Introduction to Applied Genomic Epidemiology for Outbreak Response

Paul Oluniyi, PhD, MSc

Scientist, Chan Zuckerberg Biohub

paul.oluniyi@czbiohub.org

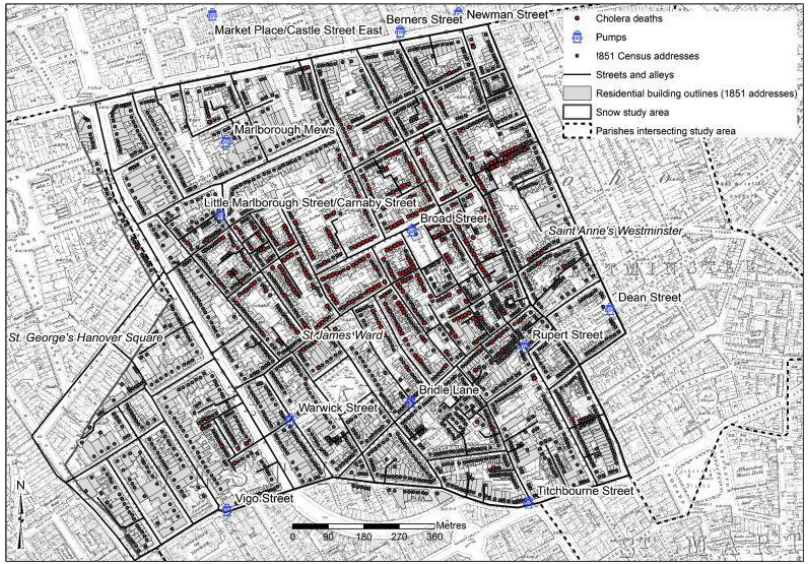
Outline

1. Theory - overlapping timescales of infectious disease transmission and pathogen evolution.

1. Different applications of genomic data for exploring epidemiologic questions.

What is epidemiology?

- Study of the occurrence and causes of diseases in a population



Newsom et al. 2006.

Questions:

- What is it?
- Has it been seen before?
- How can we fight it?
- Is it an outbreak?

Approaches include:

- Number of cases of disease in a population e.g., geographic area
- The frequency of an exposure among cases and non-cases
- Outbreak investigations

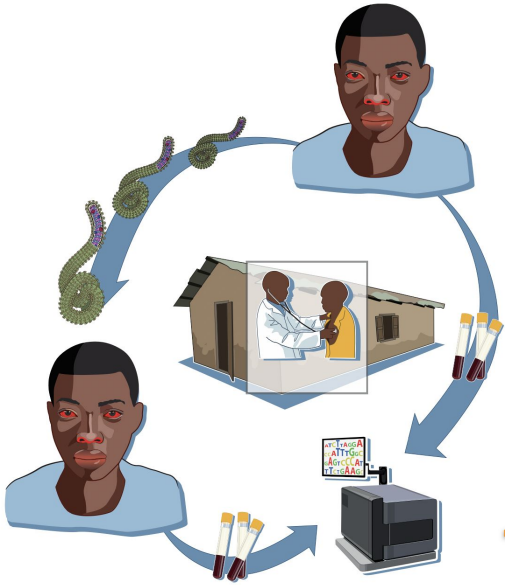
Limitations of classical epidemiology

- Observations associated with many biases and confounders
- Large populations needed for statistical power
- Labor-intensive and slow
- Often retrospective
- Limited information on pathogen circulation dynamics

What is pathogen genomic epidemiology ?

- Epidemiology of infectious diseases using pathogen sequencing

- ***Sequencing***: determining the nucleotide composition of a genome
- ***Genome***: the set of genetic information in an organism
- Genomic epidemiology is central to infectious disease public health surveillance



Andersen et al. 2015.

Genomic epidemiology is a discipline where we seek to understand the **distribution** of infectious diseases in a population and **determinants of those patterns** from **pathogen genetic sequence data**.

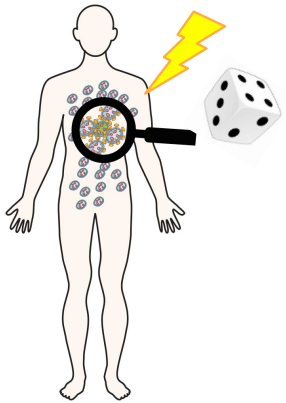
Part 1 - The overlapping timescales of pathogen evolution and infection transmission.

The ability to infer patterns of epidemic dynamics relies on a **fundamental principle**.

Pathogens evolve and **infectious diseases spread** through populations on **similar timescales.**

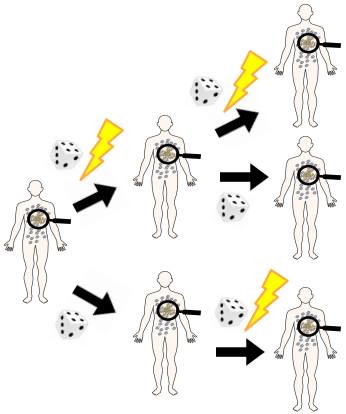
Evolutionary processes play out across multiple scales

Within hosts



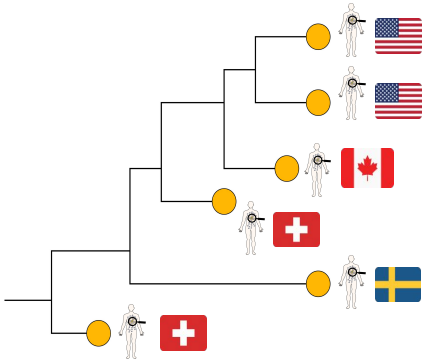
Short term evolution

Between hosts



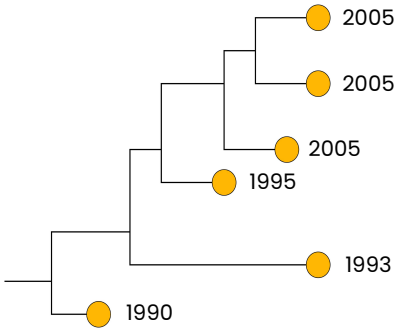
Longer term evolution

Across space



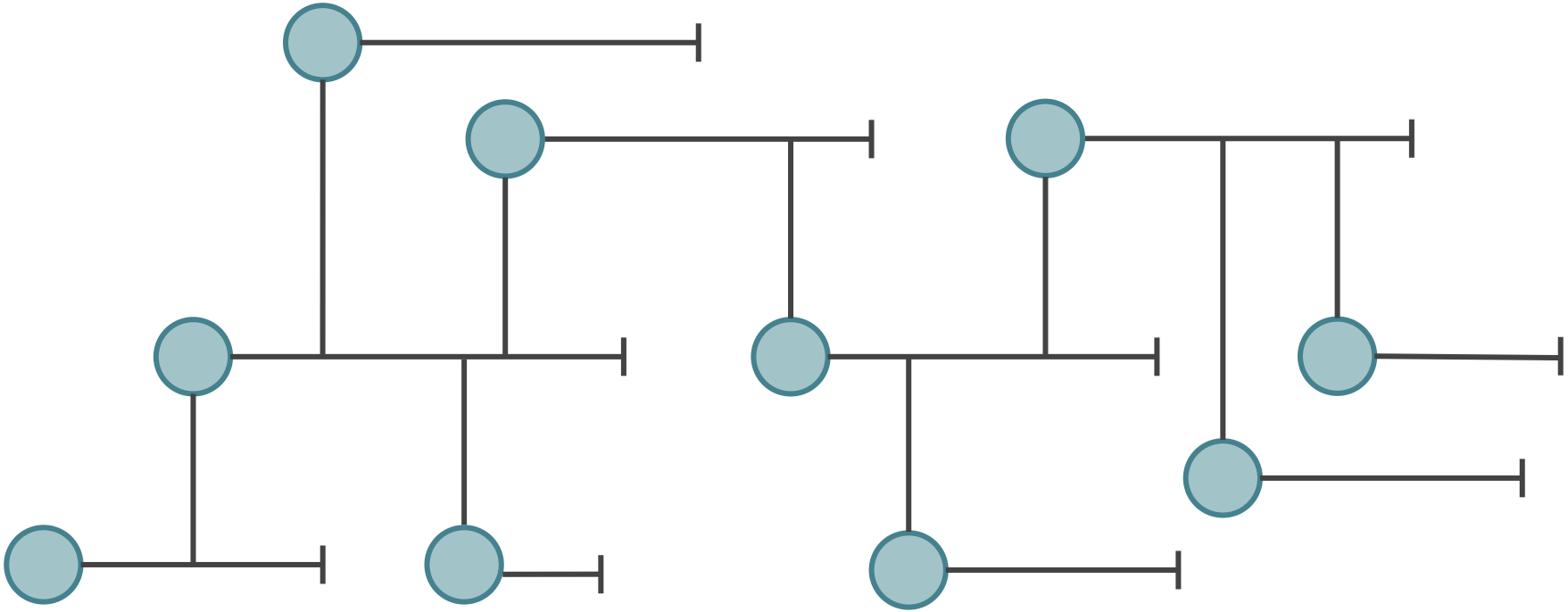
Time scale?

