AMD TRAINING LEAD *and* BIOINFORMATICS REGIONAL RESOURCE

# Demystifying Series: SARS-CoV-2 Sequencing Resources

*Joel R. Sevinsky, Ph.D.*
*WY PHL AMD Training Lead*
*MA DPH AMD Bioinformatics Resource*
*27 May 2020*

# Demystifying Series Objectives:

- Discuss a specific application of NGS data
  - Often the application is constantly evolving/improving
- Create more interactive content than static content
  - Google sheets, websites, Tableau views
- Record sessions for review
  - Unlisted youtube content available by link on www.statphb.org
- Provide slides and materials on www.staphb.org as a resource
  - Free to download slides as pdf file
- Since it is not possible to cover all aspects of a proposed topic in a single webinar, these webinars serve as a beginning to your investigation and education on the topic.
- Encourage additional discussions on Slack (contact Noah Hull, Kelly Oakeson, Erin Young, or Joel Sevinsky to be connected).

# Primary/Secondary Resources

Most of the material from this webinar can be found at the CDC Github site. This document repository is meant to serve as the start of a crowd-sourced collection of information, documentation, protocols and other resources for public health laboratories intending to sequence SARS-CoV-2 coronavirus samples. This collection is maintained and curated by Duncan MacCannell from the Office of Advanced Molecular Detection (AMD) at the Centers for Disease Control and Prevention (CDC).

The CDC Github site is very thorough and dense (the good dense). This webinar is meant to highlight aspects of this resource that are most applicable to state public health laboratories with PulseNet expertise.

# Outline for today's webinar

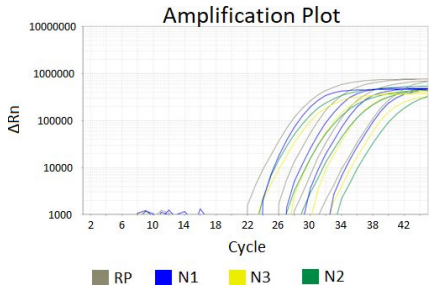
Sequencing Resources for SARS-CoV-2

- Wet bench protocols
- Bioinformatics pipelines
- Genomic epidemiology tools

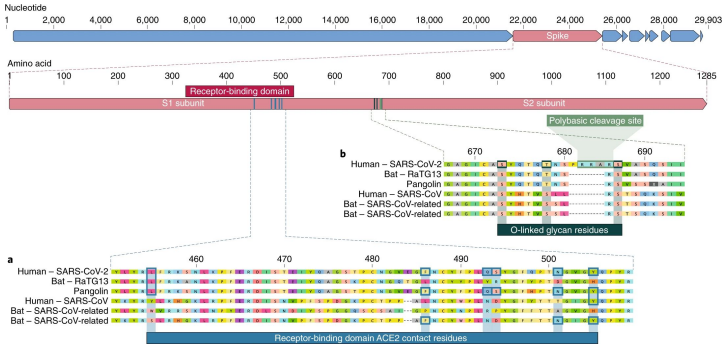
**Big Picture**: No need to reinvent the wheel, an option that will fit your lab is most likely available, as well as help setting it up.

Disclaimer: This webinar will be recorded and will be available at http://www.staphb.org/training/sars-cov-2

# SARS-CoV-2 Sequencing Workflow

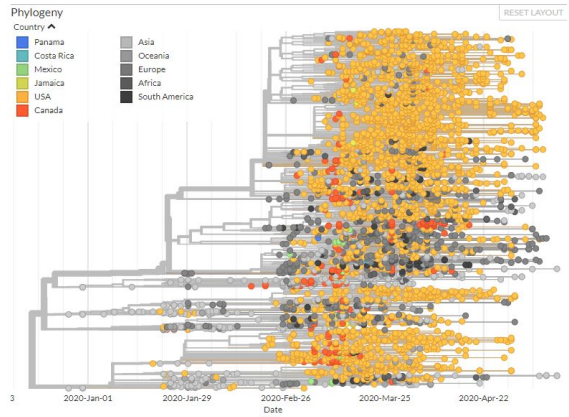



If positive, proceed to sequencing workflow.

