

Understanding Modern Genomic Epidemiology

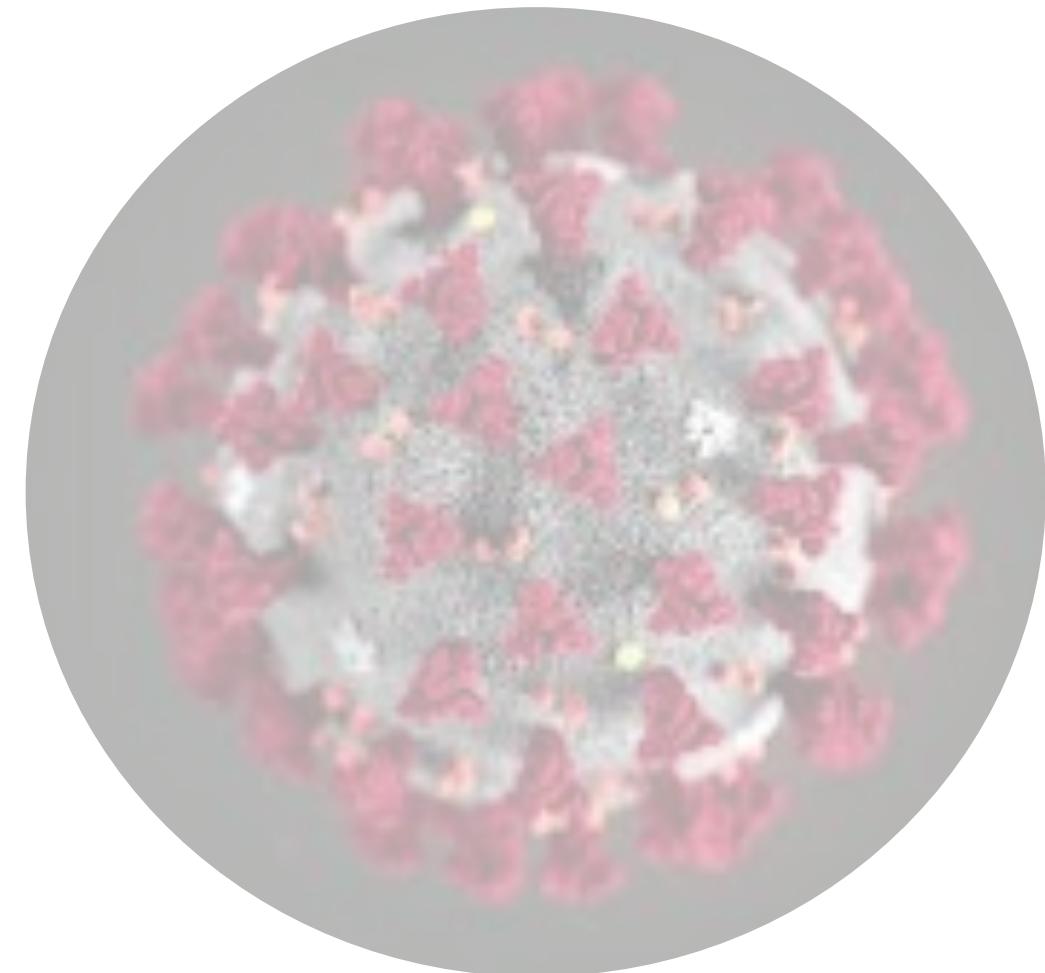
Lecture 1

JOSEPH FAUVER, PH.D.

ASSISTANT PROFESSOR

UNMC COPH DEPARTMENT OF EPIDEMIOLOGY

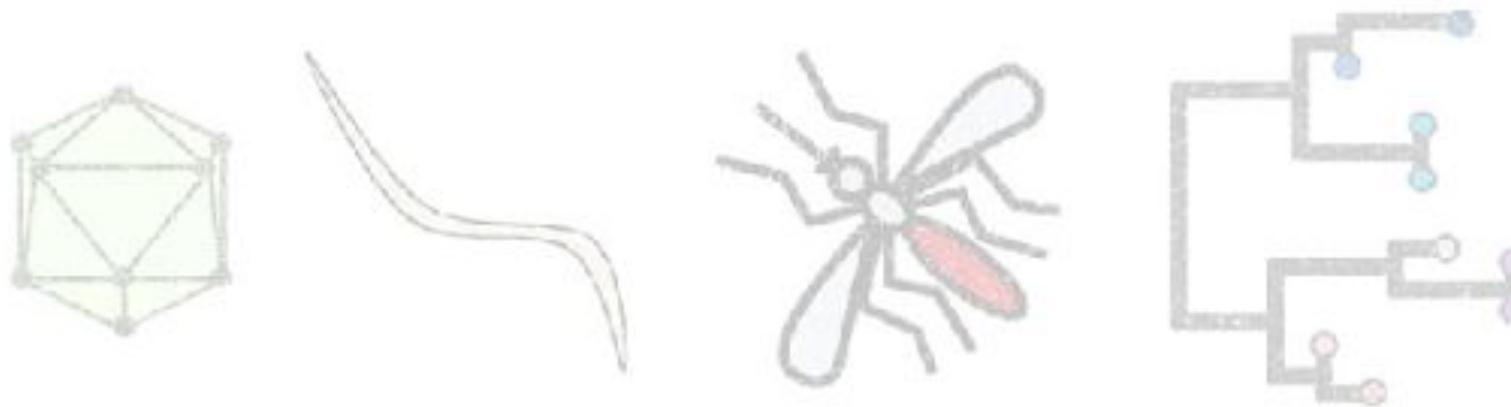
5/7/2025



Pathogens, genomics, and global health

The Fauver Lab

UNMC College of Public Health



thefauverlab.com

Global VBD and Genomics Work



What is genomic epidemiology?

What is genomic epidemiology?

Molecular epidemiology

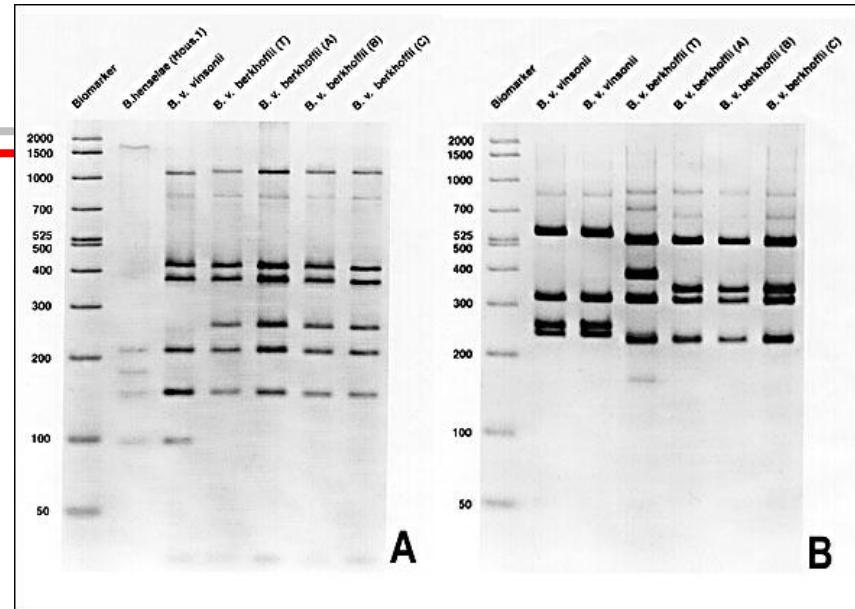
- Understanding the distribution or clustering of genetic variants, strains, serotypes, or other molecular groupings of pathogens to understand patterns of disease.

Genetic epidemiology

- Above, but with using gene or partial genome sequences.

Genomic epidemiology

- Above, but with using whole genome sequences.



Kordick & Breitschwerdt, 1998. Persistent Infection of Pets within a Household with Three *Bartonella* Species. *Emerging Infectious Diseases*



What is genomic epidemiology?

My definition: “Integrating the use of modern approaches in genomics and phylogenetics with traditional epidemiological approaches to better understand infectious disease transmission”

What need does Genomic Epidemiology Fill?

Needs:

- Earlier detection and more precise investigation of outbreaks
- More detailed knowledge of pathogen evolution and spread to inform control measures
- Rapid molecular tool development for surveillance and diagnostics
- Vaccine development

What need does Genomic Epidemiology Fill?

Needs:

- Earlier detection and more precise investigation of outbreaks
- More detailed knowledge of pathogen evolution and spread to inform control measures
- Rapid molecular tool development for surveillance and diagnostics
- Vaccine development

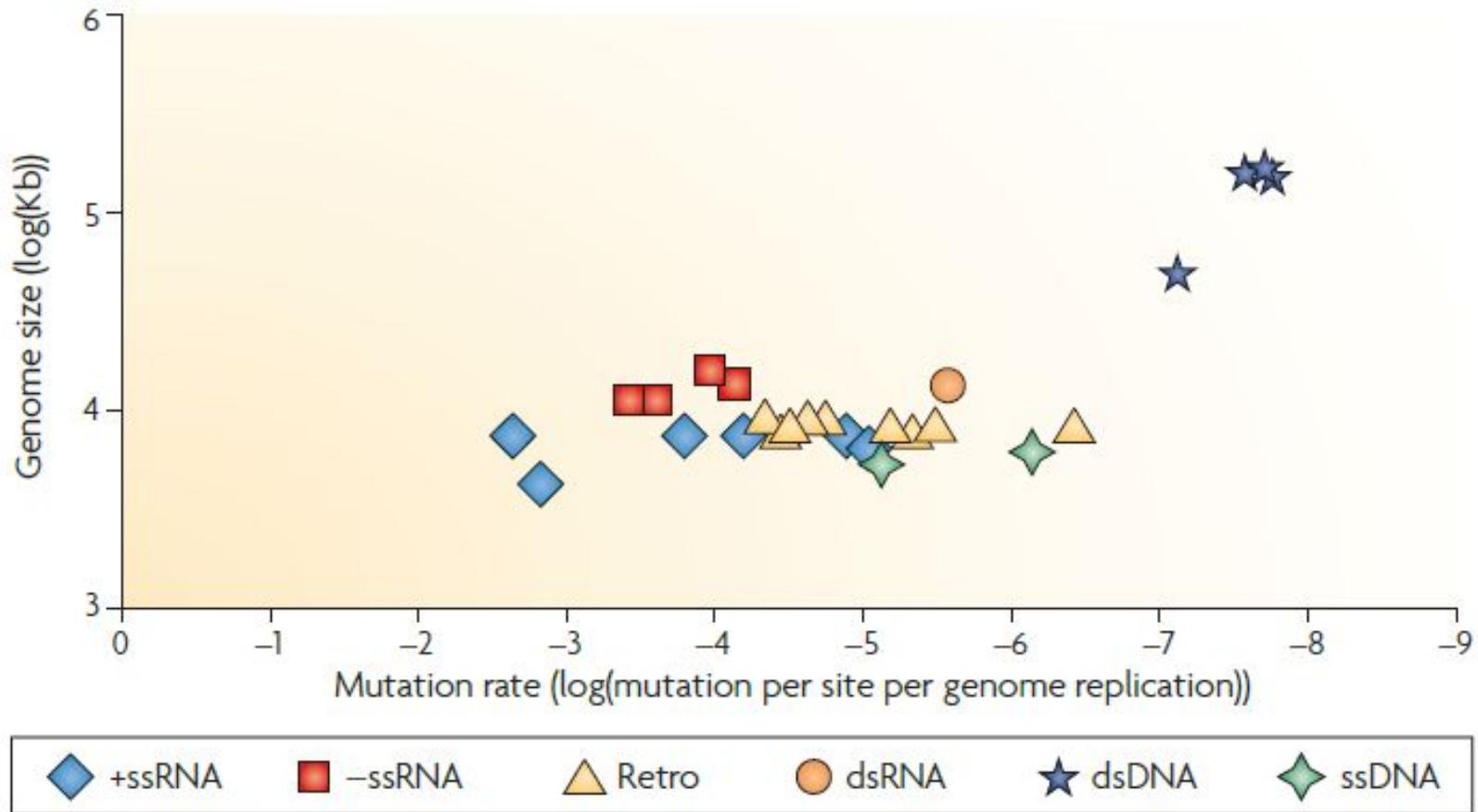
Why genomics, why now?