Across time



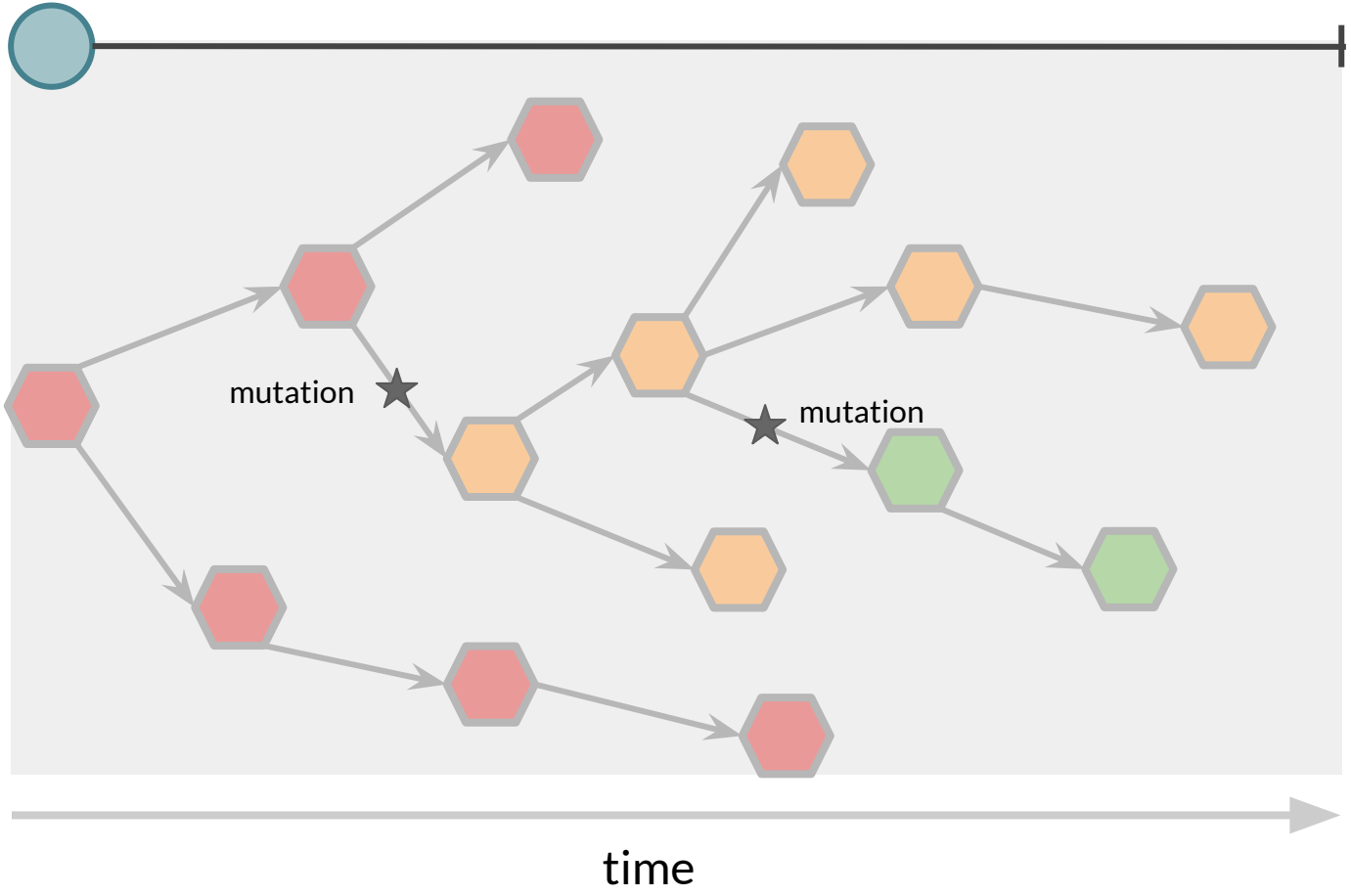
Time scale?

Outbreaks involve transmission between multiple individuals.

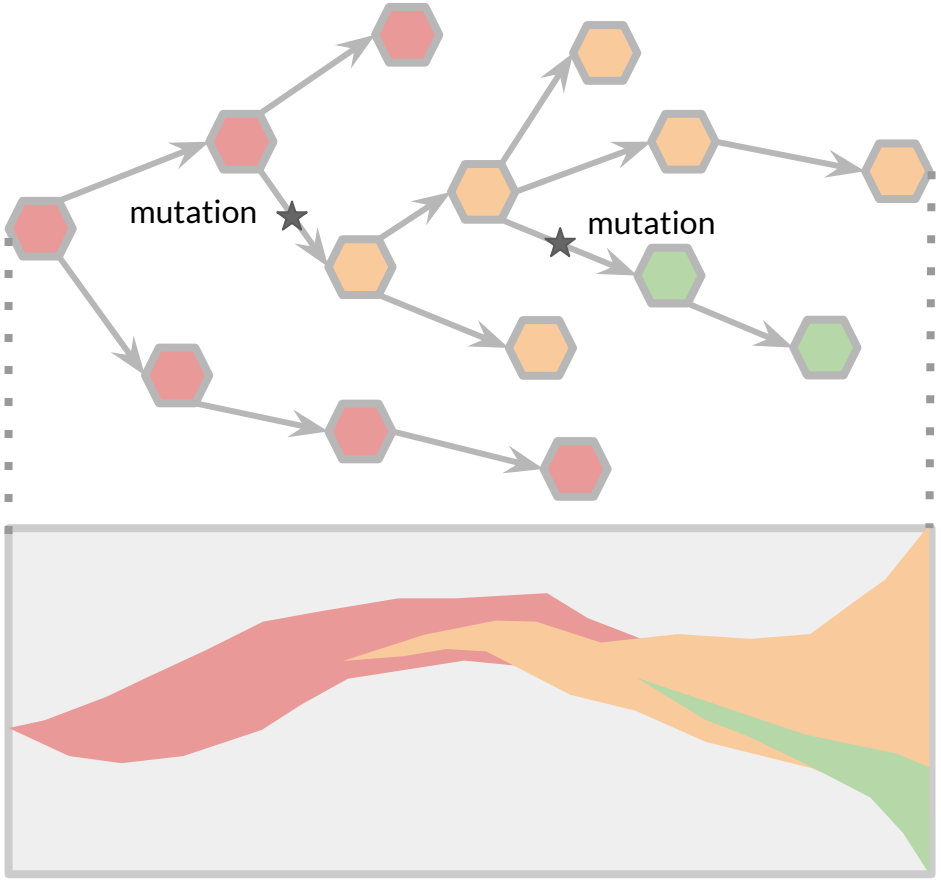


First, let's look at a **single individual** within this outbreak.

Errors occur during pathogen replication over the course of the infection.



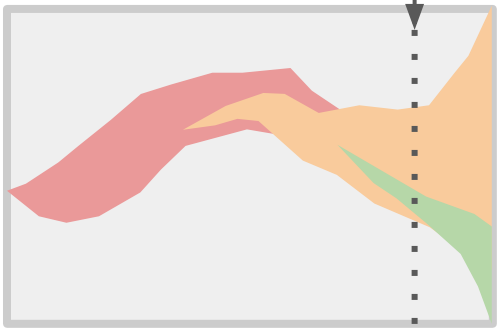
This process means that there are **various populations of pathogens** within a **single infected individual**



Single person's infection

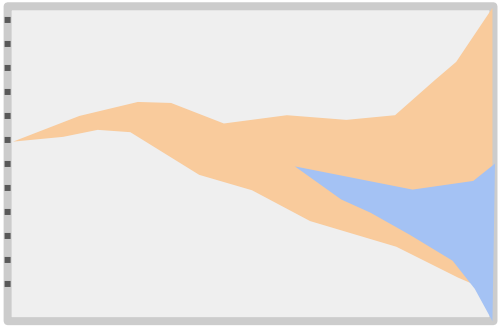
Infection event

Viral diversity accruing over infection.



