	Building trees, SNP calling, gene detection of isolates from a cluster Cross-referencing genomic data with “shoe-leather” epi data
Analytical	Is [subtype / variant] significantly associated with [outcome]?	GWAS / pan-GWAS of large datasets of sequenced pathogens with detailed clinical data
Predictive	“Can we anticipate which [subtype / variant] will circulate in the future?”	Adjusting predictive models to account for subtypes / variants

Descriptive genomic epidemiology

What [pathogen subtypes / variants] are we seeing in this population?



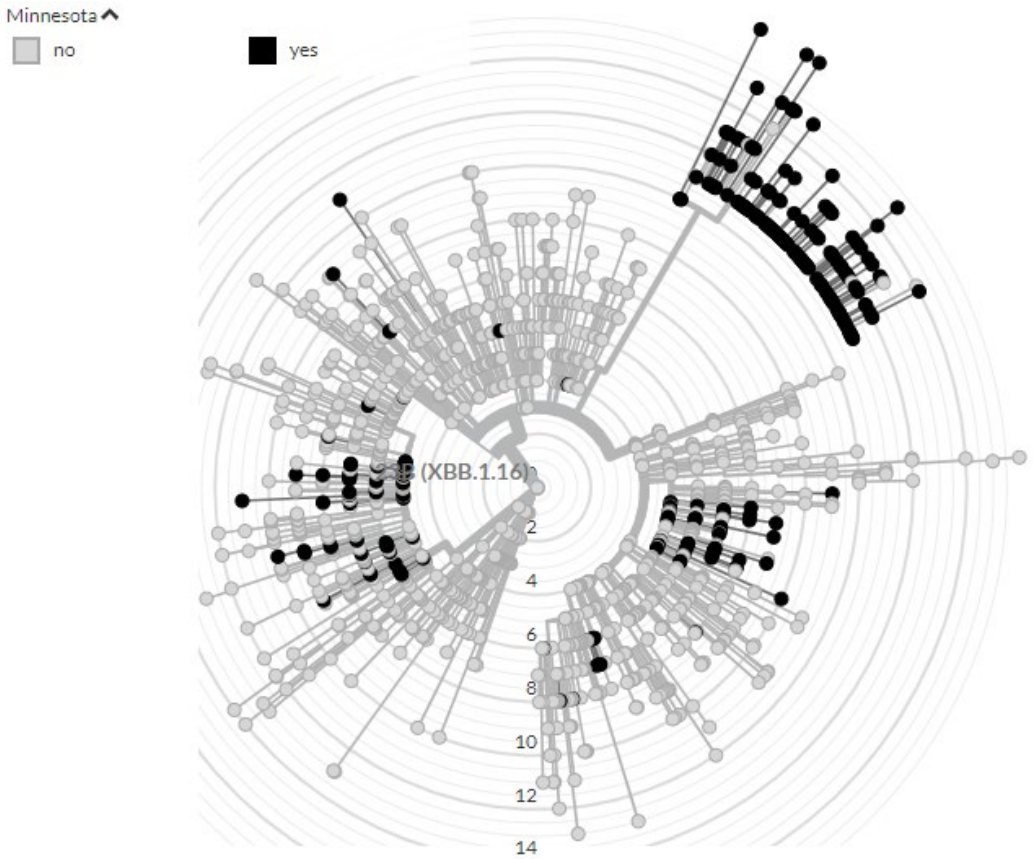


Minnesota Department of Health, unpublished data
GISAID build of influenza A/H1N1pdm genomes from MN and globally

Detective genomic epidemiology

Is [variant] in our population? Are there undetected outbreaks?

Identifying a Minnesota-specific clonal expansion of a SARS-CoV-2 variant

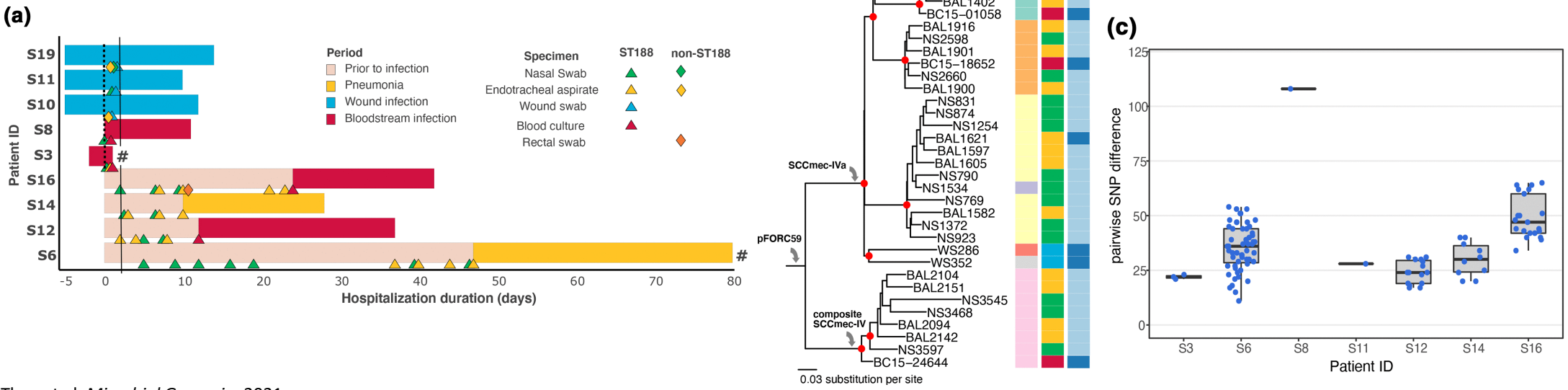


Minnesota Department of Health, unpublished data

Detective genomic epidemiology

When, and how frequently, does colonization by pathogenic bacterial strains lead to infection?

