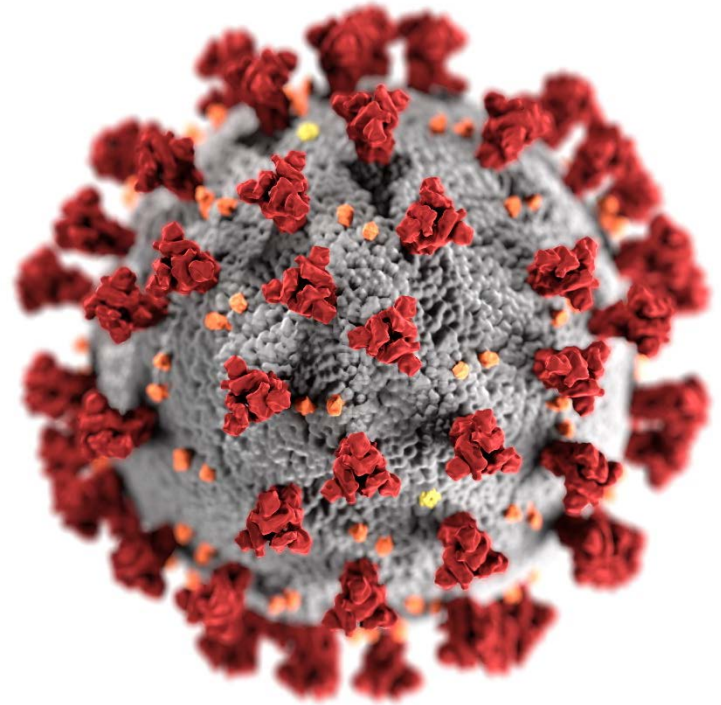


# Getting started with MicrobeTrace

## COVID-19 Genomic Epidemiology Toolkit: Module 3.2

Ellsworth Campbell, MS  
Computational Biologist  
Centers for Disease Control and Prevention



[cdc.gov/coronavirus](https://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
- 2.5 Confirming reinfection
- 2.6 Detecting & prioritizing variants
- 2.7 Wastewater-based variant tracking

## 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



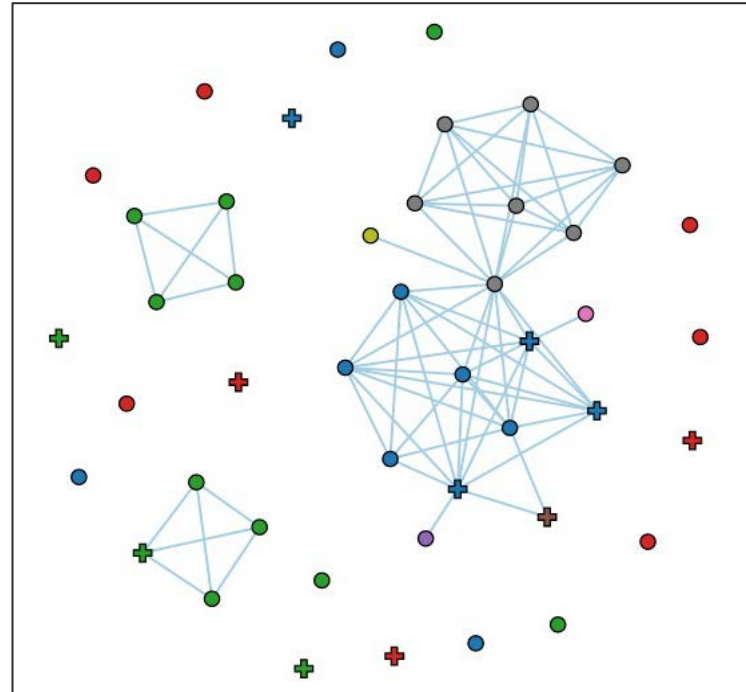
# Outline

- MicrobeTrace overview
- Transmission networks
- Types of input files
- Genomic data and MicrobeTrace
- Demo of simulated data
- Resources



# What is MicrobeTrace?

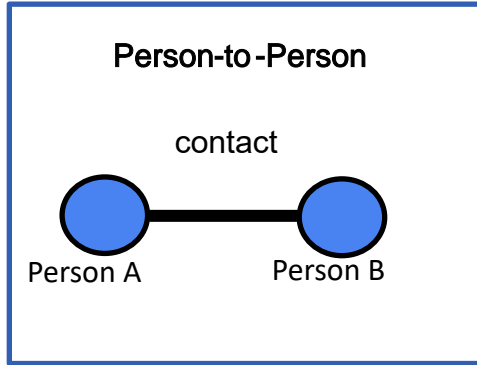
- A powerful, secure data visualization tool to map transmission networks
- Handles various data inputs
  - Genetic
  - Epidemiological
- User-friendly exploration of COVID-19 data



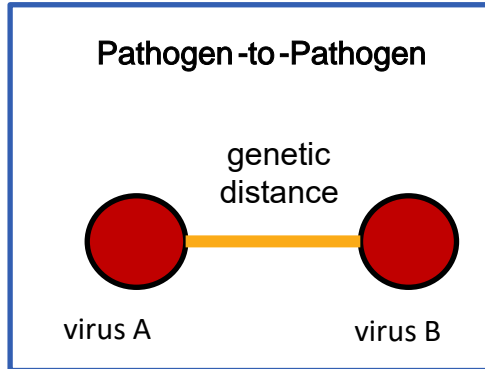
<https://microbetrace.cdc.gov/MicrobeTrace/>

