TEXAS
Health and Human
Services

Texas Department of State
Health Services

TEXAS Health and Human Services | Texas Department of State Health Services

Pathogen Genomic Epidemiology Training Series

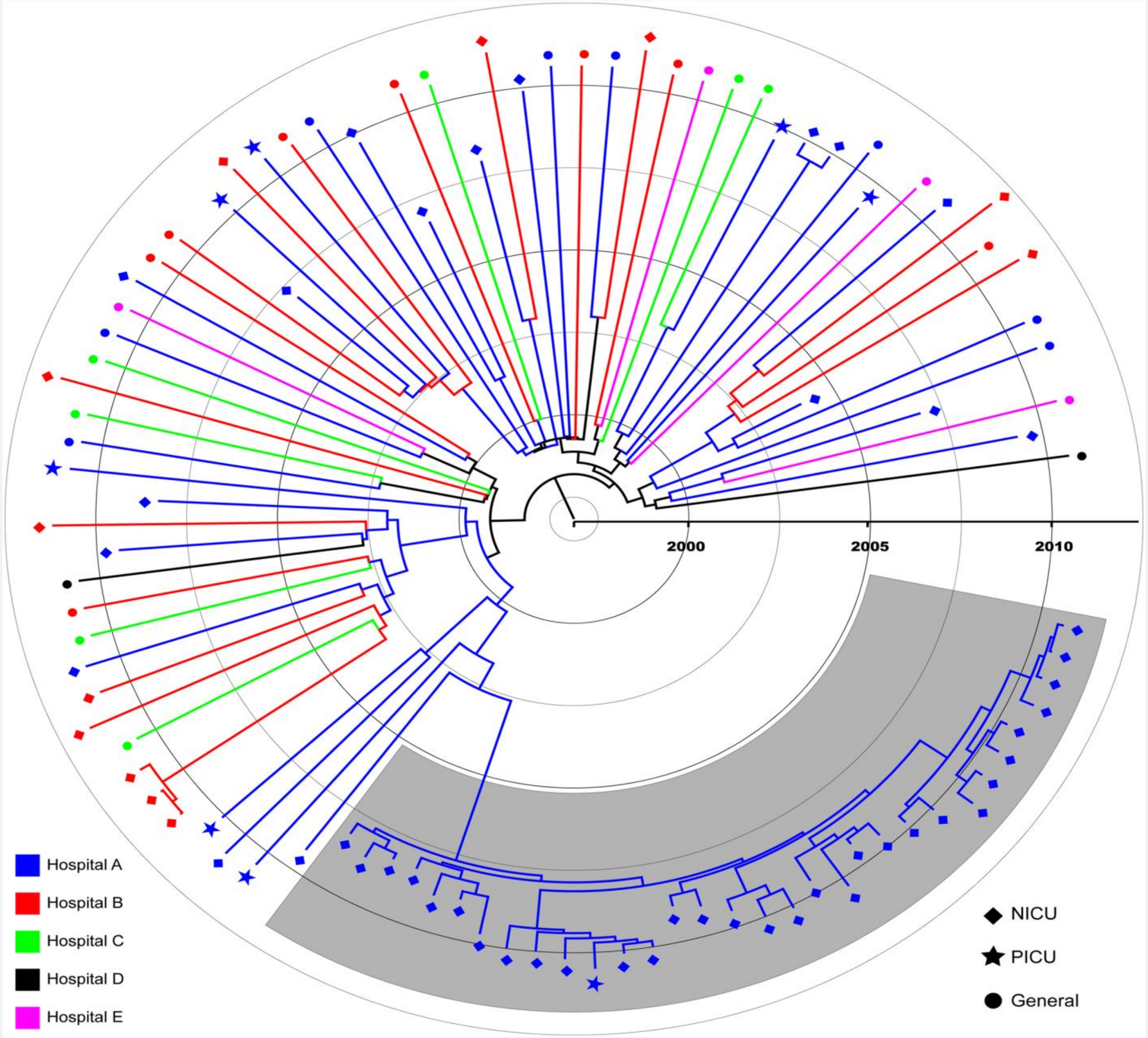


Image from doi: https://doi.org/10.1371/journal.pone.0164397.g002

Continuing Education Credit

The Texas Department of State Health Services Laboratory is approved as a provider of continuing education programs in the clinical laboratory sciences by the ASCLS P.A.C.E. program.

AMERICAN SOCIETY FOR CLINICAL LABORATORY SCIENCE ®
P.A.C.E.
JAN 1, 2023-DEC 31, 2023

This program is approved for 1.5 P.A.C.E. ® credits.

TEXAS
Health and Human Services

Texas Department of State Health Services

Learning Objectives

- Understand why genomic data is better at ruling out linkages among cases than identifying linkages
- Recognize how genomic data can help identify independent disease introductions that may not appear related from standard epi curves
- Understand how differentiating between transmission chains & their characteristics highlights how each contributes to an outbreak, and may help suggest why
- Be aware that genomic data shows it only takes one introduction for a large outbreak to arise and spread rapidly
- Understand how WGS of paired specimens can confirm pathogen reinfection in an individual

About the presenter

Anna Battenhouse, B.S., B.A.

Associate Research Scientist, The University of Texas at Austin

Center for Biomedical Research Support

Vice President for Research, Scholarship and Creative Endeavors

- Manager, Biomedical Research Computing Facility (BRCF)
- Member, Bioinformatics Consulting group (BCG)
- Staff scientist, lab of Dr. Edward Marcotte

Anna is a research scientist in the lab of Dr. Edward Marcotte, is a member of UT Austin's Bioinformatics Consulting Group, and leads the Biomedical Research Computing Facility in its mission to support the IT and computational needs of the biomedical research community. She has extensive experience working with Next Generation Sequencing (NGS) data, develops and maintains NGS analysis scripts and workflows, and teaches the Introduction to NGS Tools course in the CBRS Big Data in Biology Summer School as well as several CBRS short courses.

Anna received a B.A. in English Literature from Carleton College in 1978. After a long career in commercial software development, Anna began her "retirement career" in functional genomics in the lab of Dr. Vishy Iyer in 2006, and obtained a B.S. in Biochemistry from UT Austin in 2013.

Representative Genomic Epidemiology Case Studies

DSHS Genomic Epidemiology Training Series
Module 2.2

Module 2.2 Outline

Representative GenEpi Case Studies

1. Genomic Epidemiology principles
2. Identifying SARS-CoV-2 clusters in a SNF (skilled nursing facility)
   - https://dx.doi.org/10.15585/mmwr.mm6937a3
   - CDC COVID-19 Genomic Epidemiology Toolkit module 2-2 https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html
3. Identifying introductions and transmission in mumps outbreaks
   - https://elifesciences.org/articles/66448
   - https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000611
4. Confirming COVID19 reinfection with WGS (whole genome sequencing)
   - CDC COVID-19 Genomic Epidemiology Toolkit module 2-5, https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html
5. Other Gen Epi examples in brief

TEXAS
Health and Human Services

Texas Department of State Health Services

We gratefully acknowledge the sources below for providing background and content material for this presentation

- CDC COVID-19 Genomic Epidemiology Toolkit
  - https://www.cdc.gov/amd/training/covid-toolkit/, especially
    - Module 1.1: What is Genomic Epidemiology?
    - Module 2.2: Identifying transmission in a healthcare cluster
    - Module 2.4: Confirming SARS-CoV-2 reinfection with WGS
- The Chan-Zuckerberg BioHub COVID tracker seminar series
  - https://covidtracker.czbiohub.org/resources, especially
    - Seminar 15: Genomic Epidemiology of mumps virus
- "An applied genomic epidemiological handbook"
  - *Allison Black and Gytis Dudas, 2022-05-16* https://alliblk.github.io/genepi-book/intro.html, especially
    - Chapter 5: Broad use cases
    - Chapter 6: Case studies

Genomic Epidemiology Principles

Section 1

Why is pathogen genomic data usable in epidemiology?

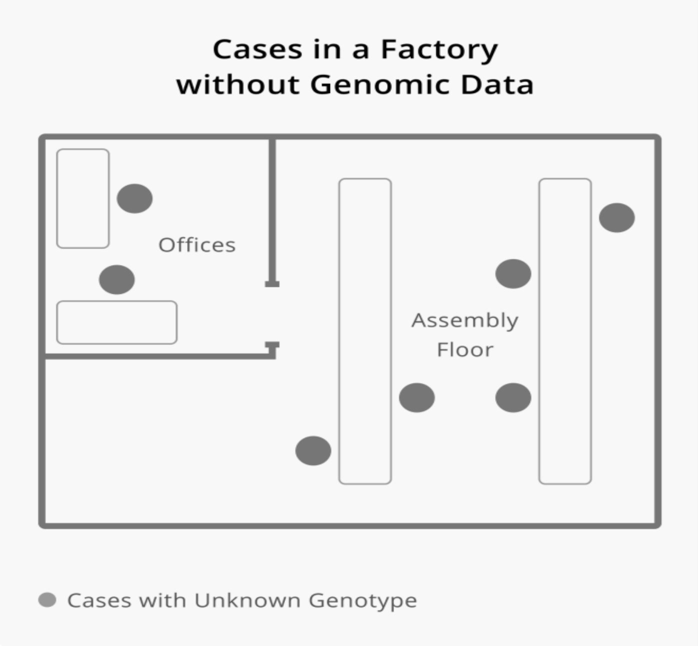
1. *Pathogens evolve on* ***roughly*** *the same time scales as they circulate through a population of hosts*

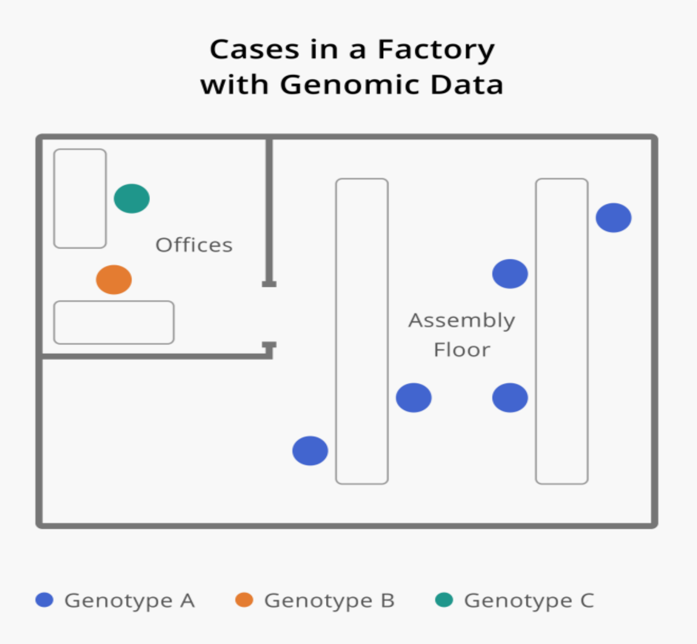
2. *Pathogens with greater genetic similarity are* ***more likely*** *to share an epidemiological association*

https://alliblk.github.io/genepi-book/fundamental-theory-in-genomic-epidemiology.html#the-overlapping-timescales-of-pathogen-evolution-and-pathogen-transmission

Genomic epidemiology applications

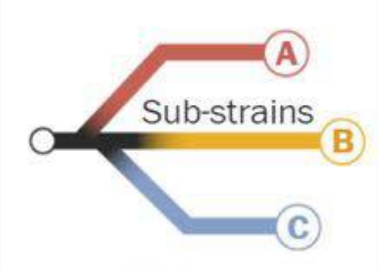
- *Surveillance*
- *Retrospective analysis*
- ***Outbreak response***
  - Classify which cases form an outbreak cluster
  - Assess linkage among cases
  - Explore relationships between cases of interest and other sequenced infections
  - Assess how demographic, exposure, and other epidemiological data relate to a genomically-defined outbreak



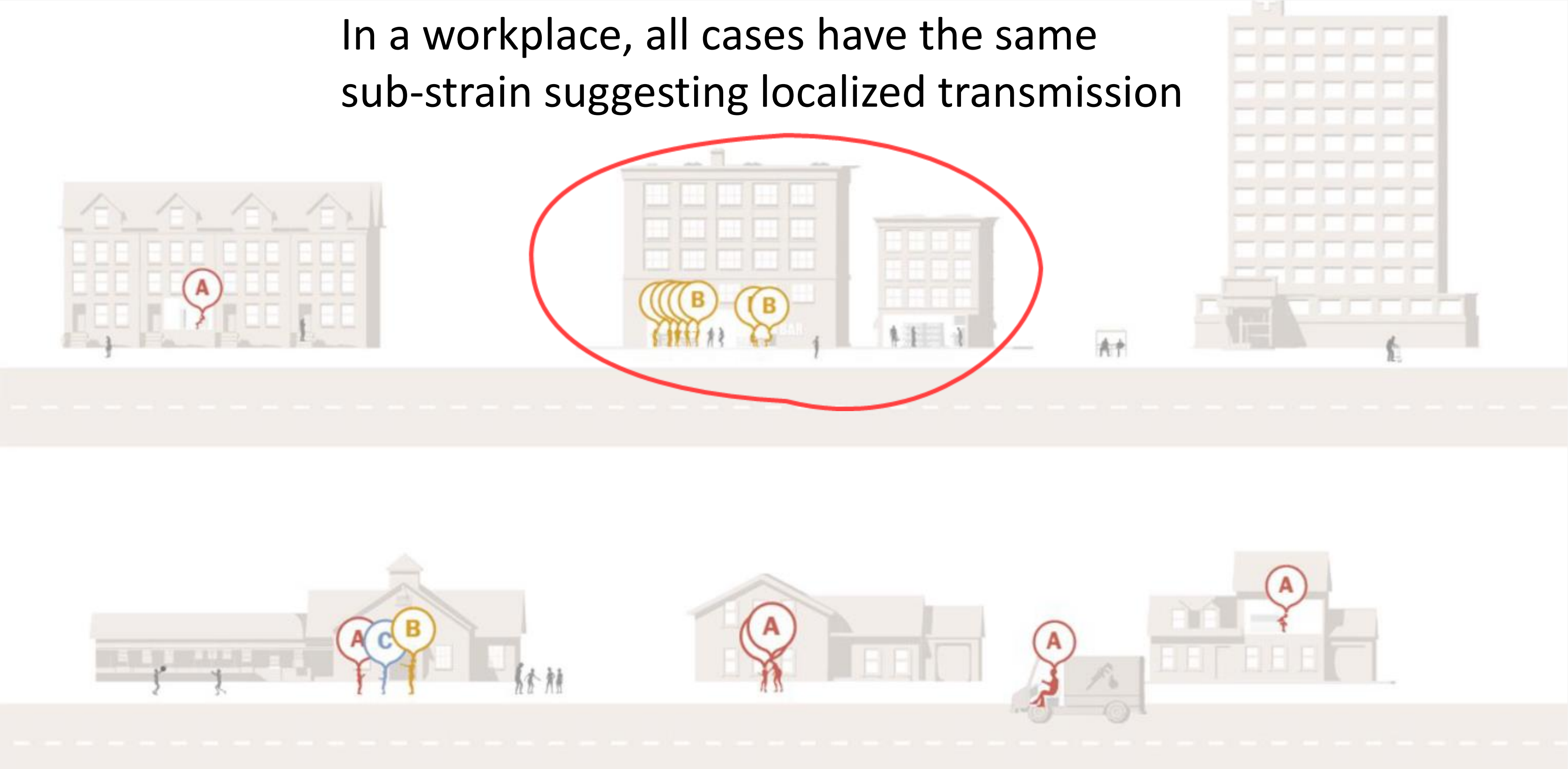


https://alliblk.github.io/genepi-book/the-value-of-pathogen-genomics-in-applied-epidemiology.html#the-value-of-genomic-epidemiology-for-outbreak-response

Clusters:
Is transmission occurring in a hotspot?

