

COVID Tracker

Genomic epidemiology for public health

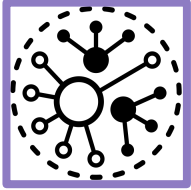
Patrick Ayscue
patrick.ayscue@czbiohub.org



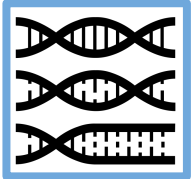
Chan-Zuckerberg Biohub
San Francisco



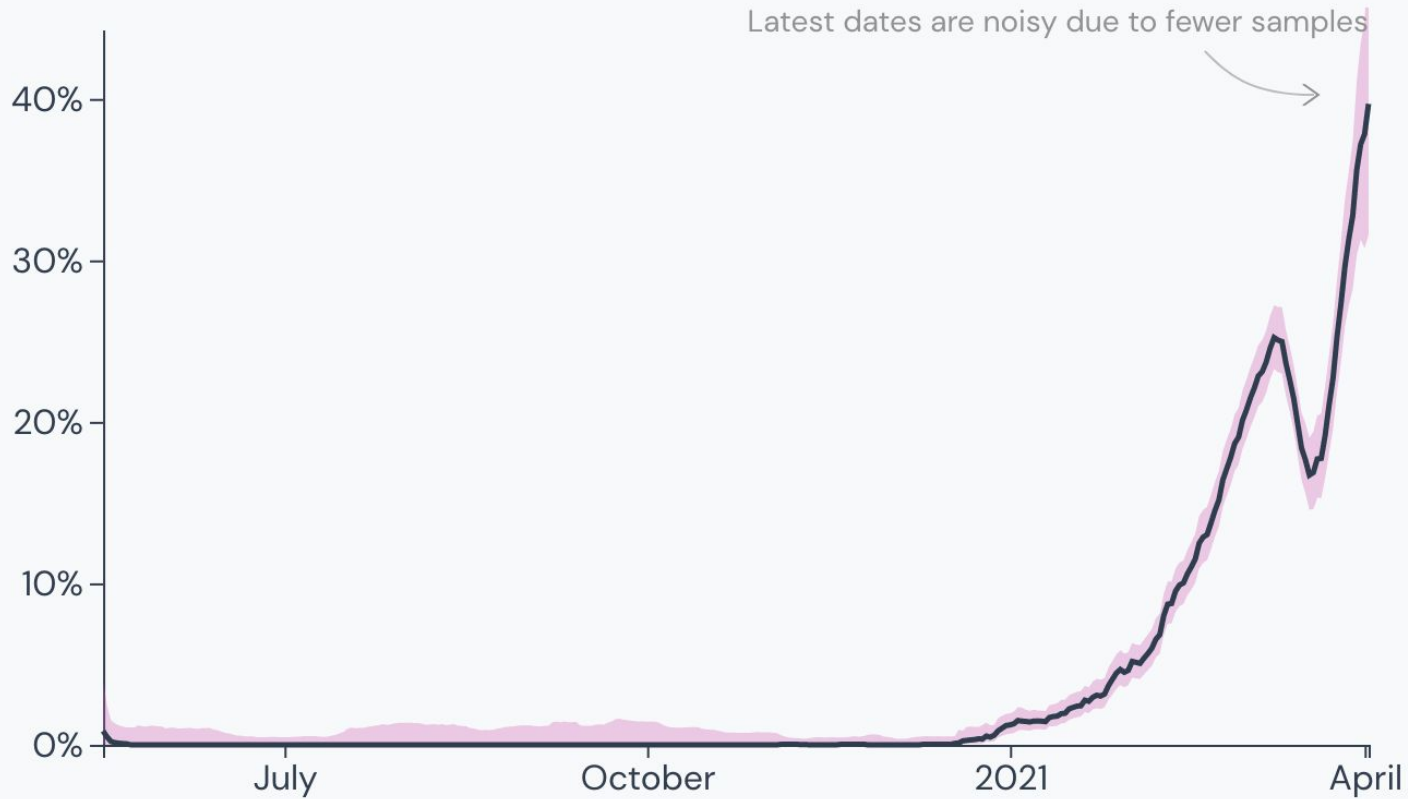
Genomic Epidemiology



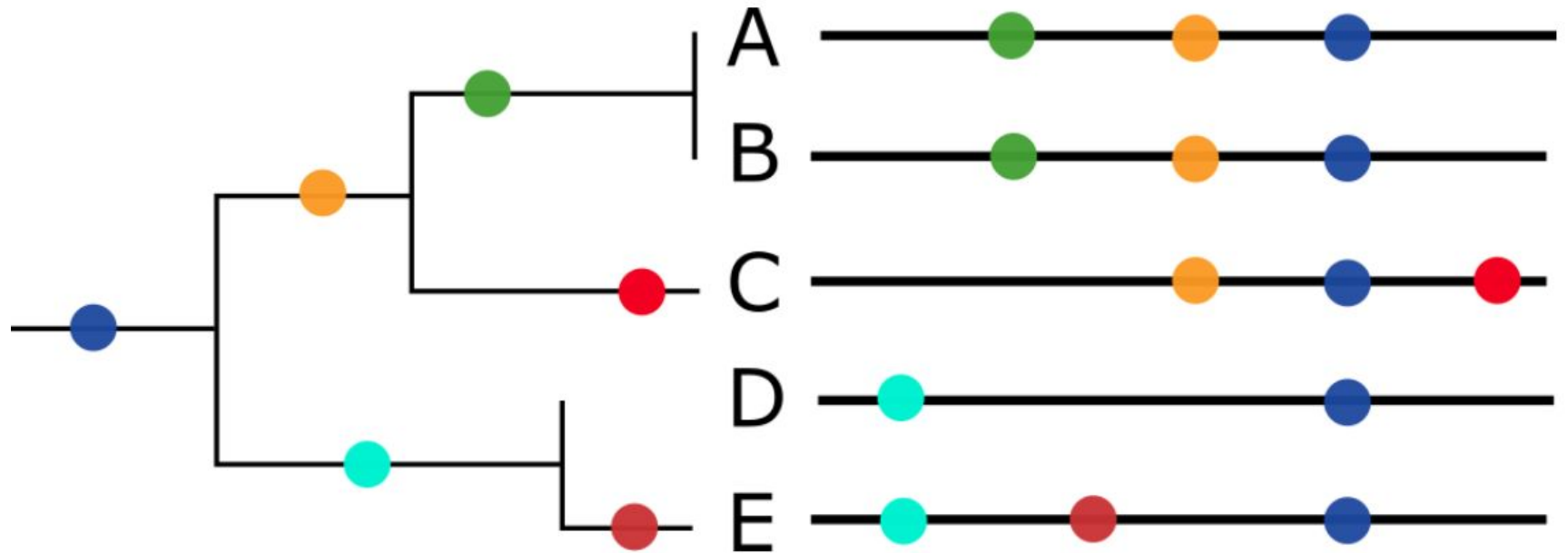
- Track how pathogens move through populations
- Relatively new discipline
 - Techniques have been prominently used in Ebola outbreaks since 2014 and Zika
 - Volume of samples, labs, analyses, users, data footprint all orders of magnitude larger for SARS-CoV-2 than at any point previously
 - Very limited capacity within US public health in Jan 2020
- ‘Sequencing is the easy part’
 - Capacity exists, challenges are in linking metadata, moving samples around, and ensuring quality and interoperability of data