# What are transmission networks?

Epi network



Genetic distance network

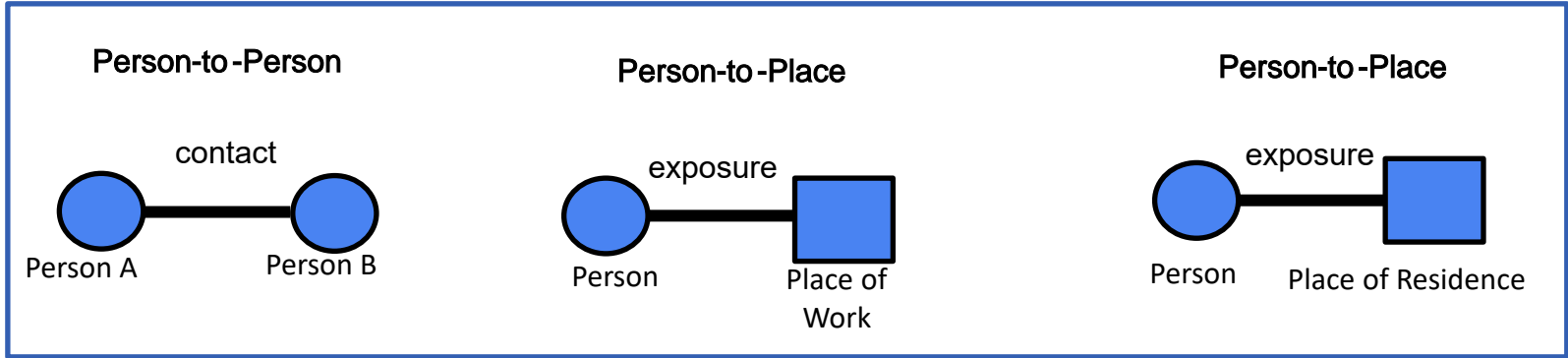


## Fine-grain resolution of transmission

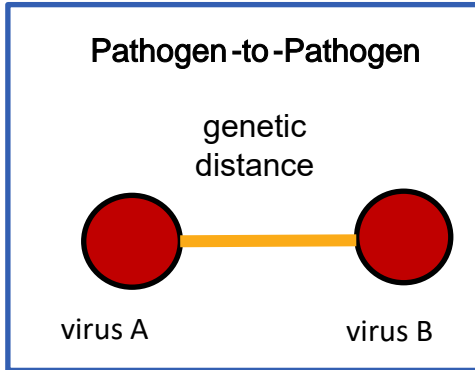
- Outbreak-specific intervention efforts
- Future prevention strategies

# Types of networks

Epi network



Genetic distance network



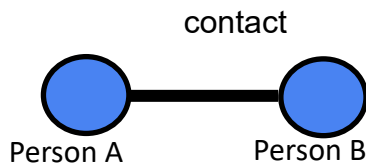
# Edge list

Edge list or contact tracing list (.csv file):

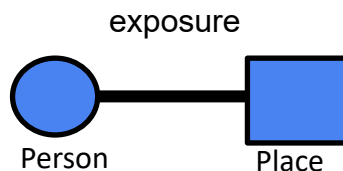
Rows = potential exposure between a case and their contact.

The edge can be from a person or a place.

