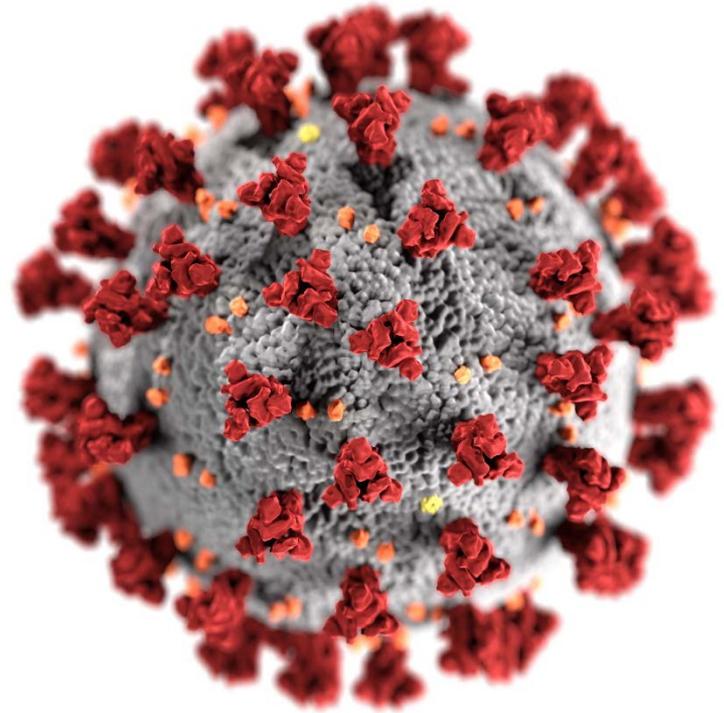


What is genomic epidemiology?

COVID-19 Genomic Epidemiology Toolkit: Module 1.1

Nancy Chow, PhD
Bioinformatics and Informatics Lead
Centers for Disease Control and Prevention



cdc.gov/coronavirus

Toolkit map

Part 1: Introduction

1.1 What is genomic epidemiology?

1.2 The SARS-CoV-2 genome

1.3 How to read phylogenetic trees

Part 2: Case Studies

2.1 SARS-CoV-2 sequencing in Arizona

2.2 Healthcare cluster transmission

2.3 Community Transmission

Part 3: Implementation

3.1 Getting started with Nextstrain

3.2 Getting started with MicrobeTrace

3.3 Linking epidemiologic data

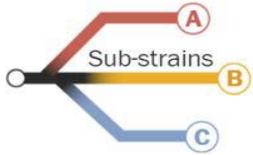
Genomic epidemiology

- **Epidemiology:** study of the distribution and determinants of health-related outcomes in a specified population *and* the application of this study to control health problems
- **Genomic epidemiology:** use of pathogen genomic data to determine the distribution and spread of an infectious disease in a specified population *and* the application of this information to control health problems