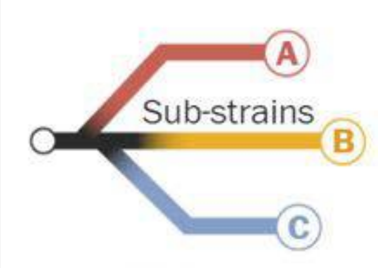


A dense, localized cluster of a single pathogen strain can indicate a hotspot

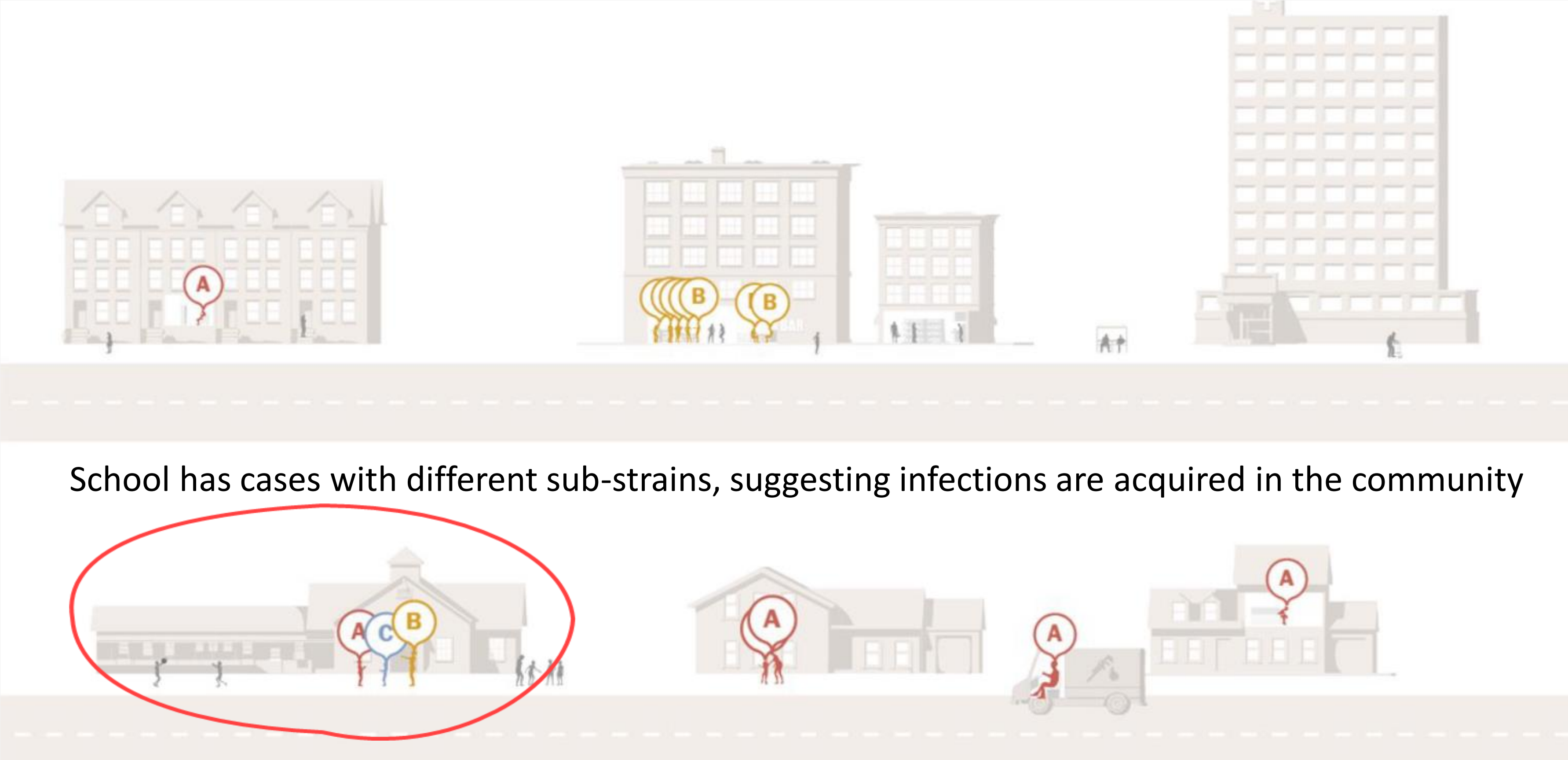

In a workplace, all cases have the same sub-strain suggesting localized transmission

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

Clusters:
Are there distinct introductions/ onward transmission?



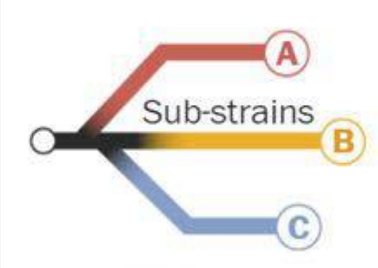
Pathogen genome data can rule out transmission within an apparent cluster and can identify multiple introductions



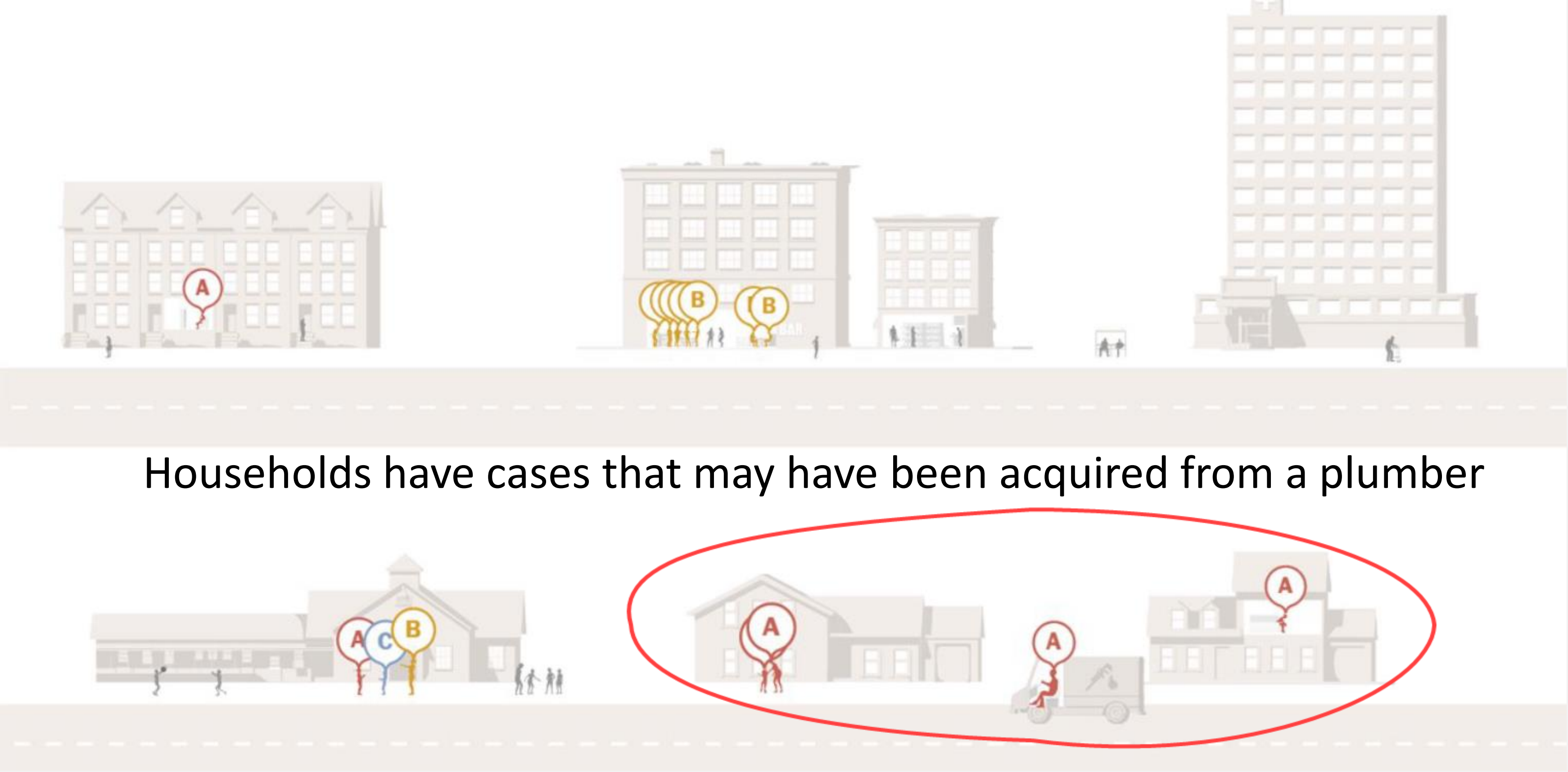
School has cases with different sub-strains, suggesting infections are acquired in the community

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

Clusters:
Are cases related?



Genomics can provide information on whether a case may have transmitted a pathogen to exposed people



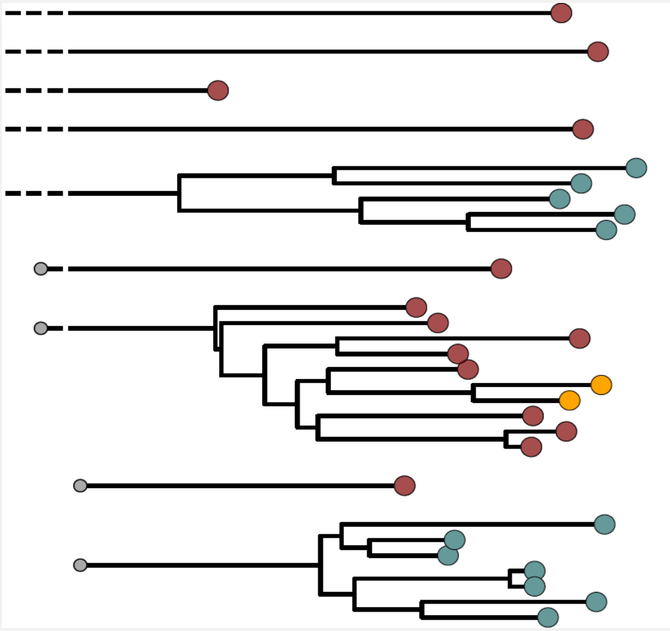
Households have cases that may have been acquired from a plumber

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

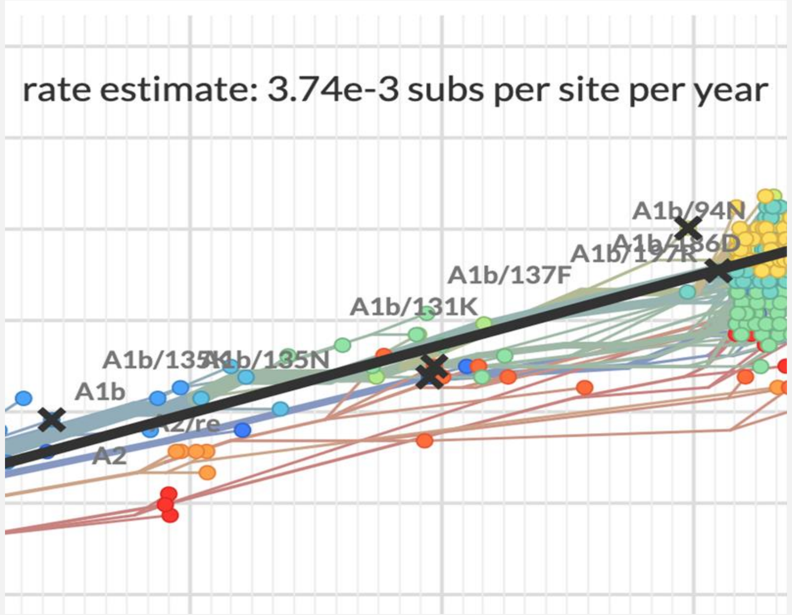
Transmission analysis

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

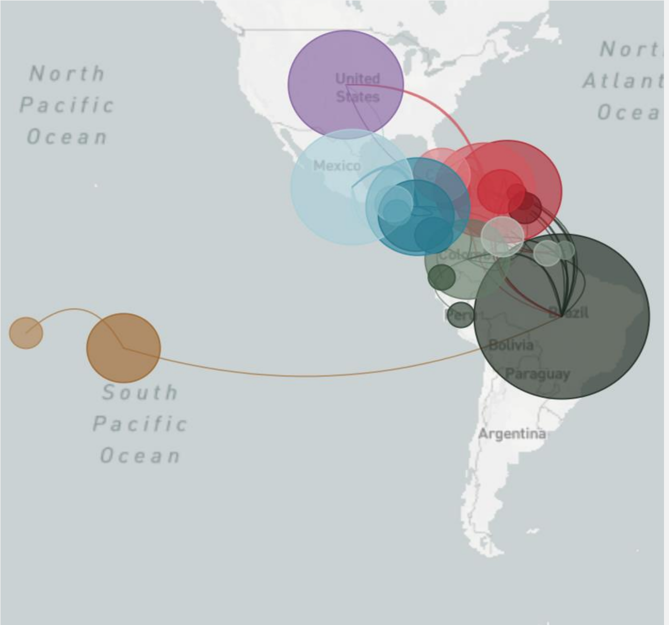
Introductions, onward transmission



Evolutionary rate



Spatial patterns and source-sink dynamics



CZBioHub COVID Tracker Seminar series https://covidtracker.czbiohub.org/resources, Week 1

Broad Genomic Epidemiology principles

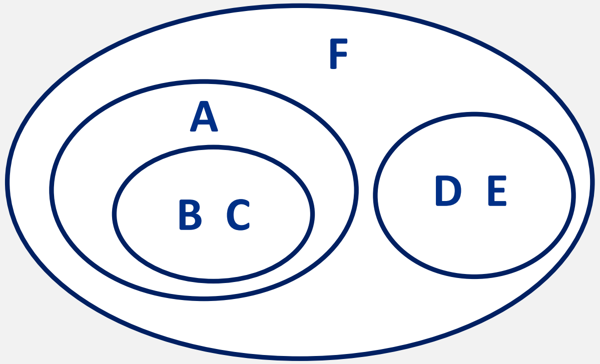
- More infections mean more generated genetic diversity
  - The amount of diversity seen is a proxy for the size of the outbreak
- Samples from cases closer in a transmission chain show similar genetic diversity
  - Cases more distantly linked will share some genetic ancestry but will usually have more unique substitutions
- The combination of shared genetic history and minimal unique substitutions usually describes closely related cases
- Observations of highly divergent populations usually indicate separate introductions from other circulating pathogen pools

Ruling out *vs* Ruling in linkages

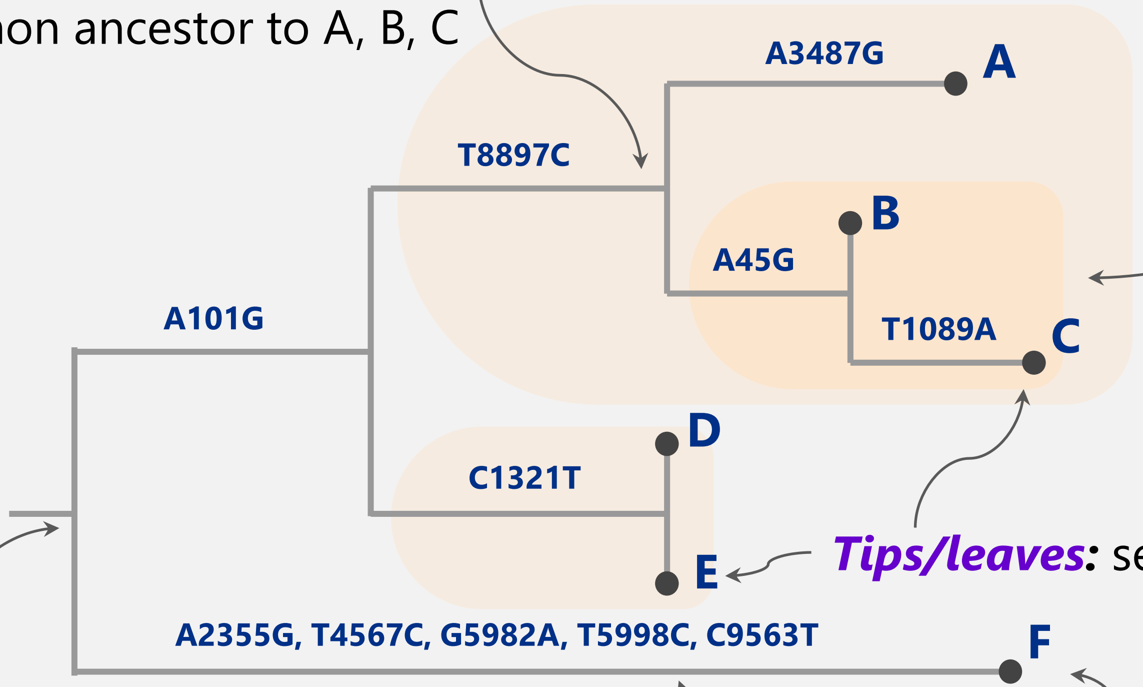
- Because of inherent limitations in genomic sequencing data (especially in sampling), Gen Epi is generally better for ***ruling out*** direct linkages than confirming them
  - While changes to pathogen consensus genomes occur on *similar* timescales to transmission, they are not fundamentally linked
- ***Example***: Two cases of COVID19 in a household, symptom onset 3 days apart
  - But sequenced consensus genomes appear quite diverged
    - Molecular clock estimates that it would take ~6 months to accumulate this amount of variation
  - ***Divergence*** between the consensus genomes allows you to easily ***rule out*** linkage between them
- ***Example***: Same household pair, but consensus genomes are identical
  - Infections are likely – but not necessarily! – related
    - Genomic data cannot resolve who infected whom
  - And was there even transmission between the 2 cases?
    - Maybe both were infected at a party they both attended
  - So cannot definitively ***rule in*** linkage, although contact tracing may help clarify

https://alliblk.github.io/genepi-book/fundamental-theory-in-genomic-epidemiology.html#why-is-sequencing-better-at-dismissing-links-than-confirming-them

Phylogenetic tree anatomy



Pairwise divergence
the **sum** of branch lengths between two tips



Internal node*:** (inferred***) common ancestor to A, B, C

Clades/lineages*:** group of closely related sequences (monophyletic*** if they share a common ancestor)

***Tips/leaves*:** sequenced samples

***Root*:** common ancestor to whole tree

***Branch*:** length represents genetic distance (# mutations) or time

***Outgroup*:** less related sequence, not part of outbreak cluster

https://alliblk.github.io/genepi-book/fundamental-theory-in-genomic-epidemiology.html#assessing-and-reading-a-phylogenetic-tree
CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 1-3