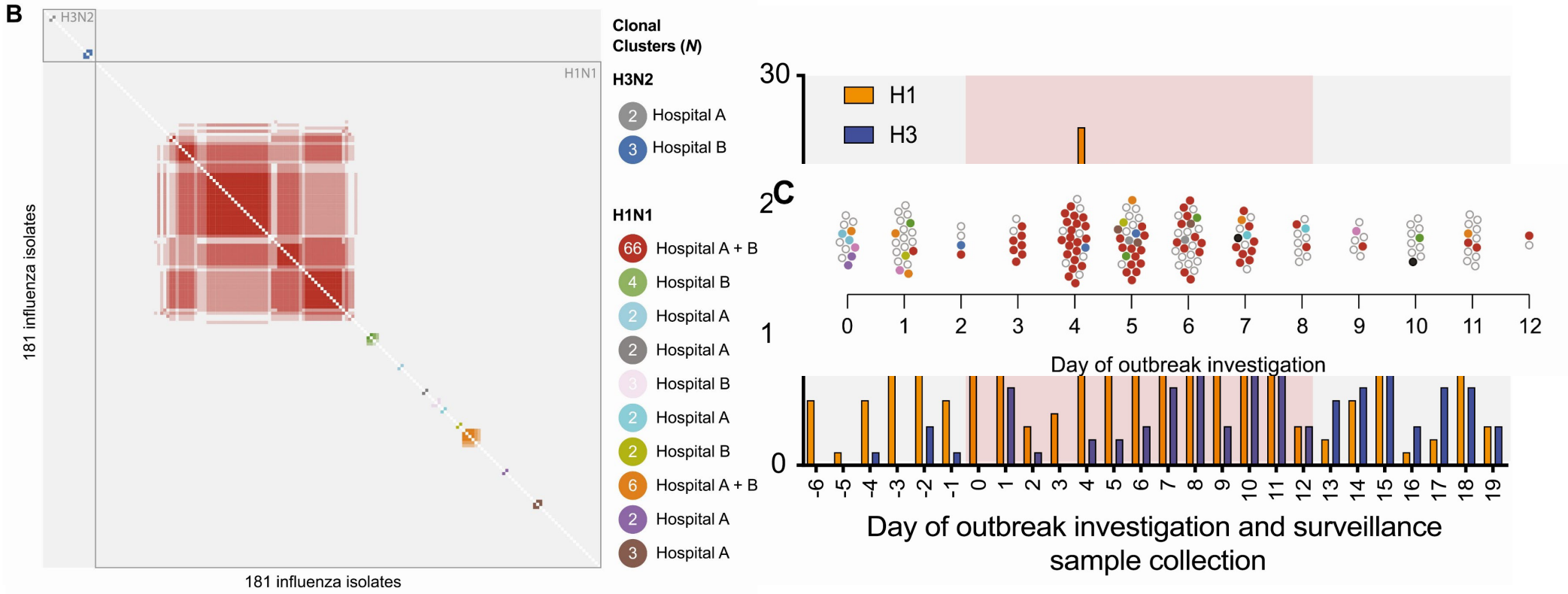
Detecting *S. aureus* ST188 colonization in the ICU and repeated testing for subsequent infections



Thuy et al, *Microbial Genomics* 2021

Investigative genomic epidemiology

Is this an outbreak? Why is this outbreak happening?

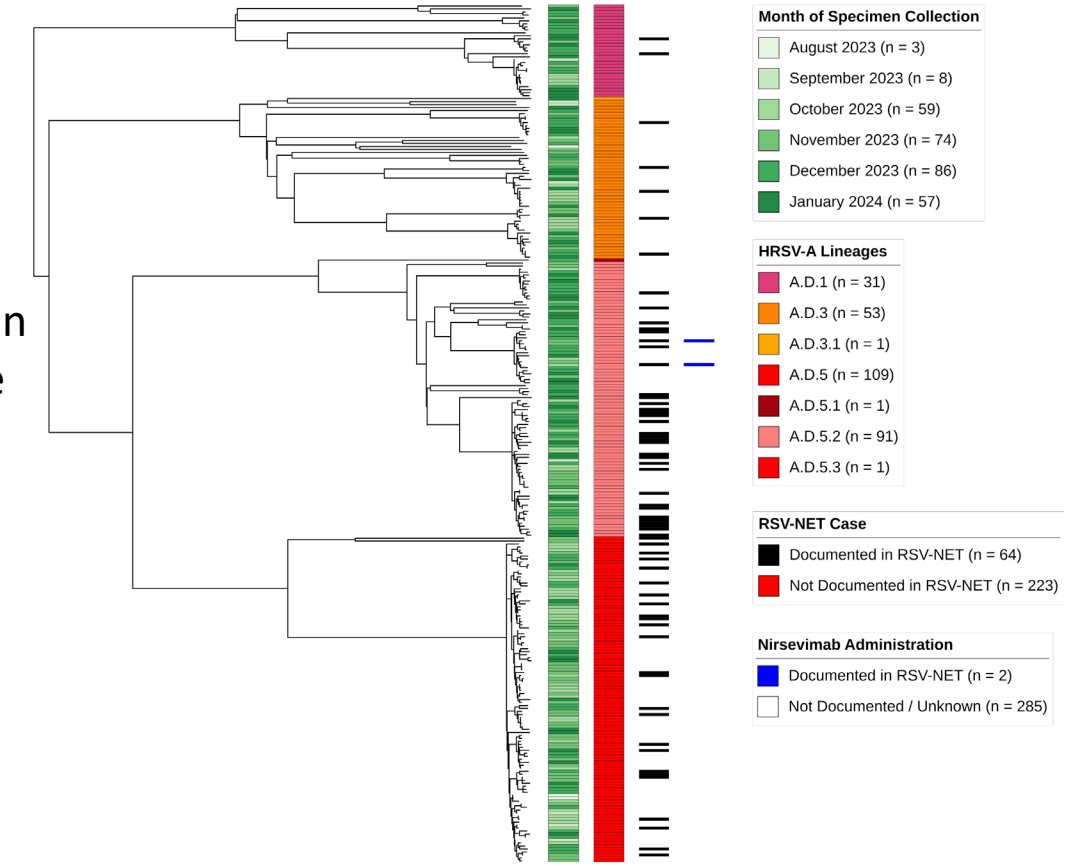


Javaid et al, *Clinical Infectious Diseases* 2021

Analytical genomic epidemiology

Is [subtype / variant] significantly associated with [outcome]?