Person-to-Person

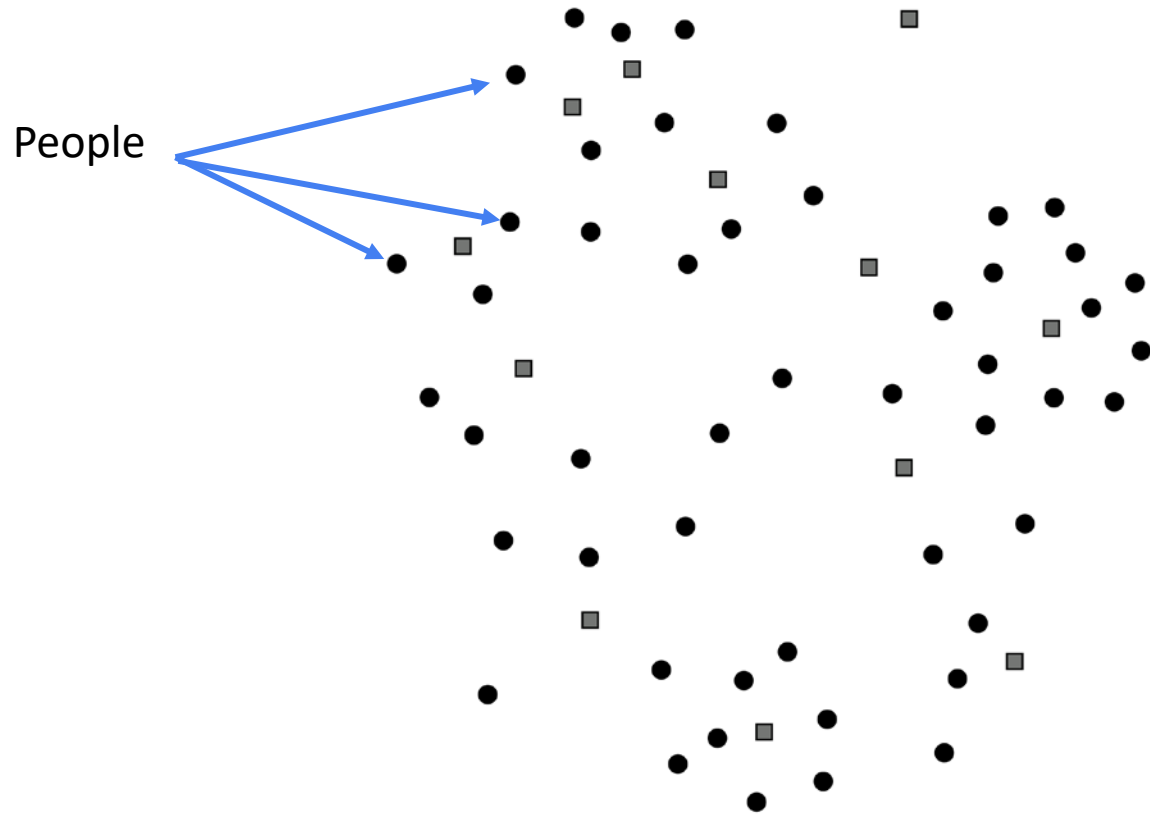


Person-to-Place



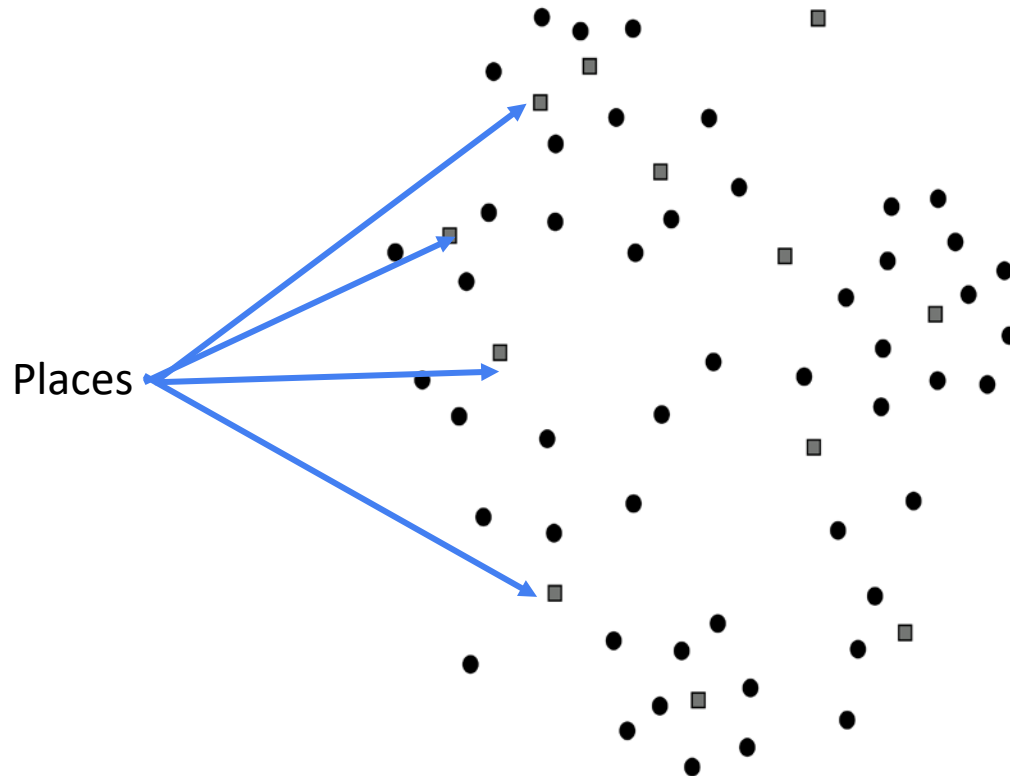
Case/Place	Contact	Contact type
36	49	locationLink
36	50	locationLink
36	51	locationLink
36	52	locationLink
31	53	contactTracing
53	49	contactTracing
31	54	contactTracing
46	54	contactTracing