The viral population selected from at transmission.

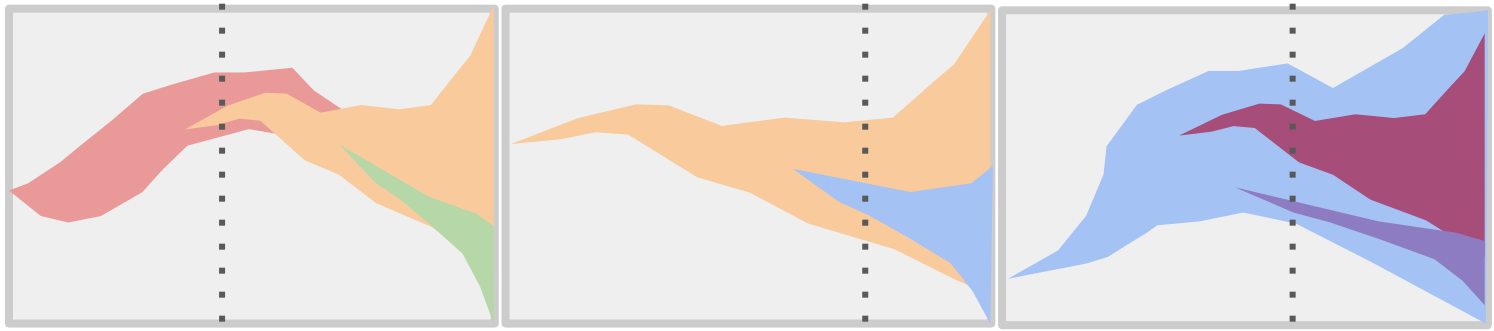
The second person's infection is founded by the orange virus, and the process happens again.



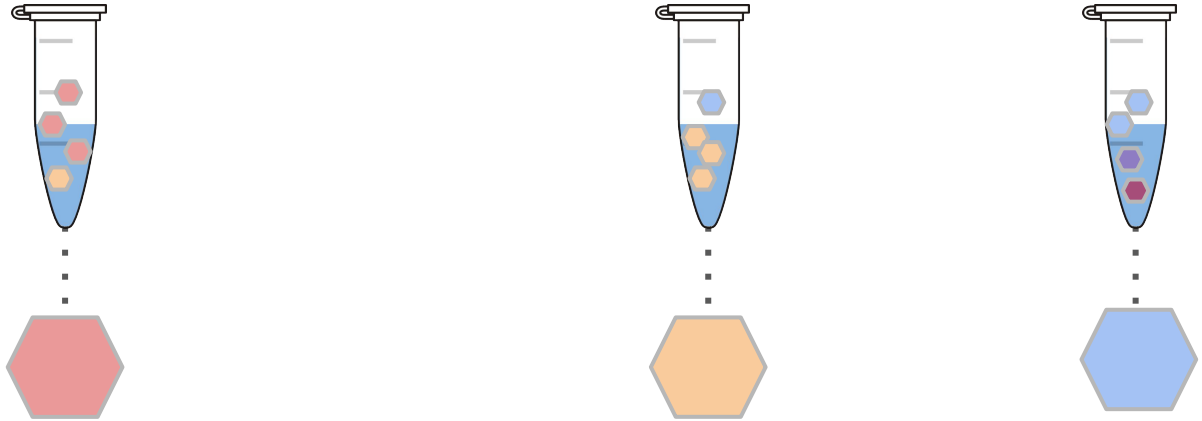
time

Typically we do not capture all of the within host diversity of an infection. Rather, we **summarize that diversity with consensus genomes.**

Within host diversity over an infection.



Diverse population of viruses sampled within a specimen.

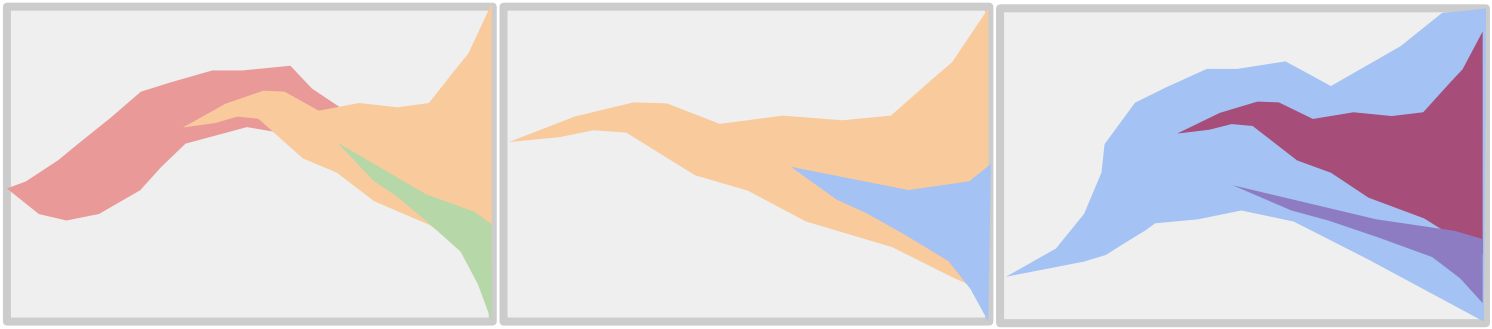


Single sequence summarizes the diversity.

The changes that we observe in **consensus genomes reflect additional processes beyond mutation**.

In nature

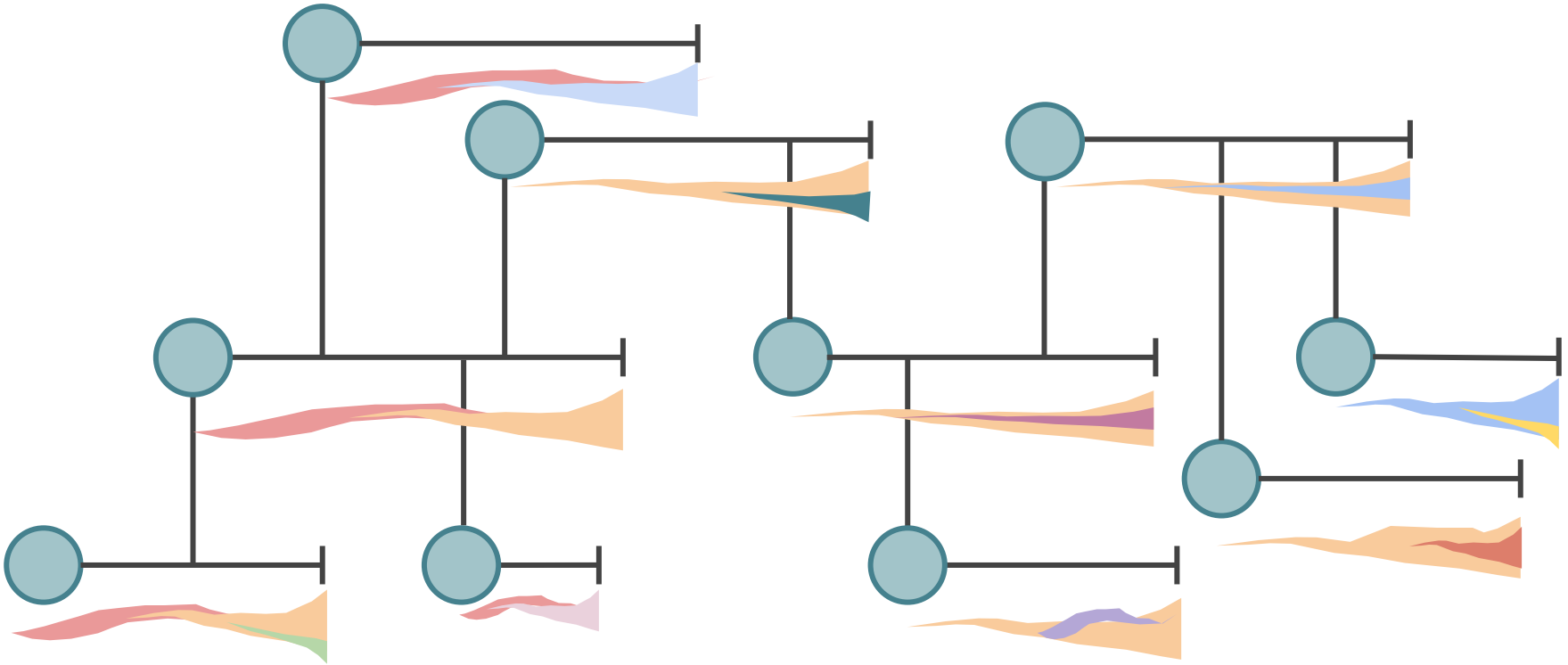
Actual errors made during the replication cycle.