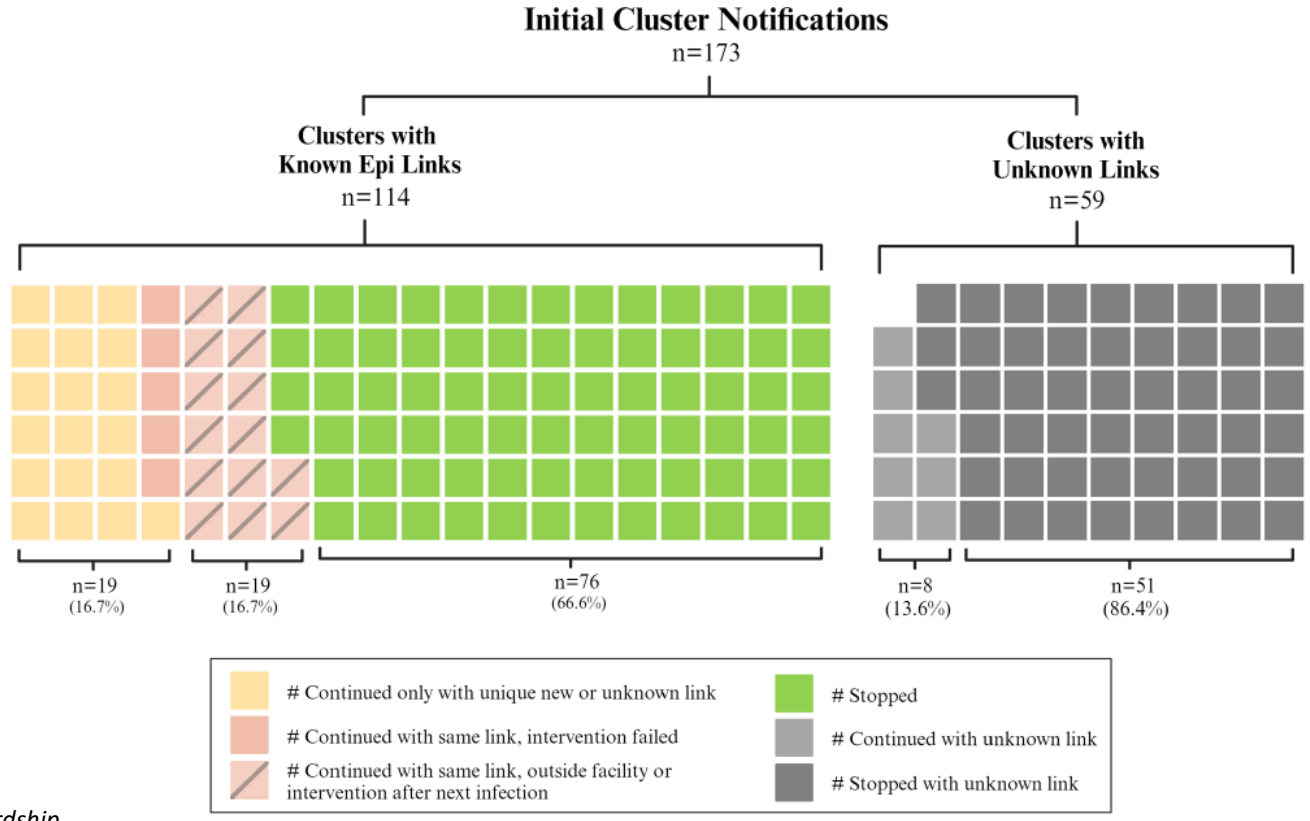
Greater incidence of hospitalization among infections of hRSV lineage A.D.5.2



Evans et al, *Emerging Infectious Diseases* 2024

Analytical genomic epidemiology

How effective is genomic surveillance for hospital infection prevention?

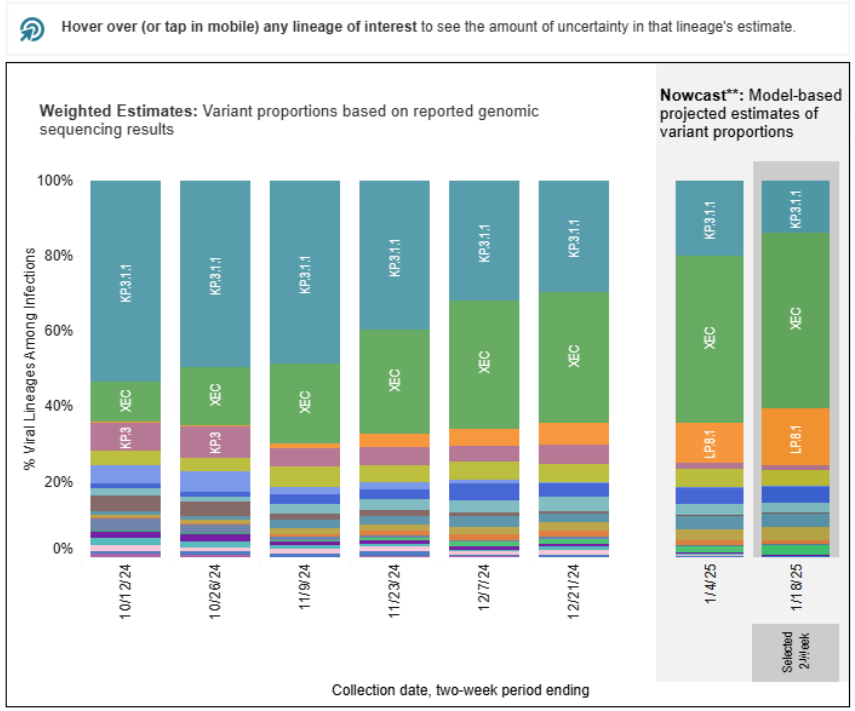


Sundermann et al, *Antimicrobial Stewardship and Healthcare Epidemiology* 2024

Predictive genomic epidemiology

Can we anticipate which [subtype / variant] will circulate in the future?