Rationale for Sequencing of SARS-CoV-2

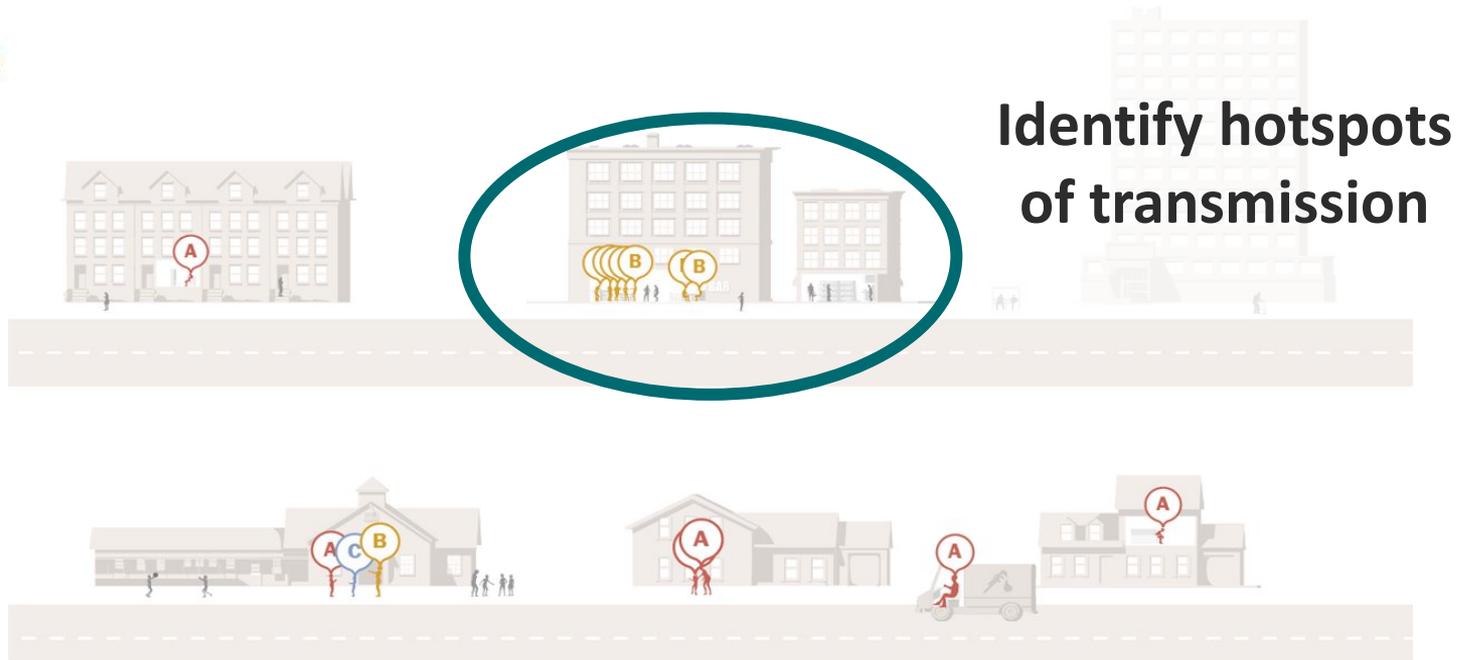
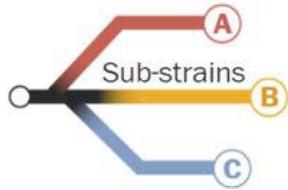
- Monitor trends at the national level
 - Monitor emergence of important new strains
 - Monitor trends after interventions such as vaccination
- Better understand epidemiology at the local level
 - Investigate clusters for transmission in a variety of settings (healthcare, workplace, bar, etc.)
 - Provide evidence for or against suspected transmission
 - Reveal important, unsuspected clusters

SARS-CoV-2 sequencing: A powerful tool for epidemiologic investigation

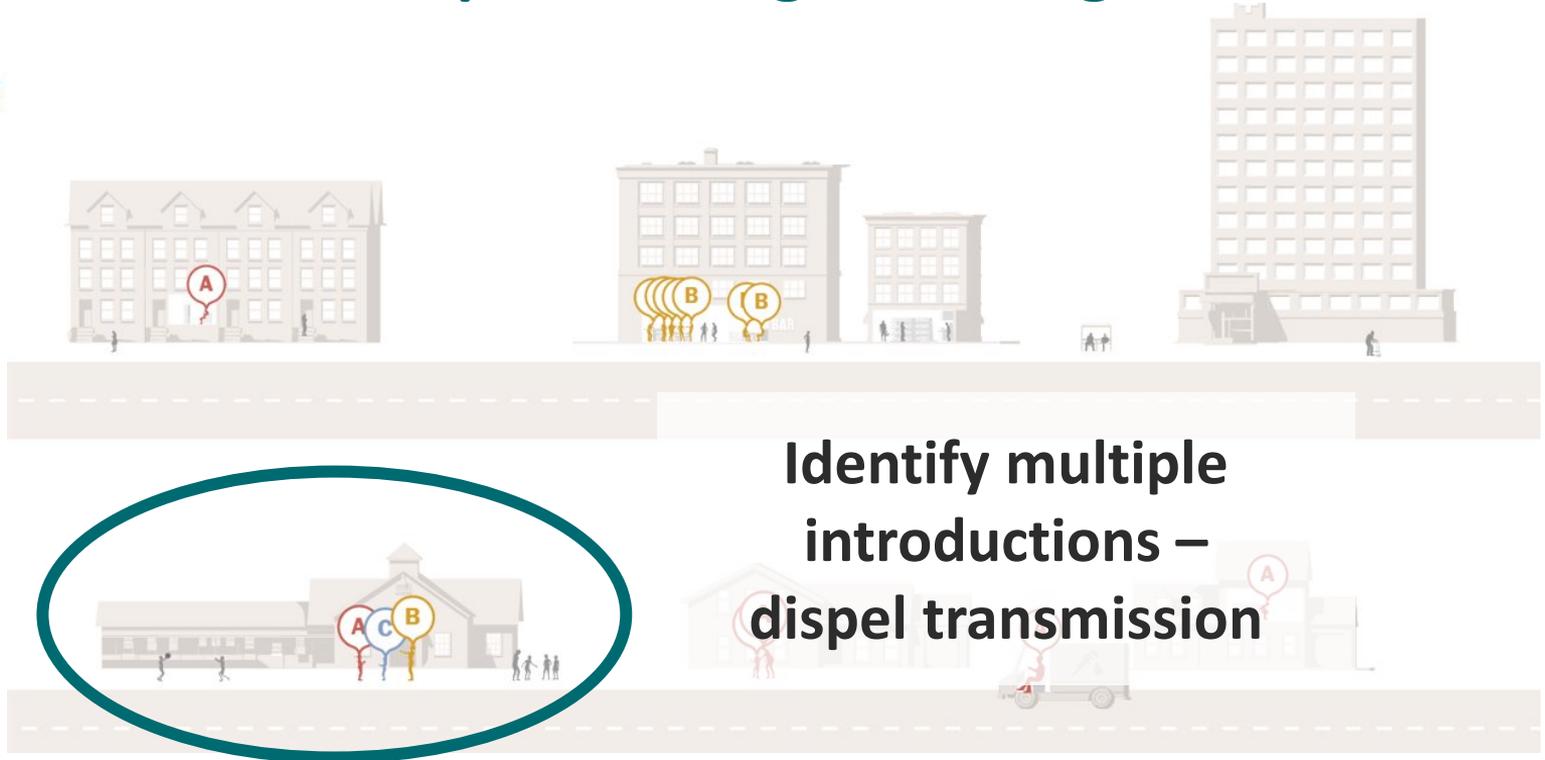
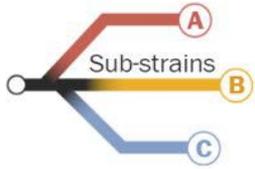