Identifying viral clusters in a SNF

Section 2

COVID-19 outbreaks at 2 skilled nursing facilities

April – June 2020

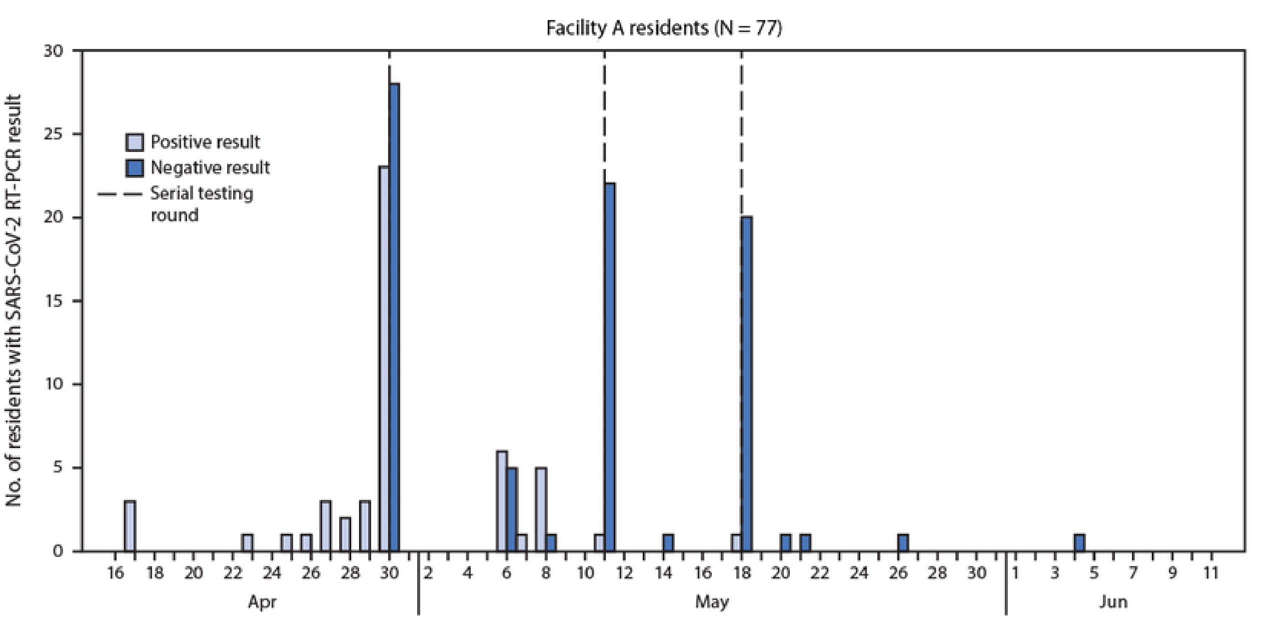
- Two SNFs contacted the Minnesota Department of Health after confirming COVID-19 cases among residents and staff
  - Facility A – 78 residents, 156 HCP (health care personnel)
  - Facility B – 183 residents, 324 HCP
- Facility-wide serial testing was implemented at both SNFs April-June 2020
  - to identify potentially asymptomatic cases of COVID-19 and inform mitigation efforts
  - identified COVID-19 cases among 64% of residents and 33% of tested HCP overall
- Genetic sequencing performed on a subset of samples
  - showed facility-specific clustering of viral genomes from HCP and residents
  - suggested transmission within each facility

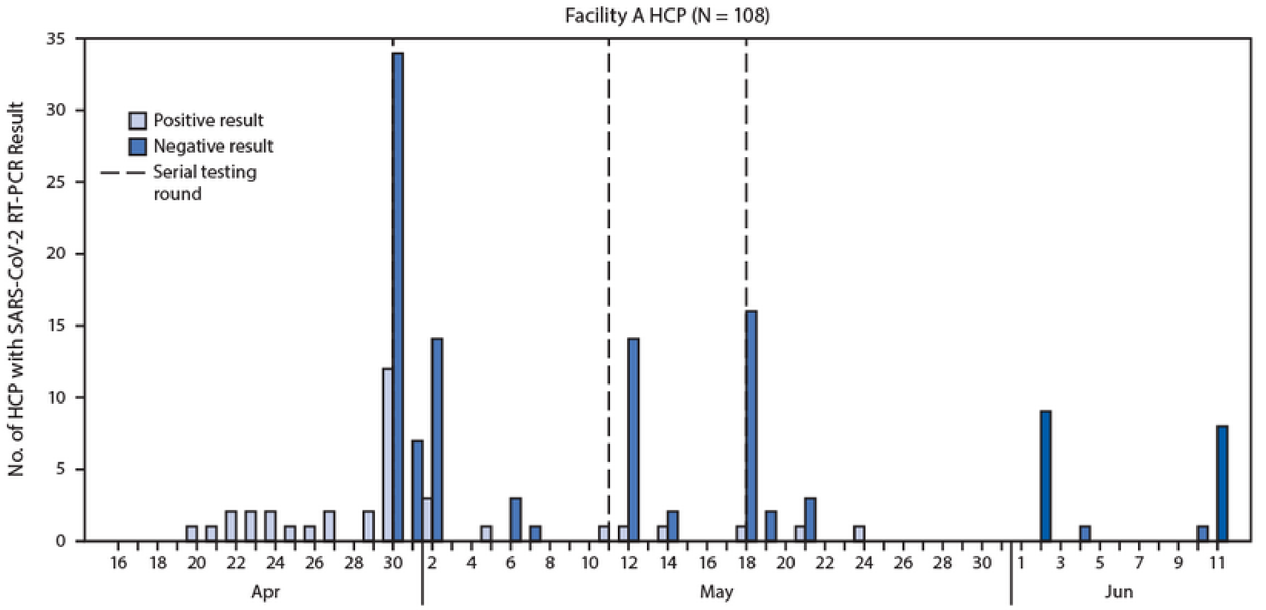
https://dx.doi.org/10.15585/mmwr.mm6937a3; CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-2

Facility A

3 rounds of serial testing

- Residents
  - 77 of 78 (99%) were tested, of which
    - 66% (N=51) were positive
    - 27% (N=14) of positives hospitalized
    - 24% (N=12) of positives died
      - a sober reminder of the heavy mortality seen in congregate care settings early in the pandemic
- HCP (testing voluntary)
  - 108 of 156 (69%) were tested, of which
    - 35% (N=38) were positive



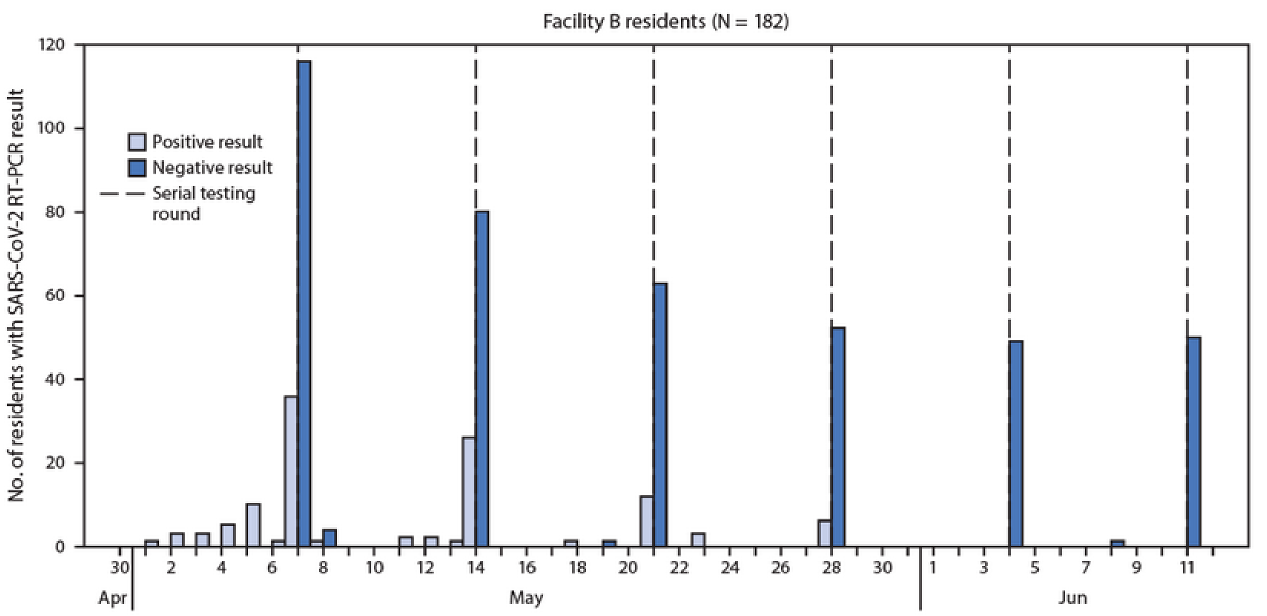


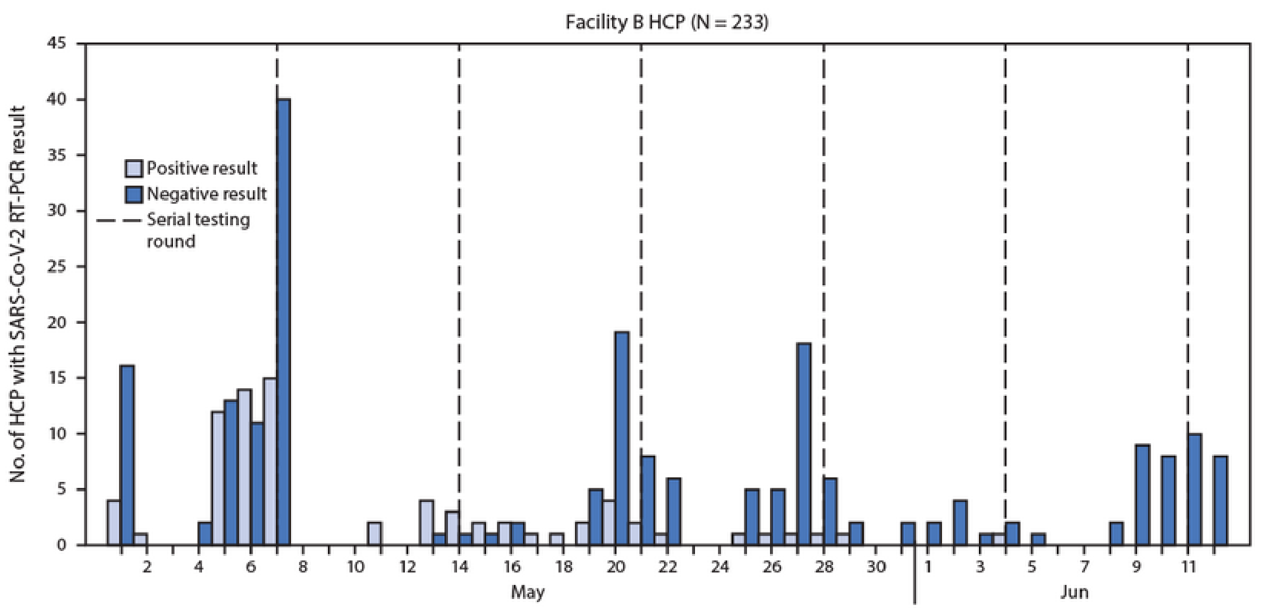
https://dx.doi.org/10.15585/mmwr.mm6937a3; CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-2

Facility B

6 rounds of serial testing

- Residents
  - 182 of 183 (99%) were tested, of which
    - 63% (N=114) were positive
    - 17% (N=19) of positives hospitalized
    - 35% (N=40) of positives died
      - a sober reminder of the heavy mortality seen in congregate care settings early in the pandemic
- HCP (testing voluntary)
  - 233 of 324 (72%) were tested, of which
    - 33% (N=76) were positive





https://dx.doi.org/10.15585/mmwr.mm6937a3; CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-2

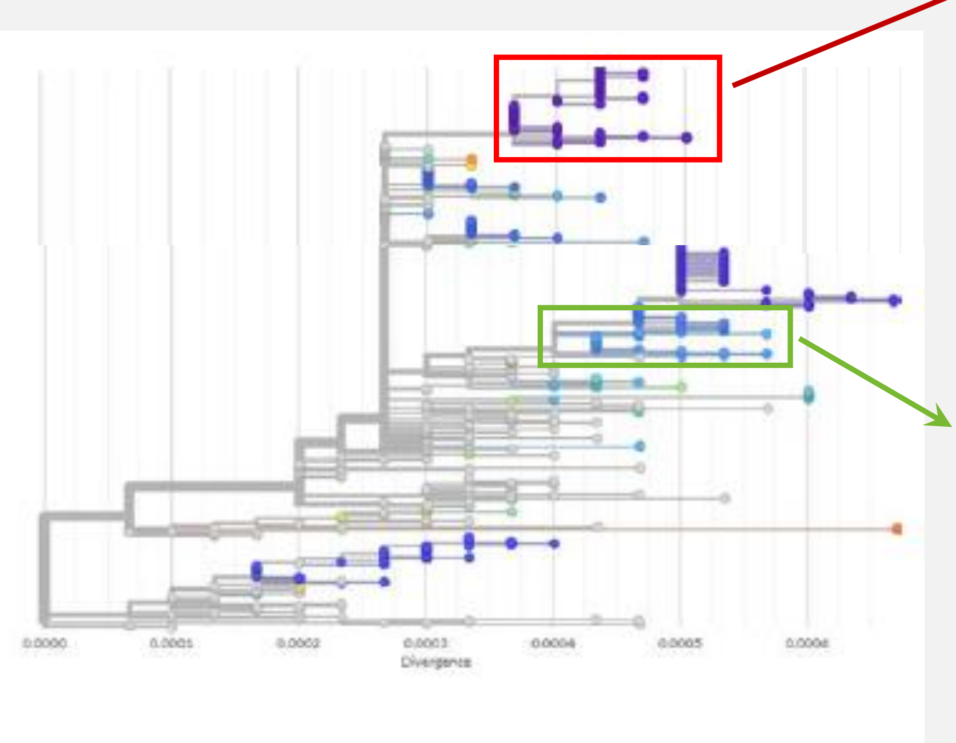
Hypotheses investigated by genomic sequencing

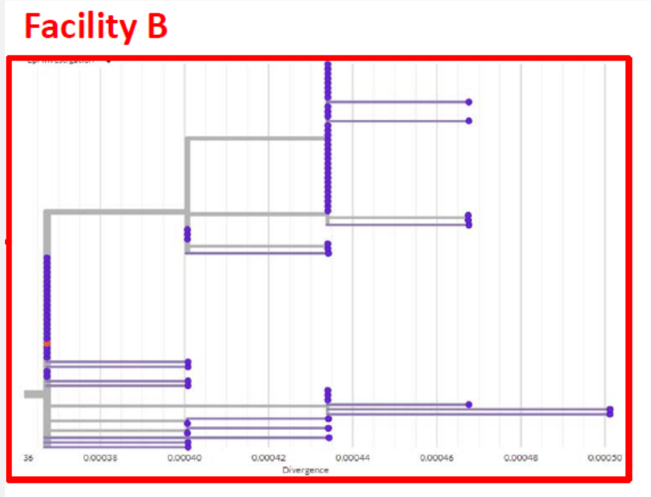
- ***Hypothesis 1***
  - Cases in Facilities A & B both result from a ***single introduction*** followed by ongoing transmission
  - Expected sequencing result:
    - SARS-CoV-2 genomes from ***all*** cases form a ***single cluster***, comprised of identical or ***closely related sequences***
- ***Hypothesis 2***
  - Cases in Facilities A & B result from ***two independent introductions*** followed by ongoing within-facility transmission
  - Expected sequencing result:
    - SARS-CoV-2 genomes form ***two distinct clusters***, each comprised of identical or closely ***related sequences***
- ***Hypothesis 3***
  - Cases in Facilities A and B result from ***multiple independent transmission events*** between each facility and the surrounding community
  - Expected sequencing result:
    - SARS-CoV-2 genomes form ***multiple clusters*** and sub-clusters with ***higher viral diversity,*** similar to the surrounding community

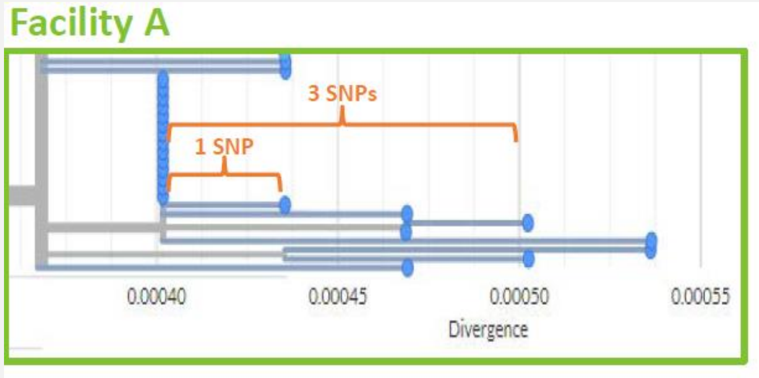
CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-2

Phylogenetic analysis

- Genomic analysis: 105 samples were sequenced (64% of positives)
  - Images here drawn by Nextstrain, including contextual samples from the region
- Observations:
  - Facility A and Facility B genomes clustered separately
  - Overall viral diversity was low
    - both inside each facility, and compared to other circulating viruses in the area
    - suggests rapid transmission within each facility
- Observations suggest independent introductions into the 2 facilities (Hypothesis #2)