# Integrated Example





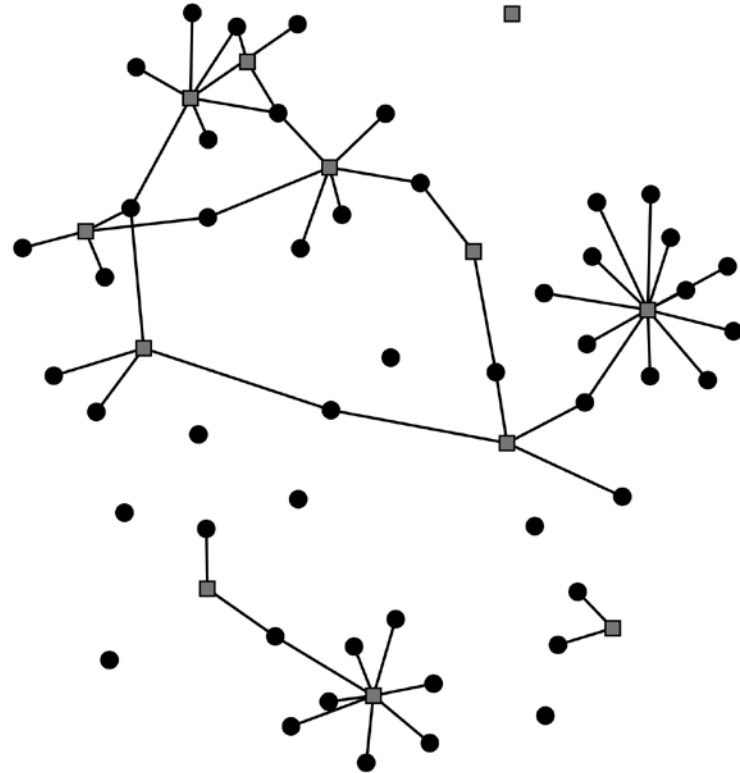
# Integrated Example 1



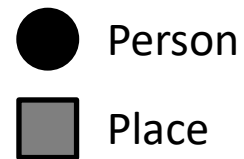
# Integrated Example 3



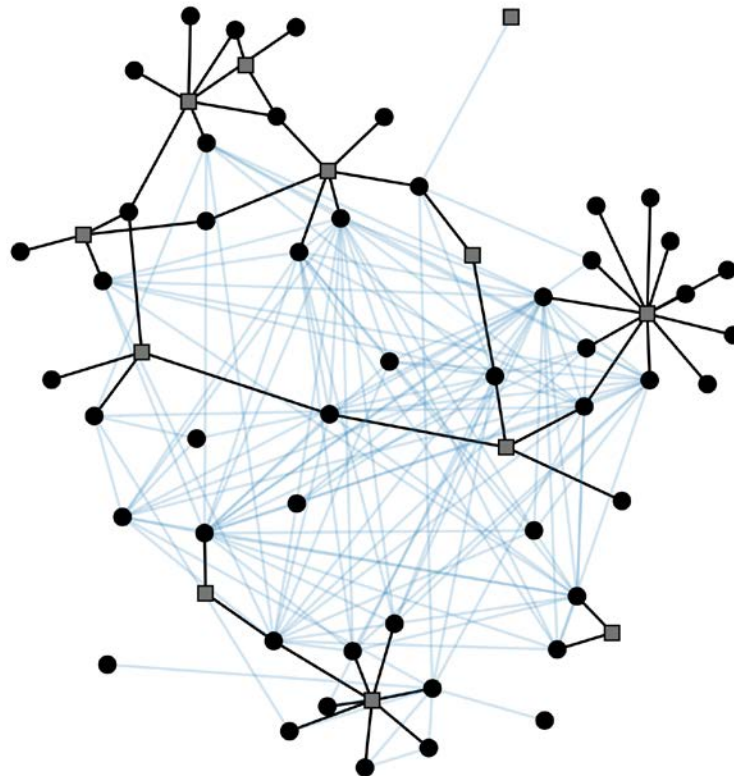
Person-to-Place  
Links



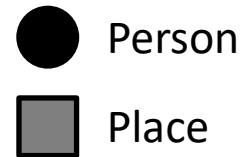
# Integrated Example 2



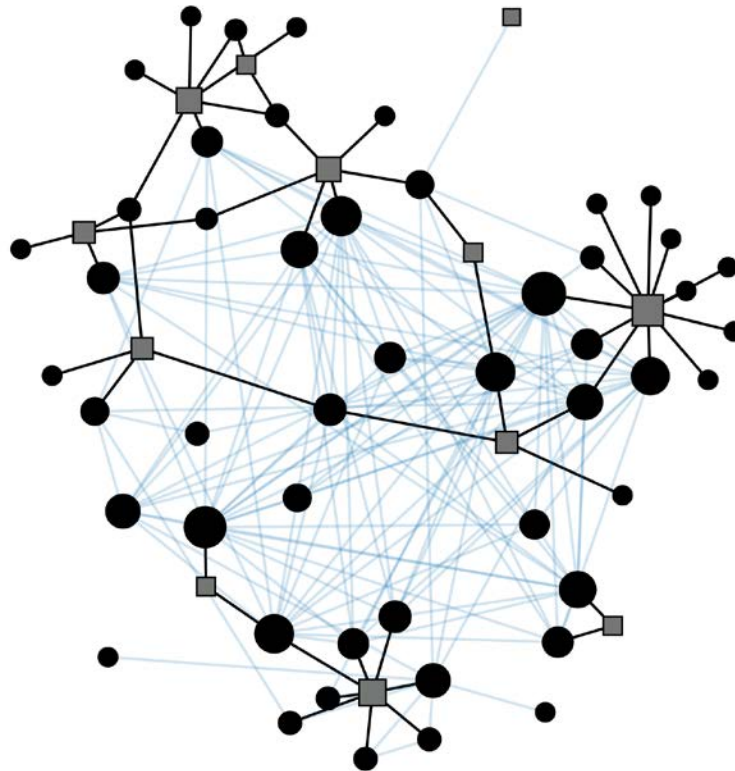
Genetic  
Links



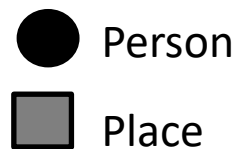
# Integrated Example 4



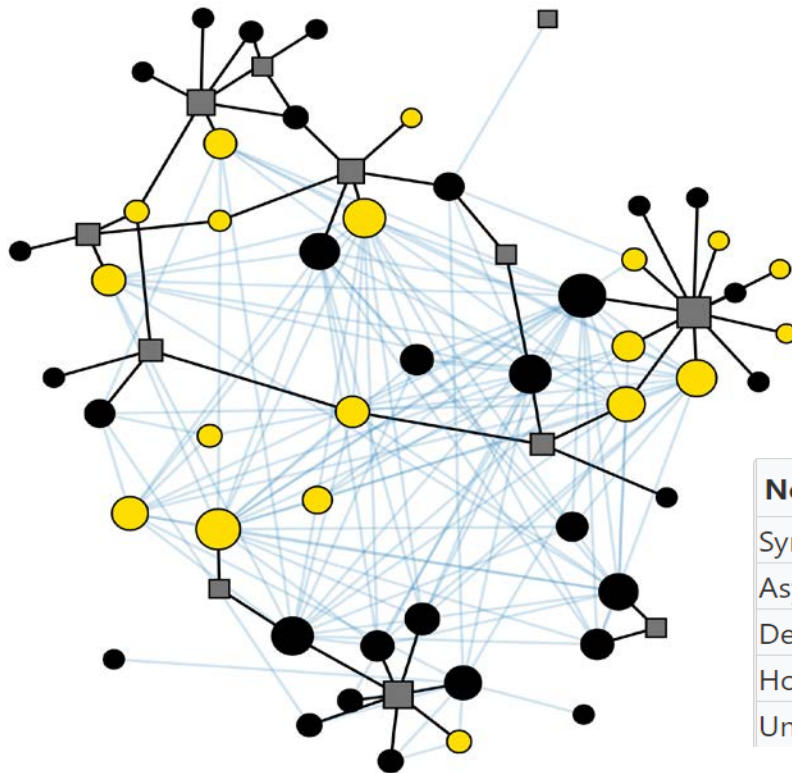
Layered  
Information



# Integrated Example 5



Layered  
Information



Node Outcome	Count	Color
Symptomatic	24	●
Asymptomatic	21	●
Death	4	●
Hospitalized	4	●
Unknown	1	●

# Genomic data inputs supported by MicrobeTrace

Aligned sequences  
(FASTA)

```
>Sequence 1  
AAATGTTATTCATGCT  
>Sequence 2  
AAATATTACTCATGCT
```

Distance matrices  
(CSV)

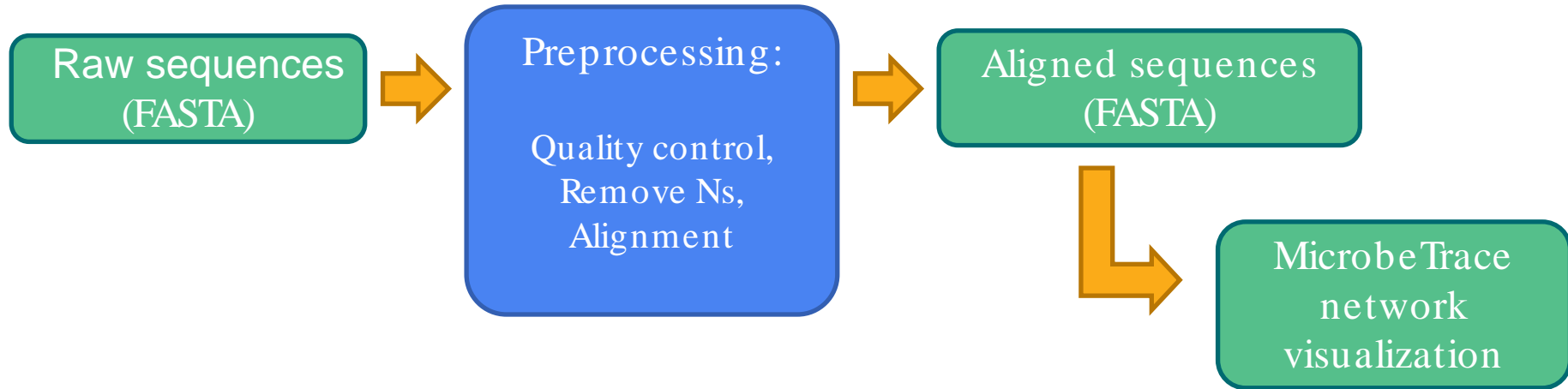
	Seq1	Seq2	Seq3	Seq4
Seq1	0	2	4	1
Seq2	2	0	1	2
Seq3	4	1	0	2
Seq4	1	2	2	0