Adapted from *The Washington Post* article: <https://www.washingtonpost.com/graphics/2020/health/coronavirus-genetic-code/>

SARS-CoV-2 sequencing: A powerful tool for epidemiologic investigation

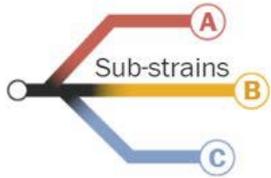


SARS-CoV-2 sequencing: A powerful tool for epidemiologic investigation



**Identify multiple
introductions –
dispel transmission**

SARS-CoV-2 sequencing: A powerful tool for epidemiologic investigation

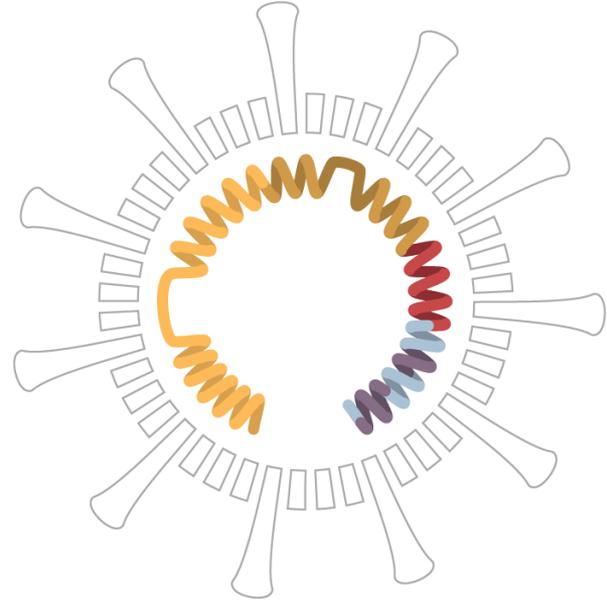


Additional ways to use SARS-CoV-2 sequencing data

- Identify introductions and sources
- Identify superspreading events
- Predict severity and size of future outbreak seasons

Utility of sequencing

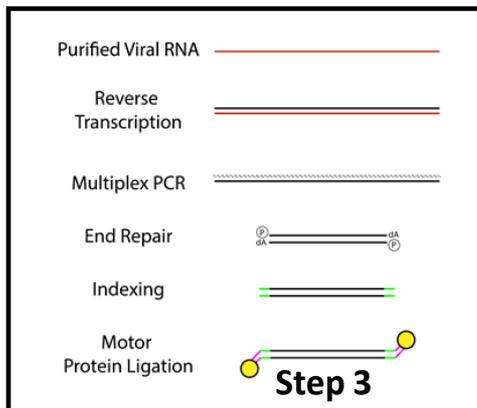
- Sequencing generates a genomic fingerprint
- Mutations demonstrate a change in that genomic fingerprint
- Sequencing allows us compare the genomic fingerprints to see how they are related



RNA whole-genome sequencing (WGS)



Sequencing Library Preparation



Common sequencing platforms

Short-read technology



Ex. MiSeq

- More accurate, short reads
- High throughput
- Relatively high up-front equipment costs
- Most widely used platform currently in US public health system