Weighted and Nowcast Estimates in United States for 2-Week Periods in 9/29/2024 – 1/18/2025



Nowcast Estimates in United States for 1/5/2025 – 1/18/2025

USA

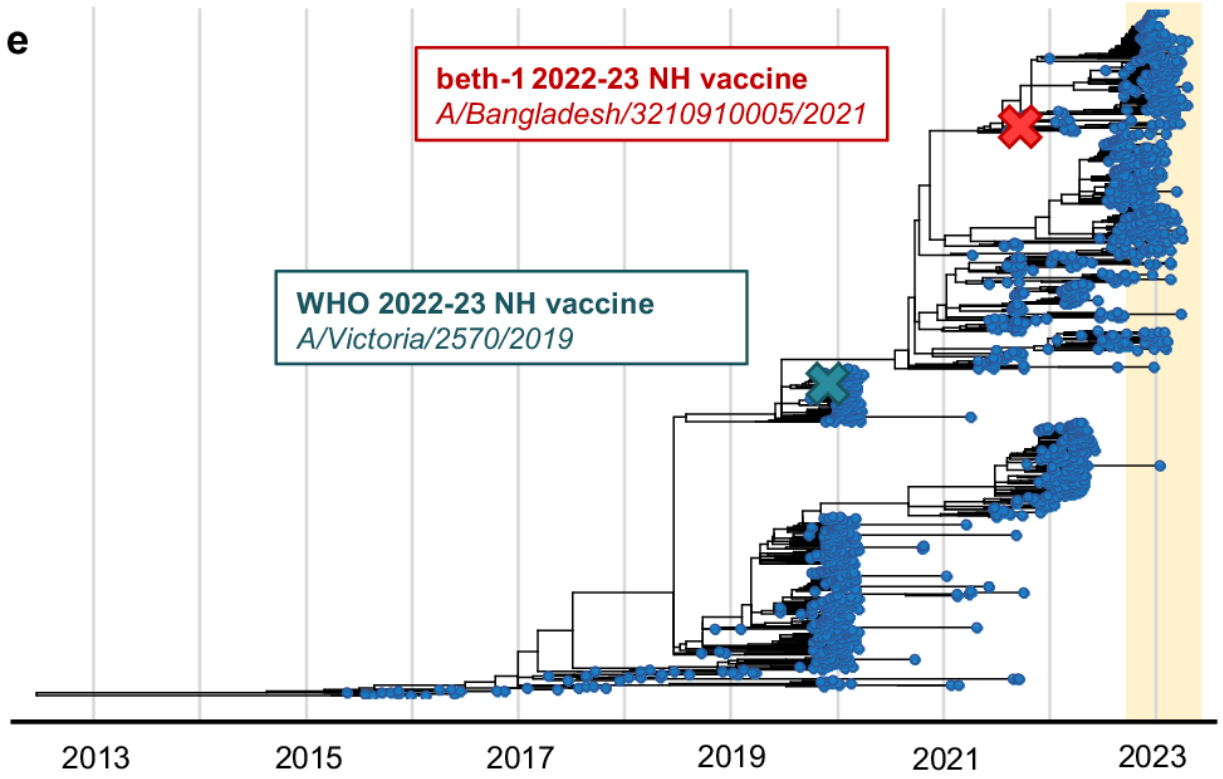
WHO label	Lineage #	%Total	95%PI
Omicron	XEC	47%	44–49%
	LP.8.1	15%	12–19%
	KP.3.1.1	14%	12–16%
	MC.10.1	5%	3–7%
	MC.1	4%	3–5%
	LF.7	4%	3–6%
	XEK	4%	3–5%
	XEC.4	3%	1–5%
	LB.1.3.1	2%	1–4%
	KP.3	1%	1–2%
	JN.1.18.6	1%	0–3%
	MC.19	1%	0–1%
	KP.1.1.3	0%	NA
	KP.2.3	0%	NA
	LB.1	0%	NA
	JN.1	0%	NA
	KP.2	0%	NA
	LP.1	0%	NA

** These data include Nowcast estimates, which are modeled projections that may differ from weighted estimates generated at later dates
# Enumerated lineages are US VOC and lineages circulating above 1% nationally in at least one 2-week period. "Other" represents the aggregation of lineages which are circulating <1% nationally during all 2-week periods displayed. While all lineages are tracke by CDC, those named lineages not enumerated in this graphic are aggregated with their parent lineages, based on Pango lineage definitions, described in more detail here: https://web.archive.org/web/20240116214031/https://www.pango.network/the-pango-nomenclature-system/statement-of-nomenclature-rules.

CDC Nowcast, January 2025

Predictive genomic epidemiology

How can we optimize the selection of influenza vaccine strains with genomics?



"Our model is better than yours…"

Lou et al, *Nature Communications* 2024

From classes to applications of genomic epidemiology

Which disease surveillance programs in your health jurisdictions use microbial genomic data?

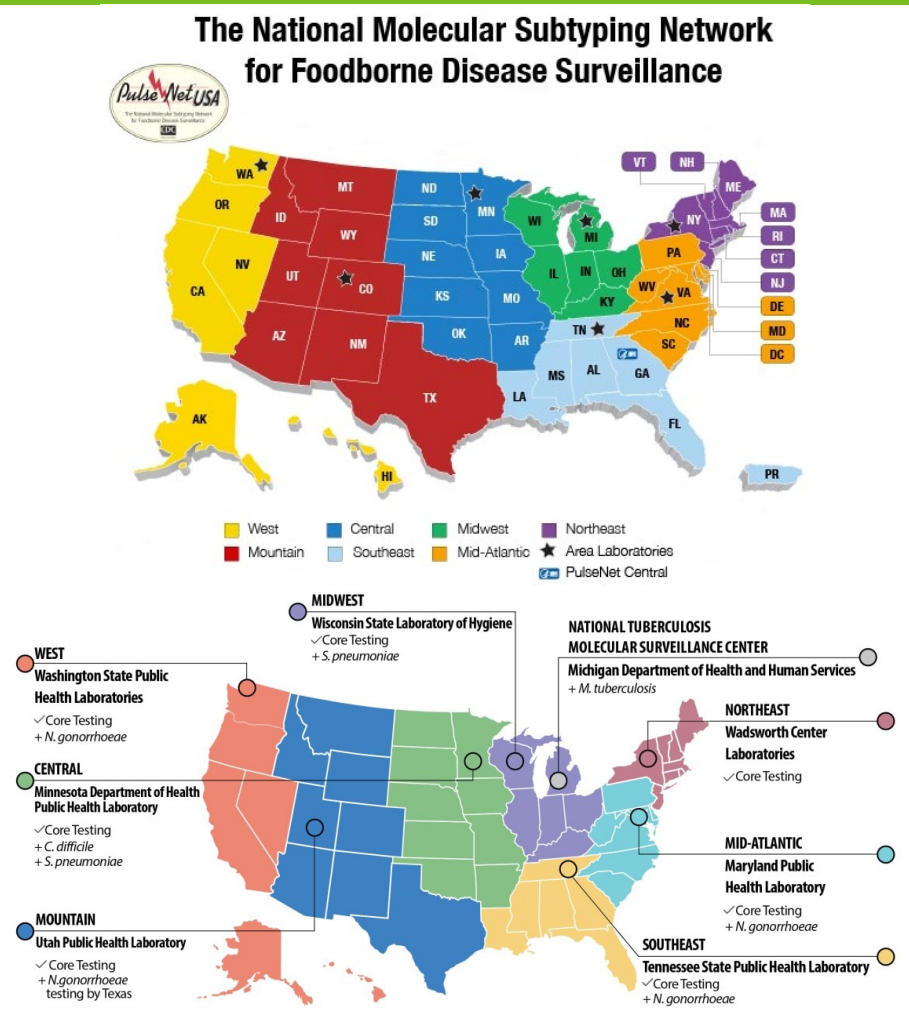
What are some regional or national surveillance initiatives in which you are involved, that use pathogen genomics?

