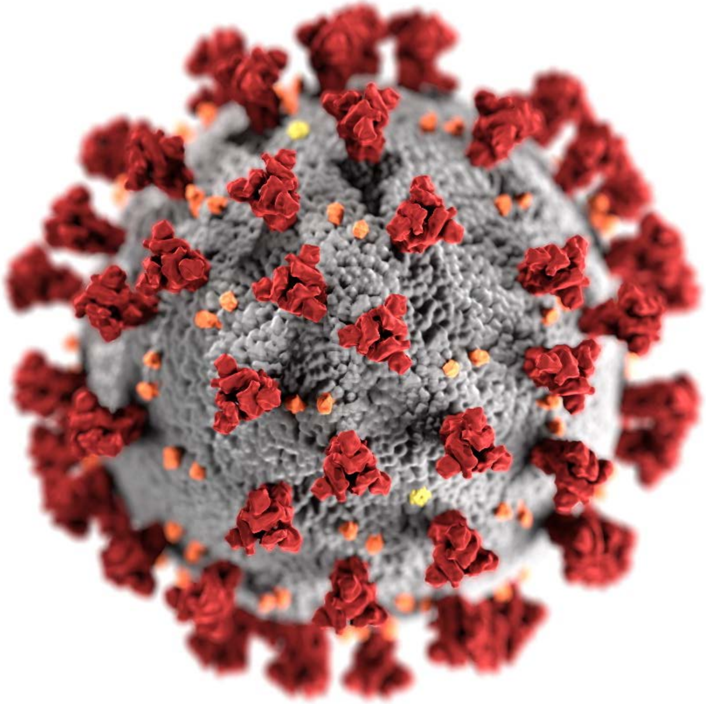
Sequencing strategies for SARS-CoV-2

COVID-19 Genomic Epidemiology Toolkit: Module 3.6

Shatavia S. Morrison, PhD
Bioinformatics Unit 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