Long-read technology

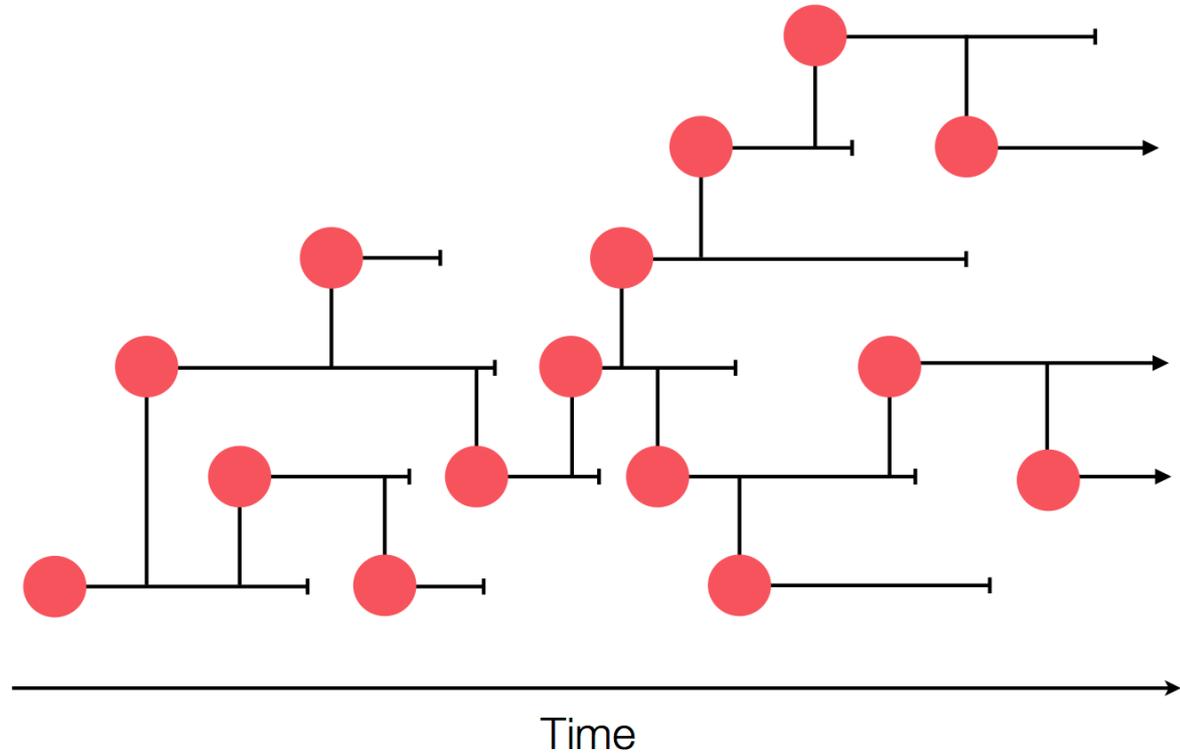


Ex. MinION

- Less accurate, long reads
- Generates data more rapidly
- Low capital costs
- Portable
- In US public health system, used for specific applications