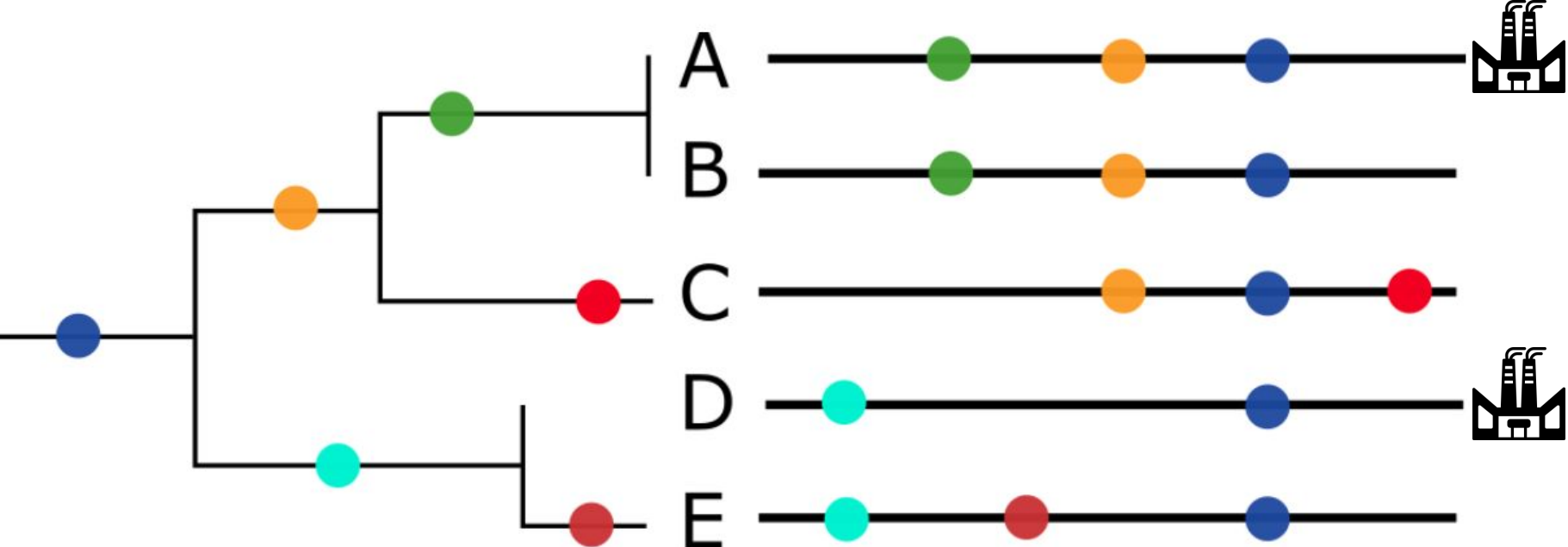
- 7 day rolling average of percent of B.1.1.7-positive sequences
- 95% confidence interval



Shared mutations indicate cases with shared transmission history



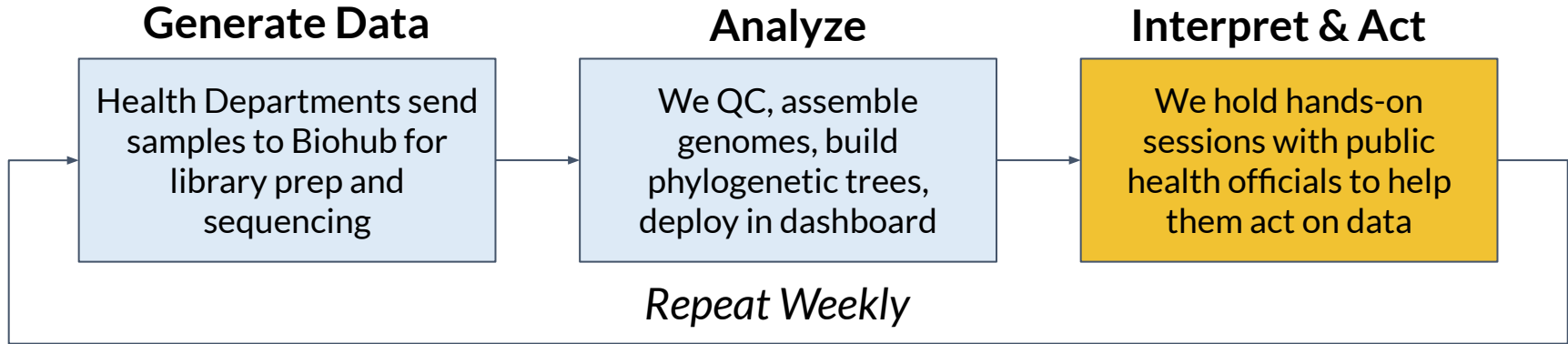
Genetic Links Direct Response



Cases unlinked → Focus interventions outside work setting

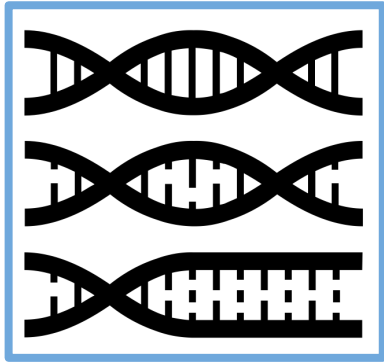
COVID Tracker Workflow

An end-to-end program to empower **local** public health to use **genomic data** in their COVID-19 response.

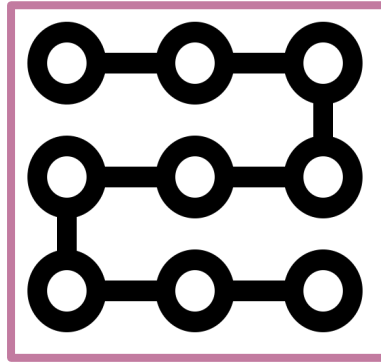


Last mile support ensures effective integration of new technology and we gain insight into user needs, use cases, and barriers to entry.