Do these programs ask for and use genomic data in the same ways, for the same purposes?

From classes to applications of genomic epidemiology

Use cases for genomic data in infectious disease surveillance will vary, based on several key factors:

- Pathogen species and infection type
- Reportable/notifiable disease rules
- Existing surveillance programs and their bureaucracies
- Perceived utility of molecular analyses
- Availability of contextual genomic data



CDC: PulseNet and Antimicrobial Resistance Laboratory Network

From classes to applications of genomic epidemiology

Most "standard" genomic surveillance and investigation in STLT health jurisdictions will fall into one of three applications:

1. An "internal" genomic surveillance program that routinely uses genomic data to detect or investigate events of concern (outbreaks, new variants, etc.).
2. A multi-jurisdictional surveillance program that routinely uses genomic data to characterize pathogens and investigate outbreaks.
3. A disease surveillance program (either internal or multi-jurisdictional) that does not routinely use genomic data to characterize or investigate.

Small group discussion with your neighbors

Divide into small groups and choose one of the three described "applications" of pathogen genomic surveillance.

1. Describe a surveillance system or example investigation that aligns with the application being discussed (routine "local", routine multi-jurisdictional, non-routine).
2. Which of the five class(es) of genomic epi (descriptive, detective, investigative, analytical, and predictive) does your example fall into?
3. What microbiological and epidemiological data would likely be available before performing any sequencing and genomic analysis?
4. What epidemiological question(s) could only be answered by sequencing, characterizing, and comparing pathogen genomes?
5. What genomic results will help answer those key epidemiological questions in #4?

Key skills in genomic epidemiology

Technical Skills

Molecular Biology

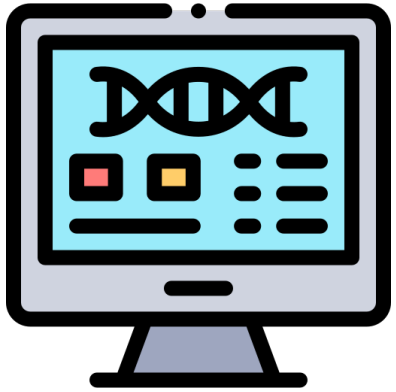
- Foundations of molecular biology and nucleic acid chemistry
- Evolutionary genomics and phylogenetics
- Interpreting phylogenetic trees
- Subtypes and variants of microbial pathogens
- Mechanisms of virulence and antimicrobial resistance

Epidemiology

- Data collection and investigation
- Design of cohort studies and sampling methods
- Using disease surveillance systems and data analytics software
- Importing and exporting case data Reducing sampling biases
- Biostatistics (e.g. significance tests, power analyses, confounding, etc.)

Genomic Data Analysis

- Limitations of genomic data types
- Figures and visualizations of genomic / molecular epi data
- Measurements of genetic similarity
- Pathogen-specific interpretations of genomics results
- Cleaning and QC'ing genomic data
- Subsampling genomic datasets



Professional Skills

Communication

- "Translating" jargon between laboratory and epidemiology
- Tailoring documents and visuals for different audiences
- Explaining genomics principles to lay audiences
- Scientific writing
- Oral presentation skills