Phylogenetic tree  
(NEWICK)

Standard text format for tree files, encodes cluster information about sequences and clades.

MicrobeTrace plots networks based on calculated genetic distance

# 1. Sequences (FASTA)

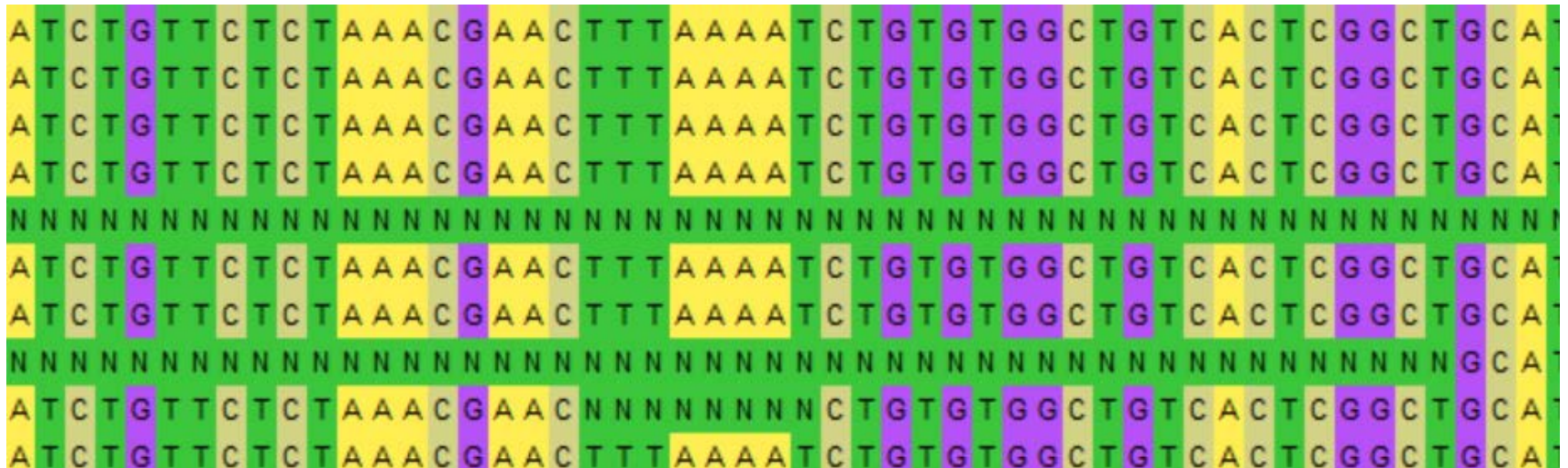


- Complete control of your sequences from beginning to end
- But data need preprocessing with external tools before importing to MicrobeTrace





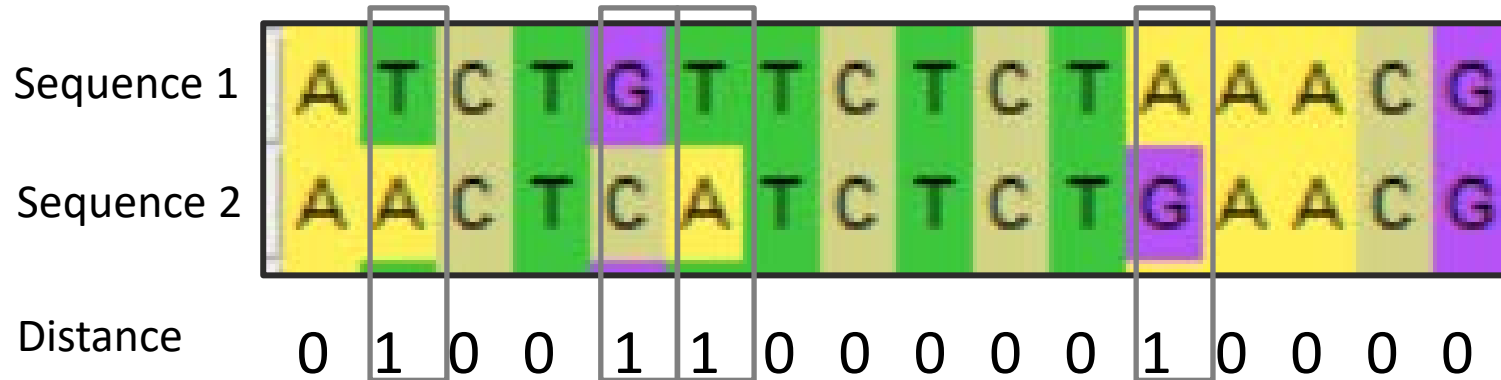
# Example alignment: poor quality, strings of Ns