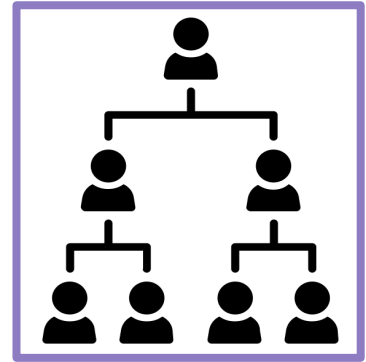
Four Primary Areas to Deliver Actionable Insights



Sequencing



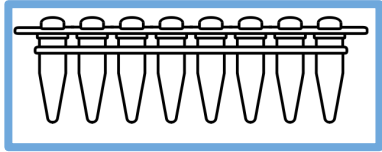
Bioinformatics



Interpretation

Data management

Sequencing



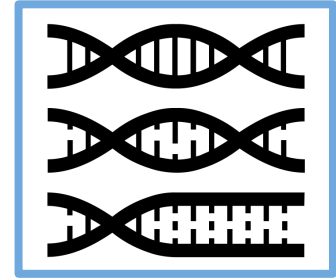
PCR Testing

- Commercial labs
- DPH labs
- Hospitals



**Find physical
sample,
extract RNA**

DPH labs

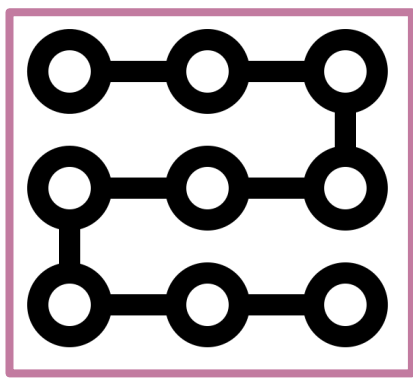
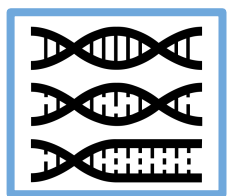