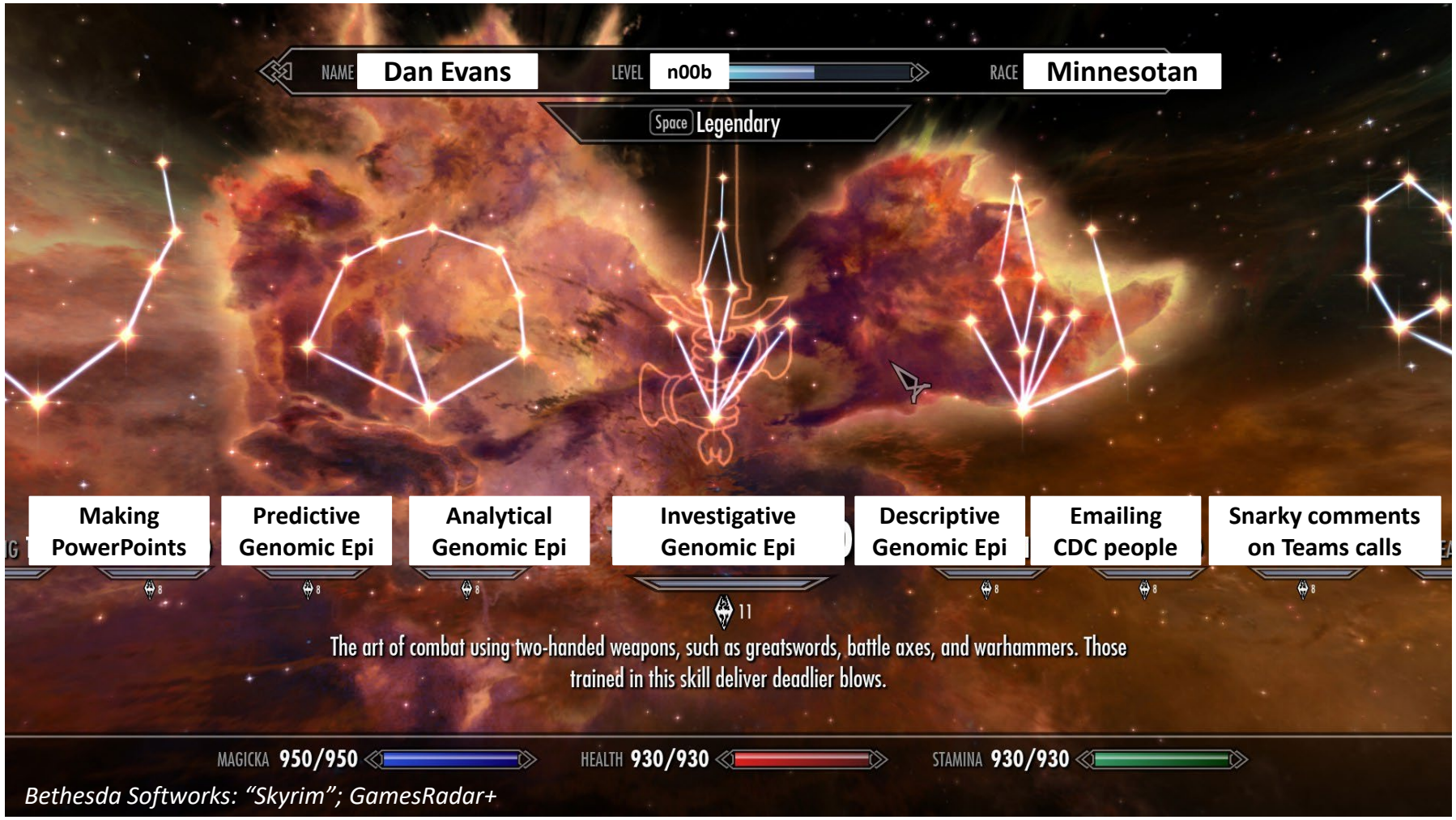
Collaboration

- Identifying different needs of stakeholders and funding sources
- Finding common ground between different teams
- Active listening and back-briefing
- Resolving disagreements between teams and team members
- Systems thinking

Problem Solving

- Critical thinking skills
- Anticipating barriers, limitations, and challenges
- Time management and organizational planning
- Creative troubleshooting
- Proactivity

Just like every other public health discipline, genomic epidemiology has “skill trees”



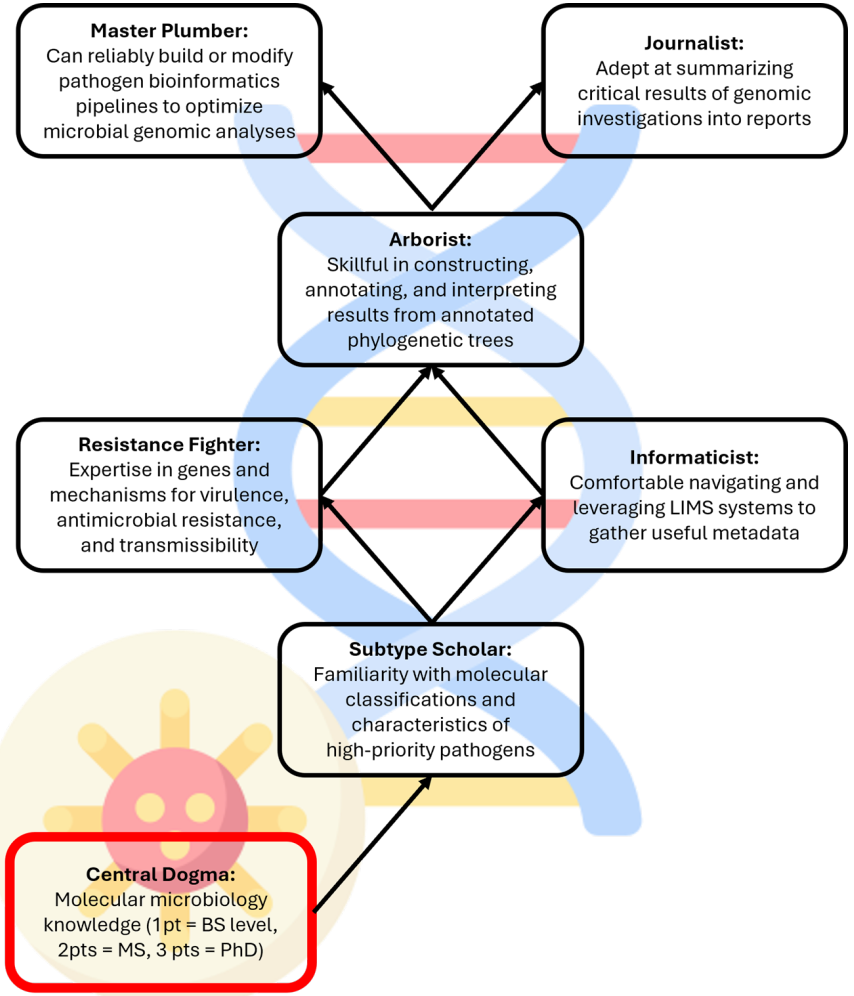
Bethesda Softworks: “Skyrim”; GamesRadar+

Let's build our genomic epi "character"!