- Growing repositories of pathogen genomic data
- Growing knowledge of genes/mutations linked to phenotypes
- Decreasing costs of high-throughput sequencing
- Increasing availability of useful analytical tools

Why does this work?

Evolution! Everything evolves. Pathogens tend to do it quicker.

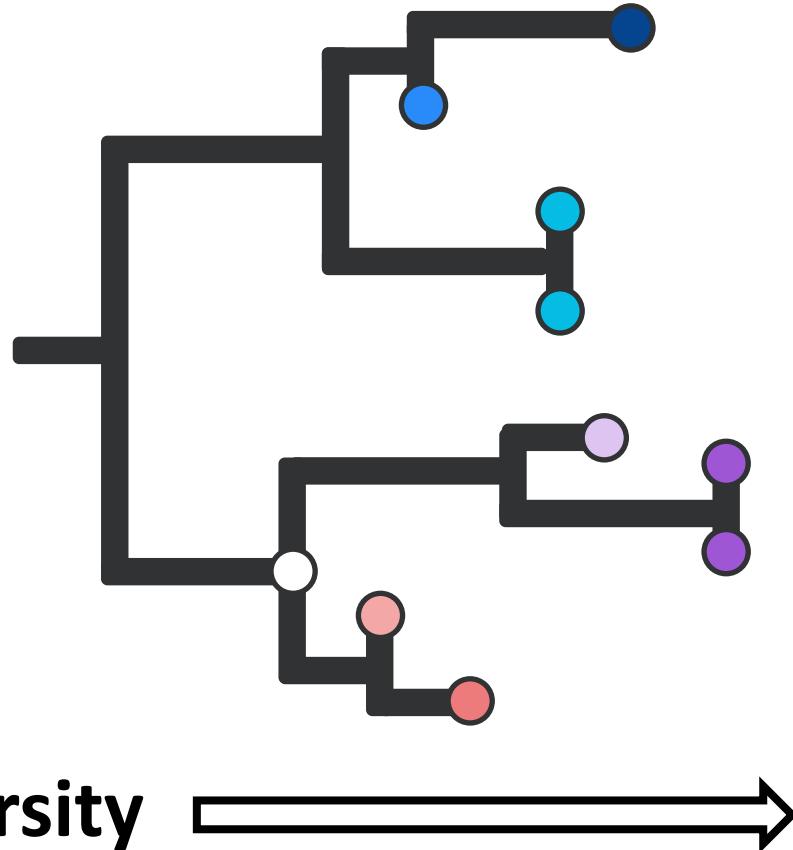
Evolution on the scale of an epidemic



Phylogenetics reconstruct evolutionary history

- Pathogens mutate!
- Infer relationships from incomplete data
- Locations and time can help infer patterns of spread
- Clusters can help identify relationships between samples