Sequencing

CZB team



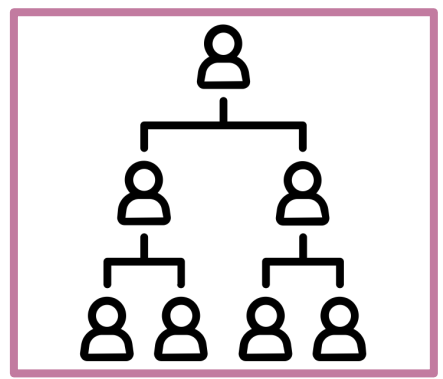
Bioinformatics



**Genome
assembly**



QC

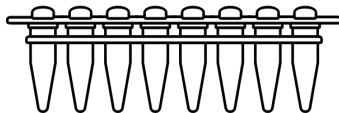


Phylogenetics

CZ Biohub data team

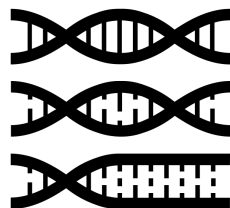


Data management



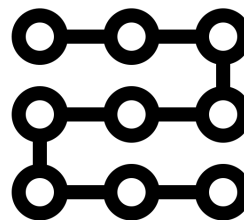
Testing

- Demographics
- Results (Ct vals)
- Source
- External identifiers



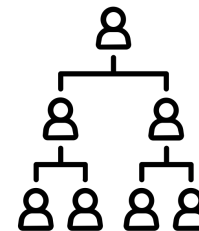
Sequencing

- Plate, well
- Run, personnel
- Internal identifiers



Assembly

- QC metrics
- Public identifiers

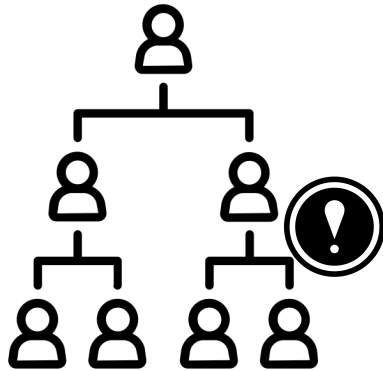


Phylogenetics

- Source
- External identifiers
- Public identifiers

CZI Engineering + CZ Biohub Data

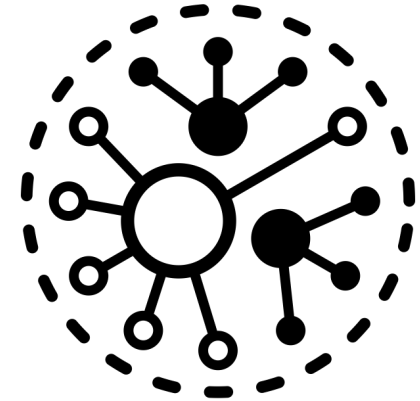
Interpretation & capacity building



Trees



Results calls,
data integration, visualizations,
tutorials, workshops



Meaning

CZB Data + CZI Data



SELECT DATASET

alameda_local_210104 ▾

Search 🔍
Restrict search to...

COLOR BY

County ▾

BRANCH LENGTH

TIME DIVERGENCE

To add additional metadata, drag-and-drop a file onto the page.

DOWNLOAD JSON

POWERED BY

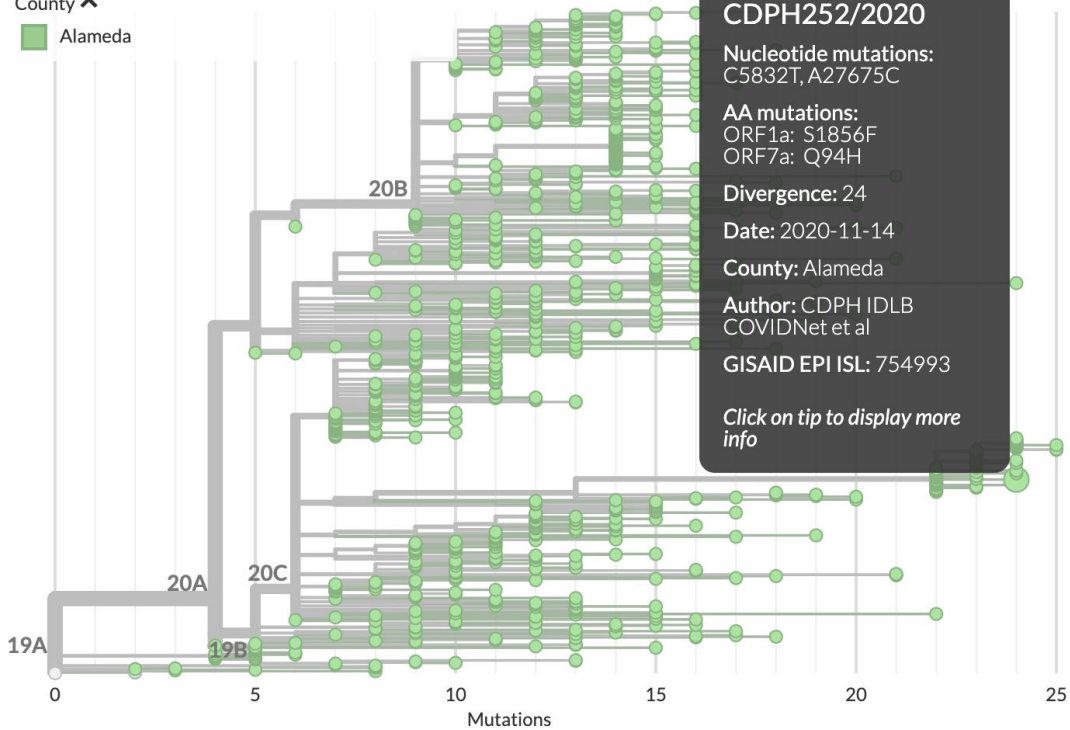


Showing 961 of 961 genomes.