1.4 Emerging variants of SARS-CoV-2

Part 2: Case Studies

2.1 SARS-CoV-2 sequencing in Arizona

2.2 Healthcare cluster transmission

2.3 Community transmission

2.4 Superspreading event

Part 3: Implementation

3.1 Getting started with Nextstrain

3.2 Getting started with MicrobeTrace

3.3 Phylogenetics with UShER

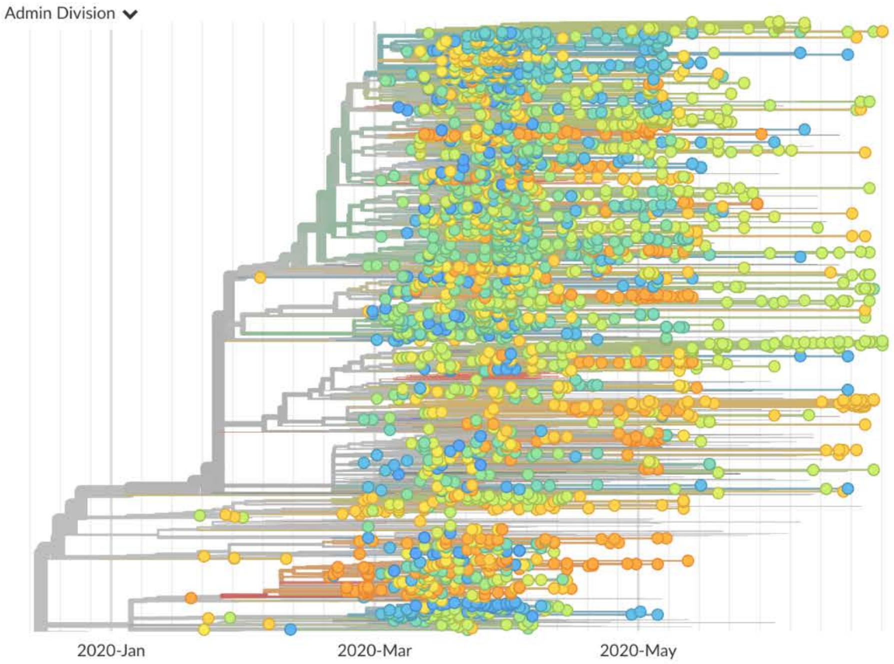
3.4 Walking through Nextstrain trees

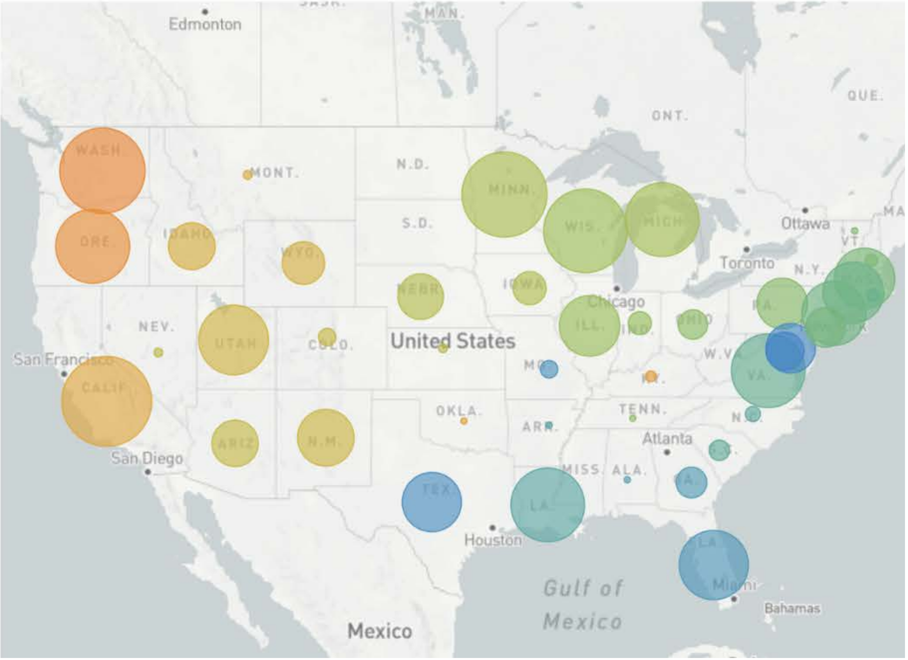
3.5 Public genome repositories

3.6 Sequencing strategies for SARS-CoV-2

National level:

- Monitor emergence of important new strains
- Monitor trends after interventions such as vaccination