# Example alignment: poor quality, misaligned

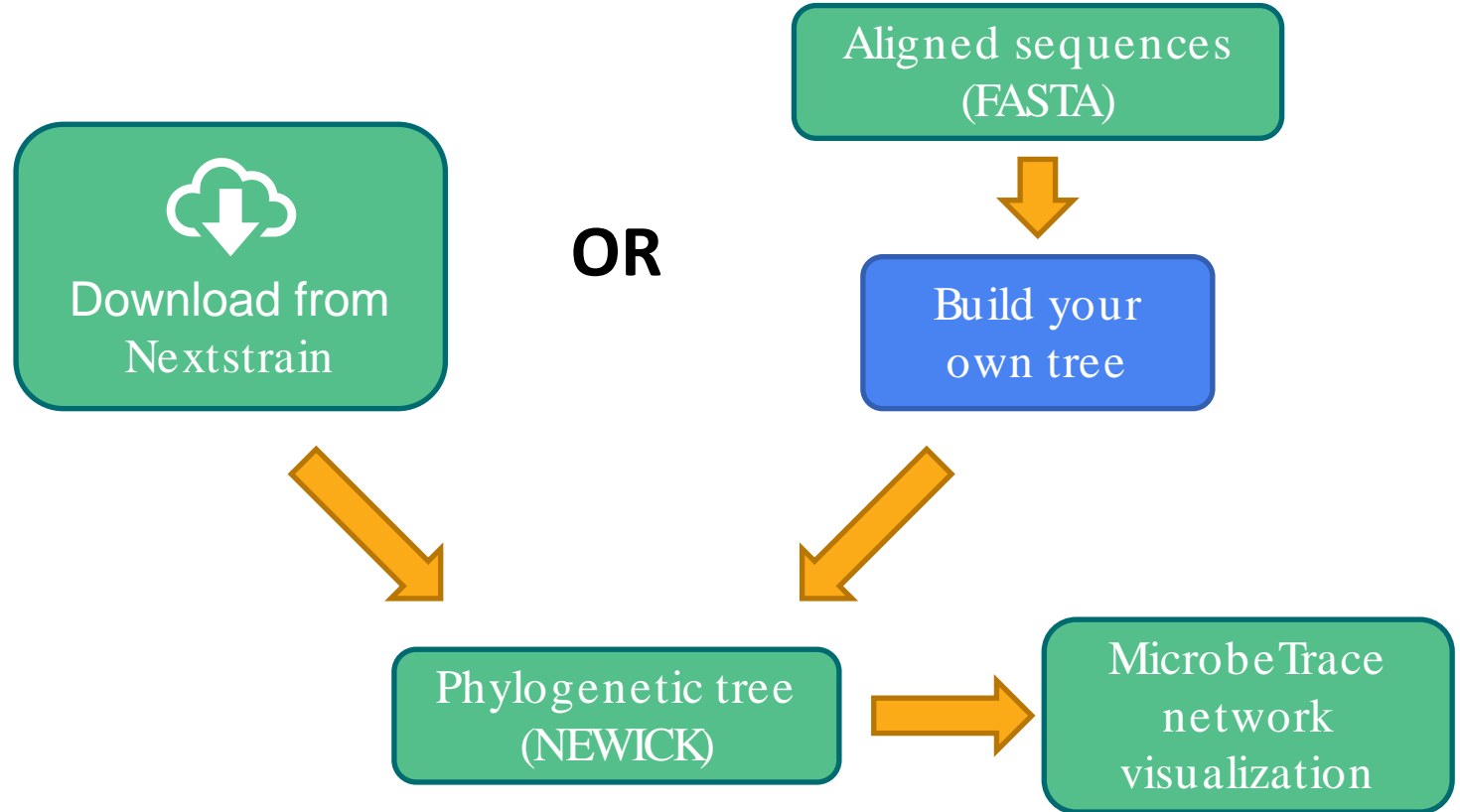
A	T	C	T	G	T	T	C	T	C	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	T	C	T	G	T	G	T	G	G	C	T	G	T	C	A	C	T	C	G	G	C	T	G	C	A
A	T	C	T	G	T	T	C	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	T	C	T	G	T	G	T	G	G	C	T	G	T	C	A	C	T	C	G	G	C	T	G	C	A	T	G
A	T	C	T	G	T	T	C	T	C	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	T	C	T	G	T	G	T	G	G	C	T	G	T	C	A	C	T	C	G	G	C	T	G	C	A
A	T	C	T	G	T	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	T	C	T	G	T	G	T	G	G	C	T	G	T	C	A	C	T	C	G	G	C	T	G	C	A	T	G	C	T
A	T	C	T	G	T	T	C	T	C	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	A	C	T	C	G	G	C	T	G	C	A	T	G	C	T	T	A	G	T	G	C	A	C	T	C
A	T	C	T	G	T	T	C	T	C	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	T	C	T	G	T	G	T	G	G	C	T	G	T	C	A	C	T	C	G	G	C	T	G	C	A
A	T	C	T	G	T	T	C	T	C	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	T	C	T	G	T	G	T	G	G	C	T	G	T	C	A	C	T	C	G	G	C	T	G	C	A
A	T	C	T	G	T	T	C	T	C	T	A	A	A	C	G	A	A	C	T	T	T	A	A	A	A	T	C	T	G	T	G	T	G	G	C	T	G	T	C	A	C	T	C	G	G	C	T	G	C	A

## Example: genetic distance



Pairwise distance = 4 single nucleotide polymorphisms or SNPs

## 2. Phylogenetic trees



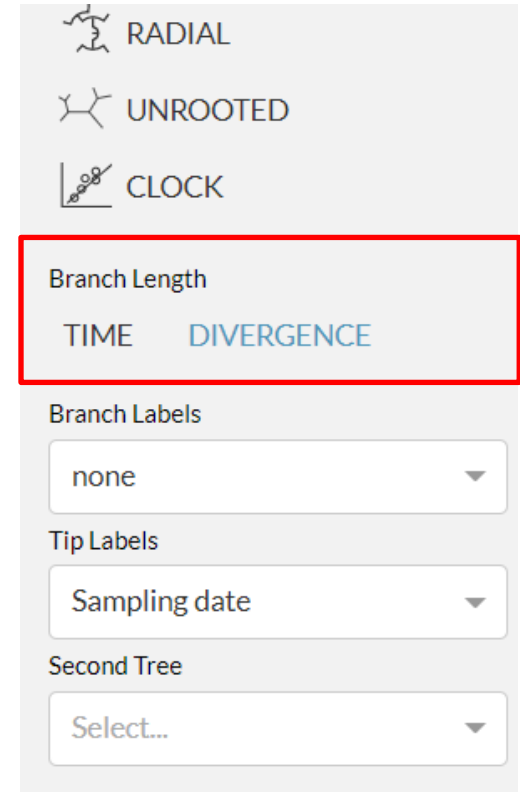
## 2. Phylogenetic trees

- Easy
- Download trees from reliable tools like Nextstrain
- Or build your own from a multiple-sequence alignment
- Nextstrain allows metadata download along with tree files

***CAUTION:*** Some web-based tree building algorithms do not clean Ns from sequences and may yield misleading results when imported into MicrobeTrace!