Phylogeny

County ^

Alameda



USA/CA-CDPH252/2020

Nucleotide mutations:
C5832T, A27675C

AA mutations:
ORF1a: S1856F
ORF7a: Q94H

Divergence: 24

Date: 2020-11-14

County: Alameda

Author: CDPH IDLB
COVIDNet et al

GISAID EPI ISL: 754993

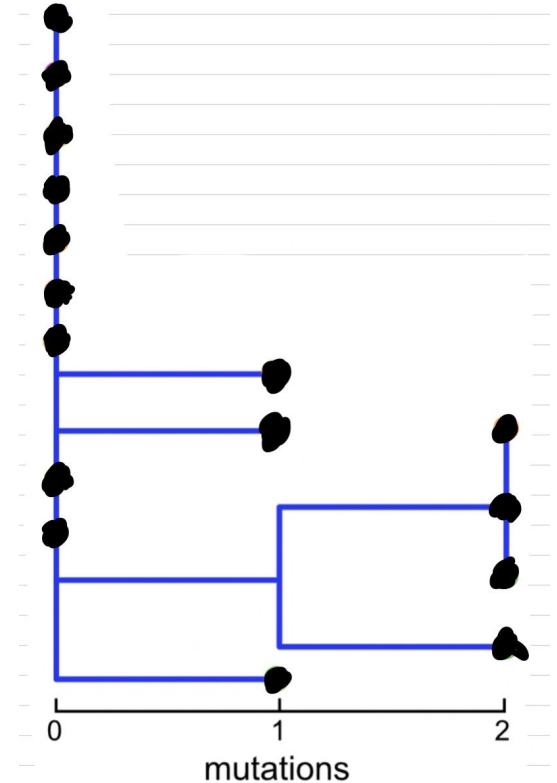
Click on tip to display more info

COVID Tracker Status

- CZB and CZI have partnered with **22 DPHs** in California and generated over **10,000 SARS-CoV-2 genome sequences** to inform pandemic response activities
- These data enable health authorities to
 - identify settings **where transmission is taking place**
 - **take action** to disrupt transmission chains
 - understand **how the virus and variants are moving** through their communities
- **Spent hundreds of hours** leading hands-on weekly training sessions with county DPHs to help interpret their data and make it actionable
- Our data & outreach have **informed local public health response in dozens of county COVID-19 investigations** in factories, farms, schools, prisons, hospitals and communities

Vignette #1: The Facilities That Share Staff

A county flagged multiple outbreaks occurring at several nursing homes.



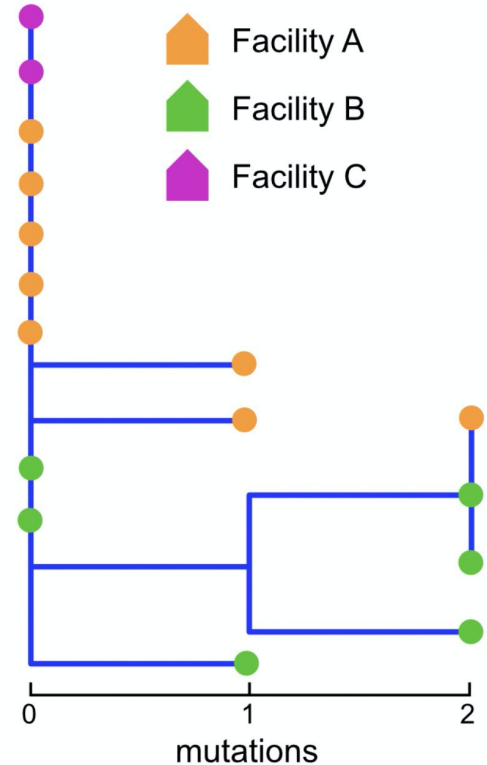
Vignette #1: The Facilities That Share Staff

A county flagged multiple outbreaks occurring at several nursing homes.

Their epidemiologists add facility metadata to the tree.