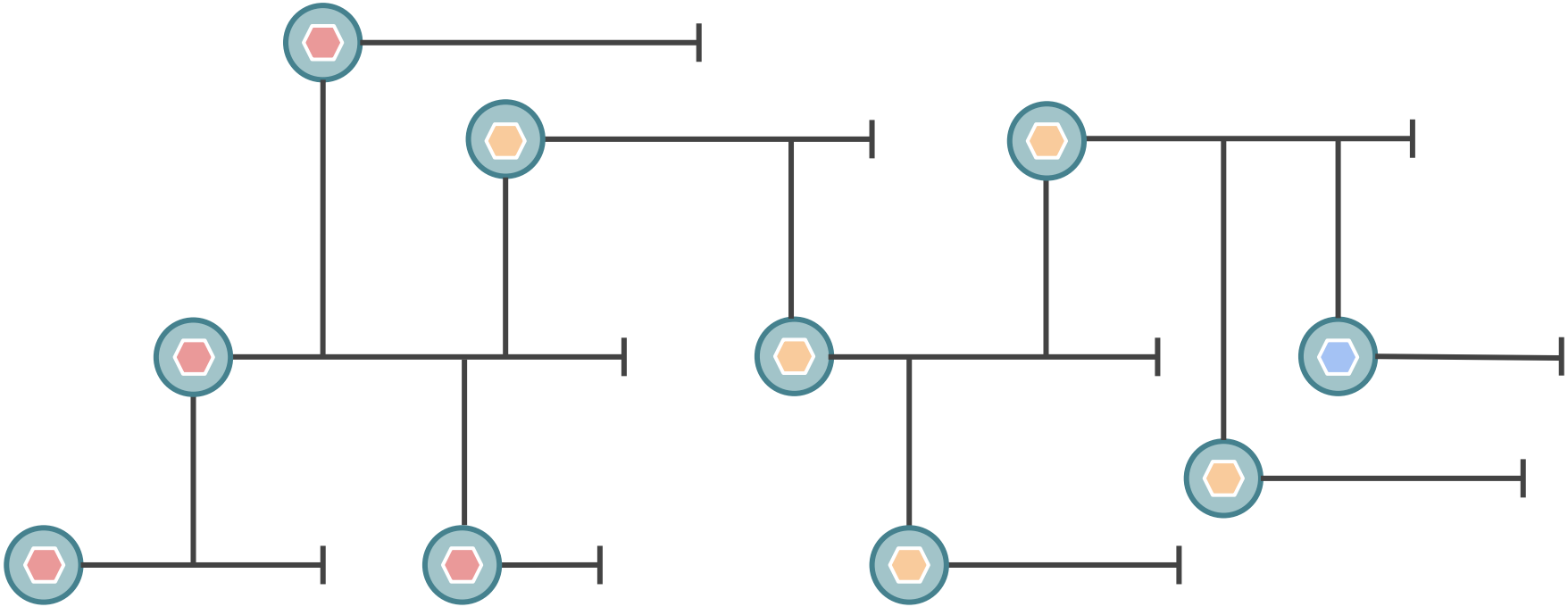
In our data

The changes to consensus sequences that we observe at the population level.

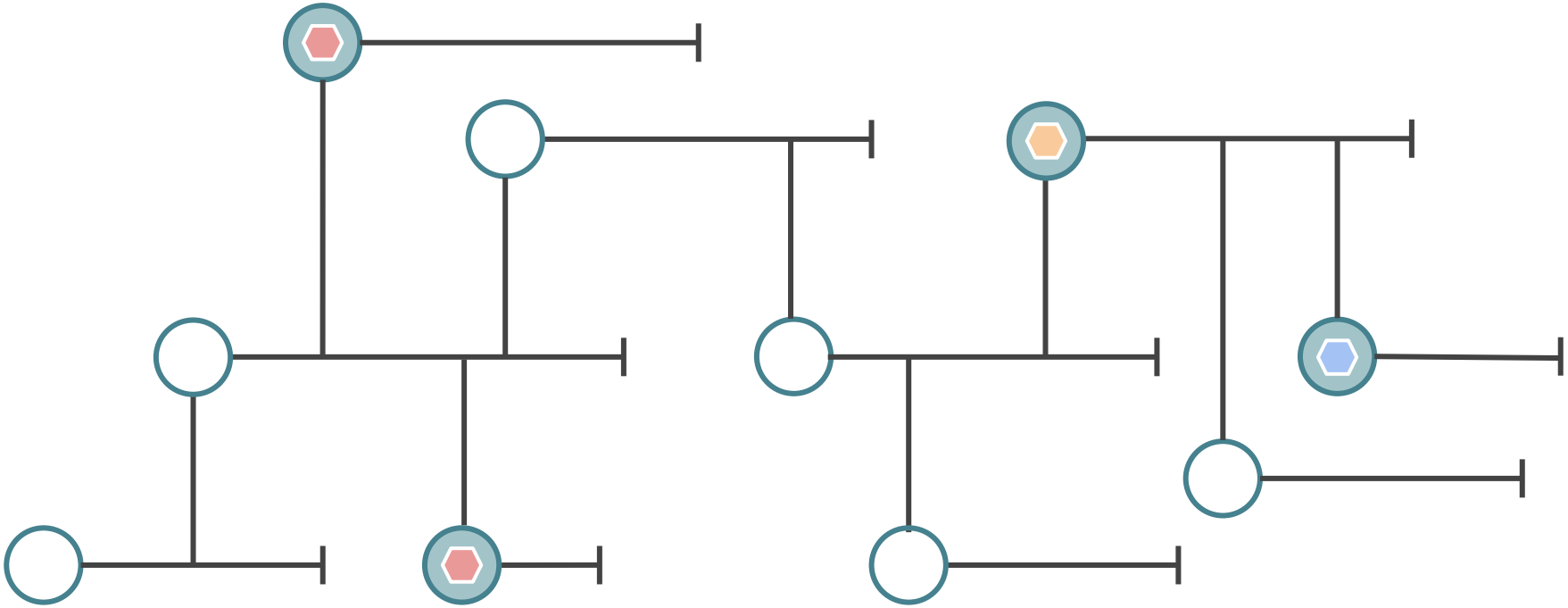
If this is the process that is **occurring in nature**...



This is the process that **we would see recorded in the data** if we sequenced every case.



And this is **typically what we observe**, because only a fraction of people are sampled



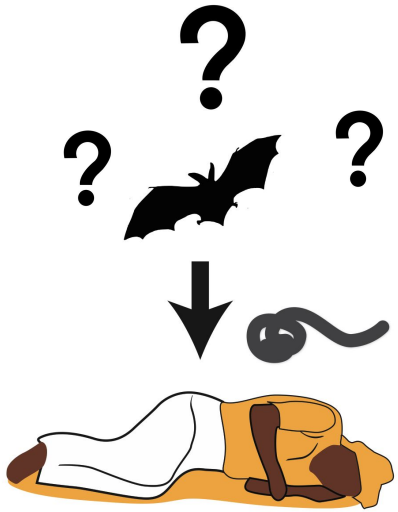
Closely-linked cases have **genetically similar** infections.

Distally-related cases have **more genetically dissimilar** infections.

Why use genomic epidemiology?

1. Identify the origin of epidemics/zoonoses

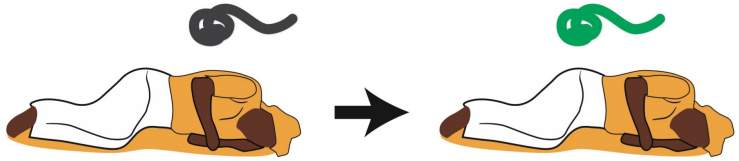
- What is the pathogen?
- What is the reservoir?
- How many introductions from reservoir?