Phylogenetic Tree



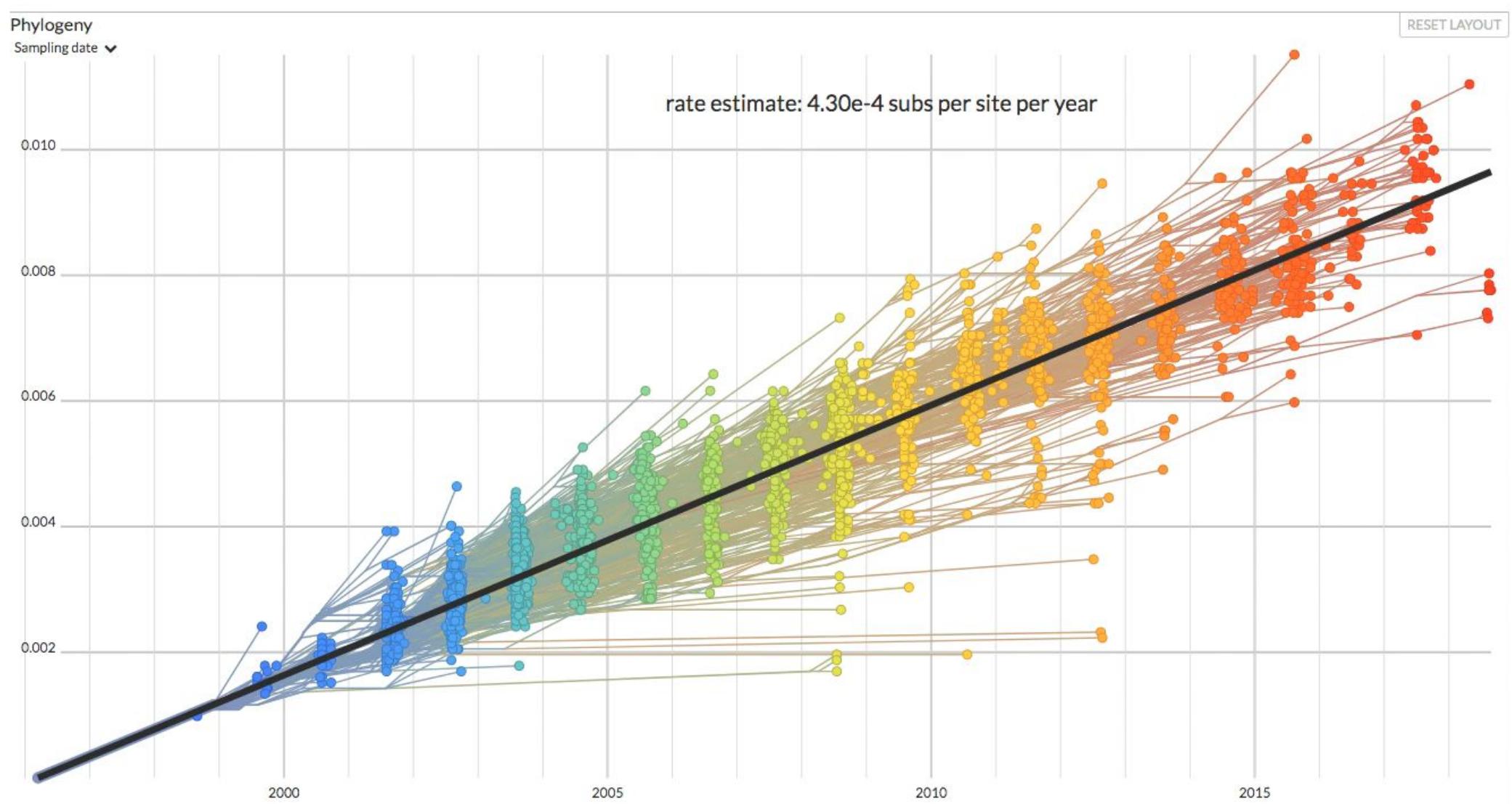
Time or Diversity



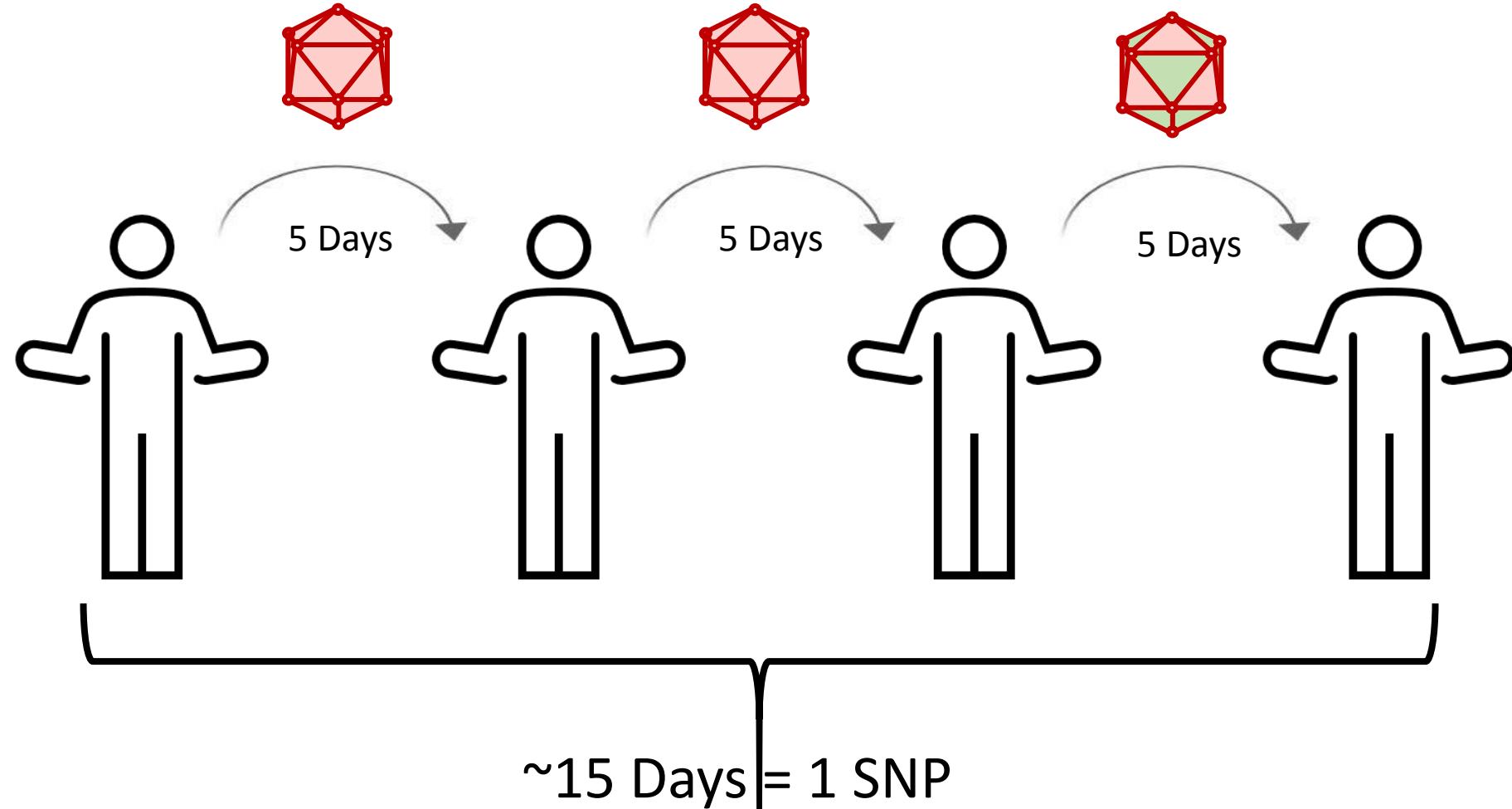
UNIVERSITY OF NEBRASKA MEDICAL CENTER™

COLLEGE OF PUBLIC HEALTH

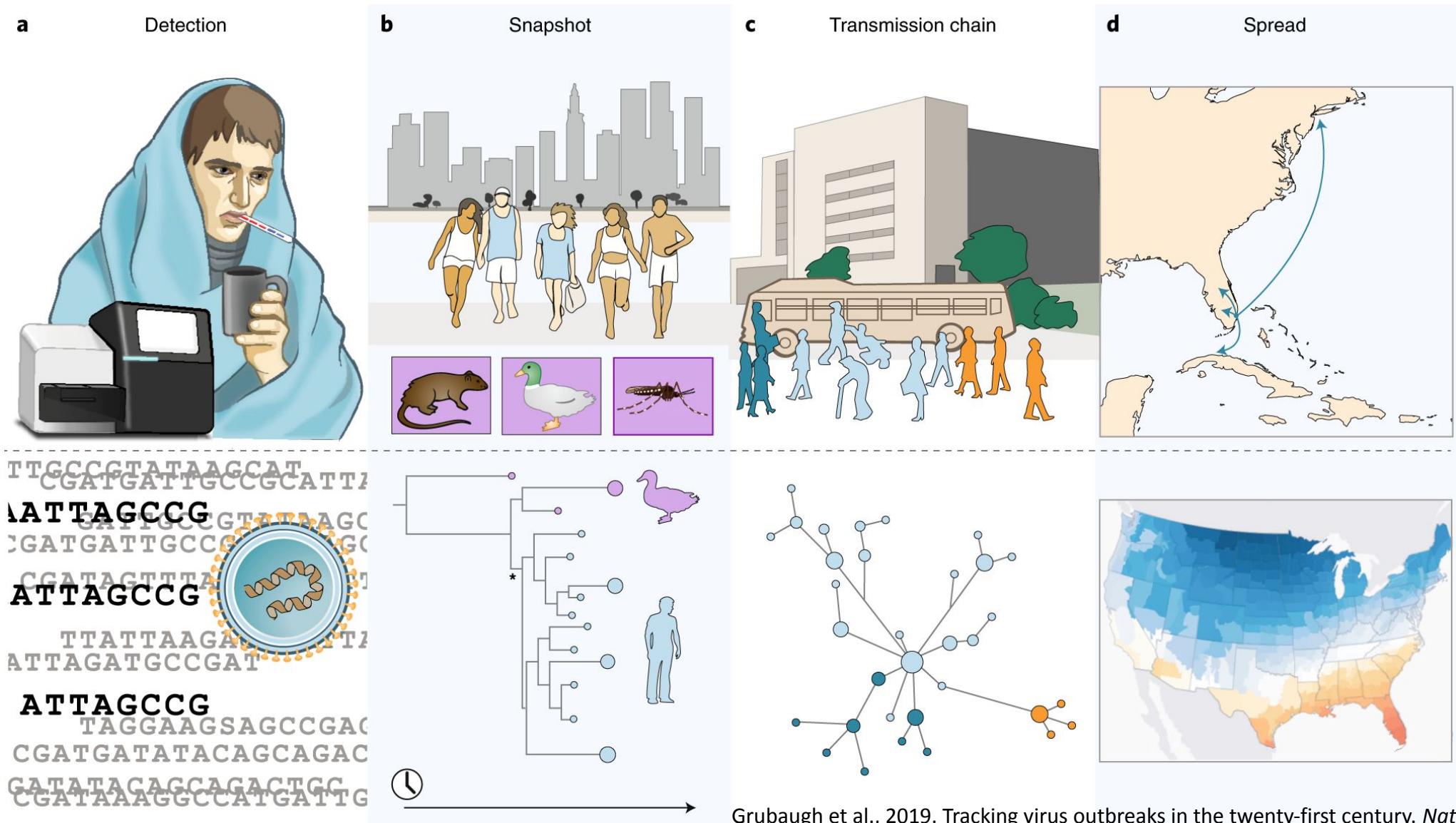
Molecular Clock



Mutation rates inform transmission reconstruction



Applications in Public Health



Applications in Public Health



Genomics as a means to detect a pathogen

- Detection of MRSA isolates associated with an outbreak in a neonatal intensive care unit (known pathogen/phenotype)
- Identification of chikungunya virus as an unrealized cause of pediatric meningitis in Bangladesh (known pathogen/unknown disease [zebras...])
- Discovery of a novel coronavirus associated with a cluster of pneumonia cases (n=59) linked to a wildlife market in Wuhan, China (unknown pathogen) (wrote this Jan 2020)

Genomics in the early days of SARS-CoV-2



Published Date: 2019-12-30 23:59:00

Subject: PRO/AH/EDR> Undiagnosed pneumonia - China (HU): RFI

Archive Number: 20191230.6864153

UNDIAGNOSED PNEUMONIA - CHINA (HUBEI): REQUEST FOR INFORMATION

Novel 2019 coronavirus genome

SARS-CoV-2 coronavirus



edward_holmes

10th January 2020

This posting is communicated by Edward C. Holmes, University of Sydney consortium led by Professor Yong-Zhen Zhang, Fudan University, Shanghai