# Using phylogenetic trees from Nextstrain

- When viewing a tree with Nextstrain, be sure to select **DIVERGENCE** so the branch lengths are scaled according to number of mutations, not time
- Export this tree in Newick format using the **DOWNLOAD DATA** button at the bottom and branch lengths will be encoded as SNP distances for analysis by MicrobeTrace



The image shows a dropdown menu for selecting tree styles in Nextstrain. The menu is light gray and contains three options: 'RADIAL' with a circular tree icon, 'UNROOTED' with a simple branching tree icon, and 'CLOCK' with a tree icon showing a scale bar. Below these is a section for 'Branch Length' with two radio buttons: 'TIME' and 'DIVERGENCE'. The 'DIVERGENCE' option is selected and highlighted with a red border. Below this are three dropdown menus: 'Branch Labels' (set to 'none'), 'Tip Labels' (set to 'Sampling date'), and 'Second Tree' (set to 'Select...').

RADIAL

UNROOTED

CLOCK

Branch Length

TIME DIVERGENCE

Branch Labels

none

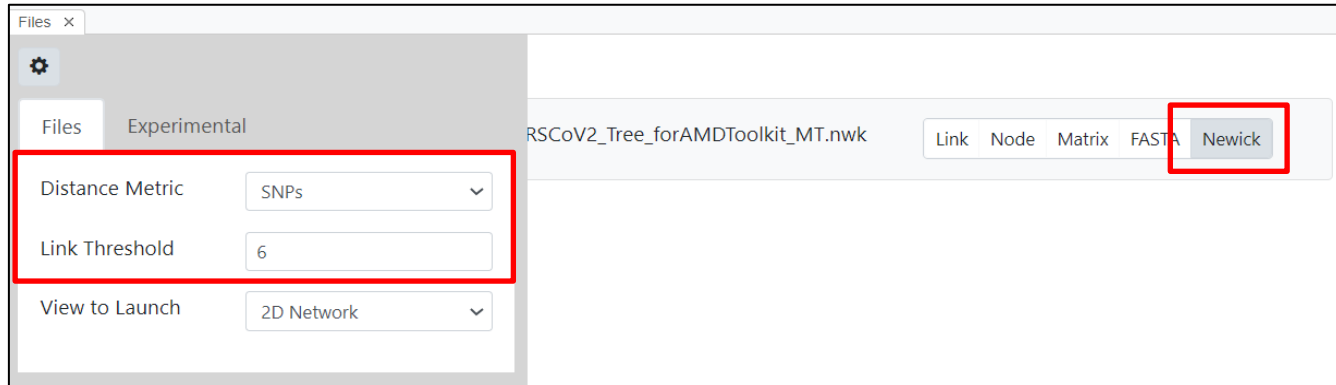
Tip Labels

Sampling date

Second Tree

Select...

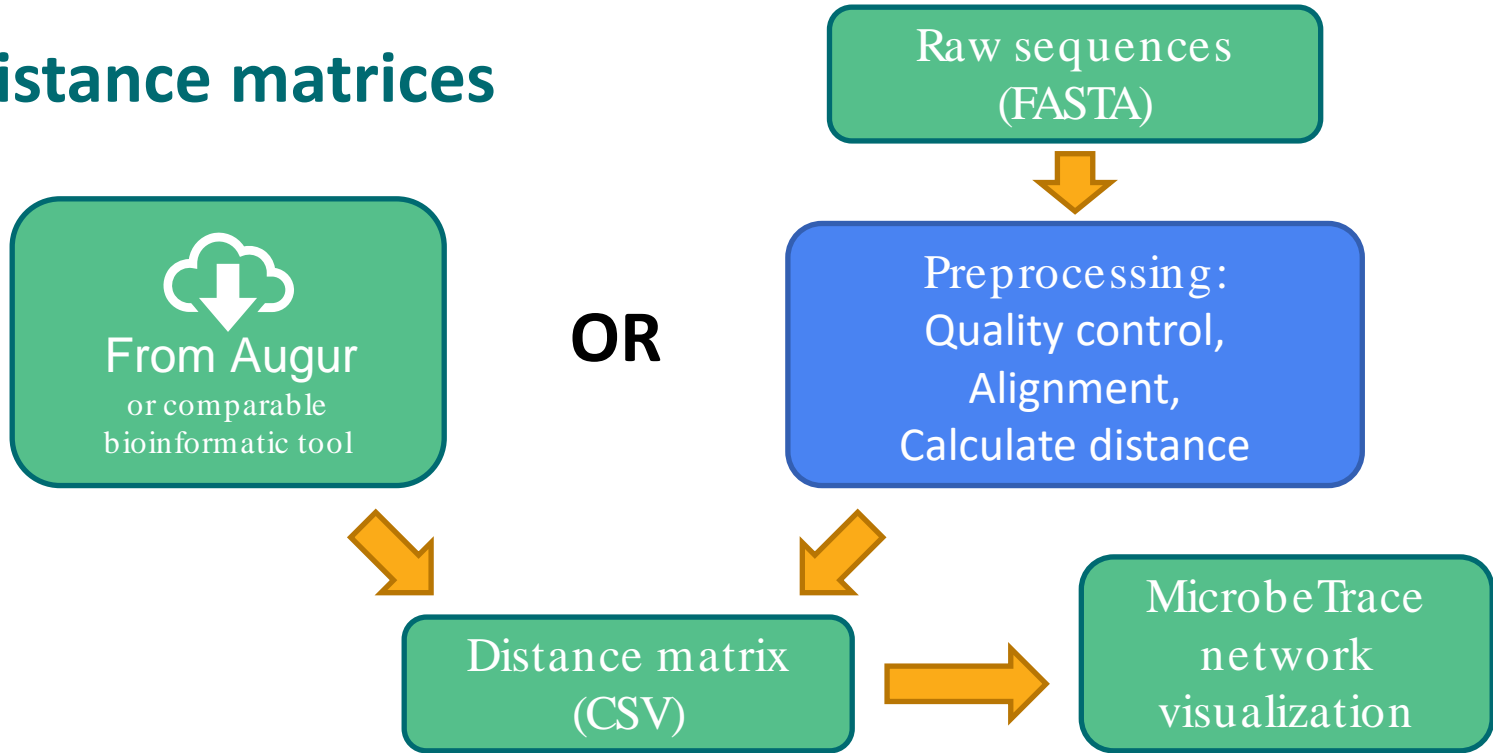
# Using phylogenetic trees from Nextstrain