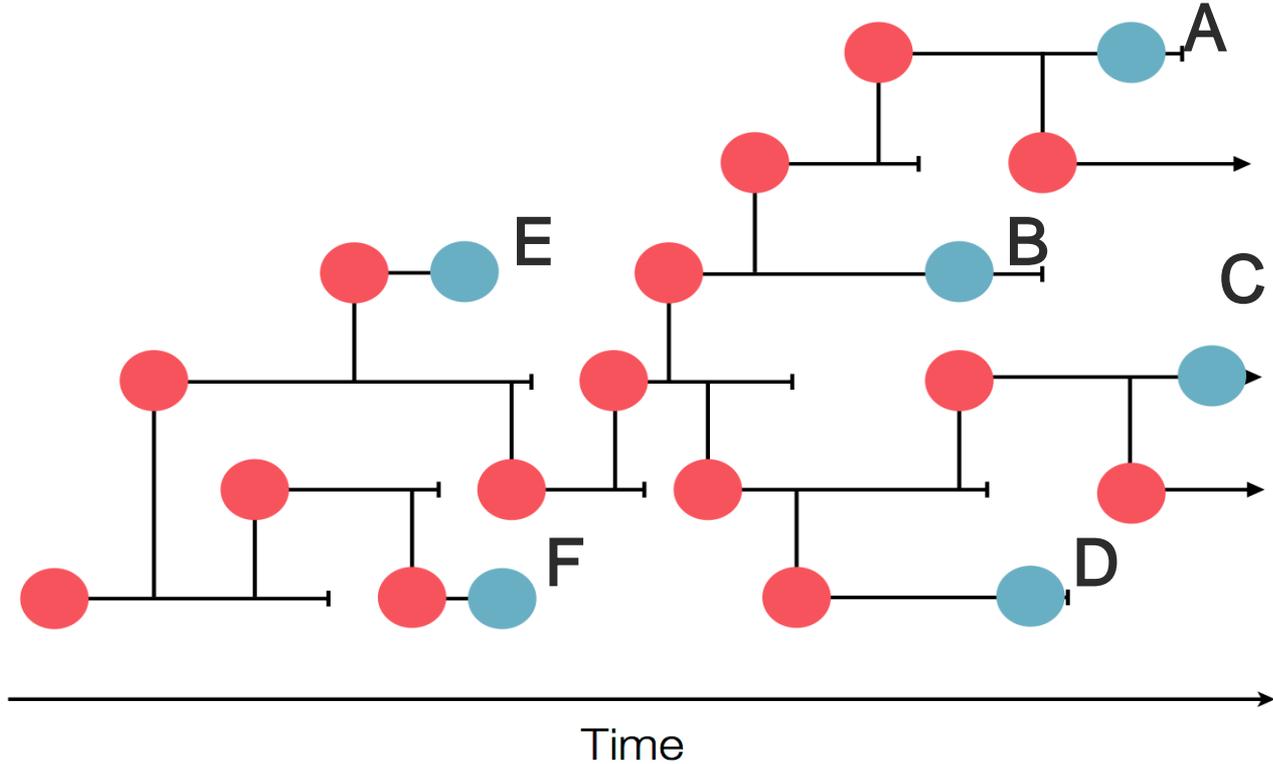
Epidemic process

All cases in transmission network represented over time



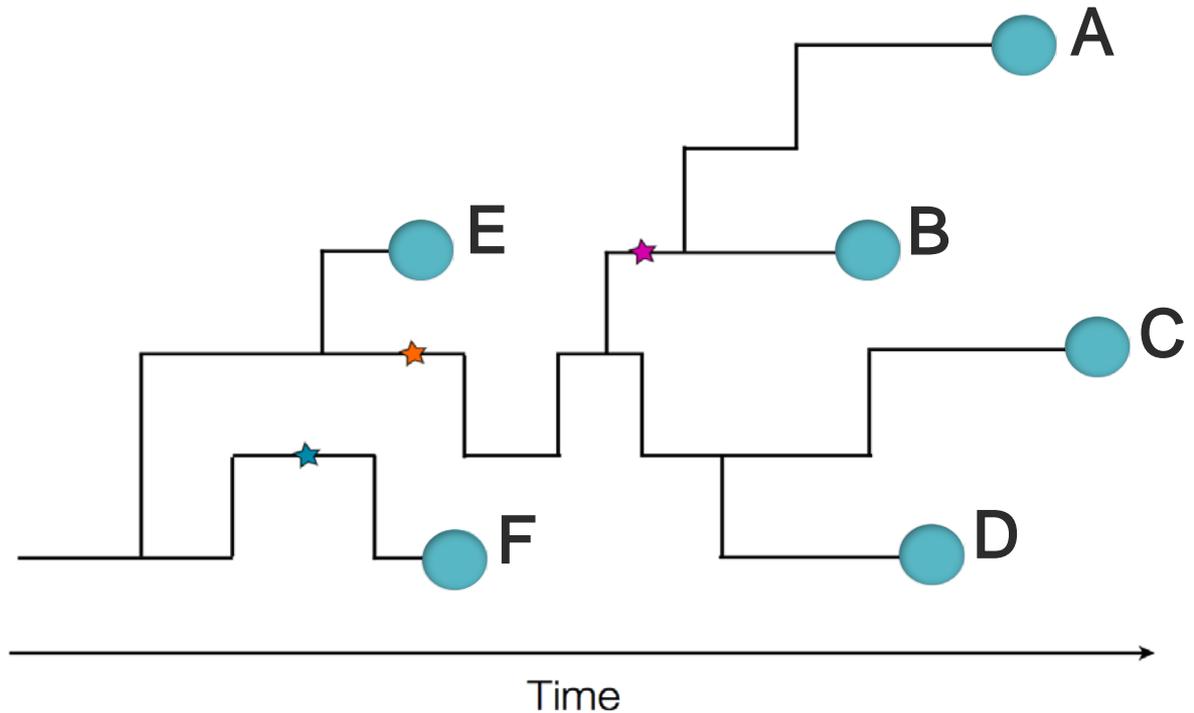
Sample individuals

Genomic sequences are available for only some cases