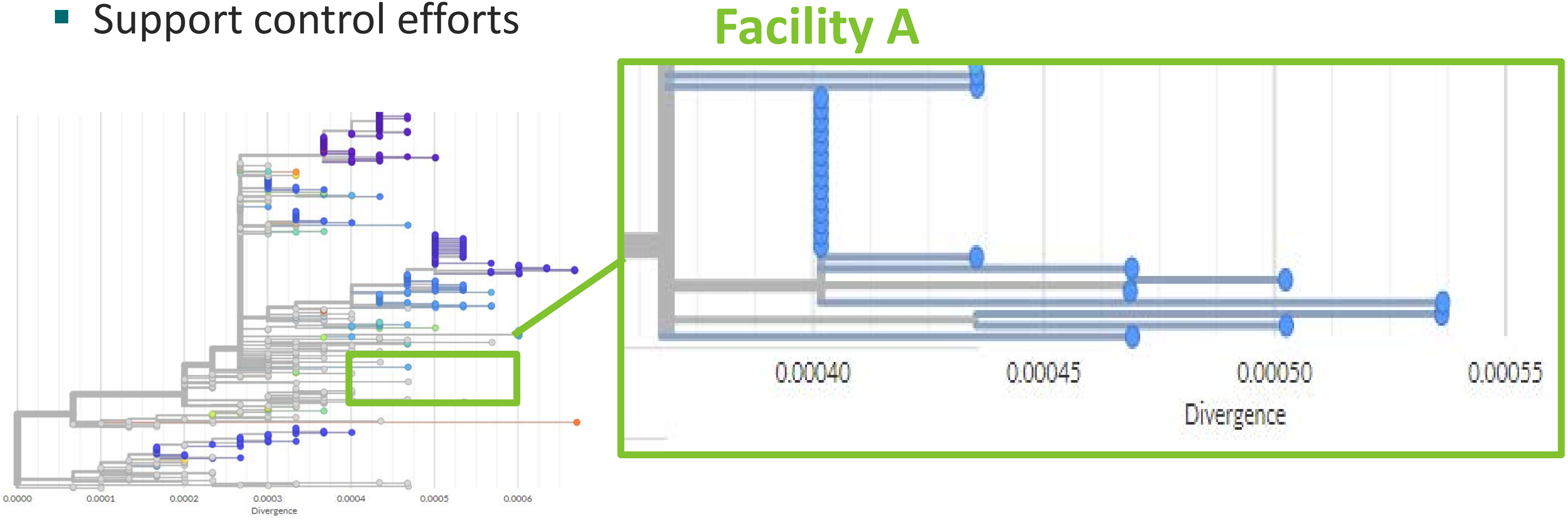




COVID-19 EPI Toolkit Module 0.   Nextstrain.org

State and local levels:

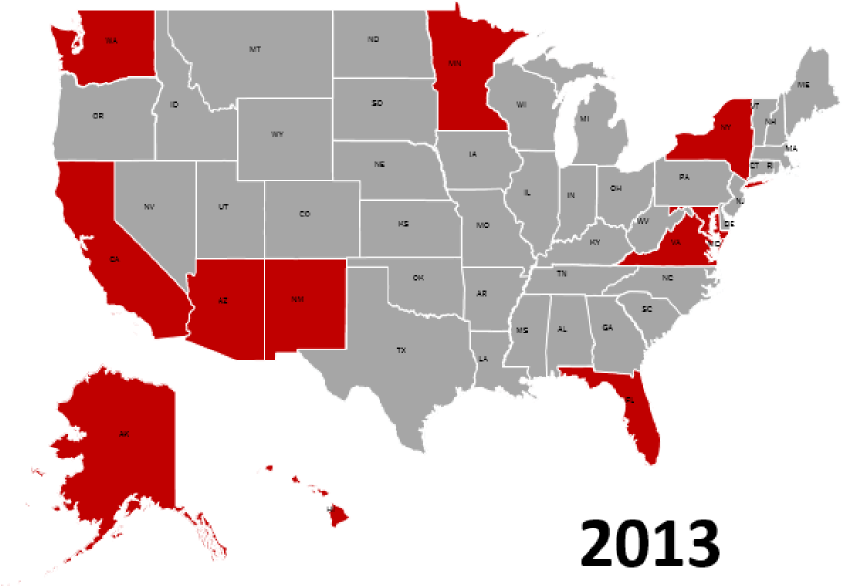
- Identify clusters
- Investigate outbreaks
- Identify Superspreading events
- Support control efforts

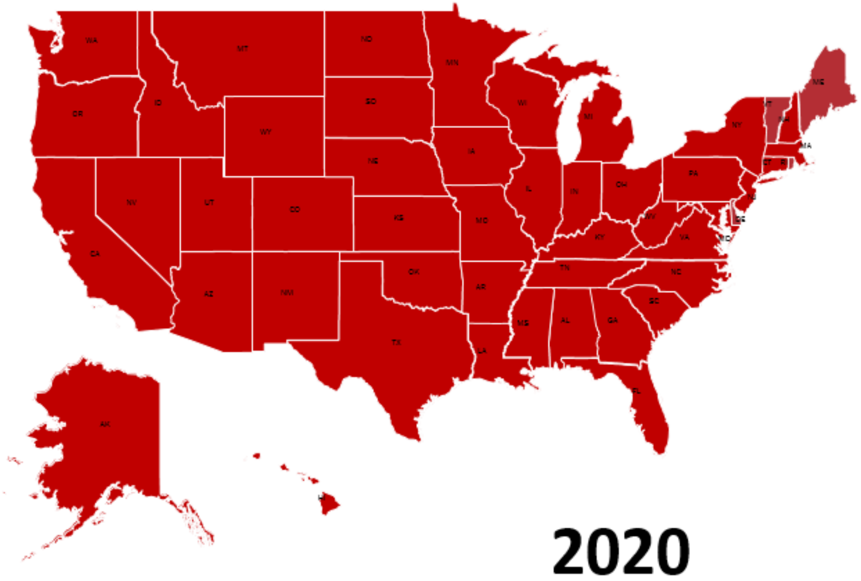


COVID-19 Genomic Epidemiology Toolkit - Module 2.2. Trees drawn with Nextstrain.org

Sequencing capacity in the United Sates

Seven years of building NGS capacity in state public health laboratories





🟥 Sequencing capacity ▨ No sequencing capacity

www.aphl.org/aboutAPHL/publications/Documents/ID_NGSSurveyReport_52015.pdf

Which specimens to sequence?