- Import the Newick tree into MicrobeTrace
- Start with a link threshold of 6 SNPs and see if the network mimics the tree visualized in Nextstrain
- Adjust threshold according to dataset or investigation
- More detailed guidance available at:

<https://github.com/CDCgov/MicrobeTrace/wiki/Nextstrain-trees-and-MicrobeTrace>

### 3. Distance matrices



- Simply import CSV into MicrobeTrace – no need for data manipulation
  - Distances measured as SNPs or percent divergence.
- Integrity of distance matrix depends on quality of sequence alignment.

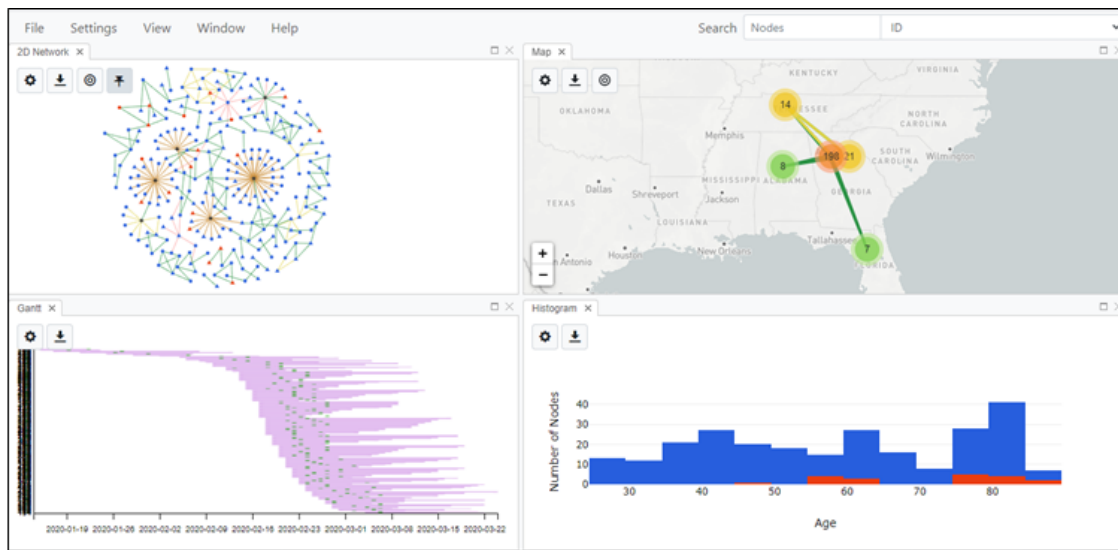


# Demo using simulated COVID genomic data



# Additional MicrobeTrace views

- MicrobeTrace has many views to help analyze genomic and epi data.
- Also useful for epidemiologic studies (e.g., contact tracing) without genomic data.



# Links and Resources

## Explore the tool

<https://microbetrace.cdc.gov>

## Tutorial video

<https://youtu.be/O52eeyUbplo>

## Everything you want to know about MicrobeTrace

<https://github.com/CDCgov/MicrobeTrace/wiki>

## Example data

<https://github.com/cdcgov/microbetrace>

## Support

[microbetrace@cdc.gov](mailto:microbetrace@cdc.gov)

# Acknowledgements

## Development

Tony Boyles

Ells Campbell

Jay Kim

Michael Chiu

Sergei Kniazev

## Testing, analytics, communications, and outreach

Anupama Shankar

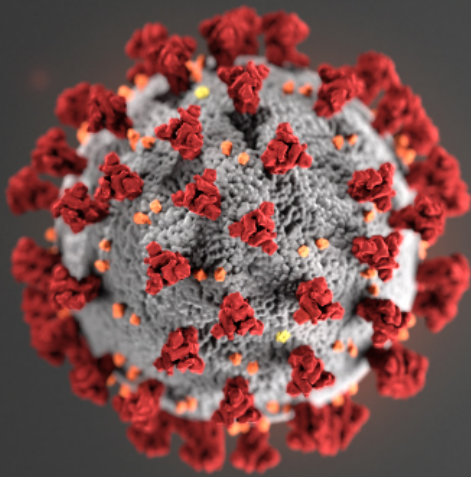
Roxy Cintron-Moret

## Team lead

Bill Switzer

## Learn more

- Next module
  - 3.3 Real-time phylogenetics with UShER
- COVID-19 Genomic Epidemiology Toolkit
  - Find further reading
  - Subscribe to receive updates on new modules as they are released
  - [go.usa.gov/xAbMw](https://go.usa.gov/xAbMw)



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



# FOR WEBSITE

## Further Reading

- Campbell et al. 2020. Biorxiv. <https://www.biorxiv.org/content/10.1101/2020.07.22.216275v1>
- Vang et al. 2021. MMWR. <https://www.cdc.gov/mmwr/volumes/70/wr/mm7001a5.htm>

## Resources

- MicrobeTrace <https://microbetrace.cdc.gov>
- MicrobeTrace Wiki <https://github.com/CDCgov/MicrobeTrace/wiki>
- MicrobeTrace tutorial <https://youtu.be/O52eeyUbplo>

## Hands-on

- Download example data <https://github.com/CDCgov/MicrobeTrace>