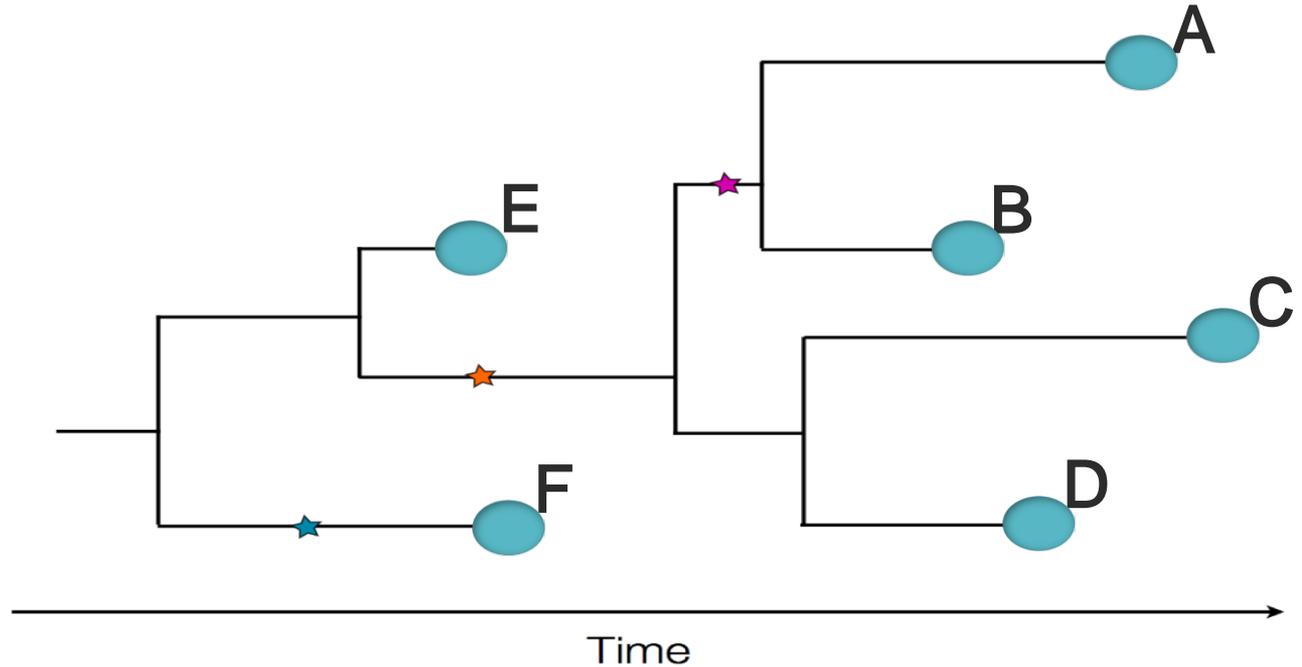
Sequence

By sequencing, connections can be made



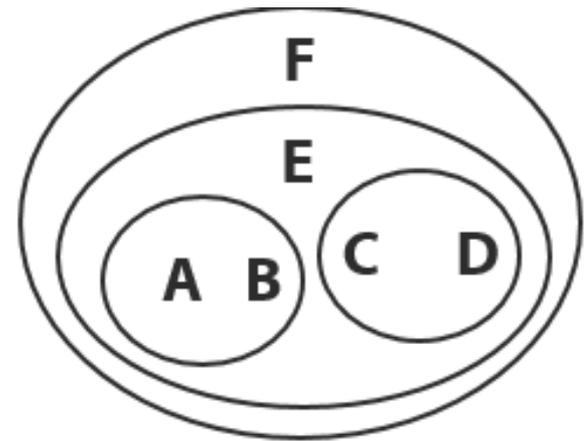
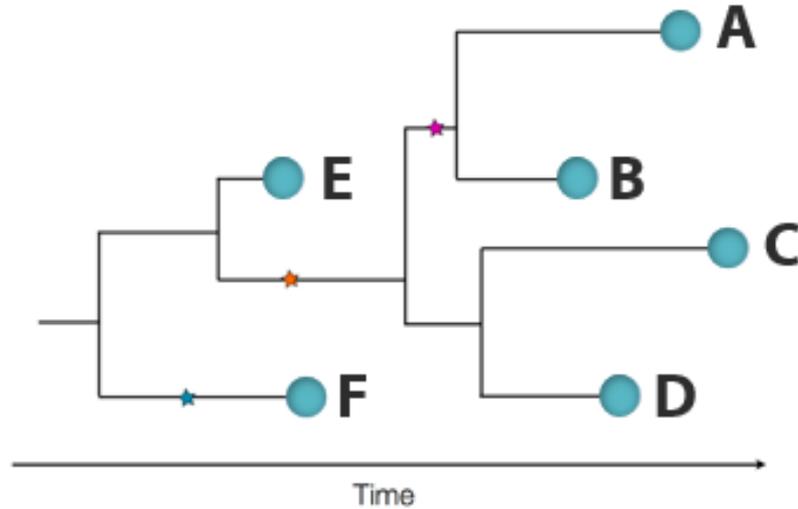
Determine phylogeny