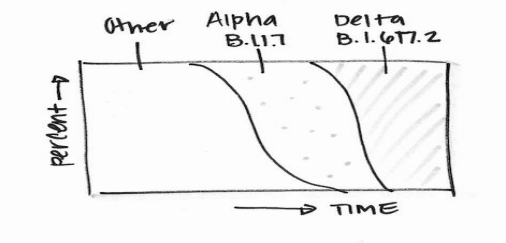
Why use genomic epidemiology?

2. Understand transmission dynamics

- Are cases linked to one another?
- Is there evidence for human-to-human transmission?
- When did the pathogen enter the population of interest?
- From where was it introduced?
- How is the pathogen evolving as it spreads?

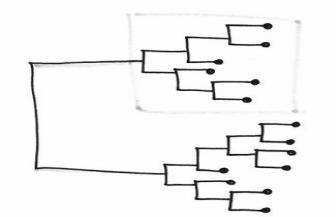


Indicator-based Surveillance



- What variants are circulating in my community?
- How are variant frequencies changing?

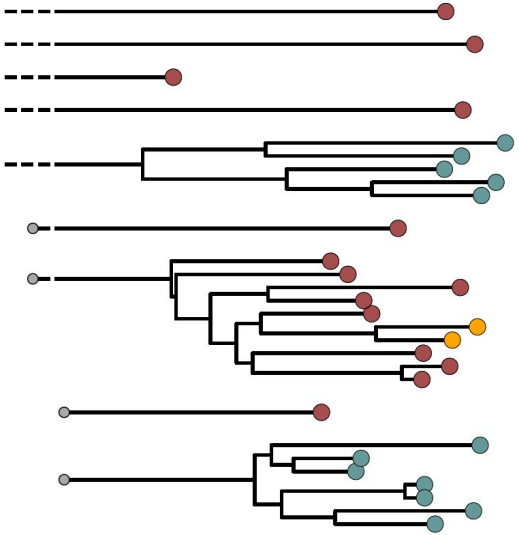
Outbreak Investigations



- Are these cases linked?
- Where is transmission occurring?
- Are my interventions working?

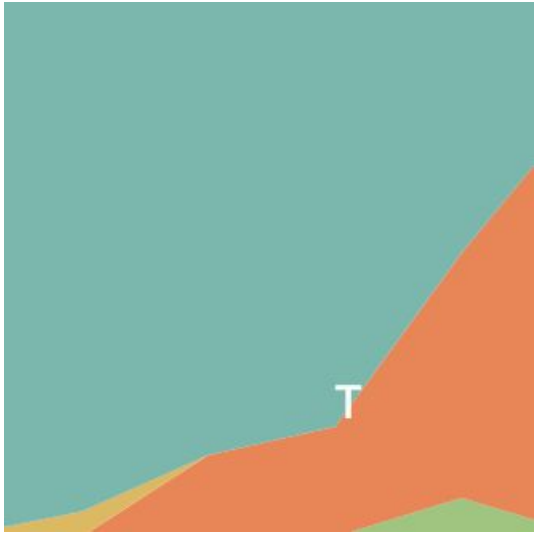
We can use genomic data and phylogenetic analyses to describe many aspects of transmission

Introductions and onward transmission



Spatial patterns and source-sink dynamics

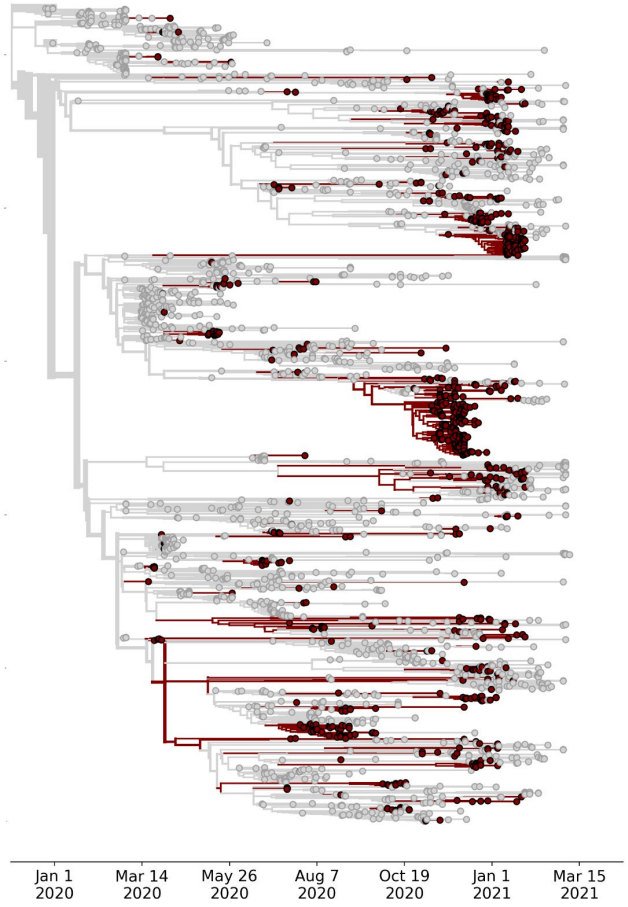
Emergence of adaptive variants

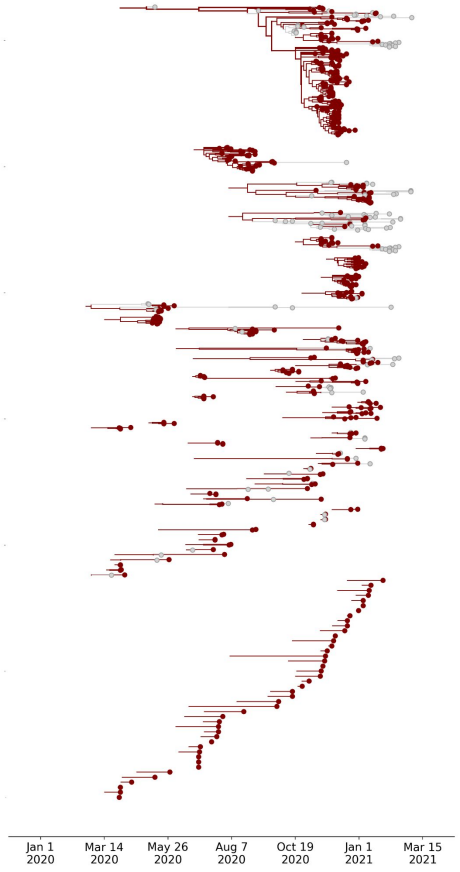


Application - *Describing patterns of introduction and spread.*


Phylogenies can help us visualize the sparks/introductions and the burn/spread.

Example: separating the tree into its distinct transmission chains.



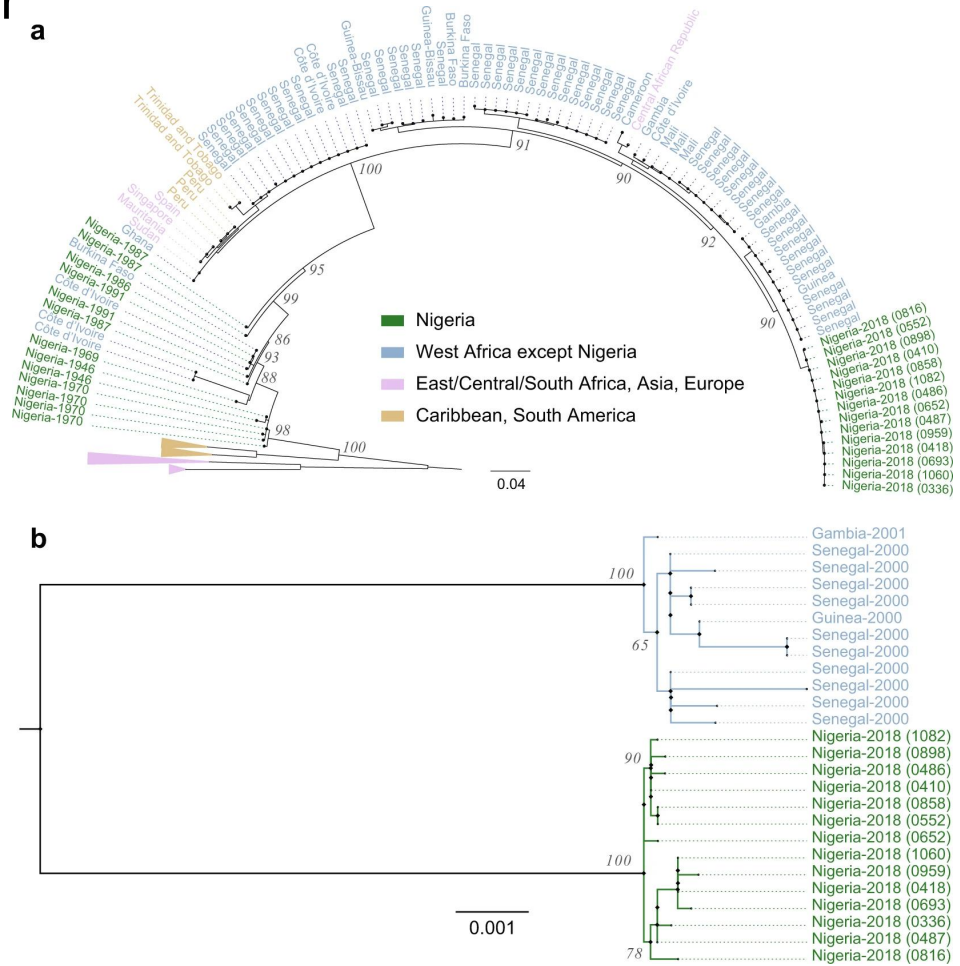


Stoddard, Black et al, 2022

Genomic Epidemiology of a Yellow fever Outbreak in Nigeria

- 2018 YFV sequences formed a tightly clustered clade.