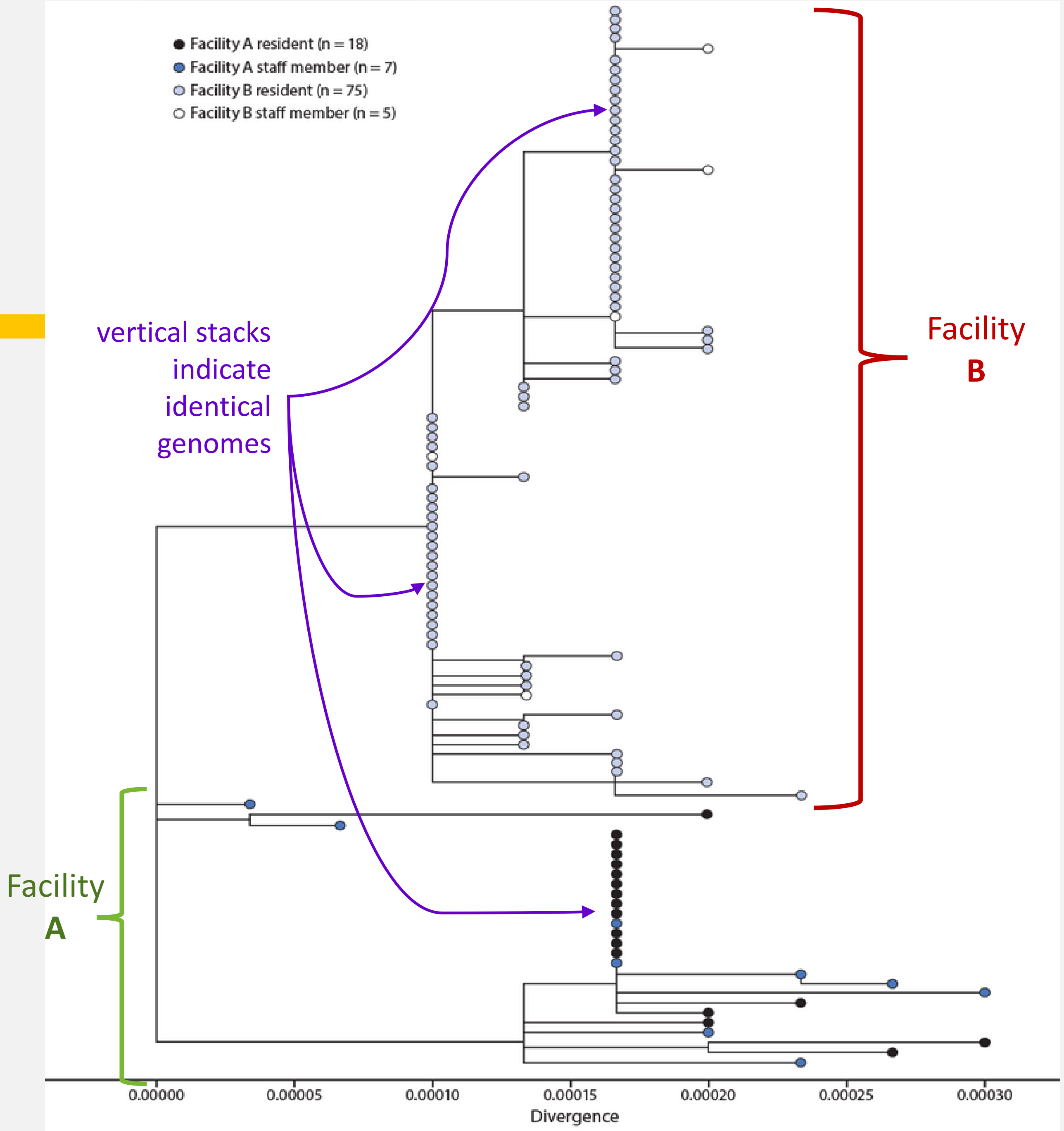


CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-2

Phylogenetic analysis

of 25 Facility A & 80 Facility B samples

- This phylogenetic tree constructed using Nextstrain’s IQ-Tree module
  - without contextual samples from the region
- Facility A and Facility B genomes again clustered separately
  - suggests independent introductions into the 2 facilities
  - ***or*** the common ancestor was not sampled
  - can’t tell which since this tree doesn’t include non-SNF samples
- Multiple vertical stacks (identical genomes) again indicate rapid transmission



vertical stacks indicate identical genomes

Facility B

Facility A

https://dx.doi.org/10.15585/mmwr.mm6937a3

Hypotheses #2 suggested by genomic sequencing results

- ***Hypothesis 1***
  - Cases in Facilities A & B both result from a ***single introduction*** followed by ongoing transmission
  - Expected sequencing result:
    - SARS-CoV-2 genomes from ***all*** cases form a ***single cluster***, comprised of identical or ***closely related sequences***
- ***Hypothesis 2***
  - Cases in Facilities A & B result from ***two independent introductions*** followed by ongoing within-facility transmission
  - Expected sequencing result:
    - SARS-CoV-2 genomes form ***two distinct clusters***, each comprised of identical or ***closely related sequences***
- ***Hypothesis 3***
  - Cases in Facilities A and B result from ***multiple independent transmission events*** between each facility and the surrounding community
  - Expected sequencing result:
    - SARS-CoV-2 genomes form ***multiple distinct clusters*** and sub-clusters with ***higher viral diversity***, similar to the surrounding community

CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-2

Other observations

- Morbidity & mortality were high among all 261 residents
  - 165 positive (63%), 33 hospitalized (13%), 52 died (20%)
    - In contrast, of all 432 HCP, “only” 4 hospitalized (1%), 2 died (0.5%)
  - Clearly resident co-morbidities contributed to resident health outcomes
- Among the 165 positive resident cases
  - Median age was 72 at Facility A, 81 at Facility B
  - 89 (78%) female; 70 (61%) asymptomatic when tested
- Among the 114 positive HCP cases
  - 56 (49%) were asymptomatic when tested
    - 30 reported working on or after symptom onset
  - 73 (64%) were nurses/assistants with direct resident contact

https://dx.doi.org/10.15585/mmwr.mm6937a3

SNF case study takeaways

- ***WGS results suggest it only takes one introduction for a large outbreak to occur!***
  - with accompanying rapid transmission, morbidity & mortality
- Recommendations for Facilities
  - Continued vigilance with infection prevention & control
    - Better access to PPE and training in its use
    - Flexible leave policies to limit transmission by infected staff
  - Regular screening of residents and staff (at least daily)
  - Universal testing of all residents and staff

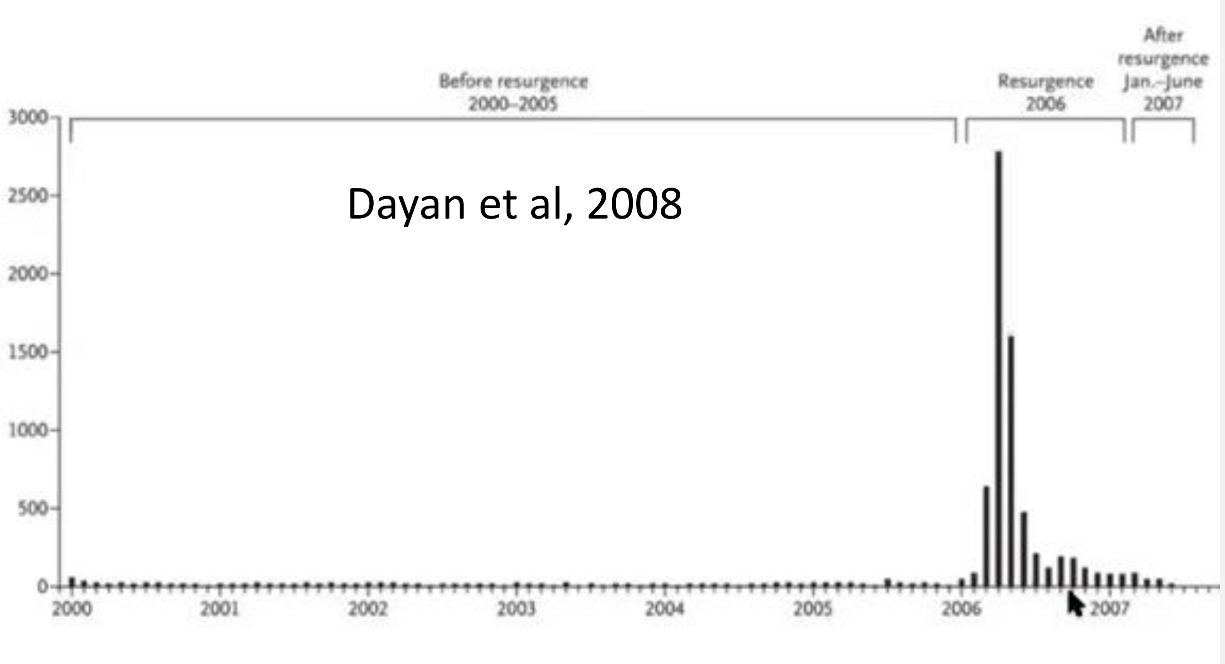
https://dx.doi.org/10.15585/mmwr.mm6937a3; CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-2

Identifying introductions and transmission in mumps outbreaks

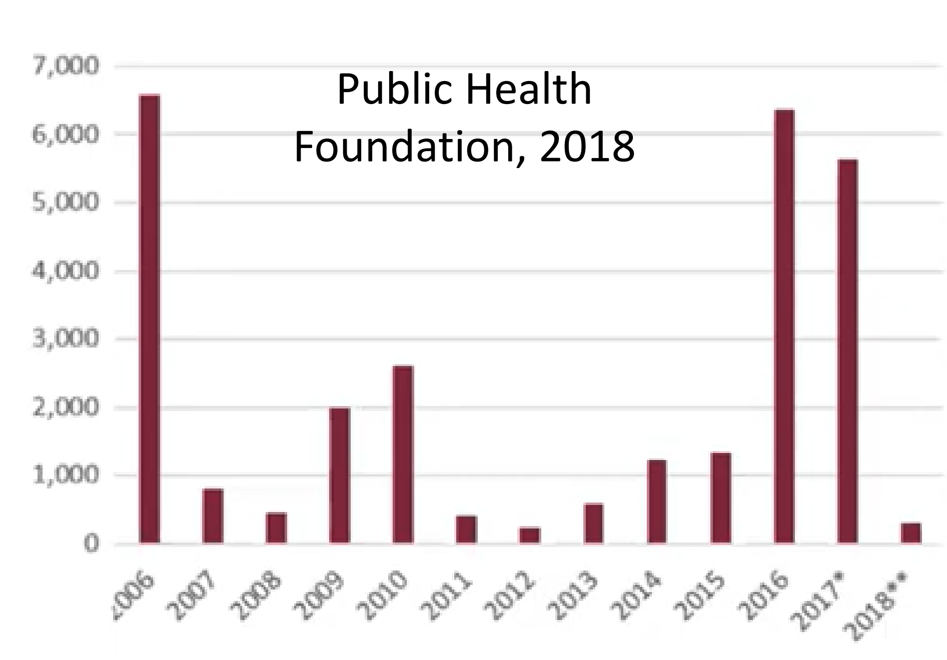
Section 3

Mumps outbreaks

- Mumps incidence in the US resurged after a long period of low incidence
  - MMR vaccine licensed 1967
  - 2nd doses started in the 1990s after uptick in 80s
  - Few cases recorded for decades
- Significant mumps outbreaks occurred in ***2006 – 2007*** then again in ***2016 – 2017***
- Genomic Epidemiology questions
  - Were outbreaks in different areas linked?
  - Was there a vaccine mismatch circulating strain?
    - “G” genotype circulates in the US, but vaccine based on “Jeryl Lynn” strain
  - What factors were driving the outbreaks?
  - Is there new epidemiology in close contact settings?



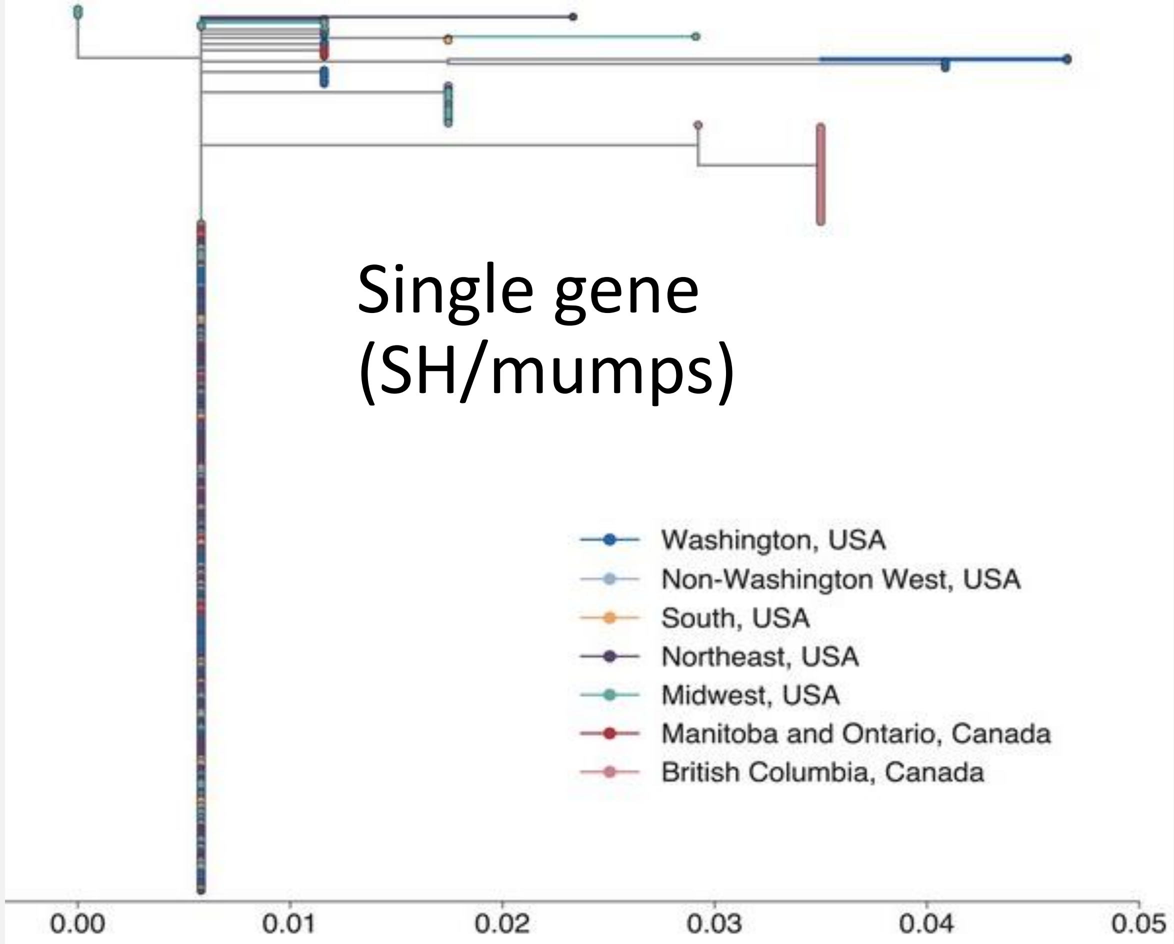
Dayan et al, 2008

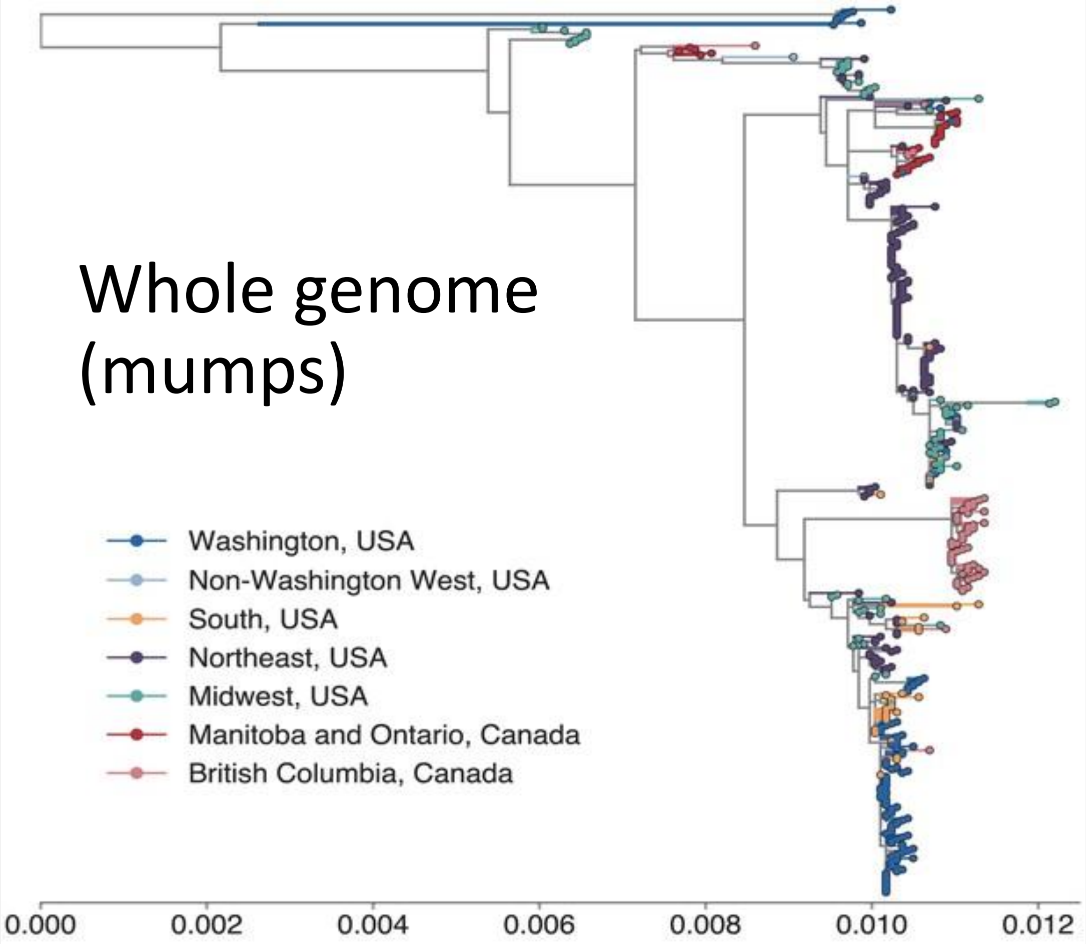


Public Health Foundation, 2018

Higher resolution obtained from full genomes *vs* single gene

- Mumps is a 15.3 kilobase, negative strand RNA virus with 7 genes
- Mumps genome has higher genetic diversity than SARS-CoV-2, but lower than influenza
  - Actual genetic diversity is not well captured by genotyping just the “standard” SH gene


Single gene (SH/mumps)