These are then represented in a phylogenetic tree



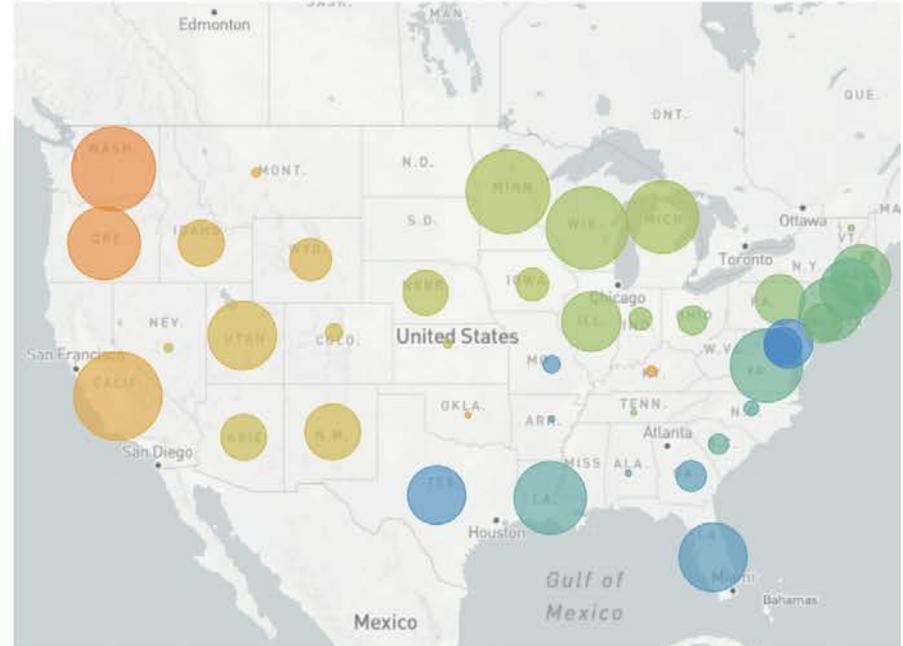
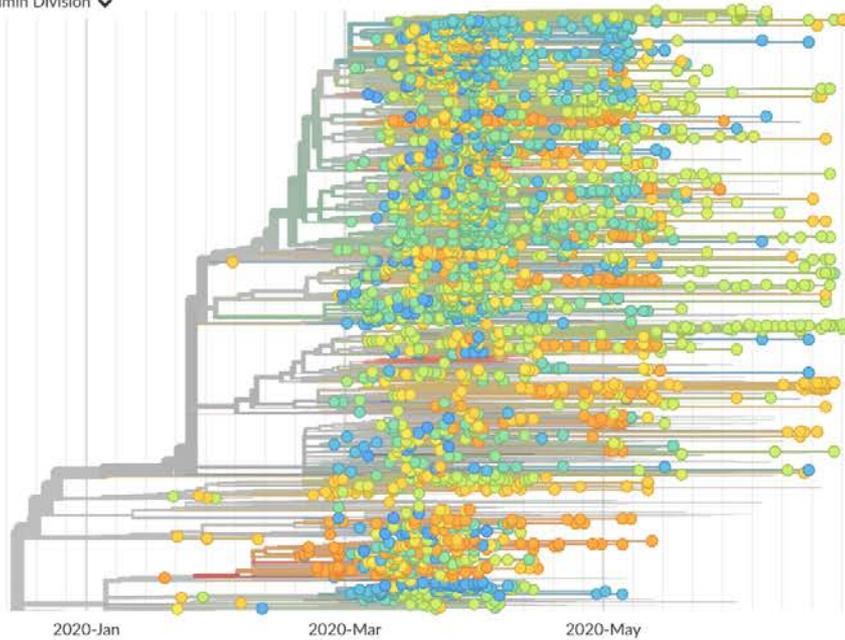
Inferring relatedness