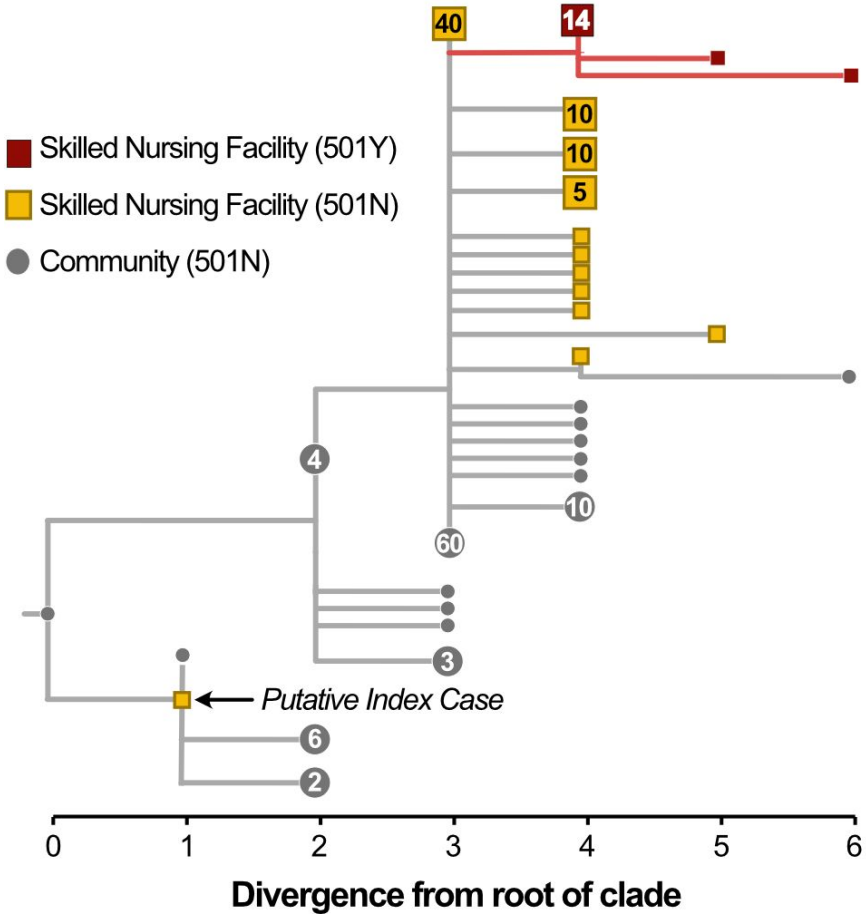
- The 2018 clade was more closely related to sequences from other West African countries (mean pairwise identity = 99·4%) than to earlier (1946–1991) Nigerian sequences (mean pairwise identity = 89·8%).



Ajogbasile *et al.*, 2020

Application - *Exploring linkage between cases and/or outbreaks.*

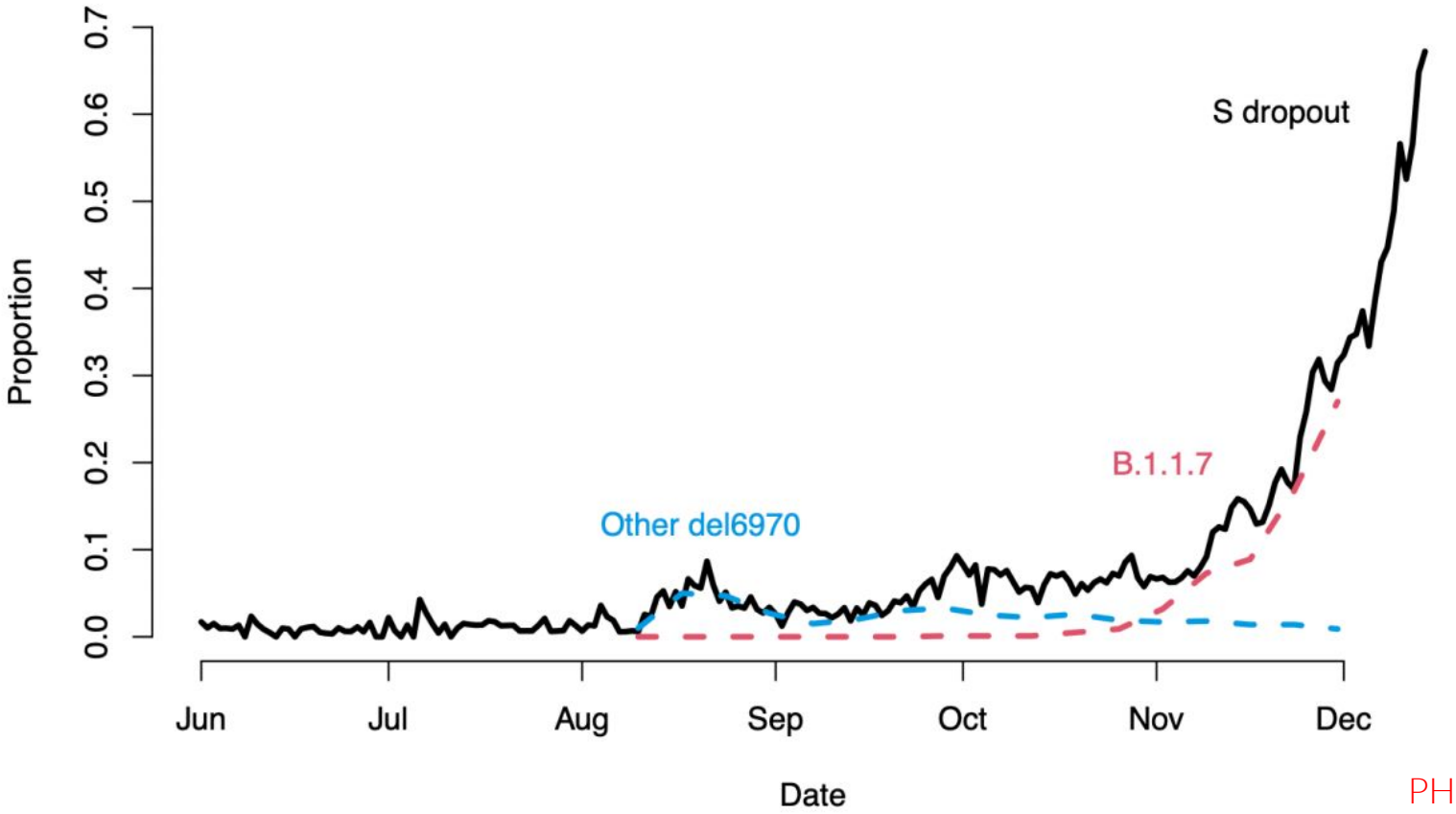
Genomics can also **rule out linkage** between cases.