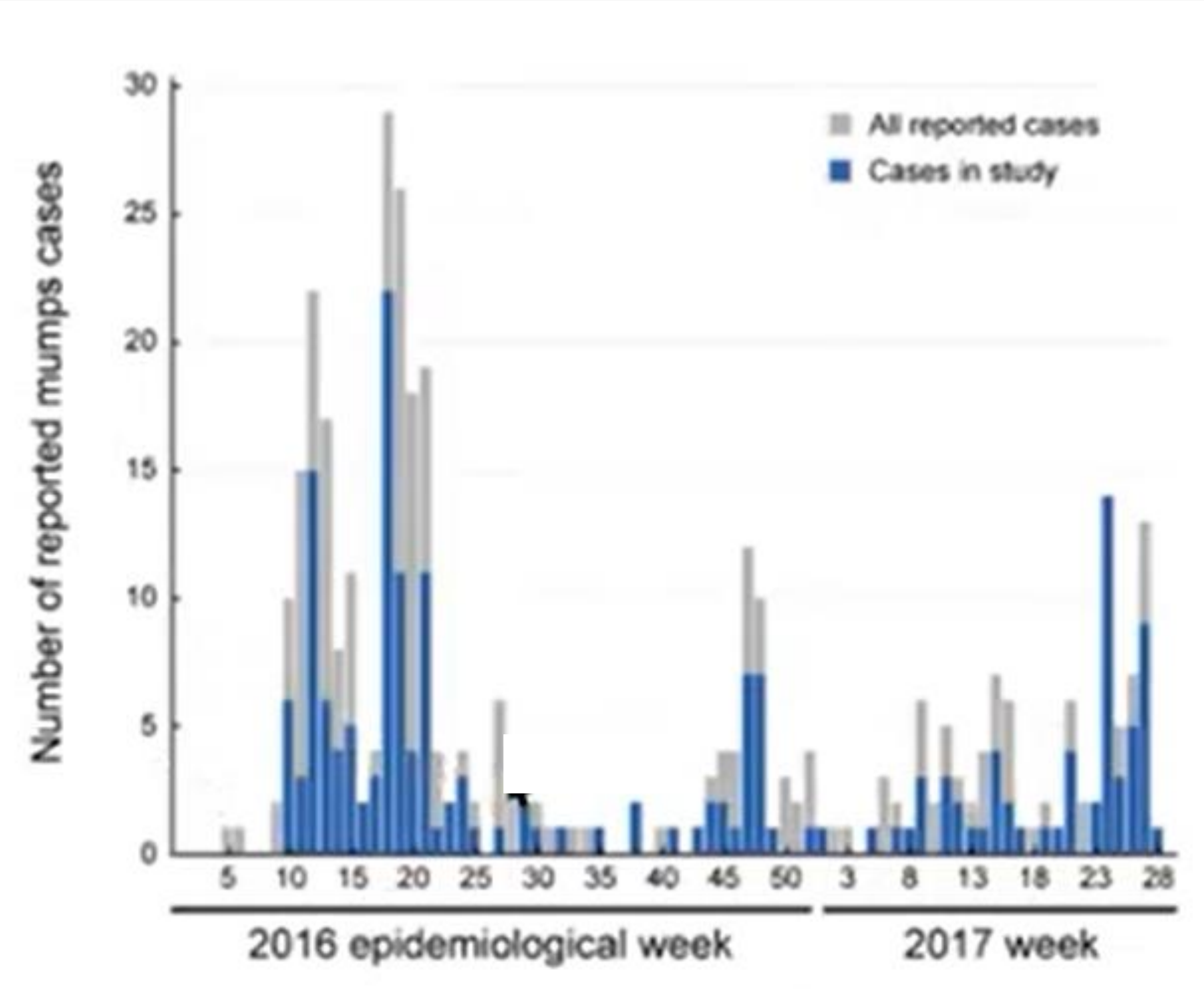

Whole genome (mumps)

Moncla et al (2021). Repeated introductions and intensive community transmission fueled a mumps virus outbreak in Washington State eLife 10:e66448. https://elifesciences.org/articles/66448

Mumps in Massachusetts

2016 – 2017

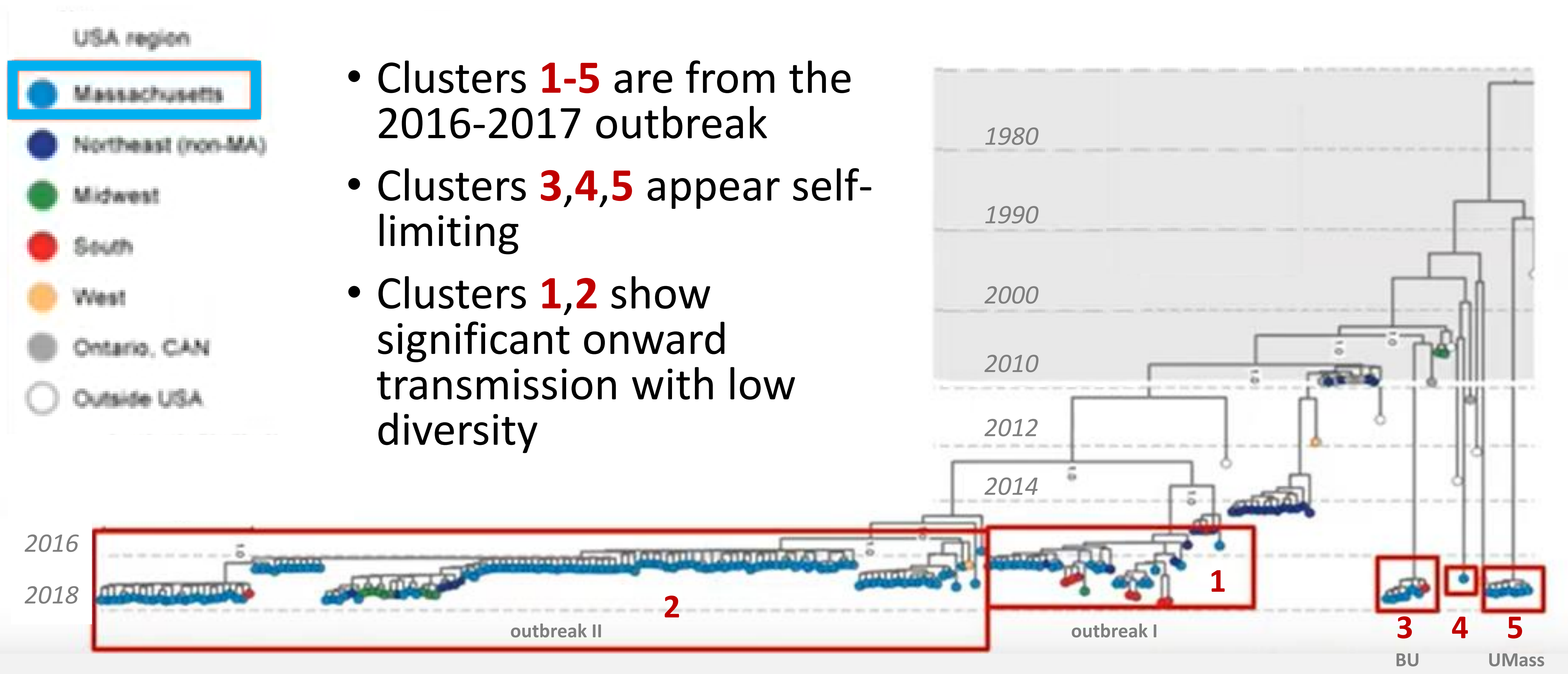
- More than 250 cases overall
  - Epi curve suggests outbreak may have resulted from one introduction, then sustained transmission
- Broad Institute genomic study, 2020
  - Sequenced 158 MA cases
    - 92 students (Harvard, Boston University, UMass Amherst)
    - 66 from community around Harvard
  - Combined with sequenced samples from other geographies
    - including historical samples



Wohl *et al*, 2020. Combining genomics and epidemiology to track mumps virus transmission in the United States. https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3000611

Phylogenetic tree shows the MA outbreak resulted from *multiple distinct introductions/transmission chains* with varying degrees of spread

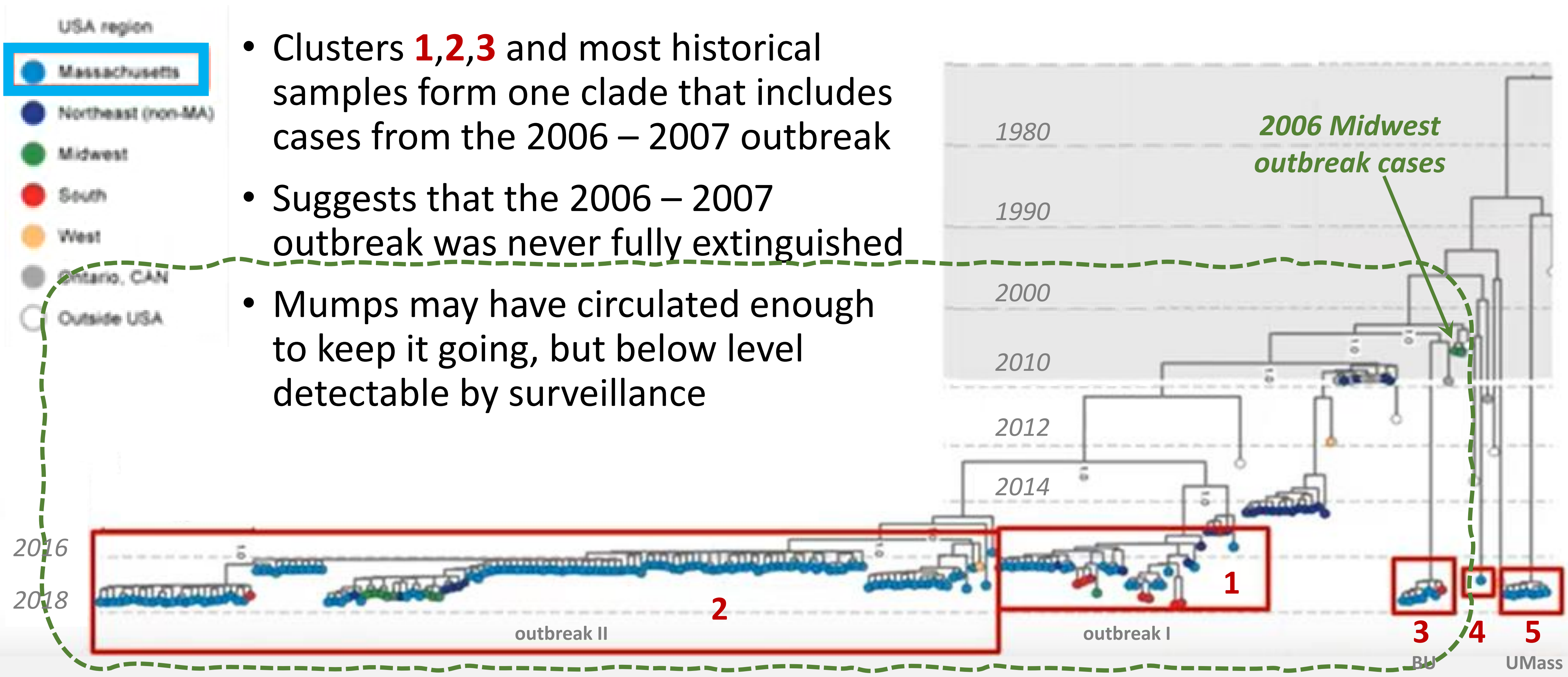
- Clusters **1-5** are from the 2016-2017 outbreak
- Clusters **3**,**4**,**5** appear self-limiting
- Clusters **1**,**2** show significant onward transmission with low diversity



CZBioHub COVID Tracker Seminar series https://covidtracker.czbiohub.org/resources, Week 15; Adapted from Wohl et al, 2020

One large clade has a common ancestor with the 2006 – 2007 Midwestern university outbreaks

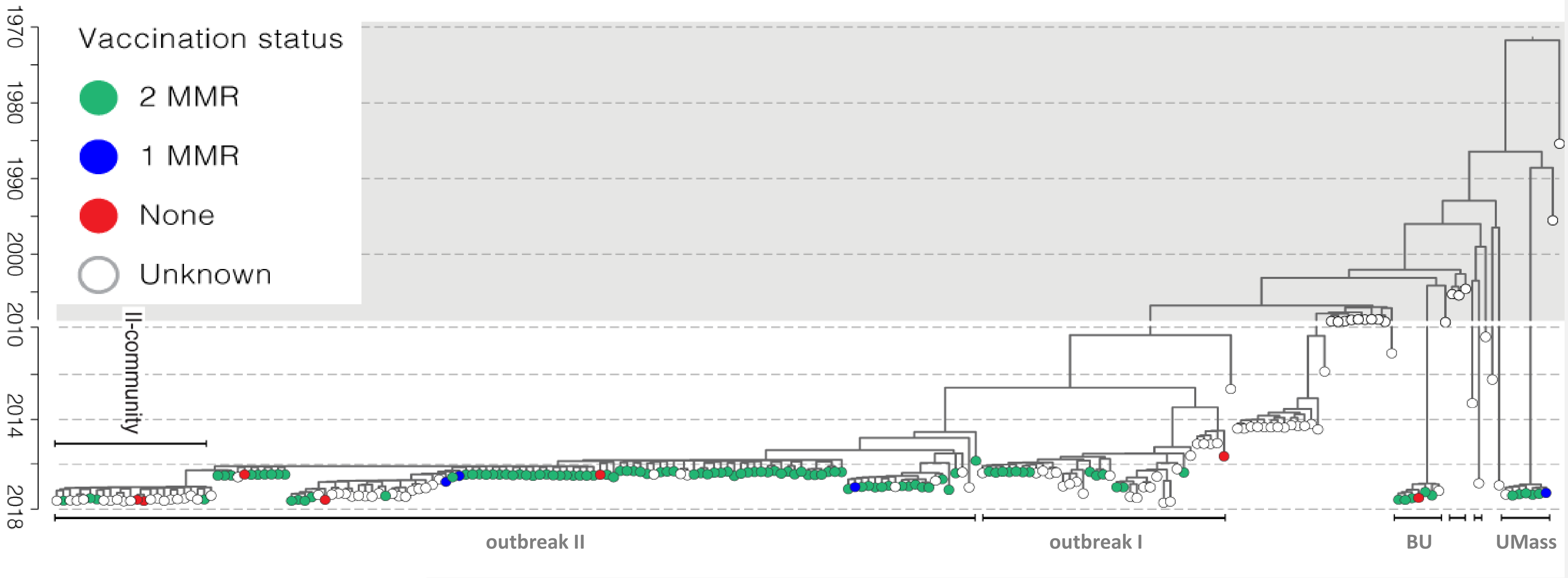
- Clusters 1,2,3 and most historical samples form one clade that includes cases from the 2006 – 2007 outbreak
- Suggests that the 2006 – 2007 outbreak was never fully extinguished
- Mumps may have circulated enough to keep it going, but below level detectable by surveillance



CZBioHub COVID Tracker Seminar series https://covidtracker.czbiohub.org/resources, Week 15; Adapted from Wohl et al, 2020

Fully vaccinated cases were distributed across all 2016 – 2017 transmission chains

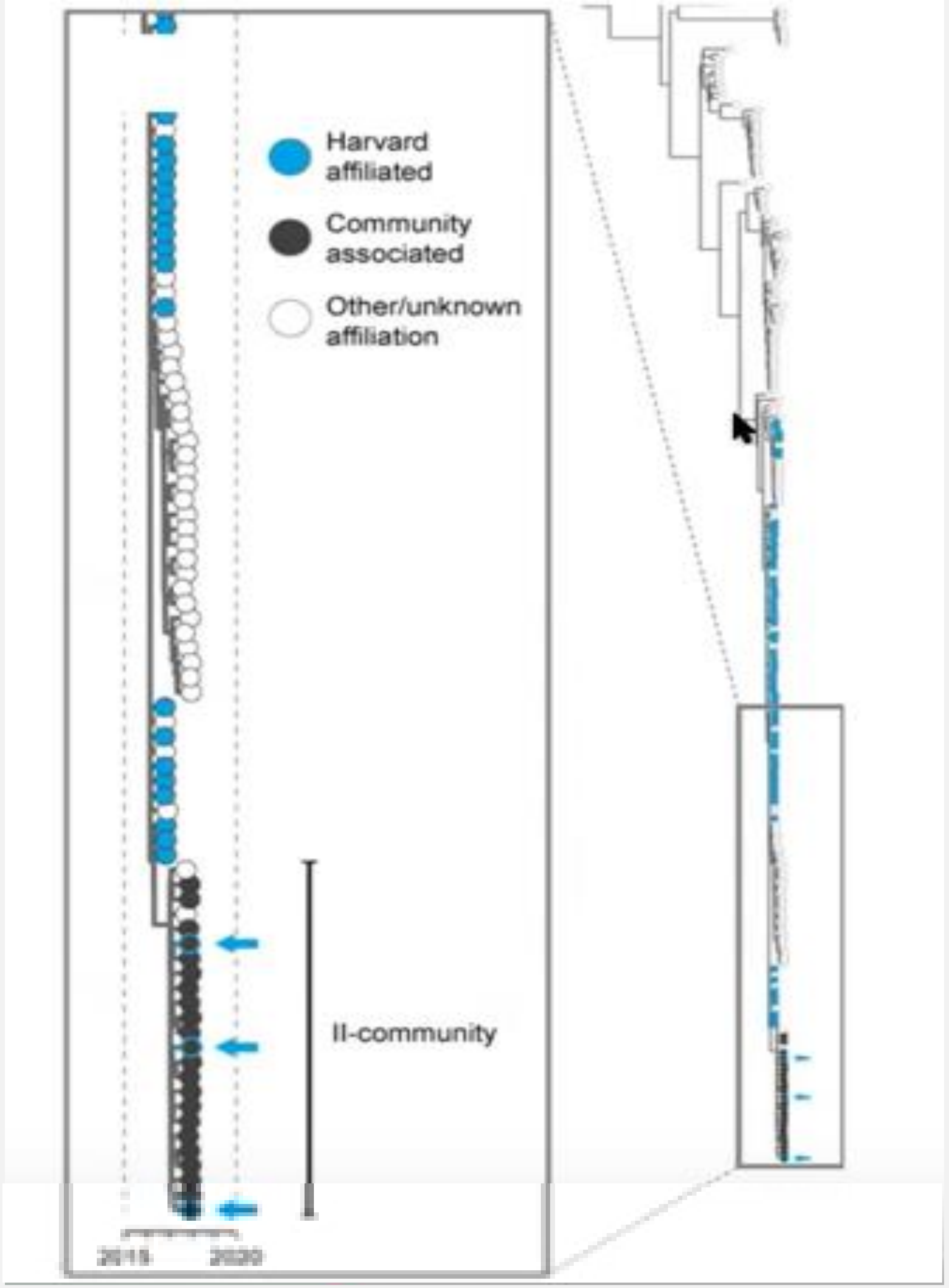
Suggests that the extent of the 2016 – 2017 MA outbreak was likely due to ***waning immunity*** rather than vaccine escape