Pneumonia of unknown cause – China

Disease outbreak news
5 January 2020



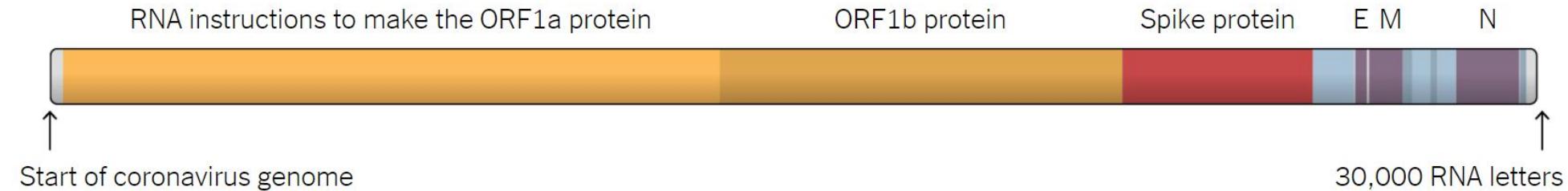
World Health Organization



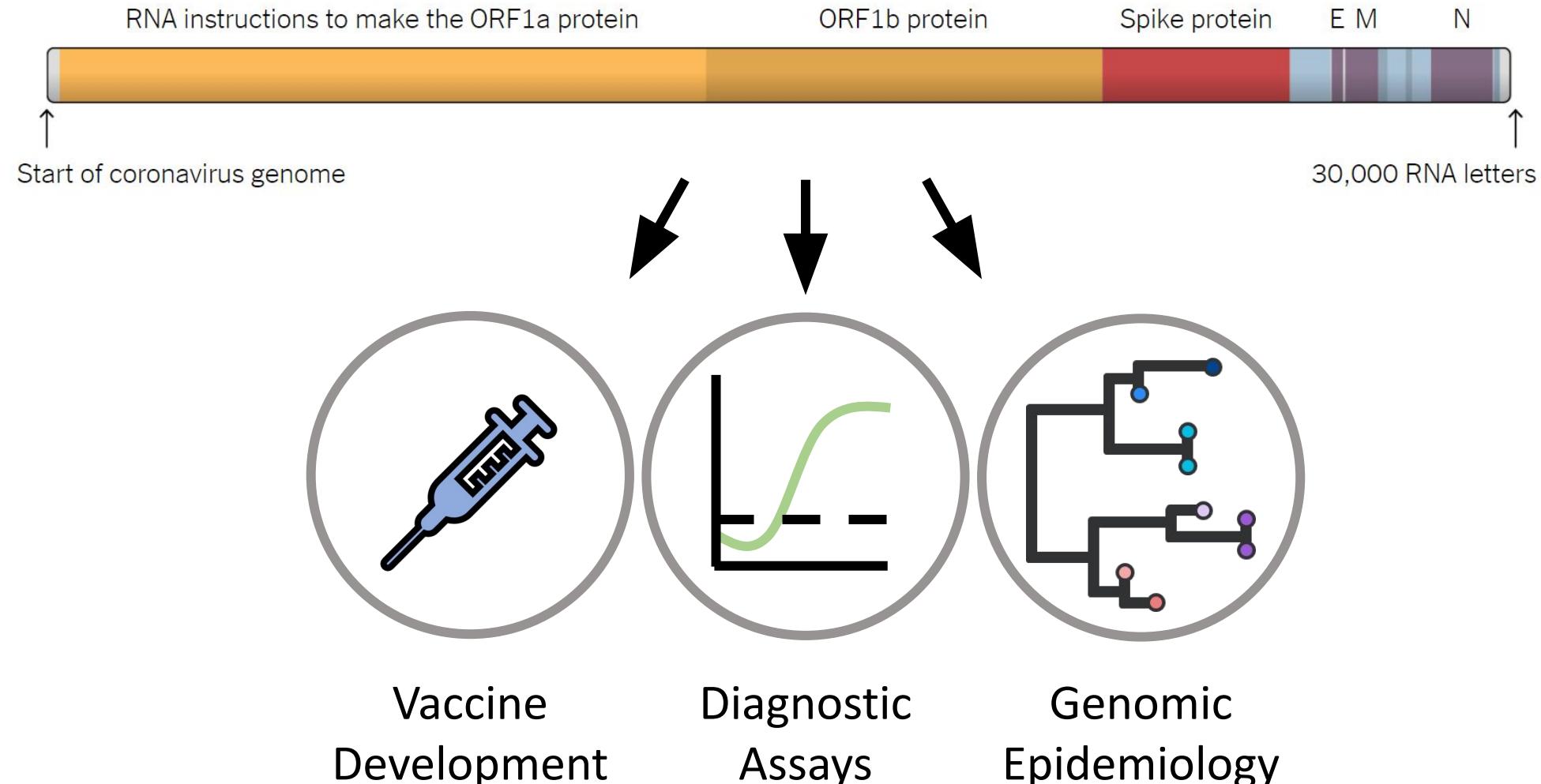
UNIVERSITY OF NEBRASKA MEDICAL CENTER™

COLLEGE OF PUBLIC HEALTH

Genomics in the early days of SARS-CoV-2



Genomics in the early days of SARS-CoV-2

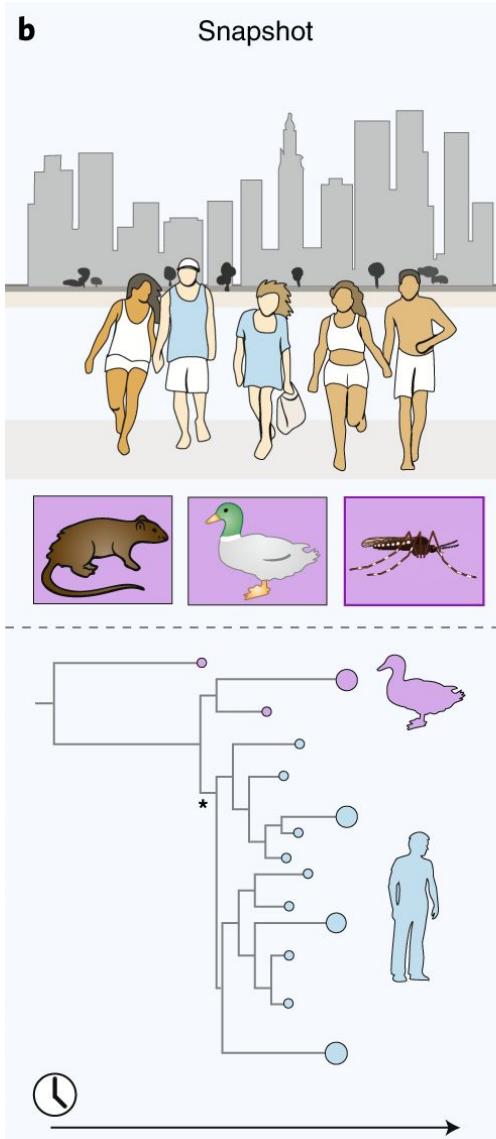


Vaccine
Development

Diagnostic
Assays

Genomic
Epidemiology

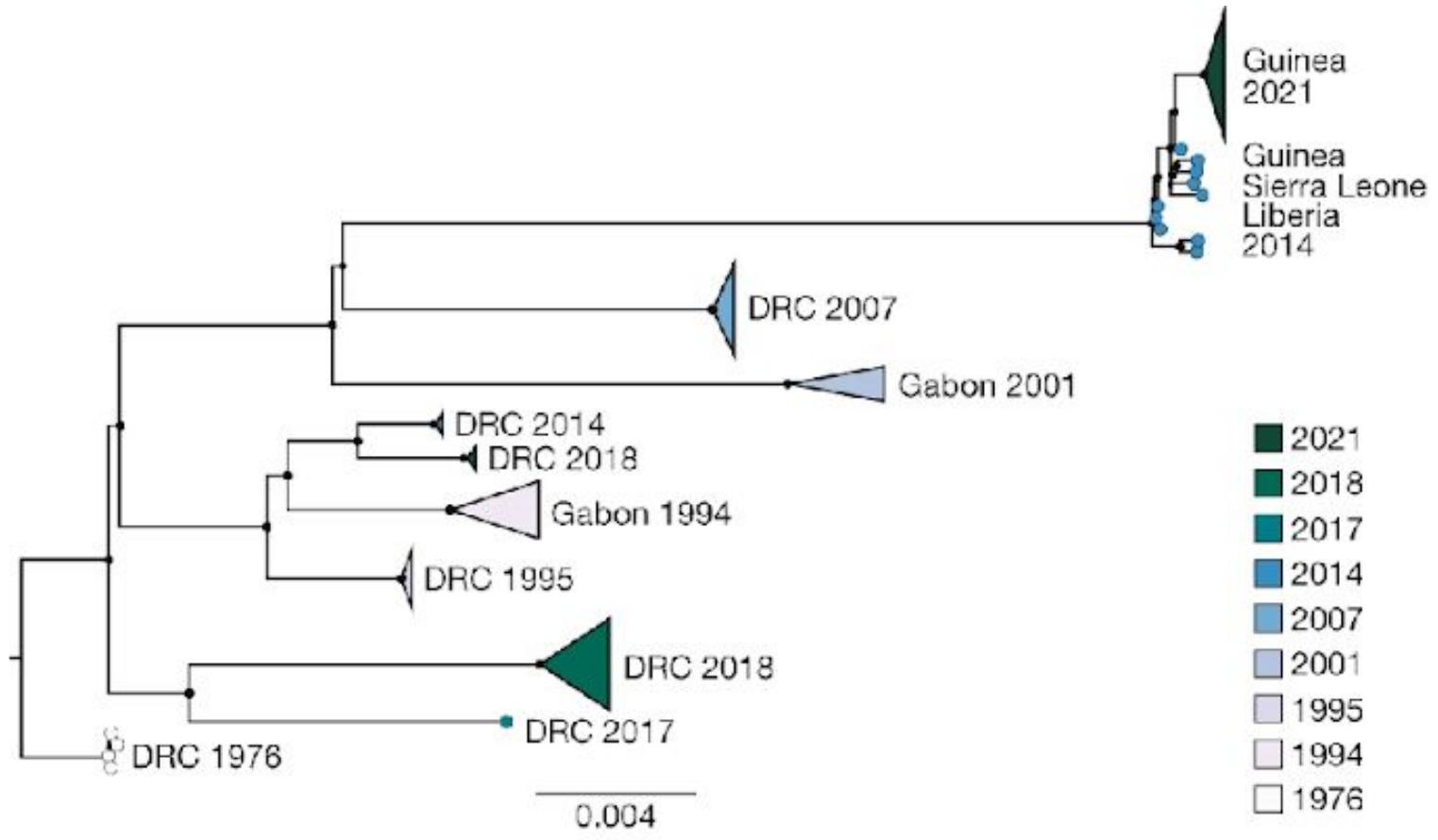
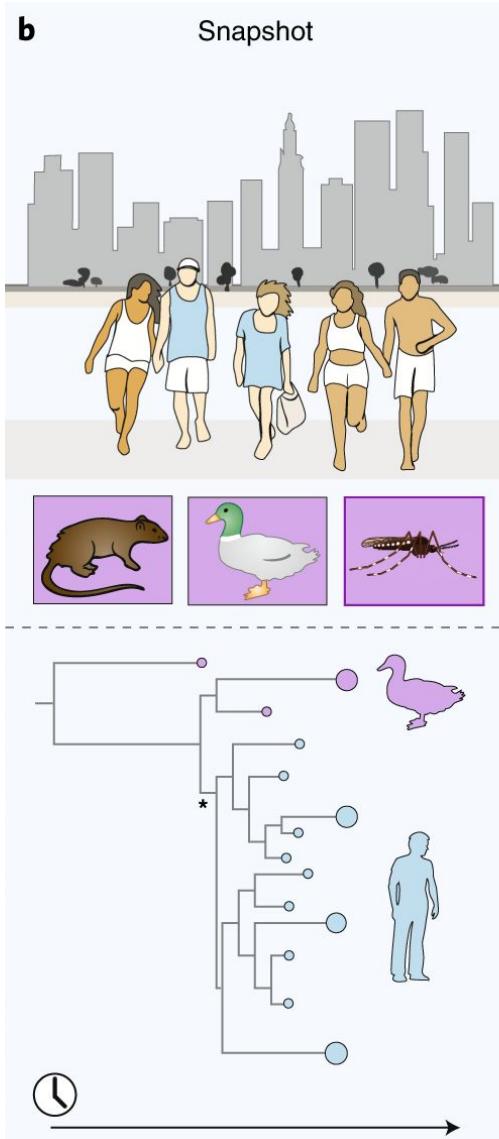
Applications in Public Health