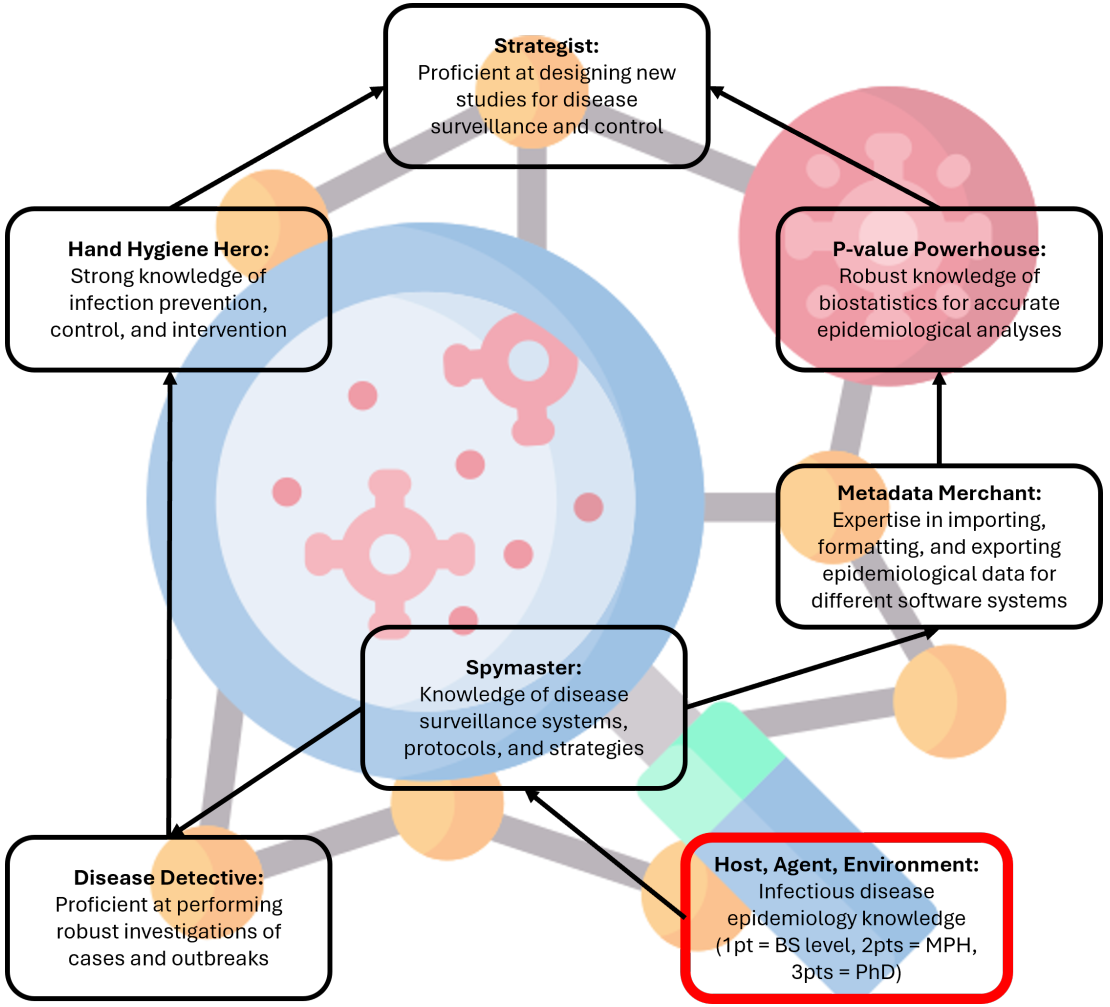
Using three "skill trees" (lab, epi, and professional skills), build a genomic epidemiologist that would best meet the needs of your health jurisdiction.

- You start the exercise with ***a limited number of points*** – 10 points for a newly created position or early-career genomic epi, or 15 points for an experienced one.
- Every skill costs (at least) 1 point to purchase. ***You can't pick all of them!***
- ***Start at the bottom of each skill tree and build upward***. You cannot "skip" a skill to get to one that is higher up on the same "branch".
- Build trees based on what your jurisdiction ***NEEDS and what it CAN HAVE***, not just what you think is the "perfect" genomic epi.