- Outbreak Investigations
  - High-risk groups, such as congregate living settings
    - eg, skilled nursing facilities, homeless shelters, correctional facilities
  - Super-spreader events
- Surveillance
  - Laboratory-based, for emerging strains and trends
    - eg, S-gene target failure, VOIs or VOCs
  - Epidemiologically defined, for cases of particular interest
    - eg, reinfection, vaccine breakthrough, travel exposure, severe COVID-19 in children

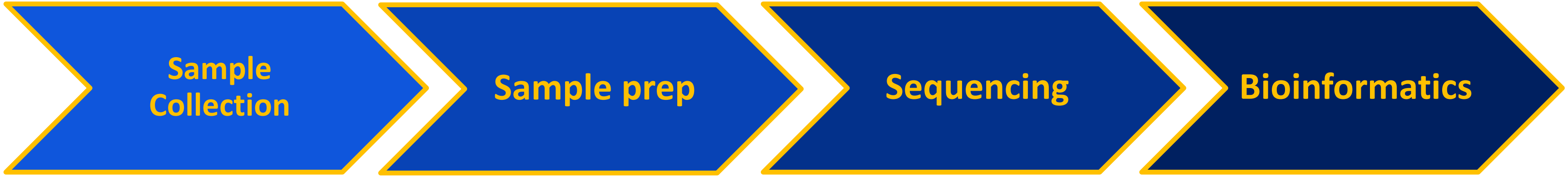
VOI – Variants of Interest, VOC – Variants of Concern : https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/variant-surveillance/variant-info.html

Toolkit case study examples

- Surveillance
  - Laboratory-based
    Module 2.1: SARS-CoV-2 trends in Arizona
- Outbreak Investigations
  - High-risk groups
    Module 2.2: outbreaks in two skilled nursing facilities
    Module 2.3: workplace and community transmission
  - Super-spreader events
    Module 2.4: super-spreading in a pre-symptomatic population

Technical considerations

- Genome Completeness
  - Is sequencing only the S-gene (spike protein) sufficient?
- Ct (cycle threshold) value
  - Ct value ↔ genome sequence recovery
- Laboratory's sequencing capacity



Sample Collection

Sample prep

Sequencing

Bioinformatics

Patient seeks COVID-19 testing, and positive confirmed

1 – 3 Days

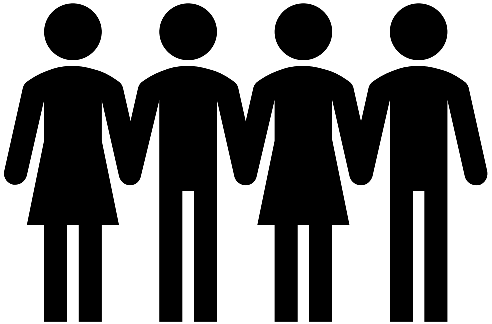
18 – 72 hours

1- 4 Days

https://www.aphl.org/programs/preparedness/Crisis-Management/Documents/APHL-COVID19-Ct-Values.pdf

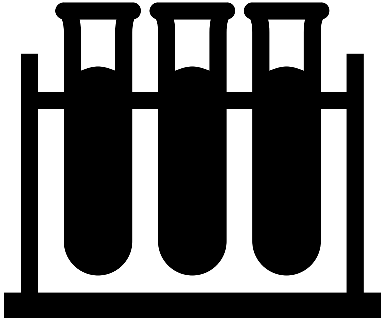
Other factors to consider for sequencing capacity

People



Places

Things



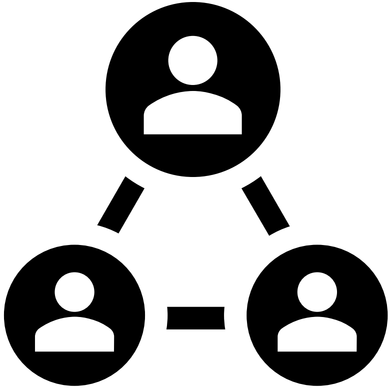
One size does NOT fit all

One size doesn't fit all – Beginner

- No or minimal in-house wet-lab or dry-lab capacity
  - Send samples to labs who have experience with wet-lab and dry-lab sequencing processes