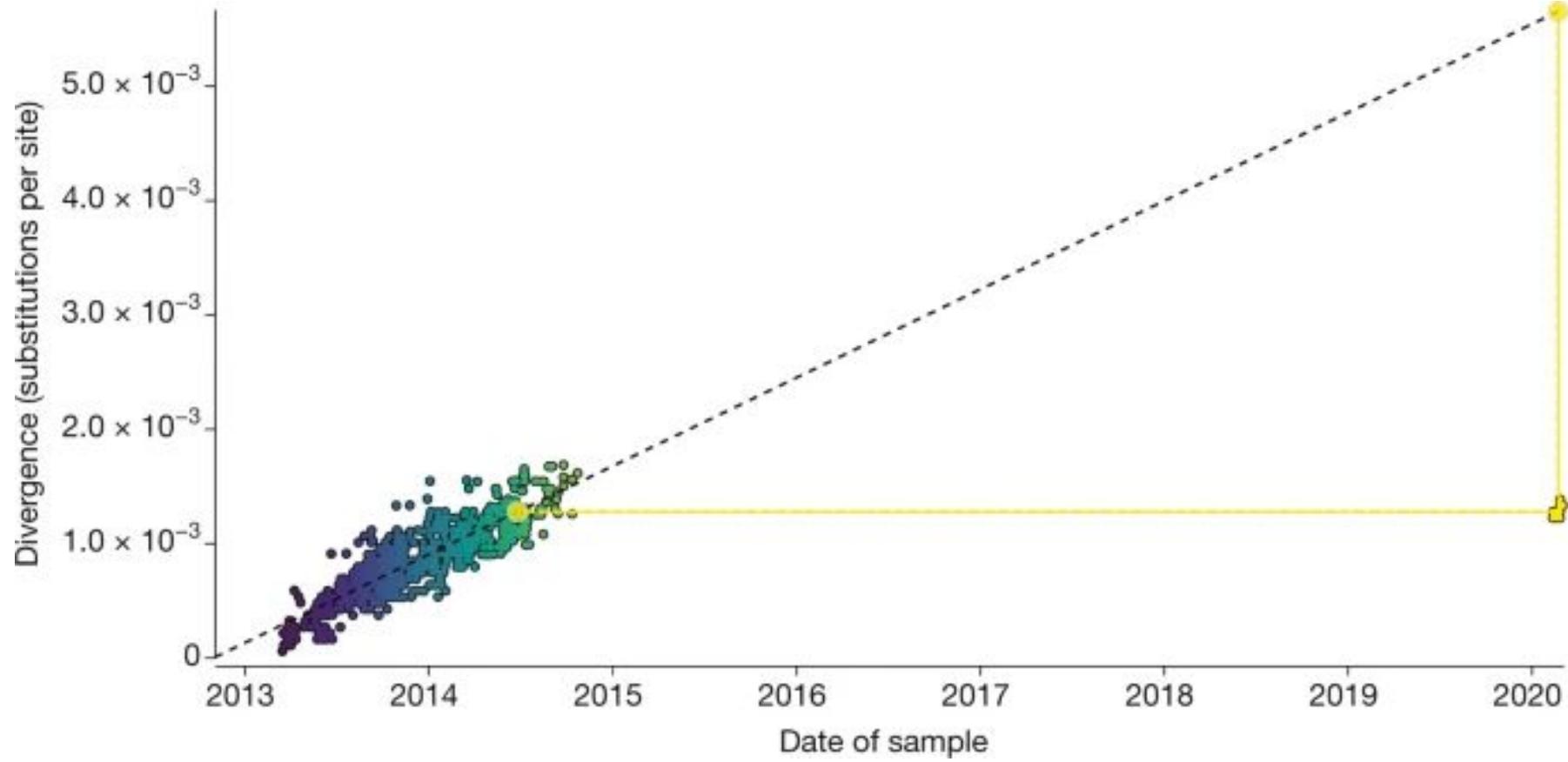
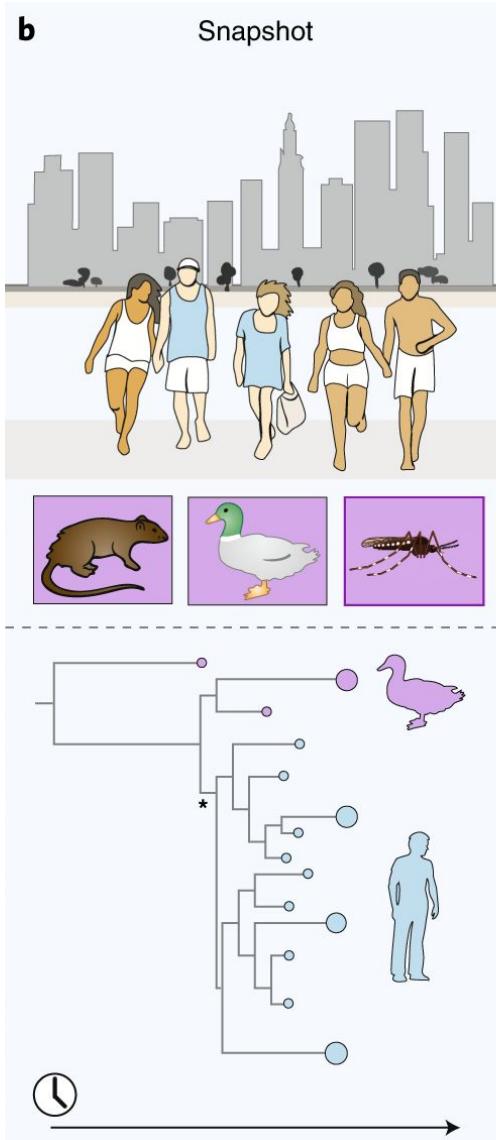
To understand a specific point in time

- Ebola relapse post-mAb treatment linked to a latent virus infection (unknown epi connection)
- Lassa fever cases linked to independent spillovers from animal reservoir, not human-to-human transmission (unknown transmission pattern)
- Congenital Zika syndrome cases in Angola linked to a virus introduction from Brazil (unknown pattern of spread)

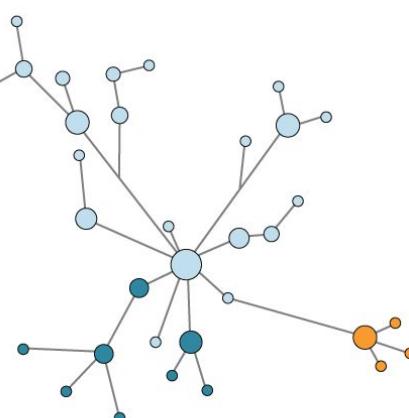
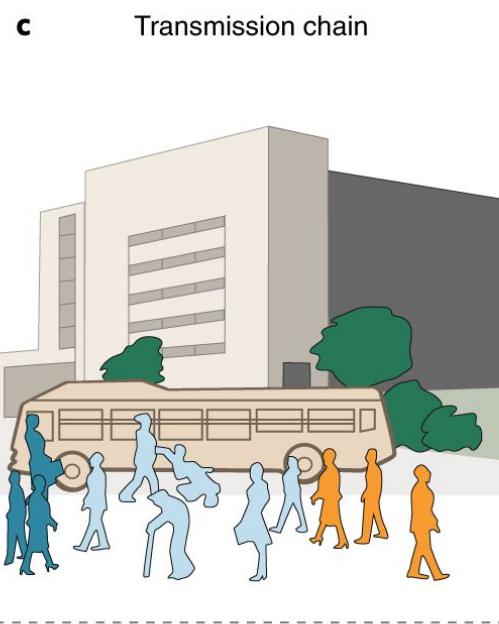
Applications in Public Health



Applications in Public Health



Applications in Public Health



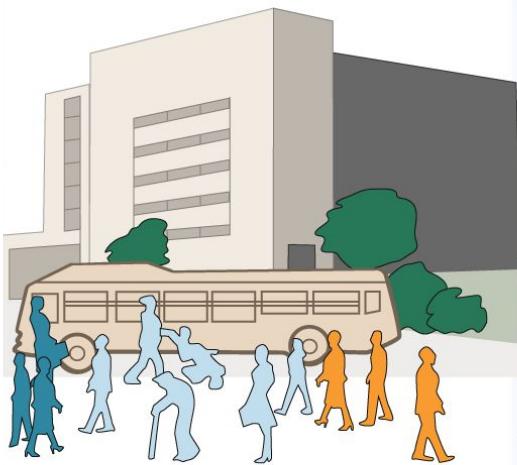
Reconstruct transmission chains and supplement contact tracing

- Mapping tuberculosis strain transmission to uncover the drivers of the MDR tuberculosis epidemic in Shanghai, China
- Near real-time monitoring of HIV transmission clusters to support public health responses in BC, Canada
- Understanding spread of SARS-CoV-2 at small scales