CZBioHub COVID Tracker Seminar series https://covidtracker.czbiohub.org/resources, Week 15; Adapted from Wohl et al, 2020

Community cases linked to Harvard

- Cases in the community occurred 5 months later than those associated with Harvard
- Phylogentic analysis show community cases cluster together, and suggests they were part of a transmission from Harvard
- Epidemiological investigations later identified 3 individuals from Harvard who could have sourced the community outbreak

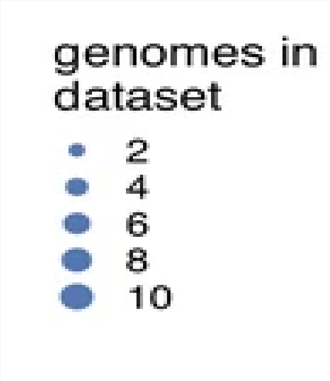


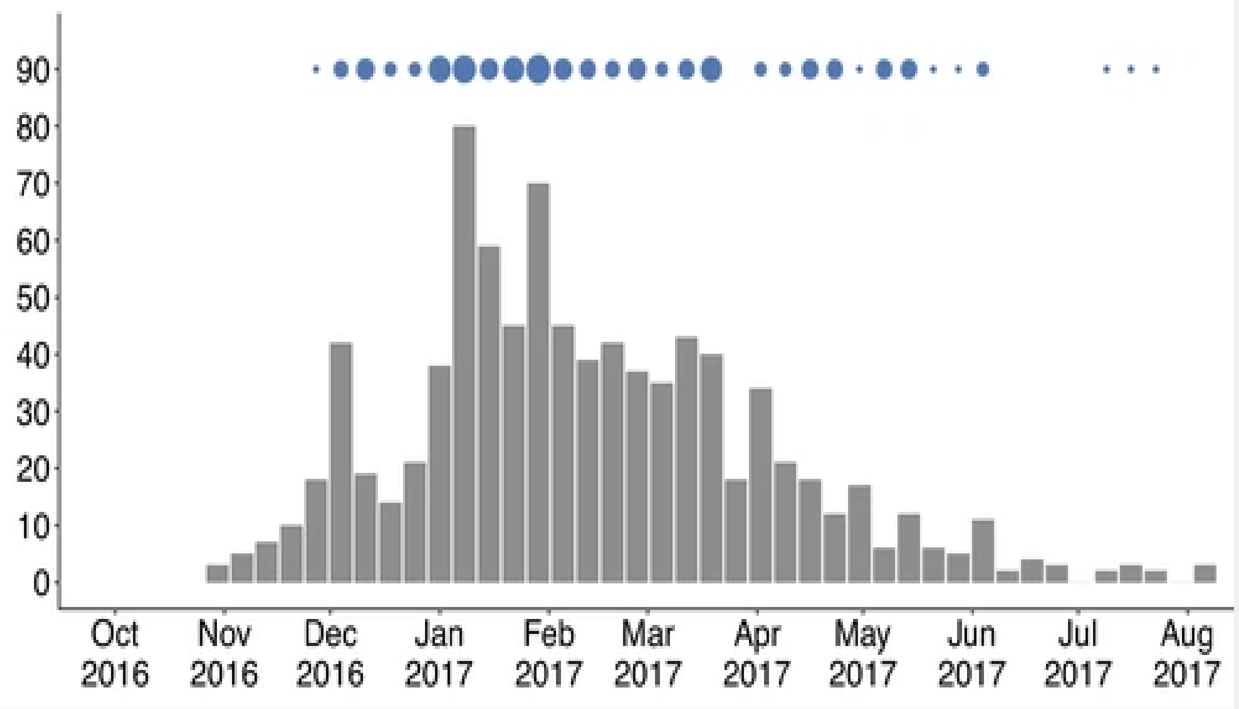
CZBioHub COVID Tracker Seminar series https://covidtracker.czbiohub.org/resources, Week 15; Adapted from Wohl et al, 2020

Mumps in Washington state

2016 – 2017

- More than 800 cases overall
  - Epi curve again suggests outbreak might have resulted from one introduction, followed by sustained transmission
- WA Dept of Health genomic study, 2021
  - 110 WA cases sequenced
    - proportional to demographic groups
  - 56 cases from 2007 – 2014
    - from other states
- ***Phylogeographic analysis*** performed to ***infer where ancestor of WA virus circulated***

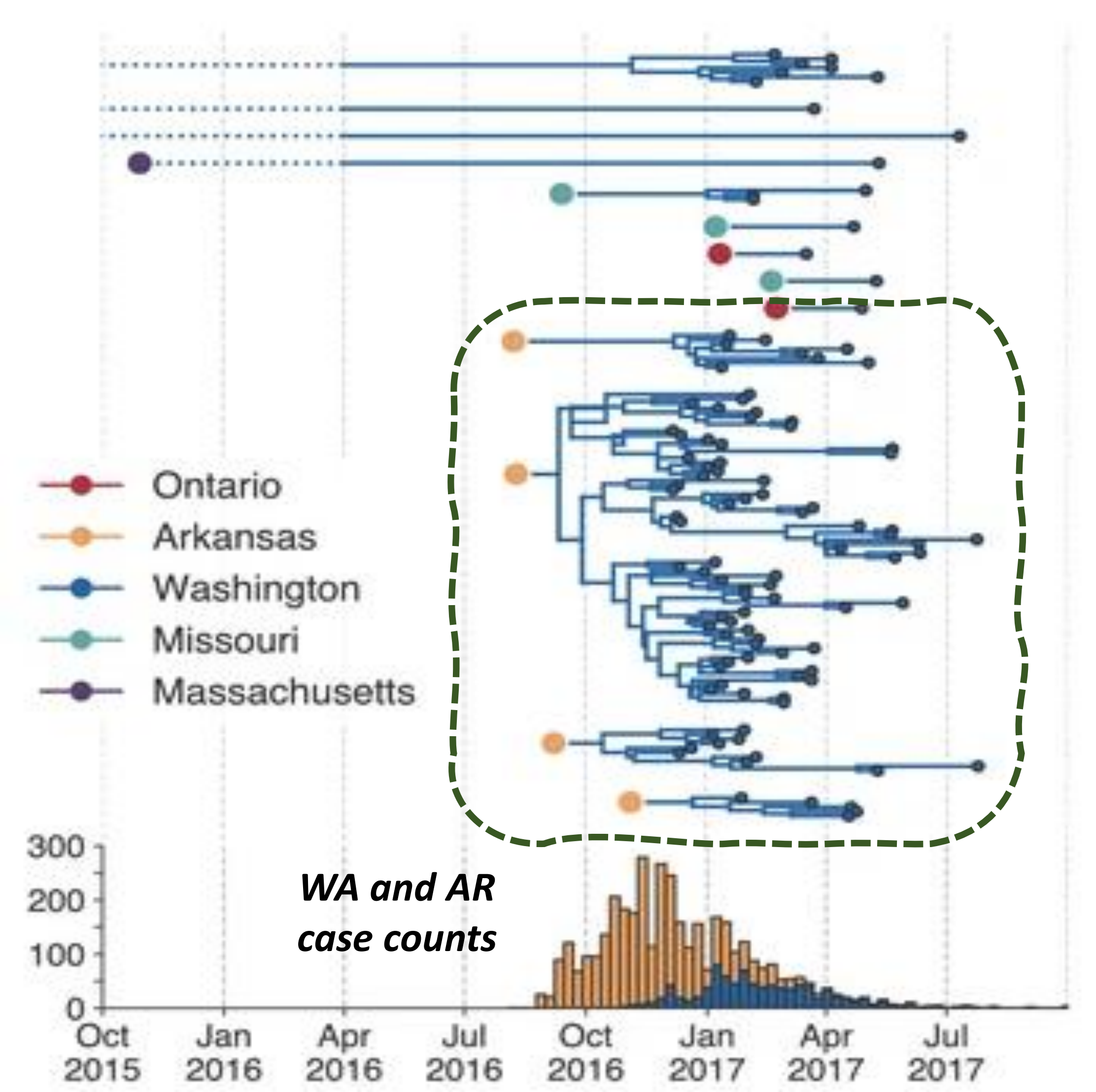




Moncla et al (2021). Repeated introductions and intensive community transmission fueled a mumps virus outbreak in Washington State eLife 10:e66448. https://elifesciences.org/articles/66448

Phylogeographic analysis suggests multiple introductions

- Phylogeographic results suggest ***multiple, independent introductions***, with ongoing sustained transmission
  - Approximately 13 introductions (estimate range of 10 – 17)
- Four clades appear to have originated in Arkansas!
  - including one particularly large clade
- Why Arkansas?

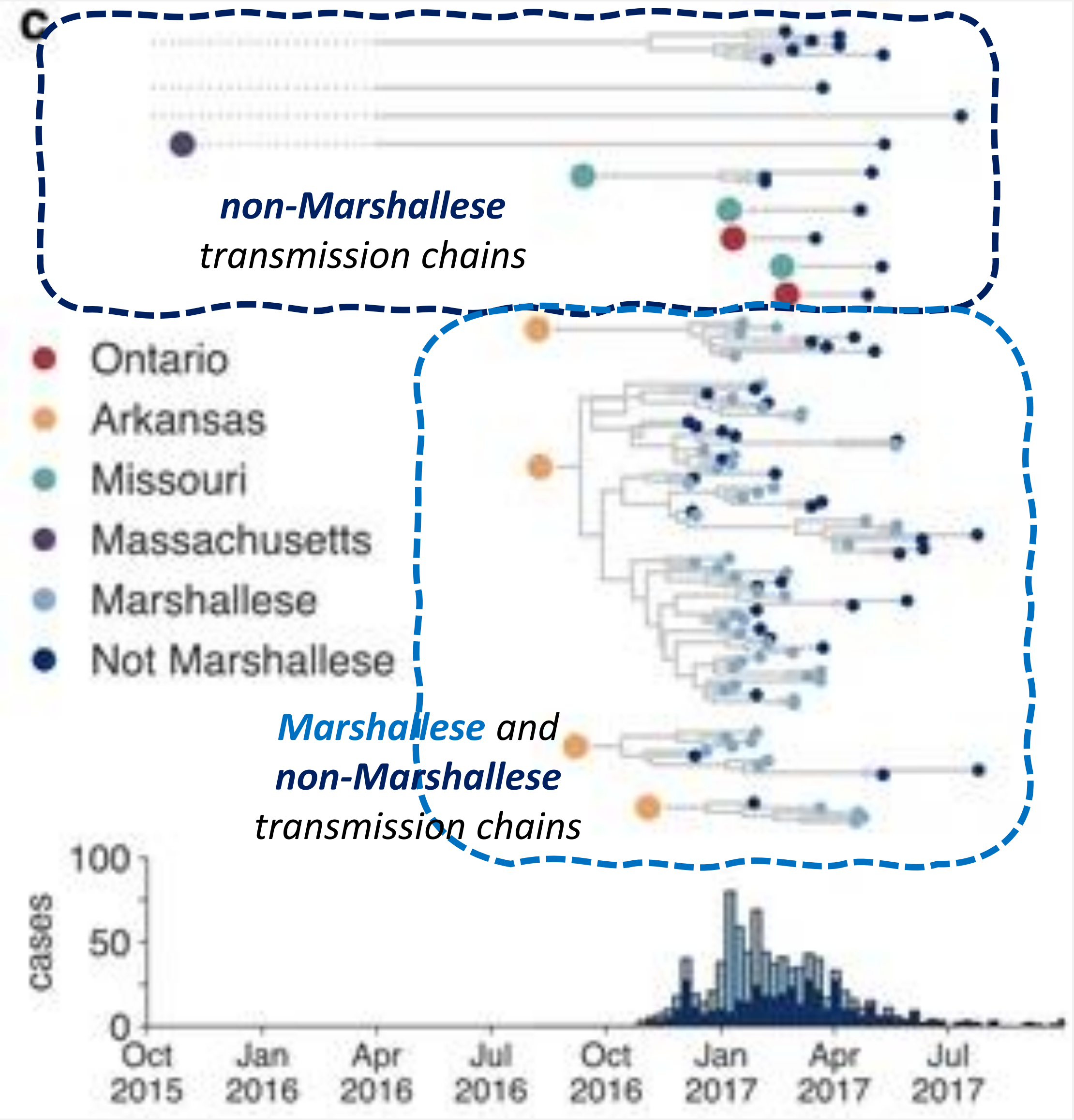


WA and AR case counts

Moncla et al (2021). Repeated introductions and intensive community transmission fueled a mumps virus outbreak in Washington State eLife 10:e66448. https://elifesciences.org/articles/66448

Marshallese individuals overrepresented

- Individuals from the ***Marshall Islands*** made up over 50% of WA mumps cases
  - but less than 1% of WA population
  - ***AR also has a relatively large Marshallese population***
- Demographic & epidemiological data were odd
  - Marshallese cases were among the young
  - Marshallese have higher-than-average vaccination rates
- How was transmission sustained?
  - Tree shows that ***non-Marshallese*** cases tend to group in small, self-limiting transmission chains
  - Suggests factors inside the Marshallese community promote spread; factors outside limit spread



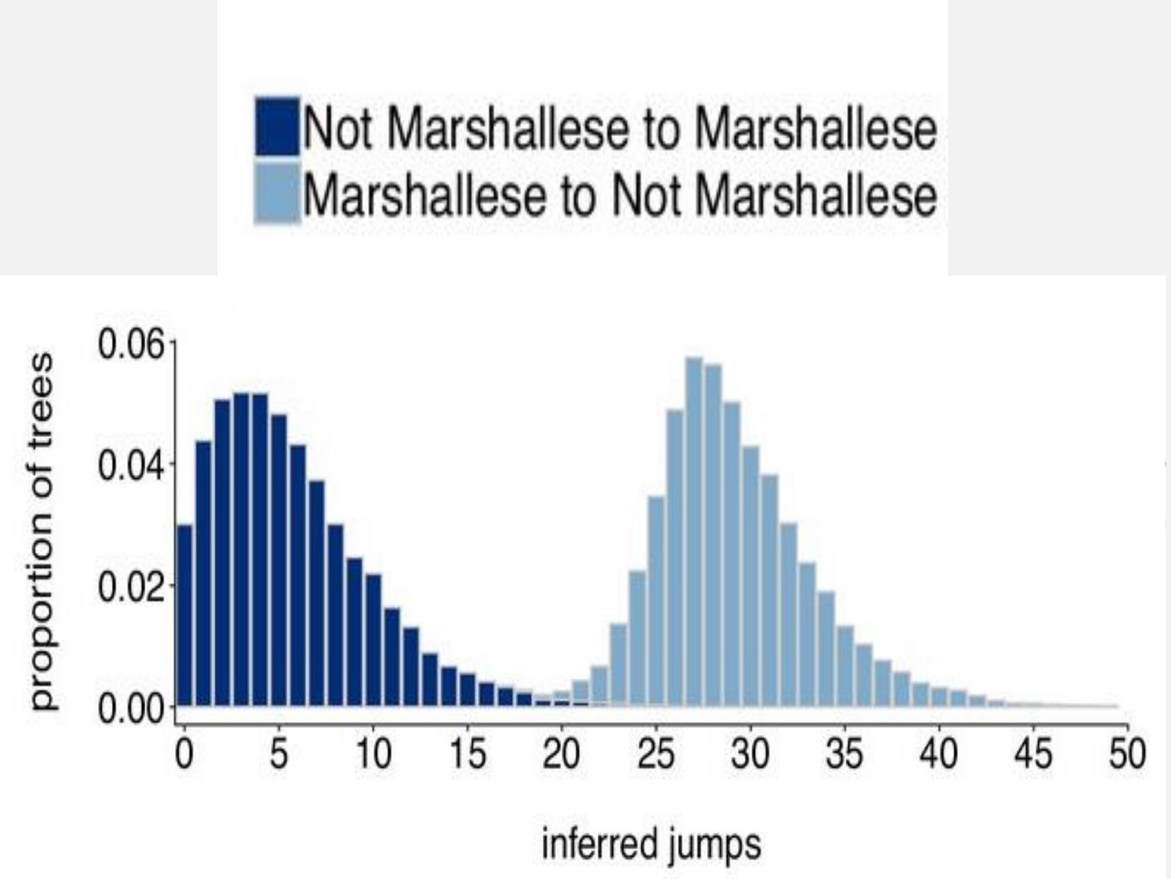
non-Marshallese *transmission chains*

Marshallese *and* ***non-Marshallese*** *transmission chains*

Moncla et al (2021). Repeated introductions and intensive community transmission fueled a mumps virus outbreak in Washington State eLife 10:e66448. https://elifesciences.org/articles/66448

Transmission occurs primarily *within* the Marshallese community

- Further analysis revealed that transmission from ***non-Marshallese*** to ***Marshallese*** occurred much less frequently than from ***Marshallese*** to ***non-Marshallese***
  - Helps explain how mumps was transmitted outside the Marshallese community, and once there, died out quickly
- Overall evidence suggest that mumps among the Marshallese appears to have ***overcome herd immunity*** from vaccination
  - Potentially due to historical disparities leading to ***dense living arrangements*** and ***contact structures***



Moncla et al (2021). Repeated introductions and intensive community transmission fueled a mumps virus outbreak in Washington State eLife 10:e66448. https://elifesciences.org/articles/66448

Takeaways from Mumps case studies

- Outbreaks that appear sustained from epi curves may be made up of multiple introductions
  - Genomic analysis can distinguish distinct introductions and their spread
- Differentiating between transmission chains & their characteristics highlights how each contributes to the outbreak and may suggest why
  - e.g. Waning immunity (MA) or Contact structure factors (WA)
- Genomic data help link outbreaks that do not appear related
  - 2006 – 2007 college outbreaks and 2016 – 2017 MA outbreaks
  - Harvard and community
  - Arkansas and Washington