Stoddard, Black et al, 2022

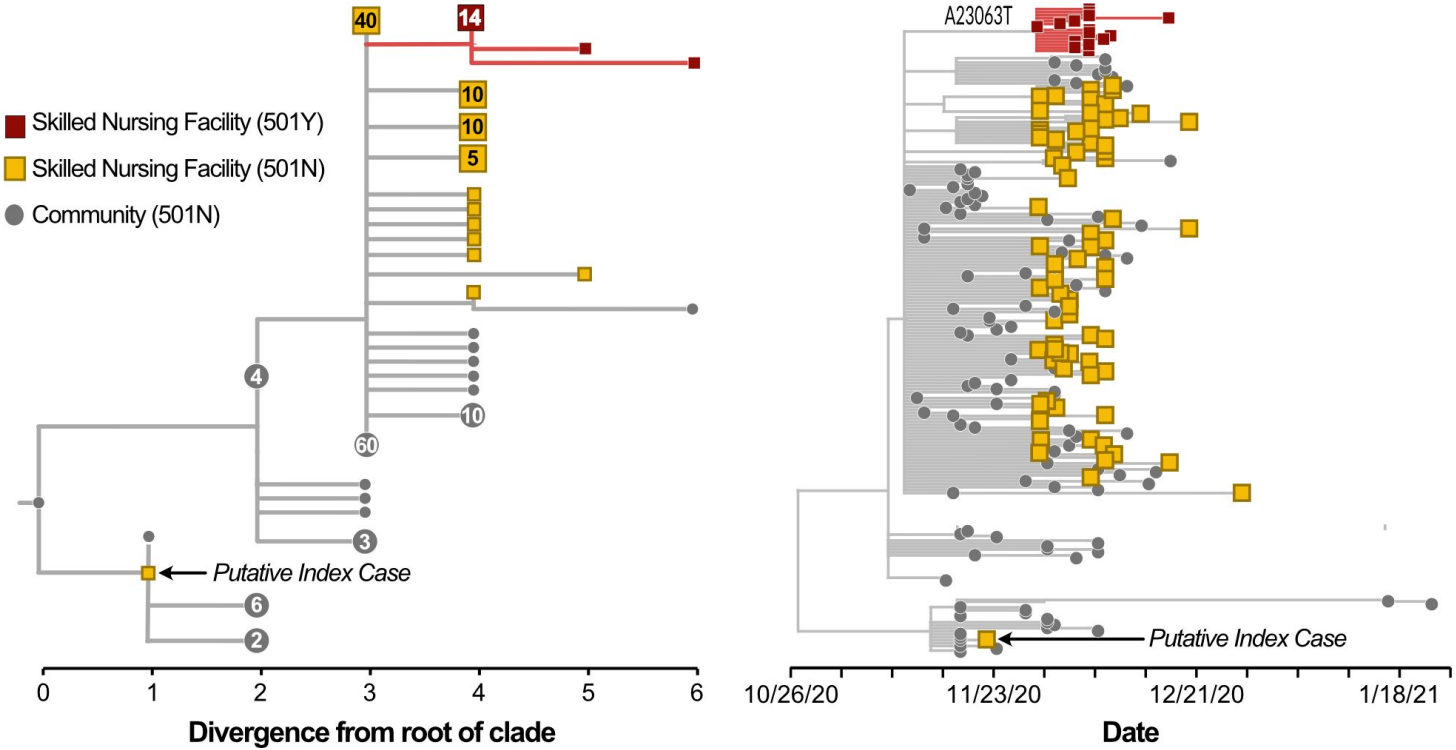
Application - *Monitoring for the emergence of adaptive variants.*

Using lineage frequency change to monitor for emergence of adaptive variants.



PHE, 2020

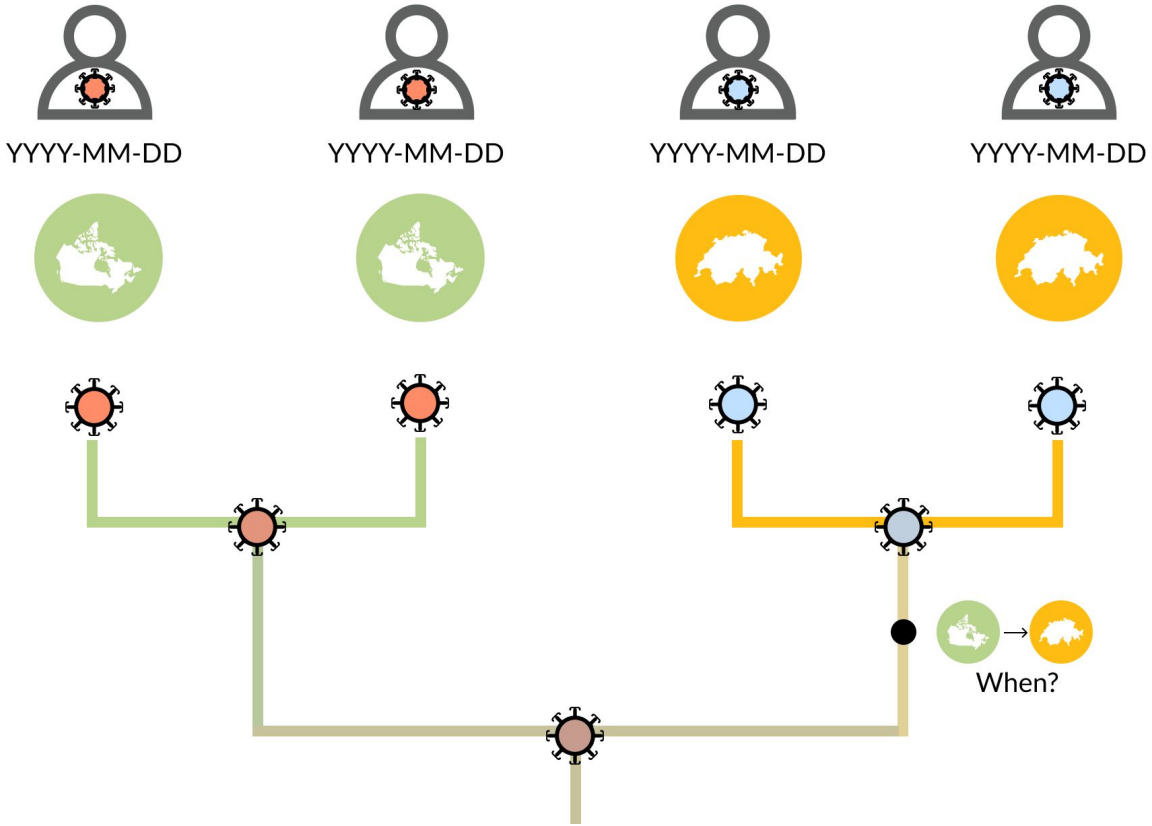
Surveillance for mutations of interest can go beyond monitoring VOC frequency.