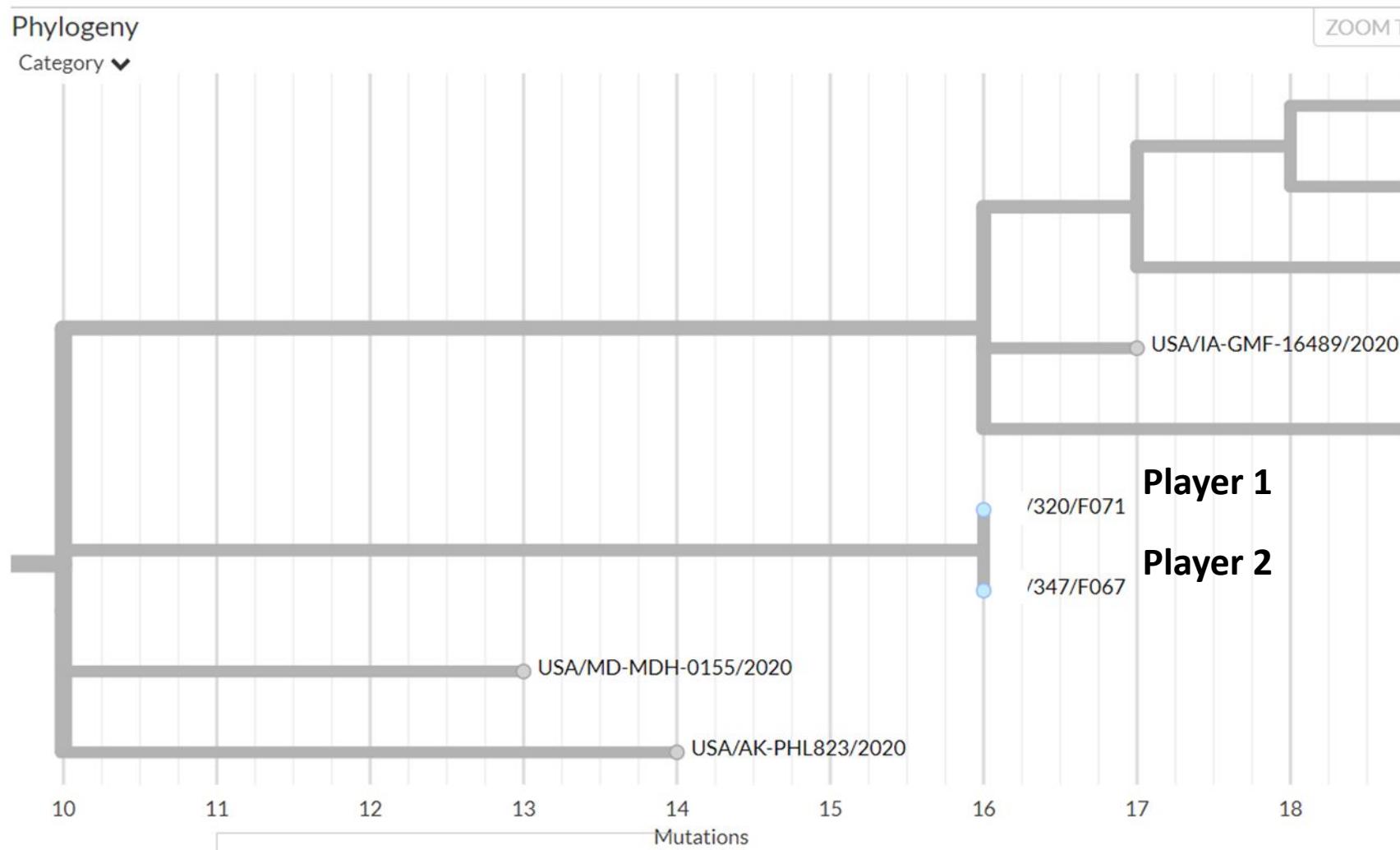
Applications in Public Health

c

Transmission chain

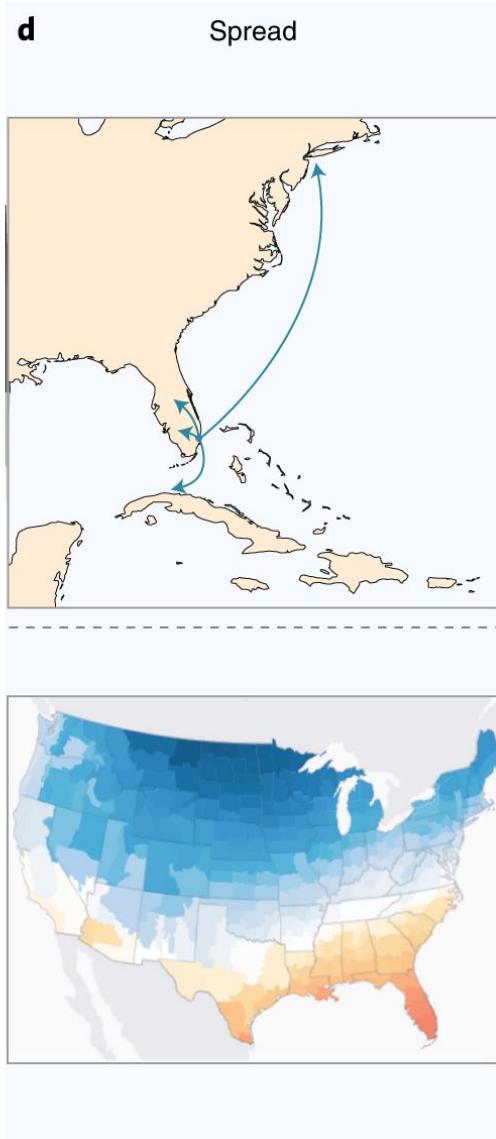


Supplementing Contact Tracing



- Plausible Epi Connection?
- Identical sequences?
- Plausible generation time?

Applications in Public Health

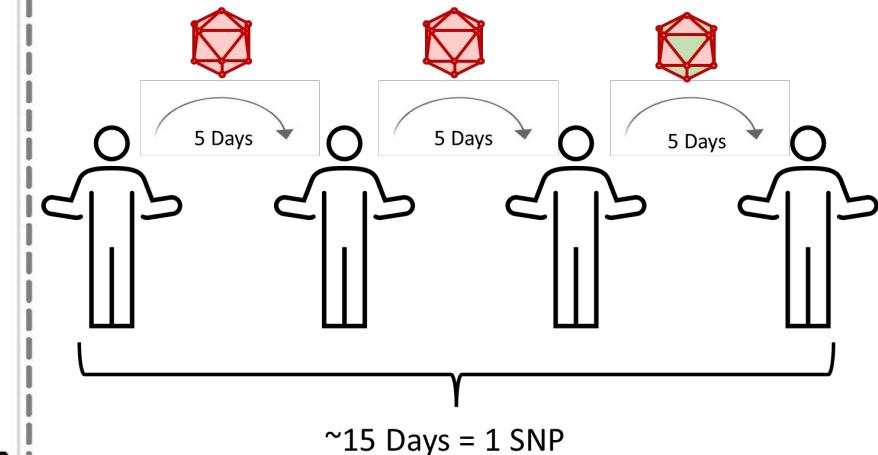
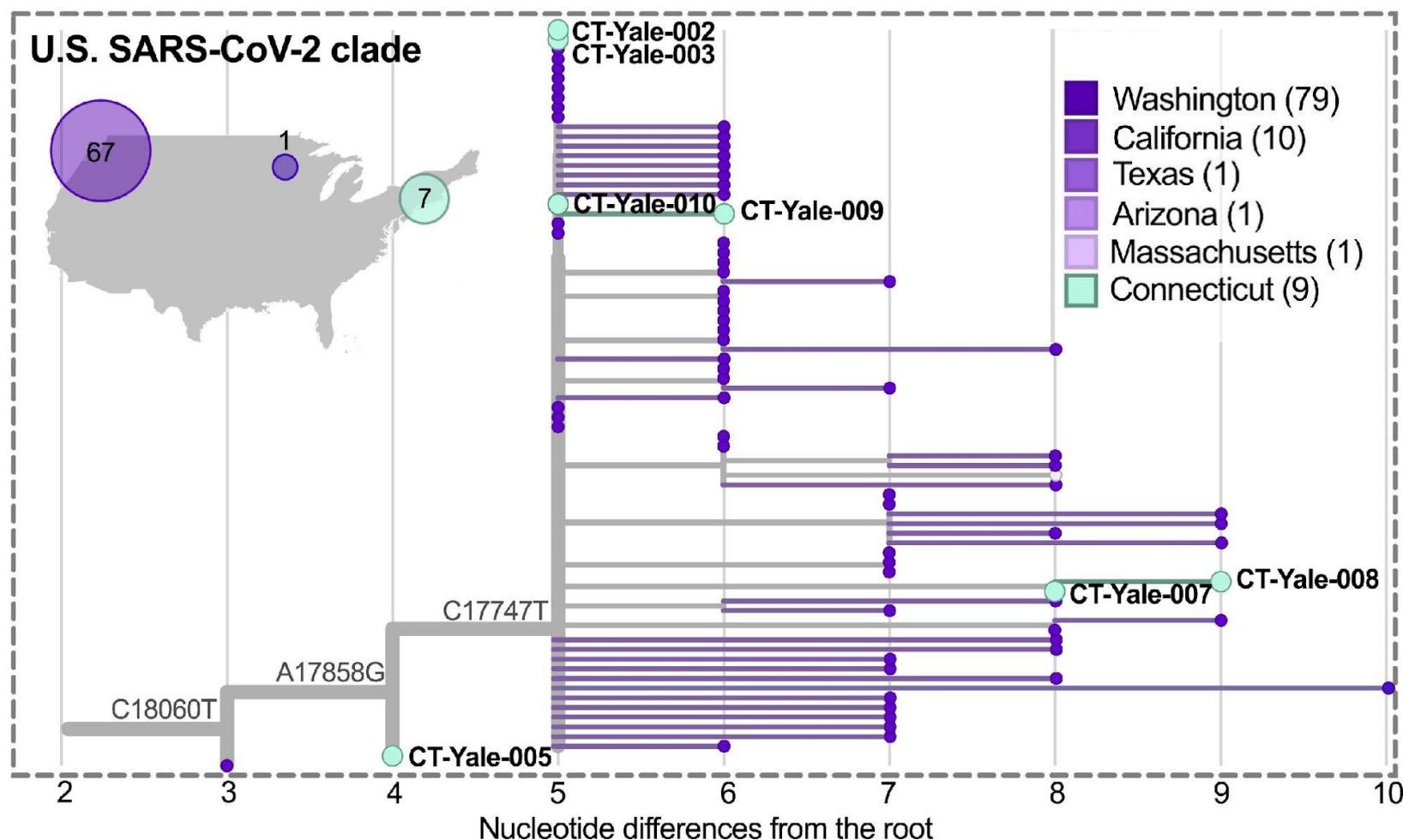


Mapping the spread of pathogens

- Mapping the global evolution and spread of artemisinin resistant malaria
- Determining the number of introductions of Zika into the U.S.
- Uncovering the epidemiological factors that spread and sustained the West Africa Ebola epidemic