One size doesn't fit all - Intermediate

- Has some wet-lab and dry-lab capacity
  - Has sequencing capacity, *however* lacks high throughput
  - Maybe focus on forming partnerships with labs that have high throughput capacity, such as academic institutions or 3rd party vendors or fellow public health laboratories



One size doesn't fit all - Advanced

- Has high throughput wet-lab and dry-lab capacity
  - Most likely has multiple sequencing instruments
  - Access to scientific computing environments
  - Experience with genomic epidemiological based investigations

Food for thought once sequencing data are available

- Submit your data to a public sequence repository
  - May supplement your local investigation and surveillance with background context from public repository

    Module 3.5: Public genome repositories for SARS-CoV-2

- Perform and apply genomic epidemiology with generated sequence data
  - May provide additional discriminatory power on introduction event of SARS-CoV-2 in outbreak or identify the most prevalent strain circulating within your population

    Module 1.3: How to read a phylogenetic tree

    Module 1.4: Emerging variants of SARS-CoV-2

    Module 2.3: Investigating workplace-community transmission

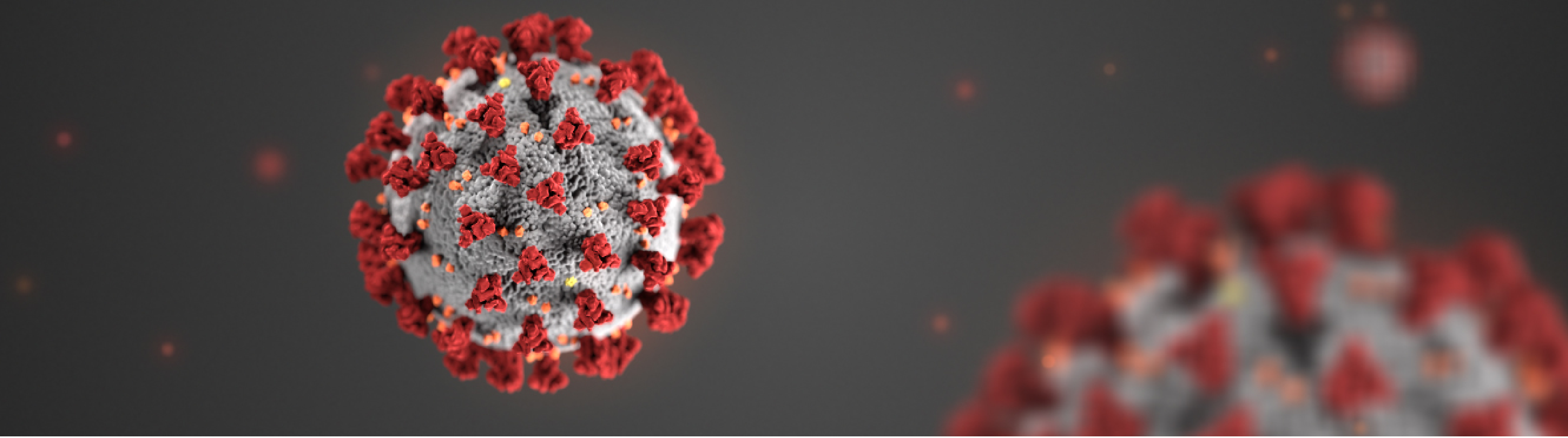
Summary

- Know your motivation for sequencing
  - Are we sequencing outbreaks and/or for surveillance activities?
- Technical considerations
  - What is our sequencing capacity and how it will impact the tasks we need to complete?
- Understanding your capacity
  - How can we get the maximum return for our effort?

Learn more

- Other modules
  - Getting started with MicrobeTrace – Module 3.2
  - Phylogenetics with UShER – Module 3.3
  - Walking through Nextstrain trees – Module 3.4
  - Public genome repositories for SARS-CoV-2 – Module 3.5

- COVID-19 Genomic Epidemiology Toolkit
  - Find further reading
  - Complete a feedback survey
  - 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.