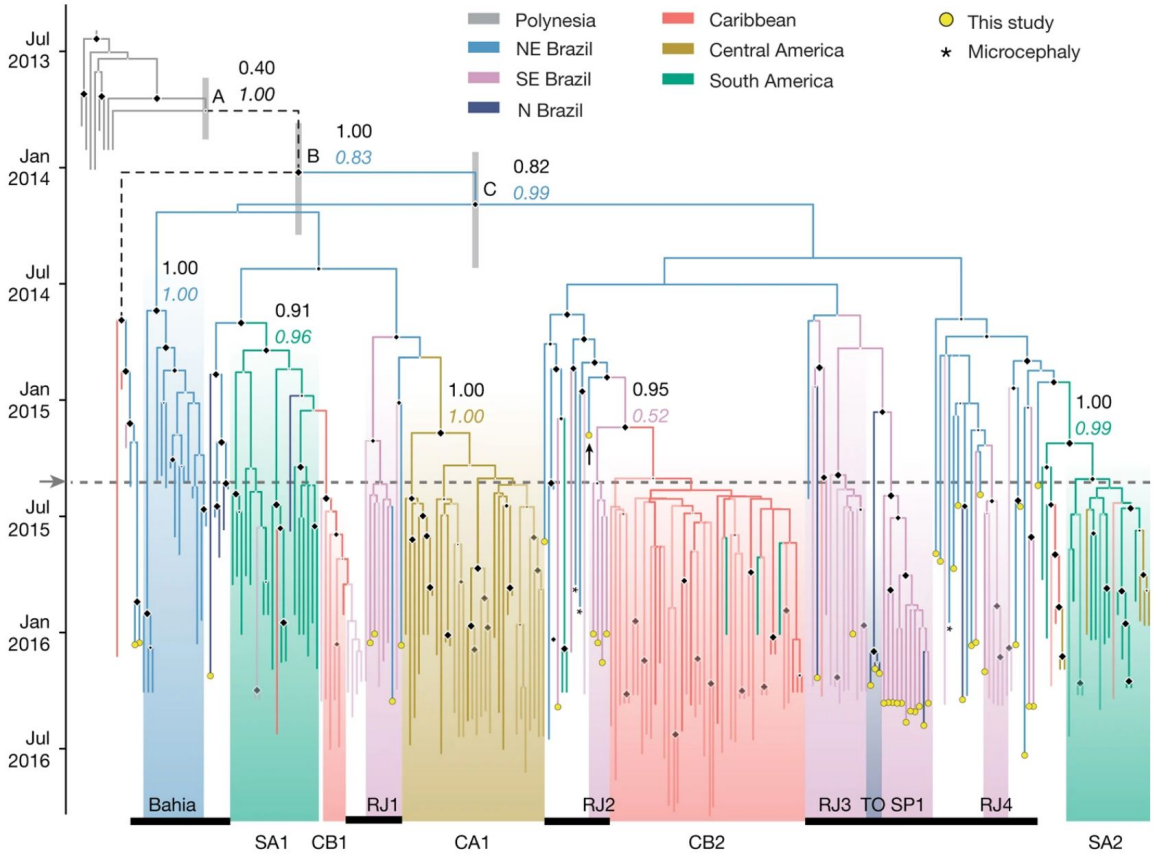
Stoddard, Black et al, 2022

Application - *Describing spatial dissemination of infectious diseases.*

Incorporating “metadata” allows us to scale phylogenies in terms of time and look at geographic transmission patterns.



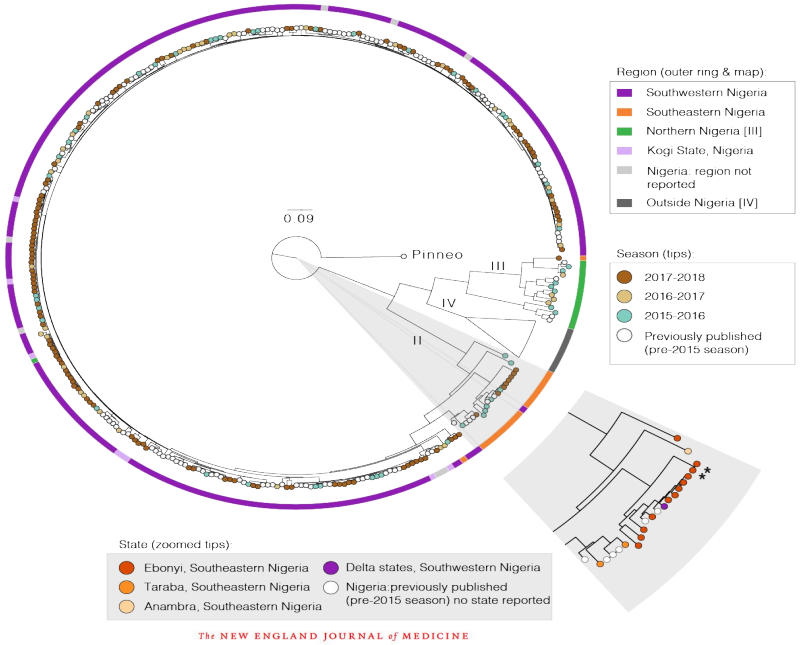
Example: estimating when Zika was introduced to Brazil, and where it went from there.



Faria et al, 2017

Application - *Animal-to-human spillover events.*

Genomic Epidemiology improves understanding of Lassa virus populations across Nigeria

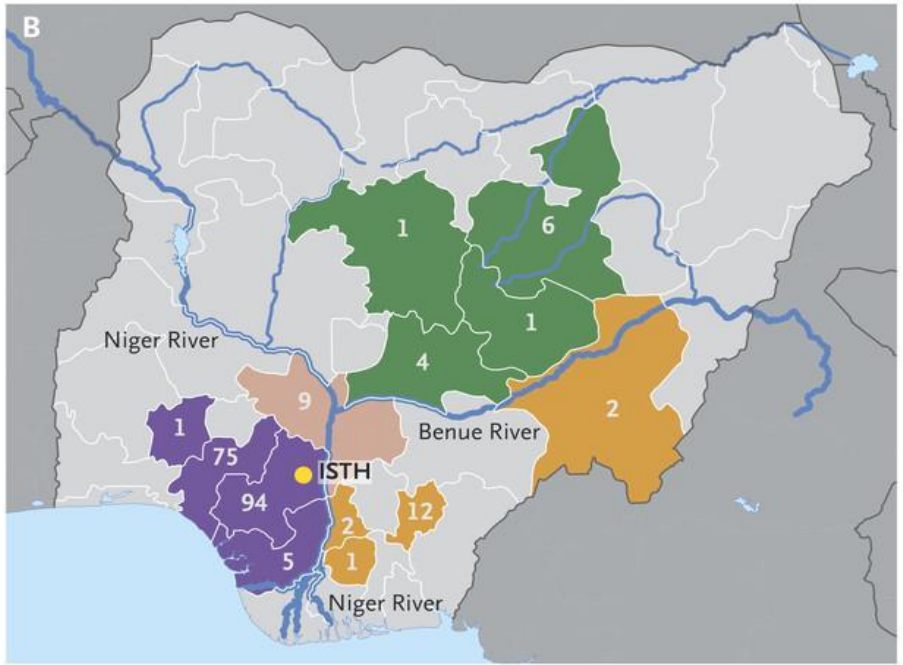


The NEW ENGLAND JOURNAL of MEDICINE

BRIEF REPORT

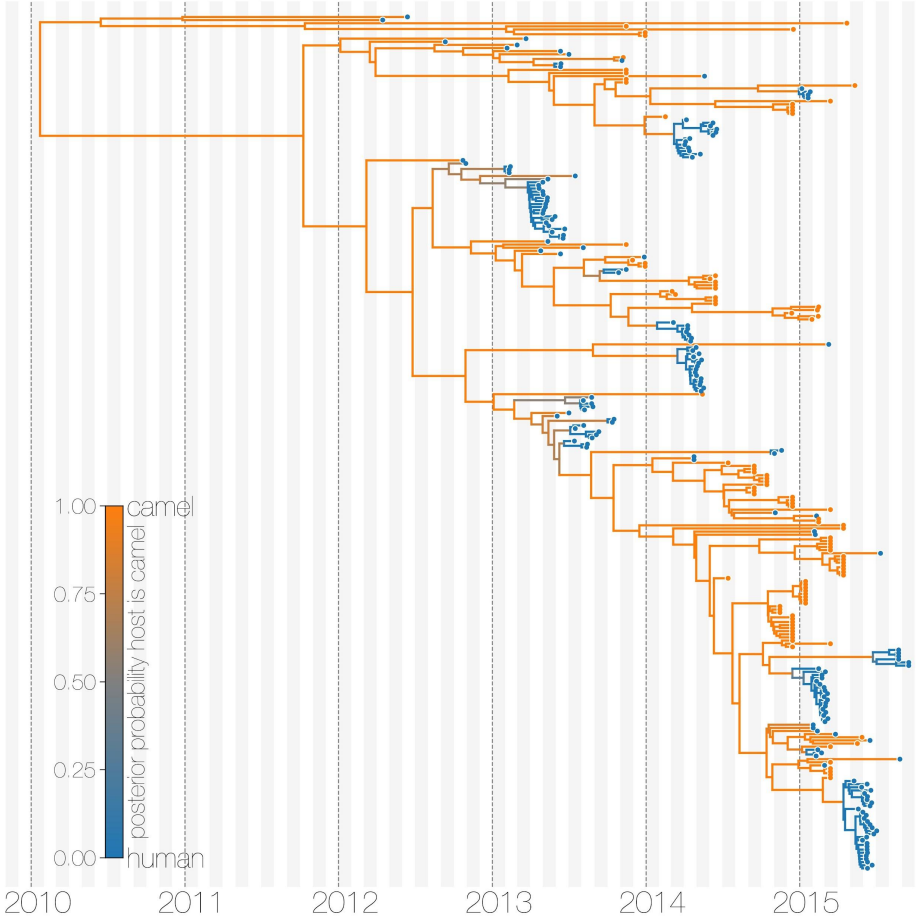
Siddle et al., 2018

Genomic Analysis of Lassa Virus during an Increase in Cases in Nigeria in 2018



Lassa virus genetic diversity in Nigeria is structured geographically, following major rivers

MERS camel-to-human spillover events



(Dudas et al, 2018).

Genomic epidemiology can inform public health action.

- Helping us understand the contributions of different epidemic processes to outbreaks.
- Investigating possible linkages between cases.
- Evaluating how well our interventions are performing.

Questions?