Conclusion: Transmission is occurring between these facilities

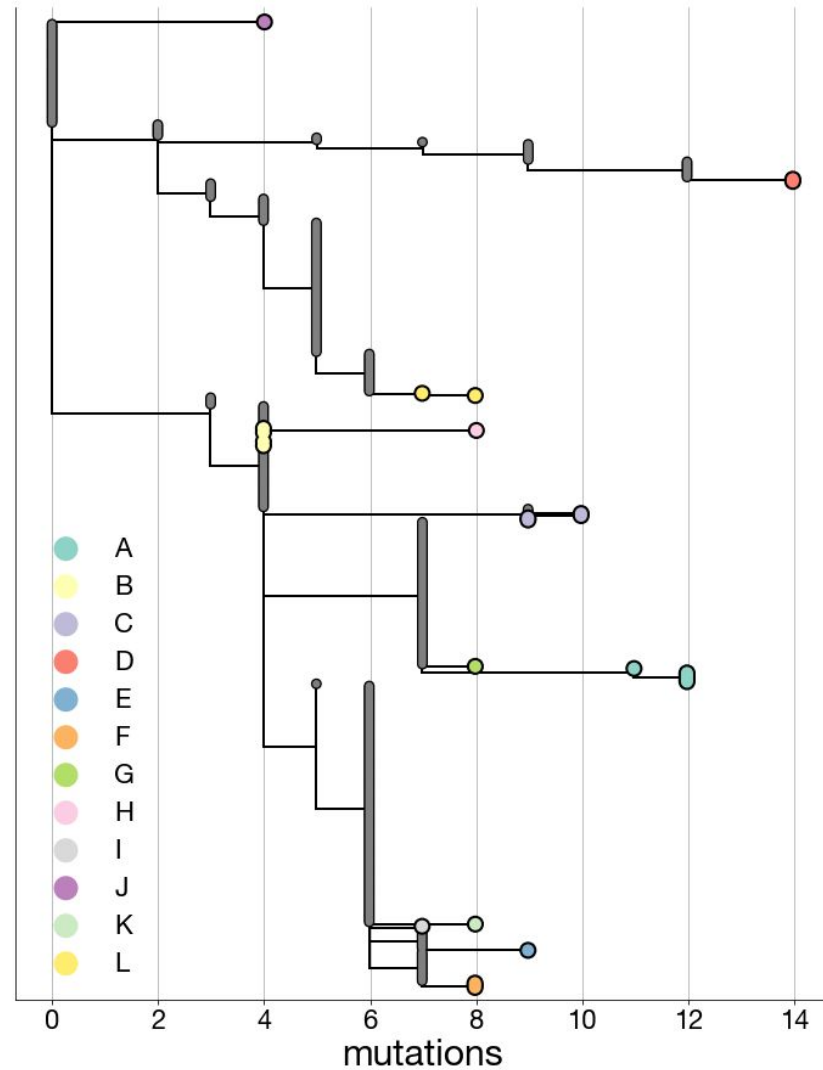
Action: Hard evidence for these private facilities to change practice



Vignette #2: Importations drive cases in rural county

Each case cluster is genetically distinct.

Action: Focus interventions on travelers.