Confirming SARS-CoV-2 reinfection with Whole Genome Sequencing

Section 4

COVID-19 disease recurrence

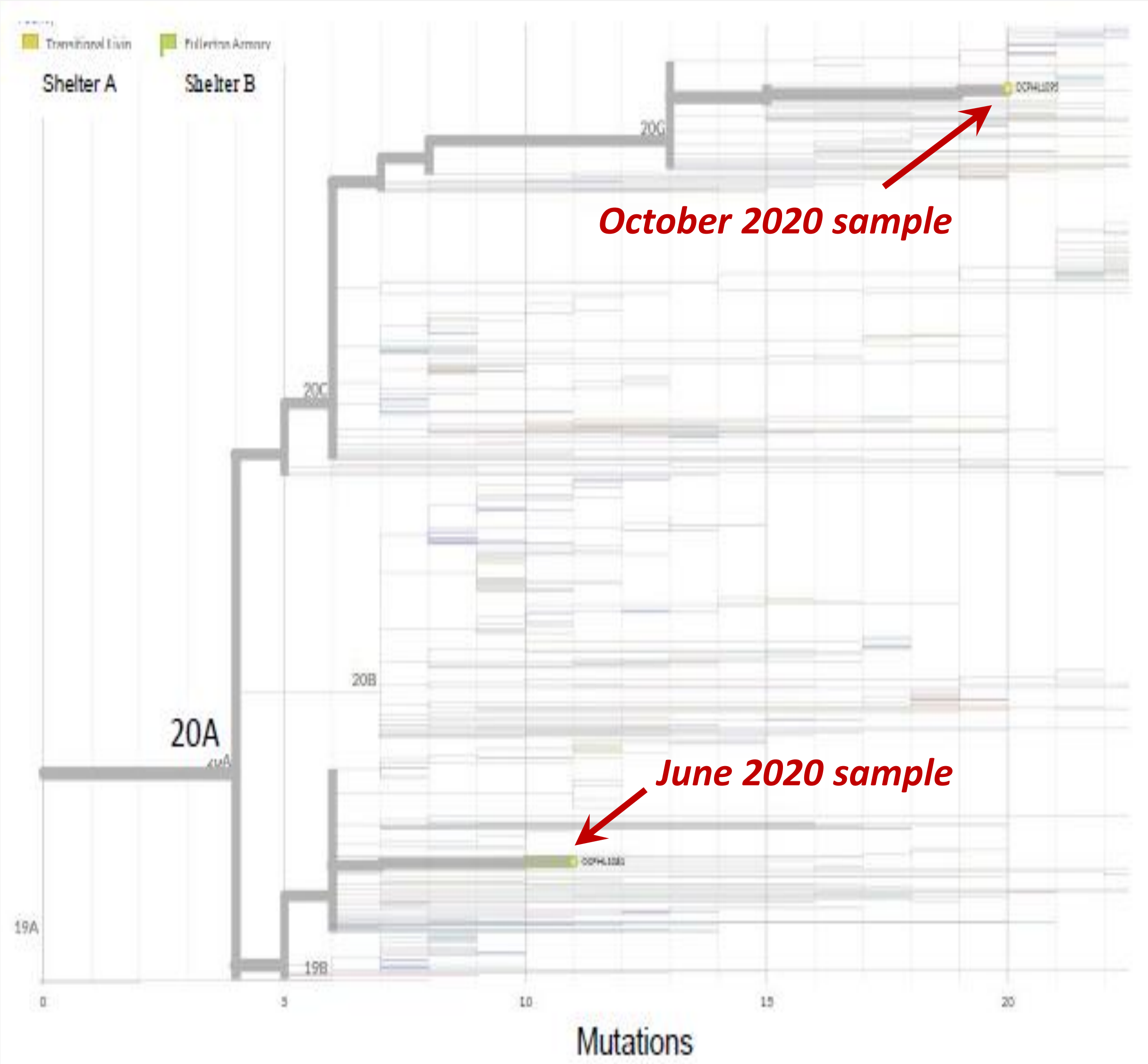
- Recurrence can be due to
  - ***Re-emergence of latent, original virus***
    - Expect the viral genomes to be similar, with later genome having the earlier genome as ancestor
  - ***Re-infection by a different viral strain***
    - Expect viral genomes to be diverged, with a distant common ancestor
- CDC protocol for investigating suspected SARS-CoV-2 reinfection
  - COVID-19-like symptoms 45-89 days after initial infection
  - Positive test with or without symptoms 90+ days after initial infection
  - Analyze genomes of paired respiratory specimens, one for each infection episode
- Reinfection once thought rare, but now seen more often post-Omicron

Abu-Raddadet al. (2020) Assessment of the risk of SARS-CoV-2 reinfection in an intense re-exposure setting, https://www.medrxiv.org/content/10.1101/2020.08.24.20179457v2
CDC Common Investigation Protocol: www.cdc.gov/coronavirus/2019-ncov/php/reinfection.html
CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-5

Reinfection case

32-year-old person experiencing homelessness

- ***June 2020*** – Shelter A
    - Tested in response to staff with positive test
    - Experienced fever, sore throat, cough, headache
- ***October 2020*** – Shelter B
    - Tested in response to resident with positive test
    - Experienced general cold symptoms, reported feeling very ill
- Recovered both times without hospitalization
- Specimens taken 138 days apart
    - Whole Genome Sequencing performed
    - Results included in a larger phylogenetic analysis
    - Genomes appear in different tree clades



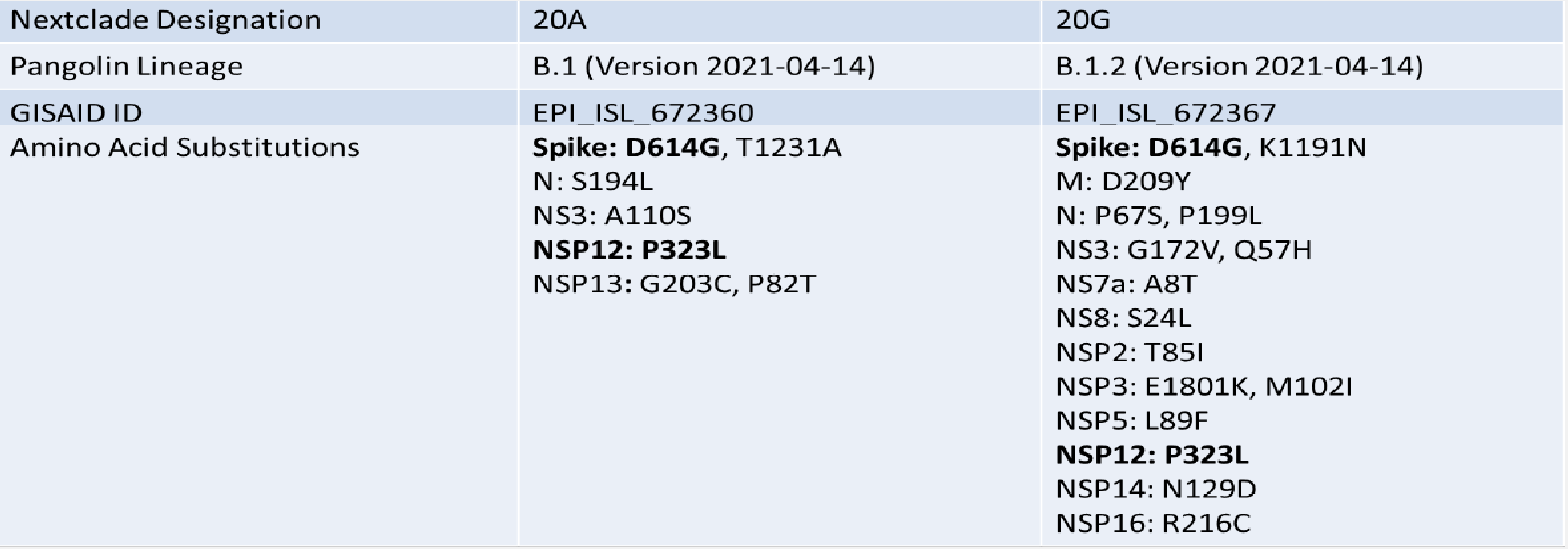
October 2020 sample

June 2020 sample

CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-5

Genotype differences suggest reinfection

- Different Nextclade clades, Pangolin lineages
- Share some early Spike (D614G) and NSP12 (P323L) mutations but otherwise quite different



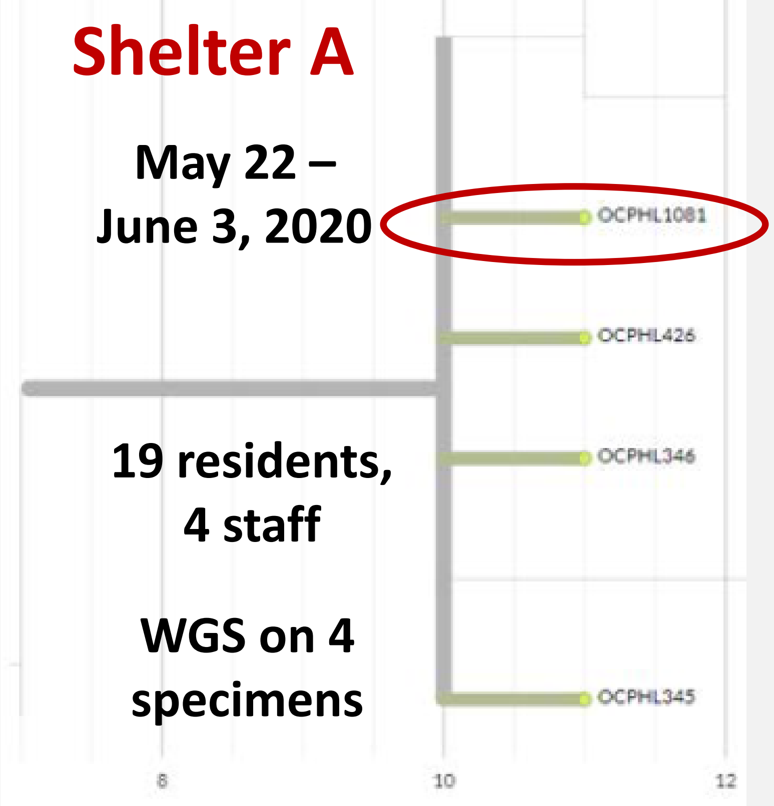
	June 2020	October 2020
Nextclade Designation	20A	20G
Pangolin Lineage	B.1 (Version 2021-04-14)	B.1.2 (Version 2021-04-14)
GISAID ID	EPI_ISL_672360	EPI_ISL_672367
Amino Acid Substitutions	**Spike: D614G**, T1231A N: S194L NS3: A110S **NSP12: P323L** NSP13: G203C, P82T	**Spike: D614G**, K1191N M: D209Y N: P67S, P199L NS3: G172V, Q57H NS7a: A8T NS8: S24L NSP2: T85I NSP3: E1801K, M102I NSP5: L89F **NSP12: P323L** NSP14: N129D NSP16: R216C

CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-5

Confirming reinfection findings

Patient was part of separate, larger outbreaks

Patient's sequenced genomes clustered
with others from the same facility in both outbreaks

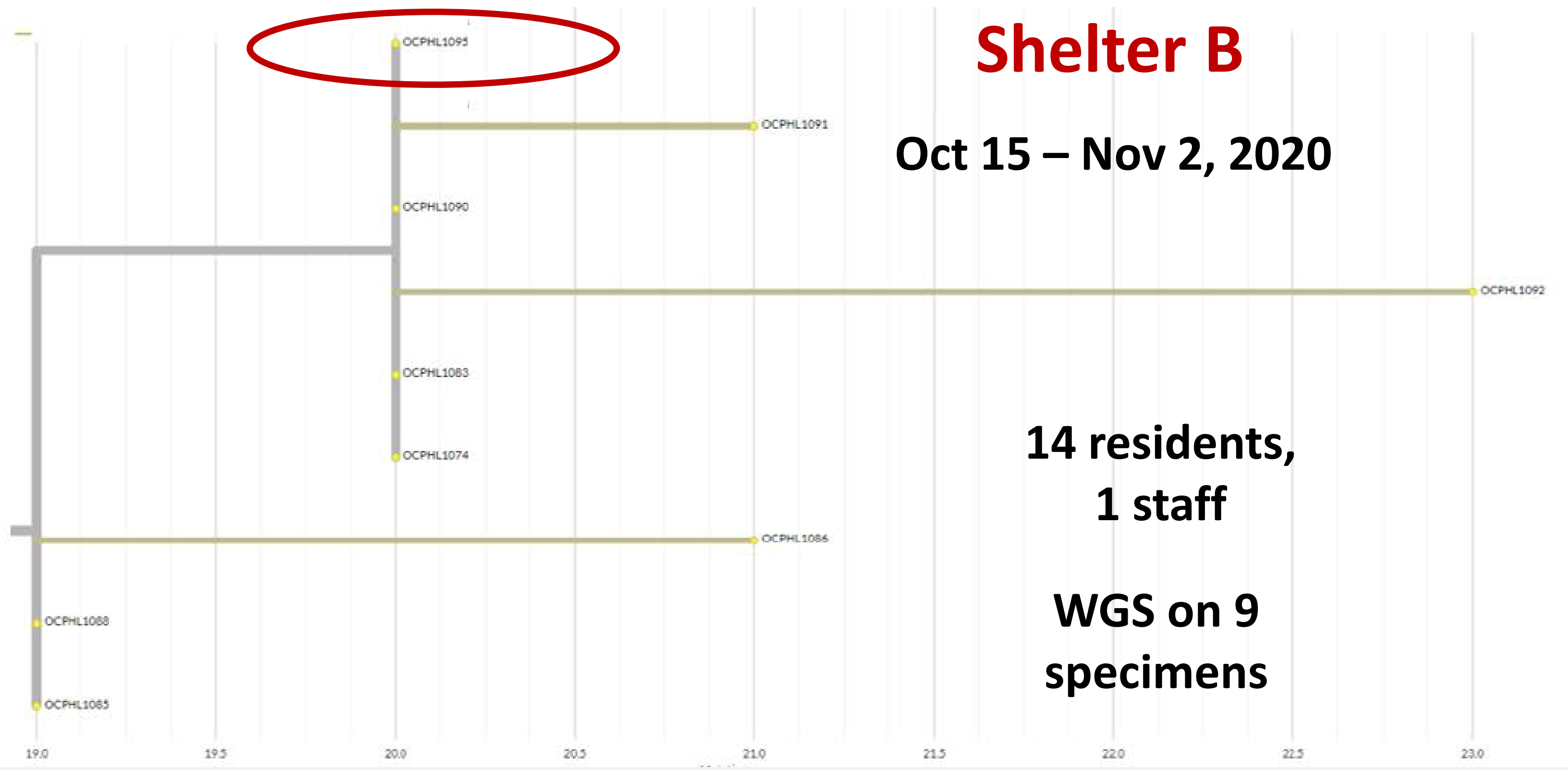


Shelter A

**May 22 –
June 3, 2020**

**19 residents,
4 staff**

**WGS on 4
specimens**



Shelter B

Oct 15 – Nov 2, 2020

**14 residents,
1 staff**

**WGS on 9
specimens**

CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-5

SARS-CoV-2 reinfection case study takeaways

- Previous SARS-CoV-2 infection does not necessarily confer immunity against a different variant
  - although recent research indicates that ***both natural infection and vaccination confer highly protective immunity against serious disease, hospitalization and death***
- WGS of paired specimens can confirm reinfection
  - via distinct clade/lineage assignments and specific mutation patterns indicating distant common ancestry
- Additional epidemiological data can inform the GenEpi analysis
  - e.g. tracking/sequencing of associated outbreak clusters
- Highlights the benefit of ongoing genomic sequencing!