Applications in Public Health

B



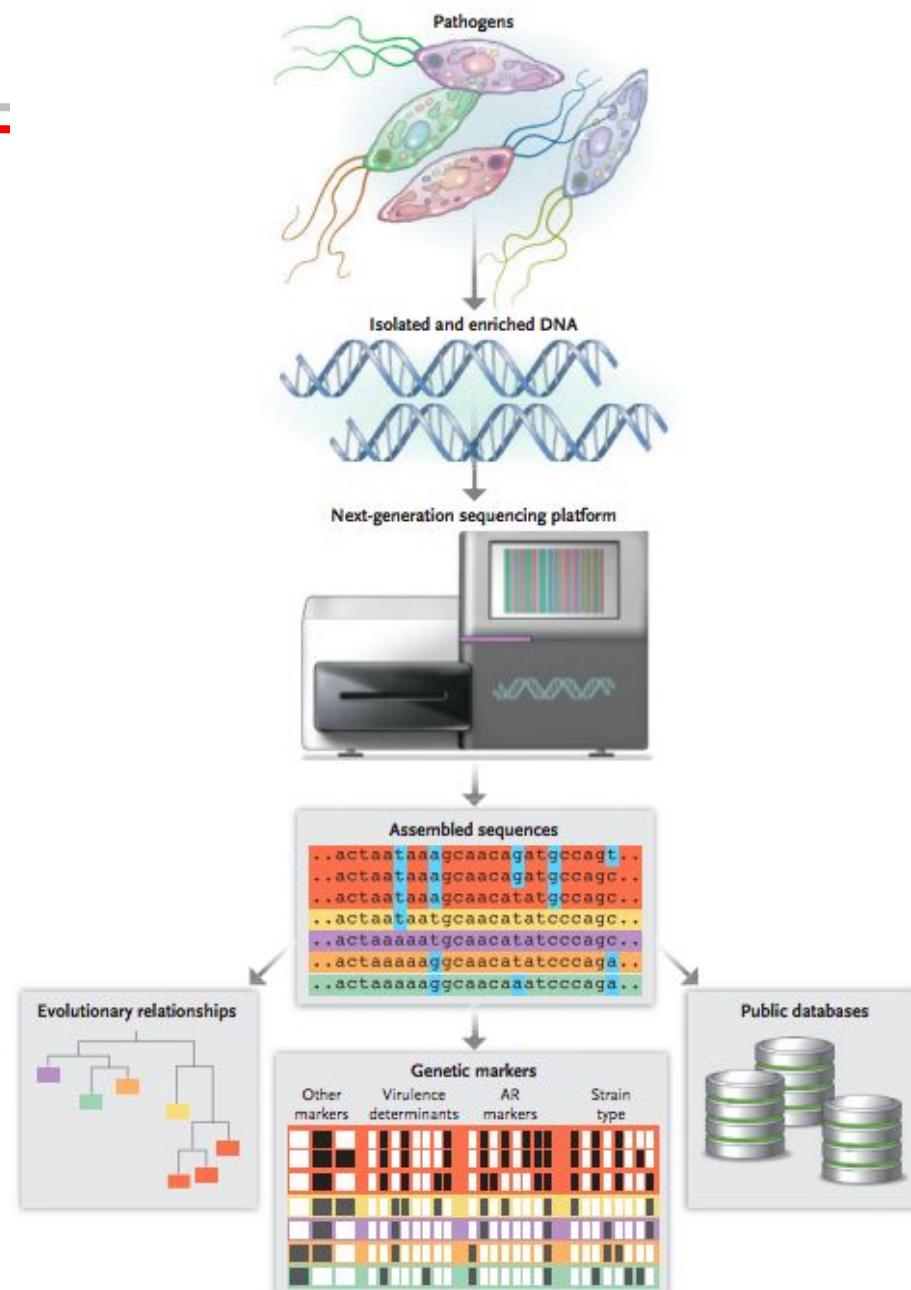
What does this look like?

“Wet” lab:

- Clinical samples/Microbial isolation
- RNA or DNA extraction
- Library preparation
- Next-Generation Sequencing

“Dry” lab:

- Data processing
- Primary genomic analyses
- Data interpretation
- Submission to repositories



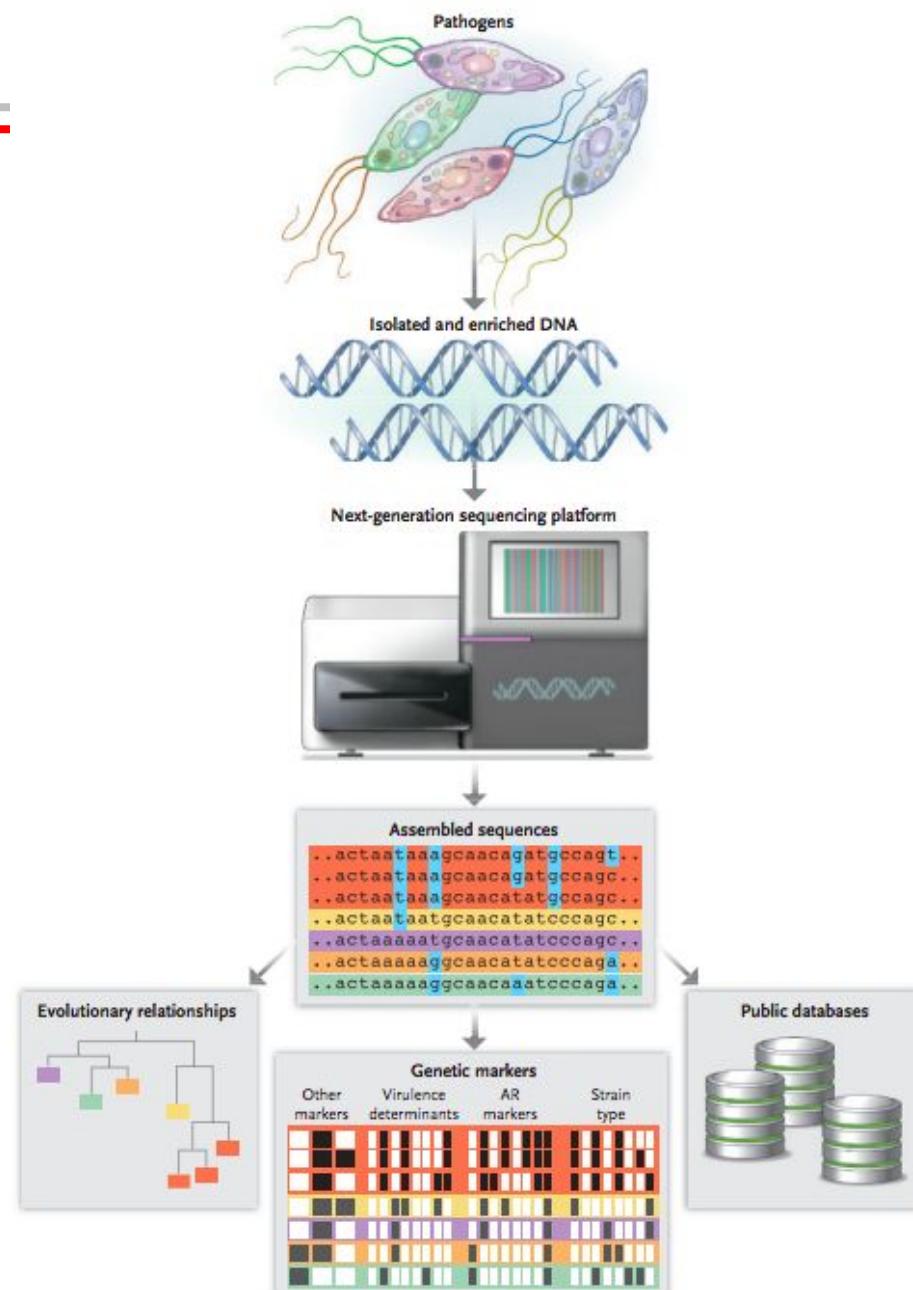
What does this look like?

"Wet" lab:

- Clinical samples/Microbial isolation
- RNA or DNA extraction
- Library preparation
- Next-Generation Sequencing

"Dry" lab:

- Data processing
- Primary genomic analyses
- Data interpretation
- Submission to repositories