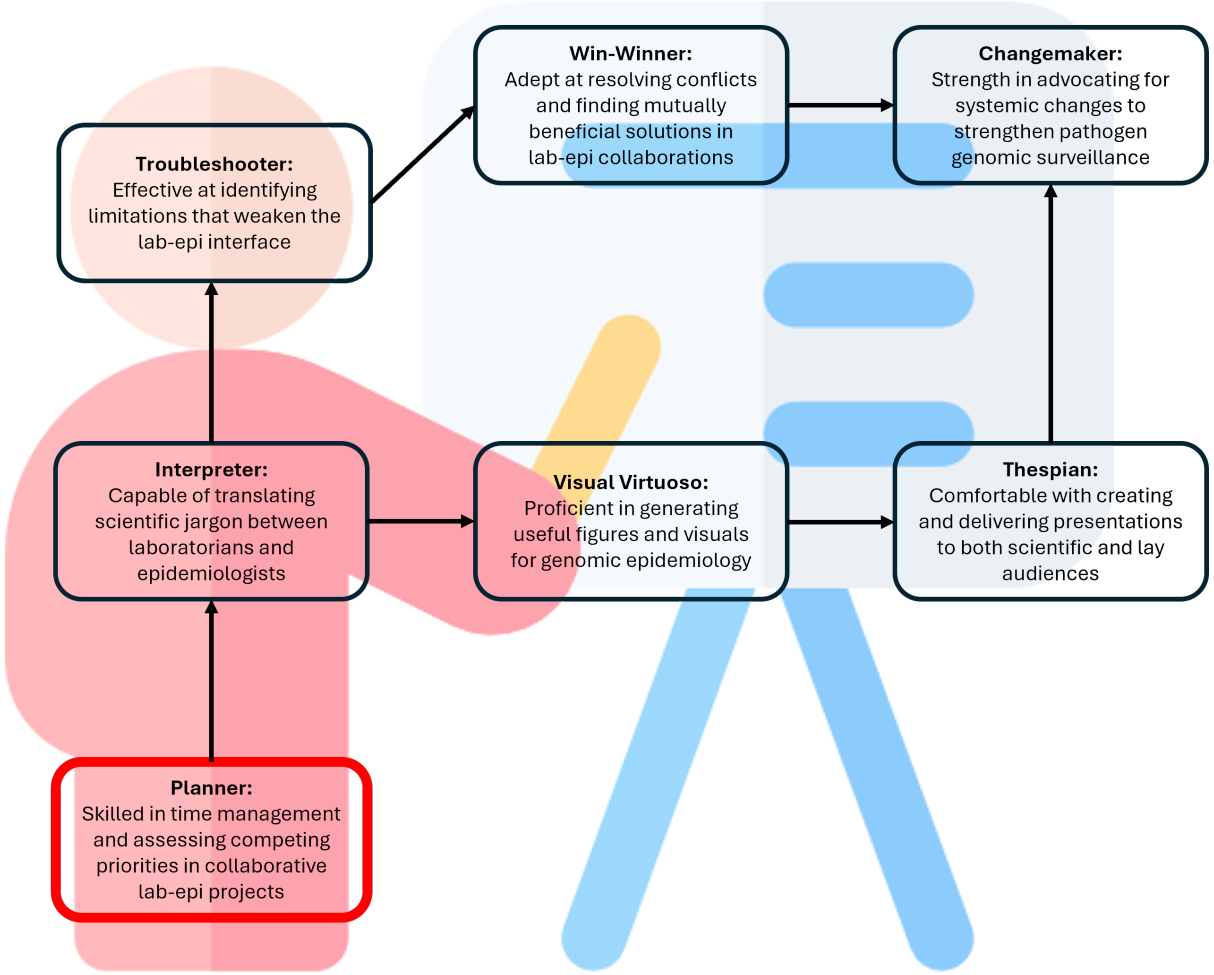
Genomics Skill Tree



Epidemiology Skill Tree



Professional Skill Tree



Debriefing questions for open discussion

1. Did you dump skill points into one tree, or spread them across all three?
2. Which skills did you (not) choose for your genomic epi “character”?
3. What skills did you want your genomic epi to have, but could not fit into your limited “budget”?
4. What skills that were “lower” on the trees’ “branches” did you need to select, to unlock the ones you wanted?
5. If other programs in your health jurisdiction built a skill tree for their “own” genomic epi, how different would it look from yours?

APHL's Genomic Epidemiologist Recruitment Guide

Public Health Laboratory Recruitment Guide:

Genomic Epidemiologist

APHL™

www.aphl.org

Summary

The genomic epidemiologist, often called molecular epidemiologists, will be responsible for leading program initiatives focusing on the integration of genomic sequencing data and epidemiologic metadata to better understand disease transmission and assist with pathogen surveillance and outbreak investigations. The genomic epidemiologist will apply a wide understanding of next generation sequencing (NGS) methodologies, pathogen genomics, data analysis and integration to generate reports and visualizations and disseminate results to relevant partners. There are various uses for sequencing data, including surveillance, outbreak investigation, and clinical diagnostics, which involve careful interpretation by genomic epidemiologists. Depending on the specific sequencing purpose, investigative questions asked, or the epidemiological application of genomic interpretations, genomic epidemiologists will work and share results with a wide variety of partners.

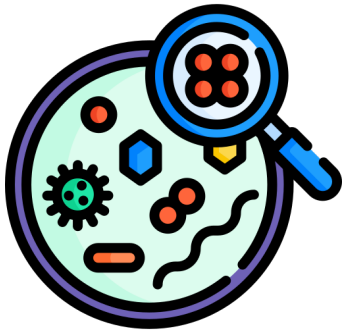
Bottom Line

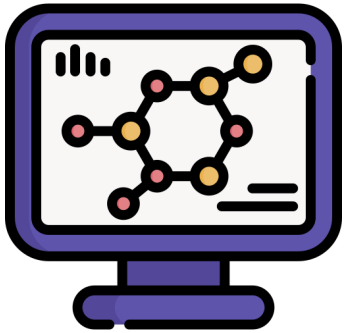
"Genomic epidemiology" broadly describes many tasks related to disease surveillance.

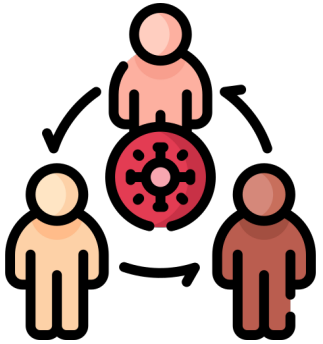
The practice of genomic epidemiology depends on state, pathogen, disease surveillance goals, and "side of the fence."

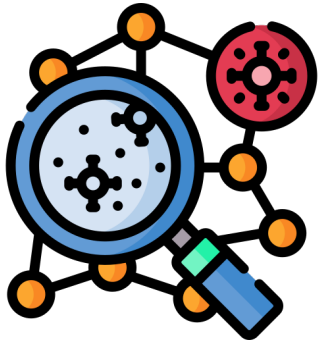
Priorities for pathogen genomic investigations will follow several general themes.

Genomic epidemiologists must develop and maintain cross-disciplinary skill sets.





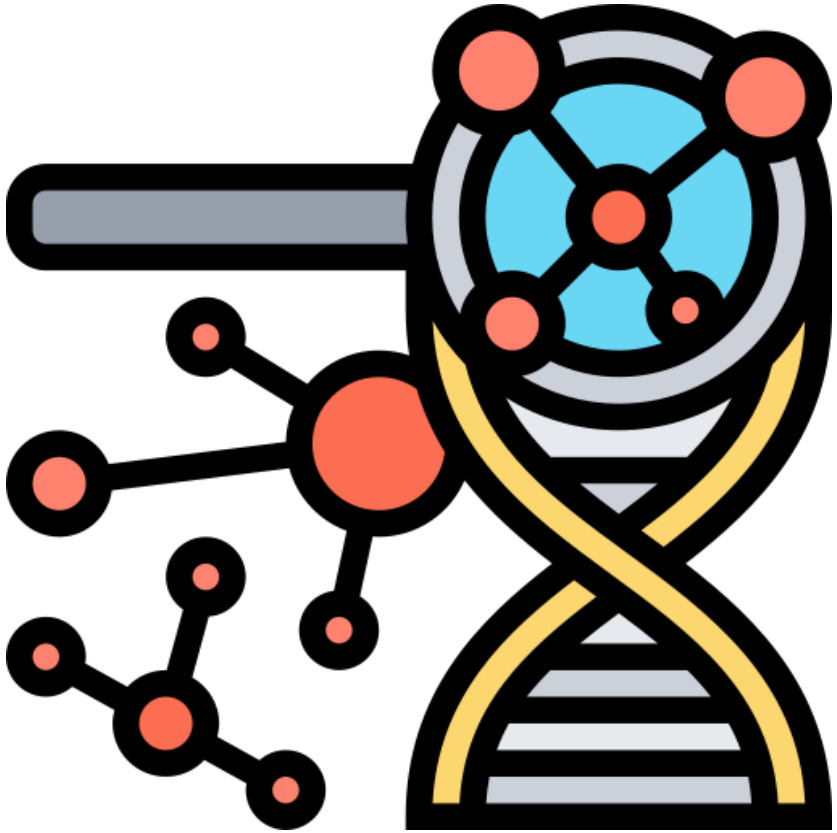




Flaticon: Freepik

Thank you for participating!

Dan.R.Evans@state.mn.us



Flaticon: Freepik