Infers relatedness among the cases



Monitor national trends

Admin Division ▼

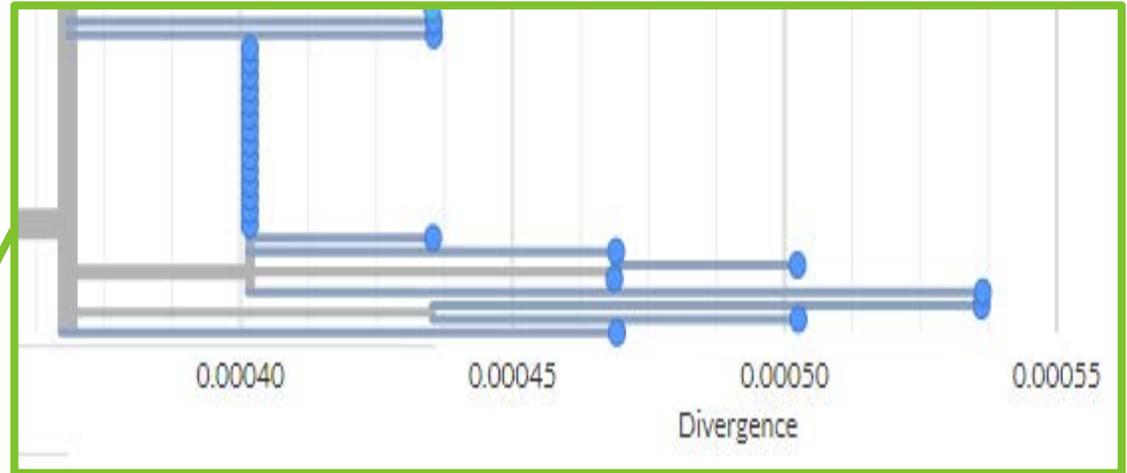
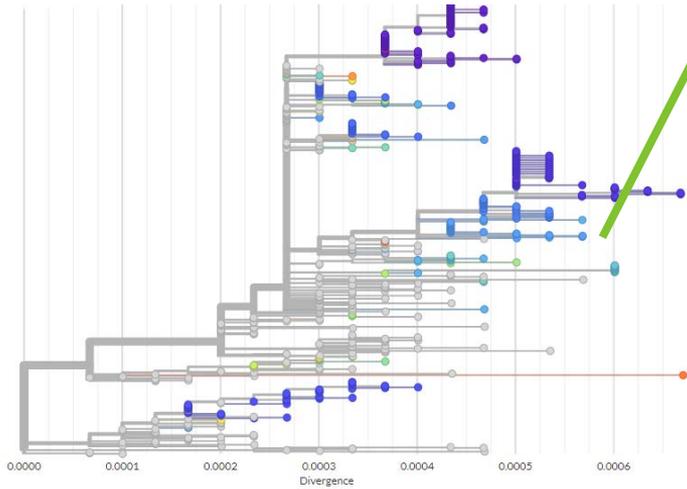


Monitor national trends – UK example

- Led by COVID-19 Genomics UK (COG-UK) consortium involving Public Health Agencies, universities, sequencing centers, & Wellcome Sanger Institute
- Since March 2020, the United Kingdom has used SARS-CoV-2 sequence data for **120** live and retrospective public health outbreak investigations
 - Captured disappearance of specific lineage after Scotland lockdown in spring 2020
 - Captured sudden expansion of a specific lineage linked to relaxation of control measures
 - Performed association studies to see if disease severity varied by lineage

Understand local epidemiology

Facility A



Lehnertz N. Minnesota Department of Health
Trees drawn with Nextstrain

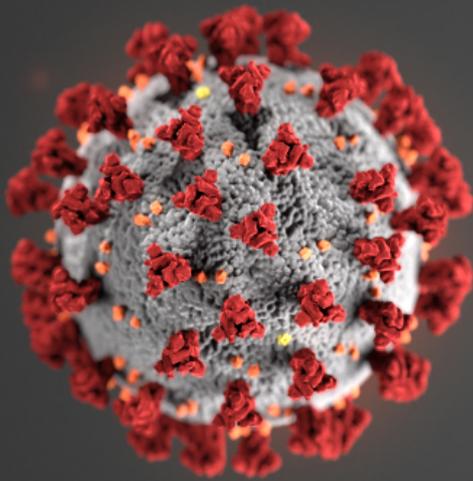
Understand local epidemiology – US examples

- In a Washington state skilled nursing facility, SARS-CoV-2 sequence data helped demonstrate¹:
 - Widespread transmission in the facility
 - Sole focus on symptomatic residents is not sufficient to prevent transmission
- Working with Major League Baseball, public health identified²:
 - A superspreading event within Team A where interactions outside of game play were the likely source of transmission

1. Arons MM, Hatfield KM, Reddy SC, Kimball A, James A, Jacobs JR, et al. Presymptomatic SARS-CoV-2 Infections and Transmission in a Skilled Nursing Facility. *N Engl J Med*. 2020;382(22):2081-90.
2. Murray MT, Riggs MA, Engelthaler DM, Johnson C, Watkins S, Longenberger A, et al. Mitigating a COVID-19 Outbreak Among Major League Baseball Players - United States, 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(42):1542-6.

Learn more

- Other introduction modules
 - The SARS-CoV-2 genomes – Module 1.2
 - How to read a phylogenetic tree – Module 1.3
- COVID-19 Genomic Epidemiology Toolkit
 - Find further reading
 - Subscribe to receive updates on new modules as they are released
go.usa.gov/xAbMw



For more information, contact CDC
1-800-CDC-INFO (232-4636)
TTY: 1-888-232-6348 www.cdc.gov

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