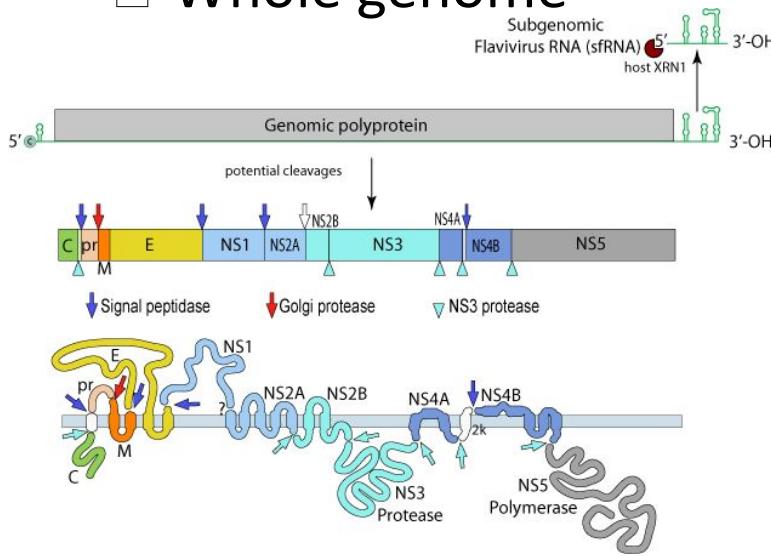


Challenges in Pathogen Sequencing

Viruses

E.g. Zika virus: 11kb

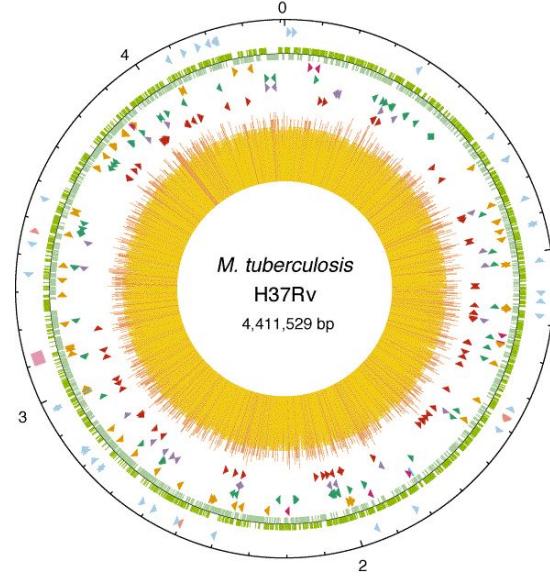
Whole genome



Bacteria

E.g. *Mycobacterium tuberculosis*: 4.4Mb

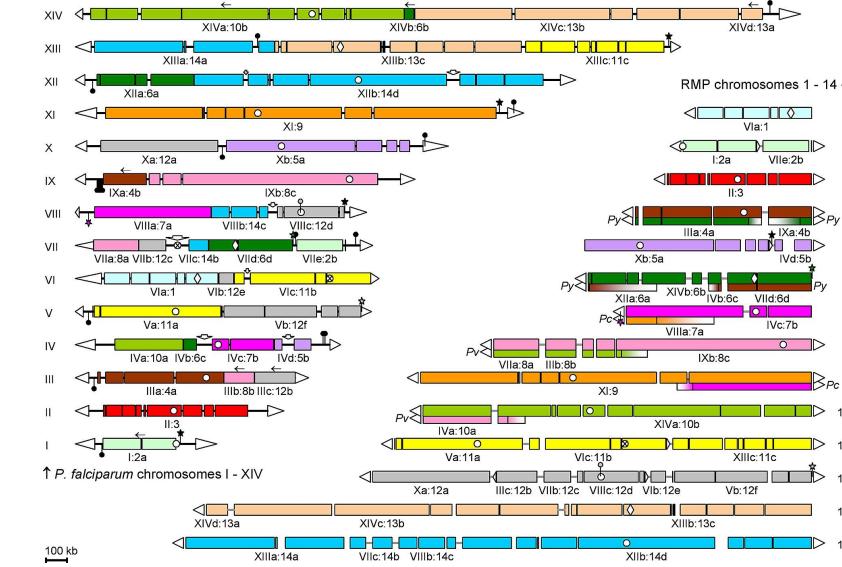
16S rRNA



Protozoa

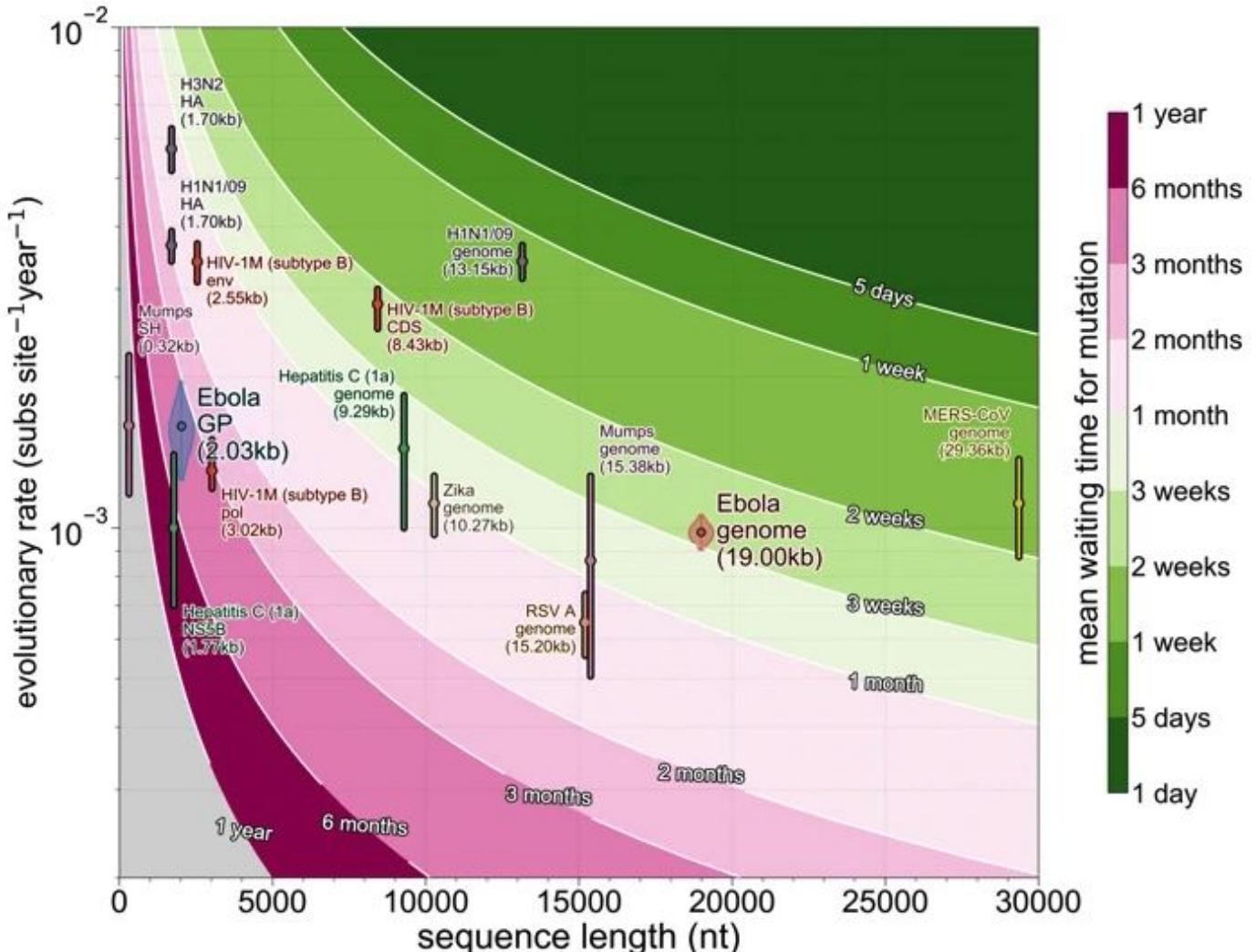
E.g. *Plasmodium falciparum*: 22.9Mb

Targeted genes



Challenges in Pathogen Sequencing

Fig. 6



Choice of Sequencing Platform/Approach

Illumina



NextSeq/NovaSeq

- Accurate sequencing, short reads
- Good for generating lots of data (NovaSeq)
- Expensive to run and maintain
- Uses:
 - Large projects (samples in home lab)
 - Accurate sequencing (intrahost, metagenomics)

MinION



- Long reads, error prone
- Good for generating quick data
- Rugged, cheap to run (reusable flowcells)
- Uses:
 - Field or limited resources
 - Lab for long reads and quick data

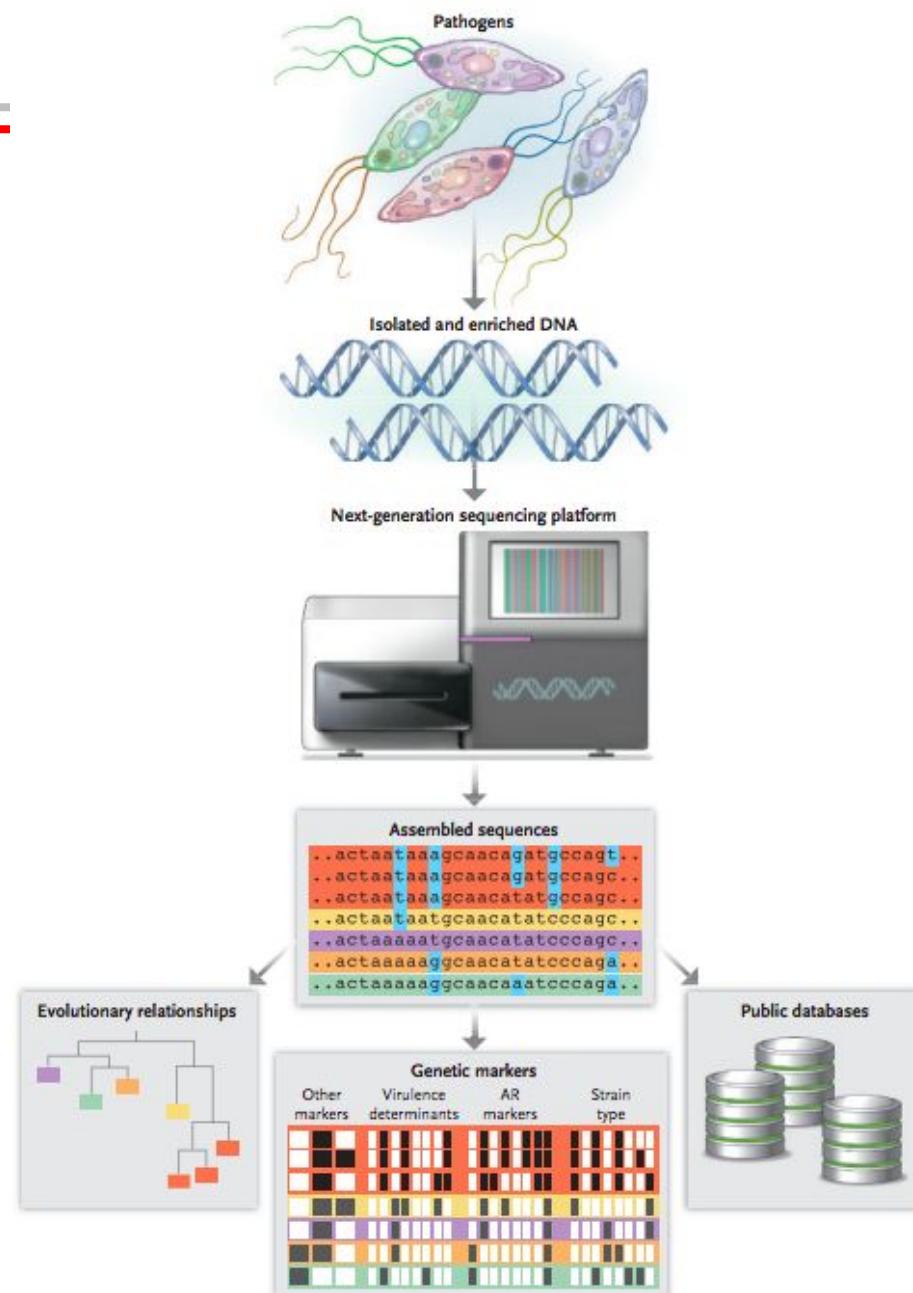
What does this look like?

“Wet” lab:

- Clinical samples/Microbial isolation
- RNA or DNA extraction
- Library preparation
- Next-Generation Sequencing

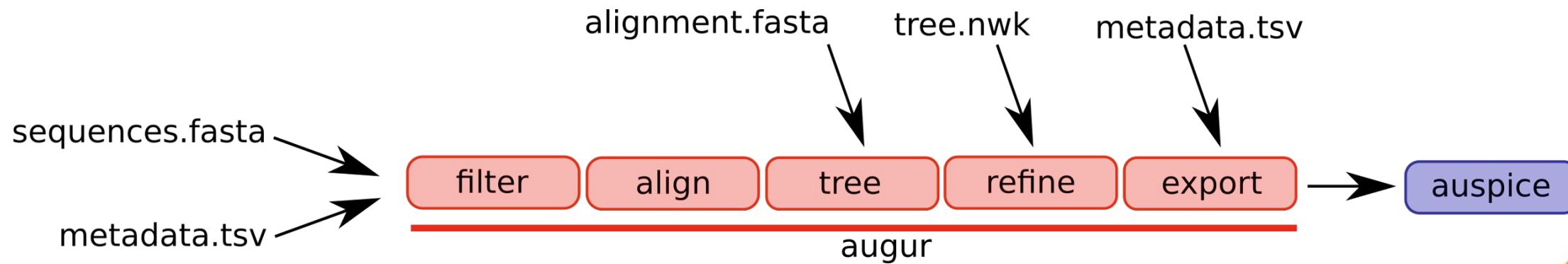
“Dry” lab:

- Data processing
- Primary genomic analyses
- Data interpretation
- Submission to repositories



What is next?

- NGS data, consensus sequence generation, data repositories
- Description of Nextstrain (what it is, what it isn't, how it works, etc.)
- Basics of phylogenetic tree interpretations



Nextstrain

Real-time tracking of pathogen evolution

geneious
prime



UNIVERSITY OF NEBRASKA MEDICAL CENTER™
COLLEGE OF PUBLIC HEALTH

Questions?

Reach out via email: jfauver@unmc.edu