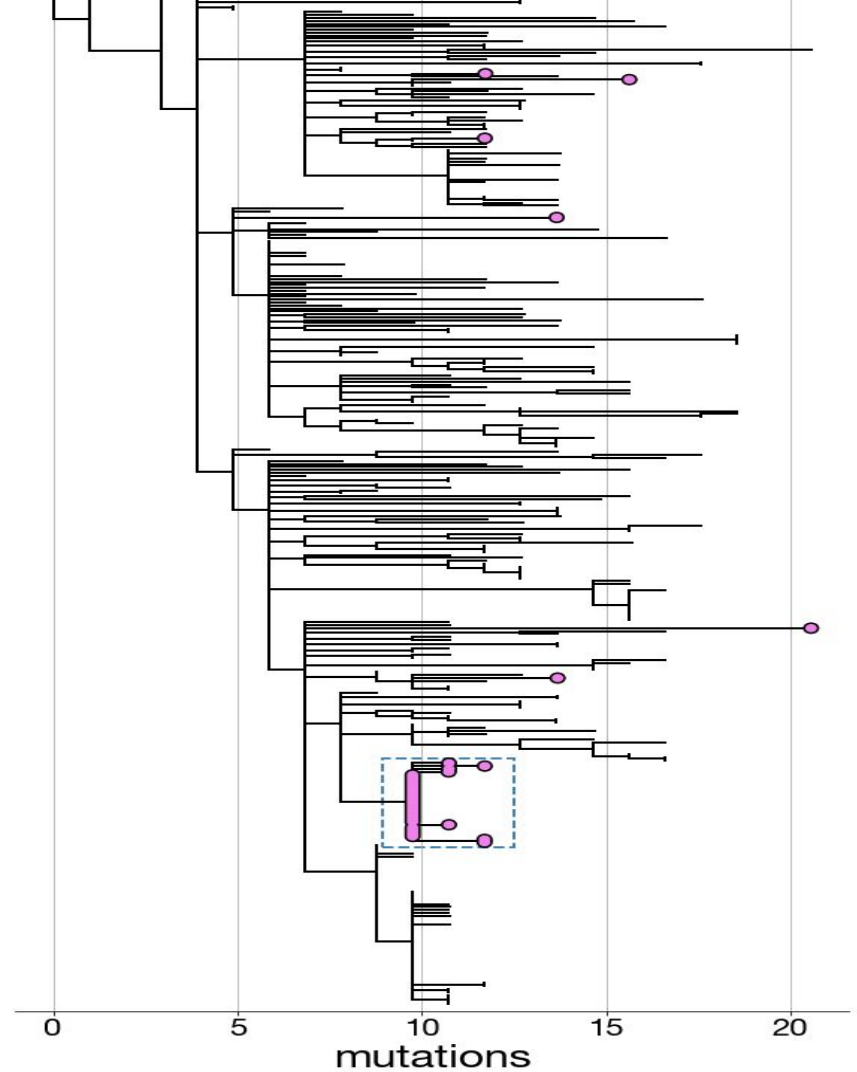


Vignette #3: Successful screening in county jail

Observation: Ongoing outbreak.
Driven by importations or internal transmission?

Inmate screening

Action: Enhance interventions within jail; continue successful intake procedures



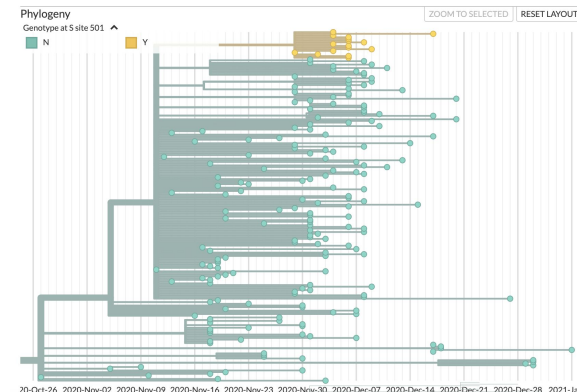
2021 goal

Enable **California Departments of Public Health** to use genomic epidemiology **independently**.

Aim to do this through **technology development and capacity building**.

COVID Tracker Training Status

- Training local and state health departments to generate, analyze, interpret, and act upon sequencing data
- Ongoing assessment of gaps and needs across these groups
 - ~100% have gaps in analysis and interpretation
 - Variable gaps in ability to generate data and act on findings
- Building from lessons learned and best practices from our previous partnership with Gates Grand Challenges standing up similar capabilities in research and clinical laboratories around the world
 - Direct instruction → Inquiry-based learning model



The Importance Of Centralizing Data

Should counties sequence and analyze their own data for their own epis? *

Genomic epidemiology is most effective when a large volume of genomic data from across the entire state is jointly analyzed, compared and made available to both state and local public health officials.

Local health jurisdictions should share and compare their genomic data **across jurisdictions** to understand where locally occurring lineages came from, to quickly contact relevant jurisdictions to investigate inter-county transmission chains and more.

Sequence data and analyses should be rapidly and automatically shared state-wide, regardless of who sequenced or analyzed it.

Top 5 lessons learned



1. Public health in America is deeply fragmented. Top-down programs have had modest success at best.

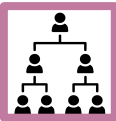
For emerging outbreaks, the front-line work is done by under-resourced municipalities and counties.



2. Long term, **genomic epidemiology is generalizable** if systems are designed accordingly. Most DPHs are also interested in genomic epi for routine practice, beyond COVID.



3. For prospective impact, need **rapid turnaround time** → most DPHs targeting small-batch, on-site sequencing.

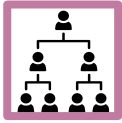


4. **No-code solutions are key.** DPHs typically don't have any computational staff and need accessible analytical and data management tools.



5. Moving positive samples (and their associated data) from **testing to sequencing** is far from trivial.

Landscape moving forward



- Lots of interest
 - Proliferation of tools
 - Federal \$\$ infusion
 - Attempts to define national, regional, state, and local systems for SC2 and beyond



Protocols and Standards emerging

- ARTIC protocols are widely in use for SARS-CoV-2 sequencing
 - <https://artic.network/>
- Nextstrain (augur) pipelines are widely used for generating phylogenetic trees
 - <https://nextstrain.org/>
- GISAID is a primary repository for SARS-CoV-2 sequence data
 - <https://www.gisaid.org/>
- Terra platform is variably used to host and run analytic pipelines
 - <https://terra.bio/>
- Public Health Alliance for Genomic Epidemiology (PHA4GE) aims to set standards in the space, with most uptake currently of metadata ontologies and conventions
 - <https://pha4ge.org/>
- Pangolin lineages are stabilizing as primary nomenclature
 - <https://cov-lineages.org/pangolin.html>

Thank you!

patrick.ayscue@czbiohub.org