CDC COVID-19 Genomic Epidemiology Toolkit https://www.cdc.gov/amd/training/covid-19-gen-epi-toolkit.html, module 2-5

Other Gen Epi examples in brief

Section 5

PulseNet

Established 1996

https://www.cdc.gov/pulsenet/

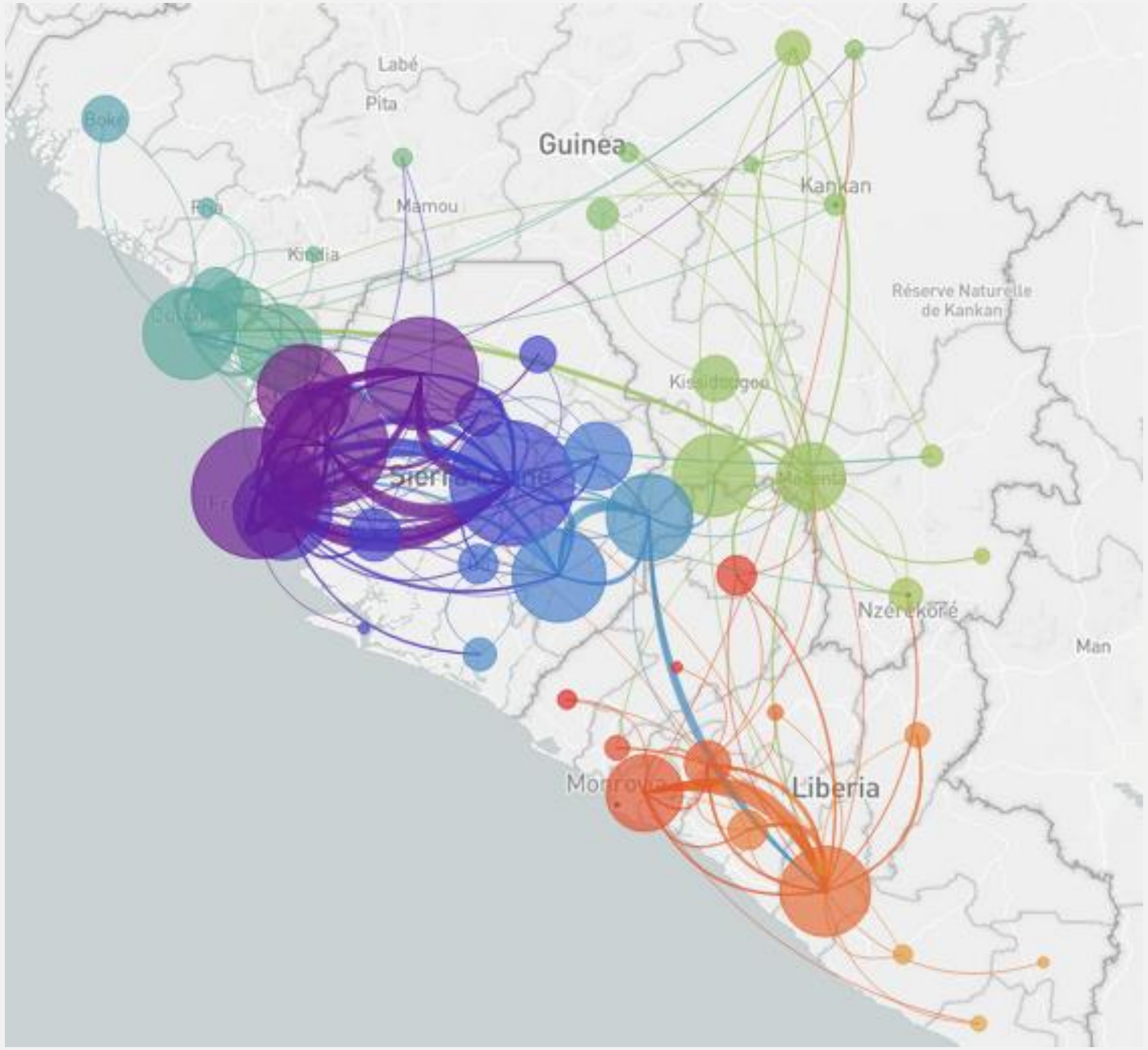
- CDC initiative involving 80+ Public Health labs and the food processing industry
  - DSHS is a participant
- Connects foodborne & waterborne illness cases to detect outbreaks
  - Adopted Whole Genome Sequencing in 2013
  - Provides protocols and maintains WGS database
- Many success stories: https://www.cdc.gov/pulsenet/anniversary/success-story.html
  - April - November 2019: 7 ***Listeria monocytogenes*** cases detected in TX, FL, SC, PA, ME
  - 4/5 reported eating egg-containing food; 3/4 identified deli salads with hard-boiled eggs
  - PulseNet DB found two genetically related samples from routine testing of Facility A in Feb 2019
  - Another genetic match found when Facility A was inspected again in Dec 2019; egg recalls initiated

https://www.aphl.org/aboutAPHL/publications/Documents/FS-2020Mar-WGS-Egg-Outbreak.pdf

Ebola in West Africa

2013 – 2016, primarily Sierra Leone, Liberia

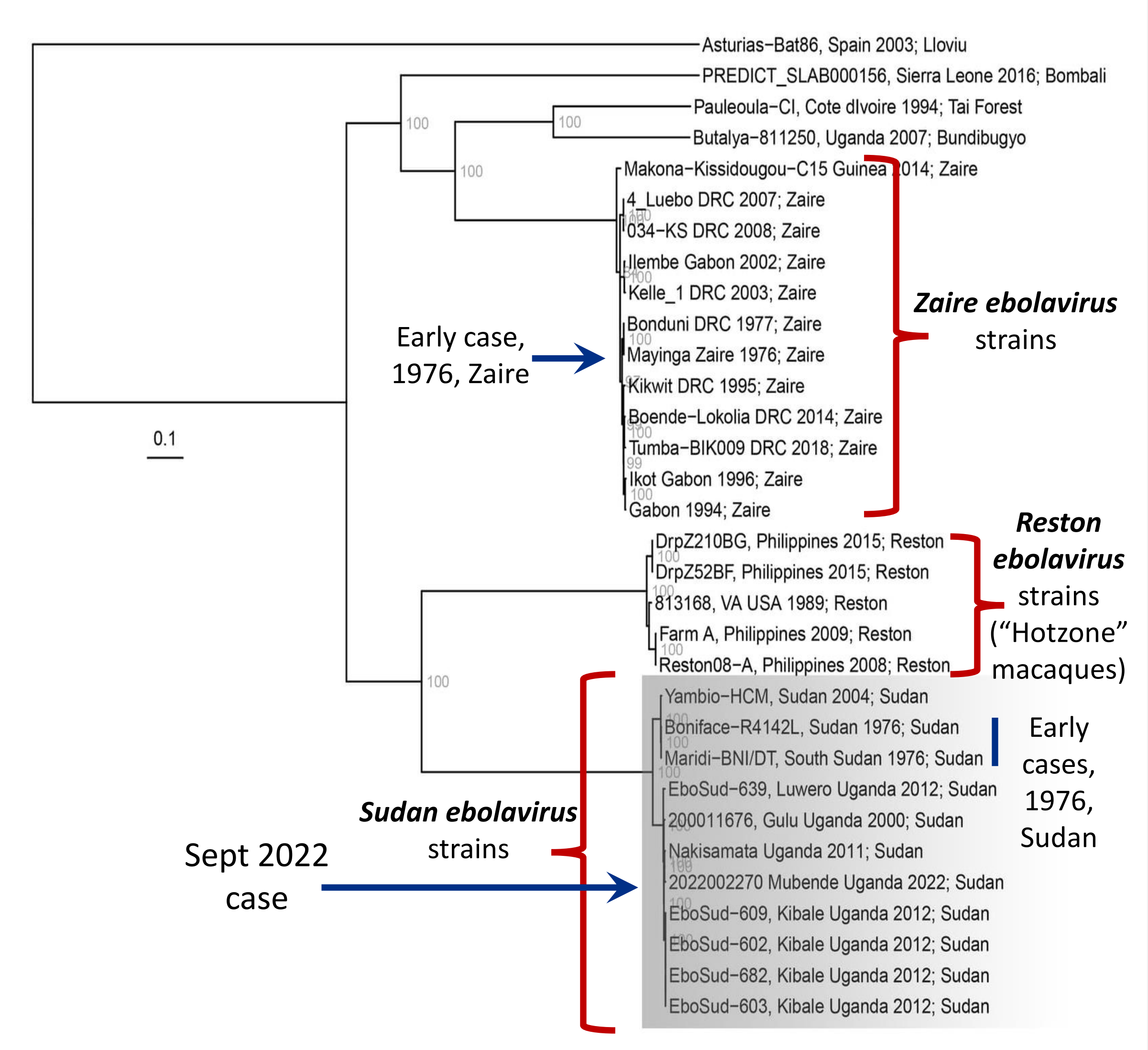
- ***Phylogeographic analysis*** used to model disease transmission dynamics over time and geographies
- Results available in a public Nextstrain build
  - https://nextstrain.org/ebola
  - Click “PLAY” under “Date Range” to view the spread



Ebola in Uganda

September 2022

- ***Ebolavirus*** was identified Sept 19, 2022 in a Ugandan individual from the Mubende District
    - DSHS Epidemiologist Rania Milleron is part of a worldwide multi-disciplinary team studying this outbreak
- Phylogentic analysis of 2022 virus placed it in the well-established ***Sudan Ebolavirus*** clade
    - Most closely related to ***Nakisamata Sudan Ebolavirus*** that emerged in Uganda, May 2011
    - Determining the viral sequence can help predict potential effectiveness of existing Ebola vaccine
- Earliest Ebolavirus cases are quite diverged
    - Suggests independent spillover events


Asturias−Bat86, Spain 2003; Lloviu
PREDICT_SLAB000156, Sierra Leone 2016; Bombali
Pauleoula−CI, Cote dIvoire 1994; Tai Forest
Butalya−811250, Uganda 2007; Bundibugyo
Makona−Kissidougou−C15 Guinea 2014; Zaire
4_Luebo DRC 2007; Zaire
034−KS DRC 2008; Zaire
Ilembe Gabon 2002; Zaire
Kelle_1 DRC 2003; Zaire
Bonduni DRC 1977; Zaire
Mayinga Zaire 1976; Zaire
Kikwit DRC 1995; Zaire
Boende−Lokolia DRC 2014; Zaire
Tumba−BIK009 DRC 2018; Zaire
Ikot Gabon 1996; Zaire
Gabon 1994; Zaire
DrpZ210BG, Philippines 2015; Reston
DrpZ52BF, Philippines 2015; Reston
813168, VA USA 1989; Reston
Farm A, Philippines 2009; Reston
Reston08−A, Philippines 2008; Reston
Yambio−HCM, Sudan 2004; Sudan
Boniface−R4142L, Sudan 1976; Sudan
Maridi−BNI/DT, South Sudan 1976; Sudan
EboSud−639, Luwero Uganda 2012; Sudan
200011676, Gulu Uganda 2000; Sudan
Nakisamata Uganda 2011; Sudan
2022002270 Mubende Uganda 2022; Sudan
EboSud−609, Kibale Uganda 2012; Sudan
EboSud−602, Kibale Uganda 2012; Sudan
EboSud−682, Kibale Uganda 2012; Sudan
EboSud−603, Kibale Uganda 2012; Sudan
100
100
100
100
0.1

Early case, 1976, Zaire

Zaire ebolavirus strains

Reston ebolavirus strains (“Hotzone” macaques)

Early cases, 1976, Sudan

Sudan ebolavirus strains

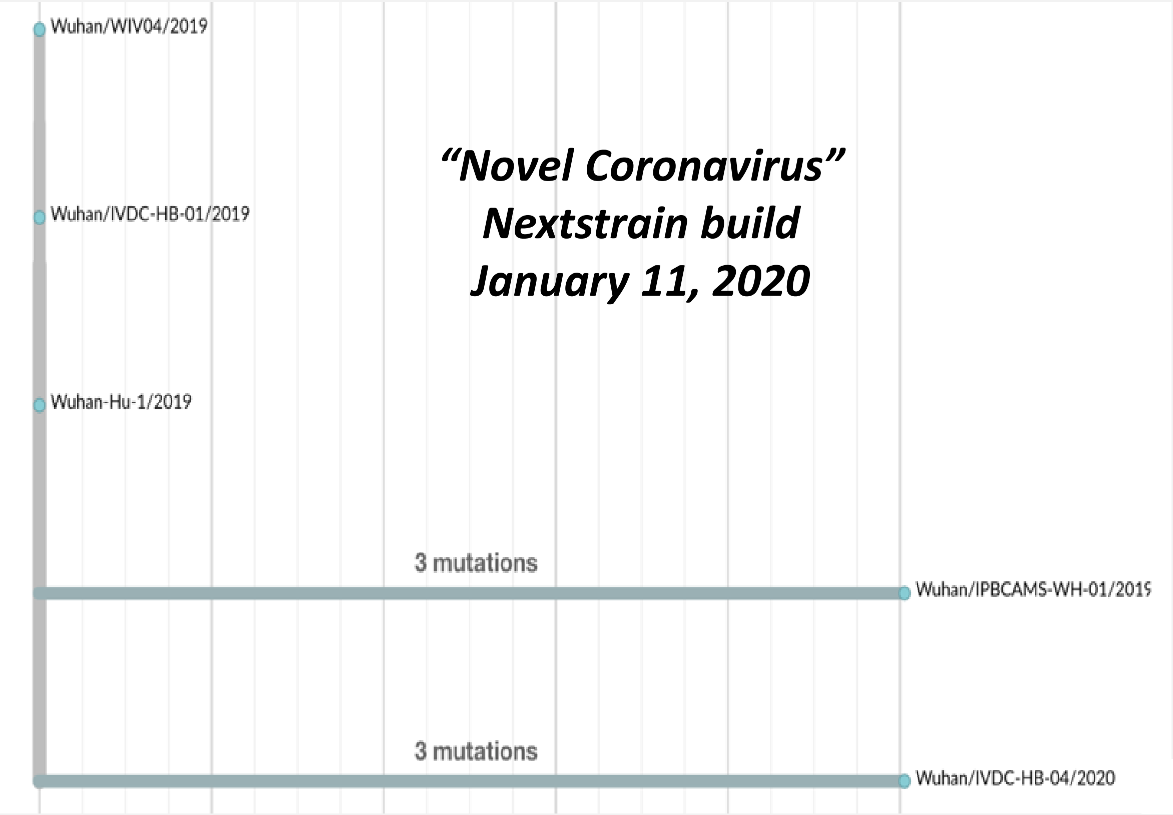
Sept 2022 case

https://virological.org/t/september-2022-sudan-ebola-virus-disease-outbreak-in-uganda/902

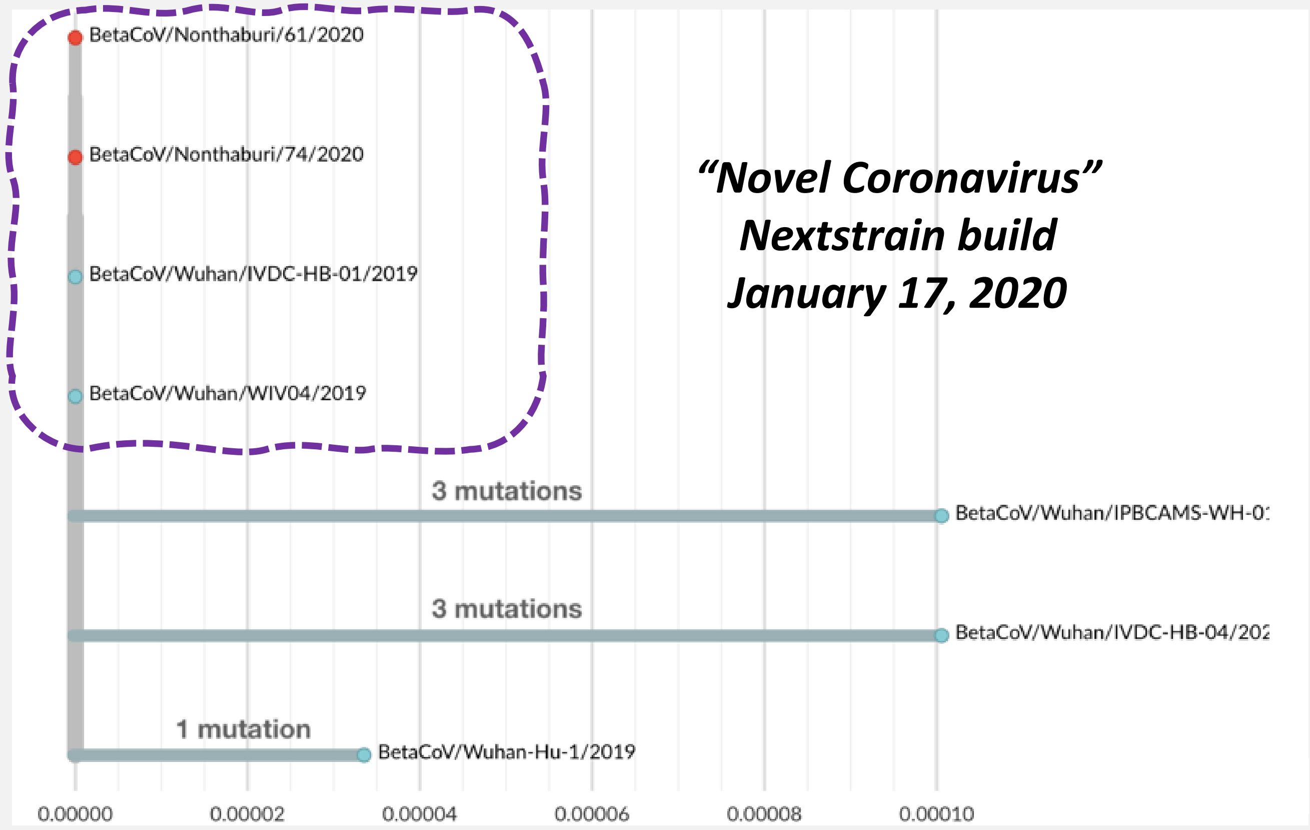
Genomic evidence for human-to-human transmission of SARS-CoV-2

Specimens sequenced early in the pandemic showed little variation, unlikely the result of separate spillover events

A week later, specimens from different cities appeared identical



"Novel Coronavirus" Nextstrain build January 11, 2020



"Novel Coronavirus" Nextstrain build January 17, 2020

CZBioHub COVID Tracker Seminar series https://covidtracker.czbiohub.org/resources, Week 1

Module 2.2 Summary

- Case studies can illustrate the many uses and benefits of genomic epidemiology
- Genomic epidemiology analyses suggest it only takes one introduction for a large outbreak to occur!
    - with accompanying rapid transmission, morbidity & mortality
- Outbreaks that appear sustained from epi curves may be made up of multiple introductions
    - Genomic data can help link outbreaks that do not appear related
- Differentiating between transmission chains & their characteristics highlights how each contributes to the outbreak and may suggest why
- WGS of paired specimens can confirm pathogen reinfection
    - via specific mutation patterns indicating distant common ancestry

Thank you!

DSHS Genomic Epidemiology Training Series

Module 2.2

Representative GenEpi Case Studiens

For further questions, contact

- Anna Battenhouse <abattenhouse@utexas.edu>

DSHS Genomic Epidemiology Training modules

Comprehensive training that can be accessed ***à la carte*** according to user needs

Group 1: **Introduction**

1.1 General Overview of Pathogen Genomic Epidemiology

1.2 Whole Pathogen Genome Sequencing

1.3 Understanding Phylogenetic Trees

Group 2: **Outbreak Investigation**

2.1 General Considerations and Approaches in Outbreak Analysis

2.2 Representative Case Studies

2.3 DSHS Outbreak Workflows with Genomic Epidemiology

2.4 Gen Epi Tools in Different Analysis Contexts

Group 3: **Surveillance to Action**

3.1 The SARS-CoV-2 Genome and its Variants

3.2 Surveillance and Retrospective Analysis

3.3 Communicating Genomic Epidemiology Insights