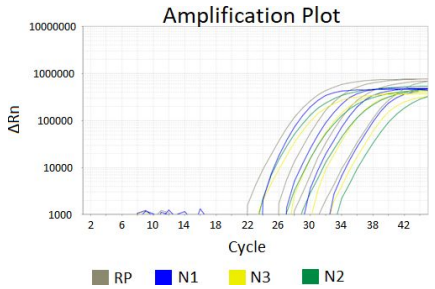



*Nature Medicine* **26**, 450–452(2020)

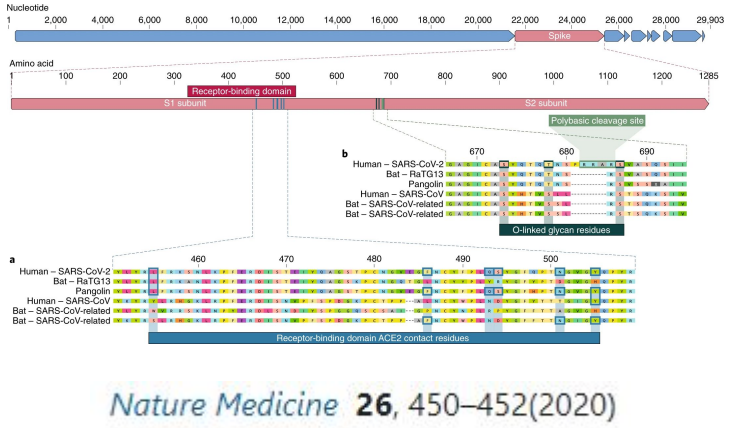
nextstrain.org



# SARS-CoV-2 Sequencing Workflow

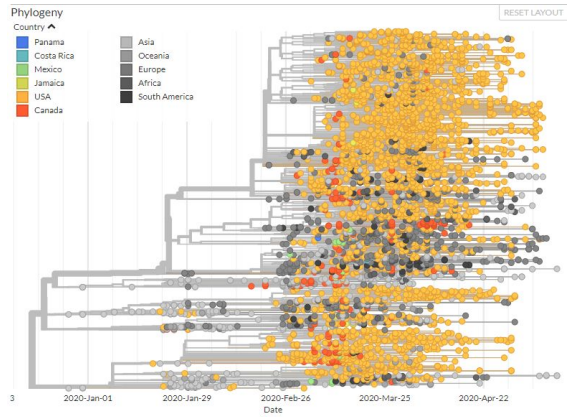



If positive, proceed to sequencing workflow.

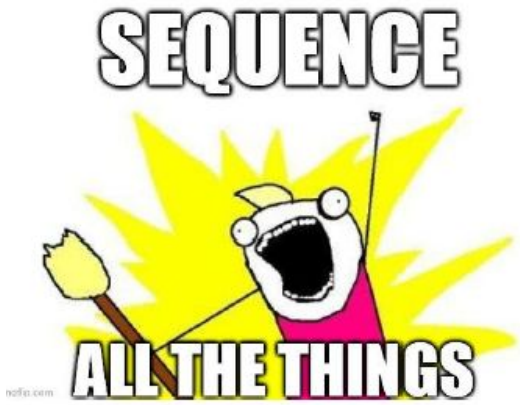



*Nature Medicine* **26**, 450–452(2020)

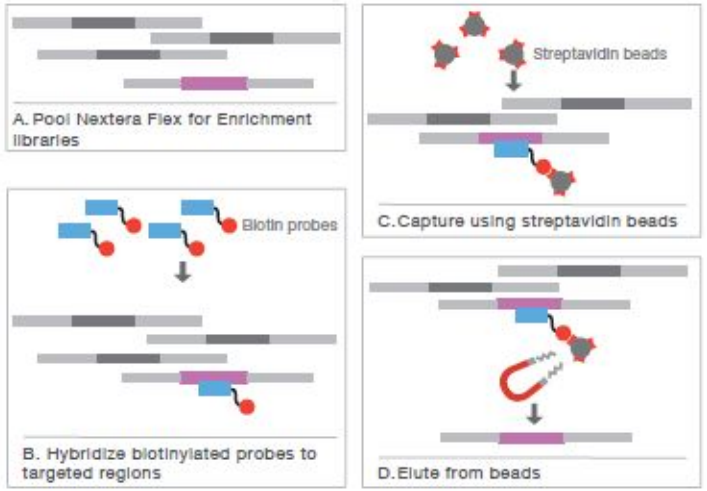
nextstrain.org



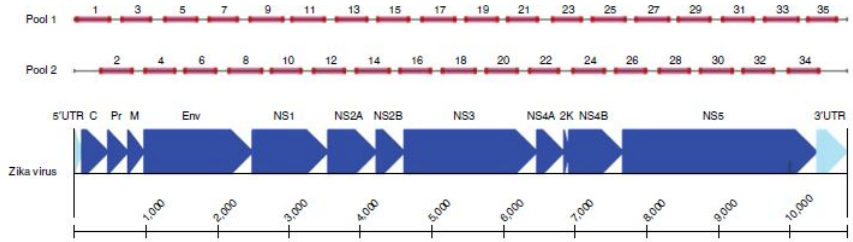
# Sequencing - Four Methods and Counting




Metagenomics



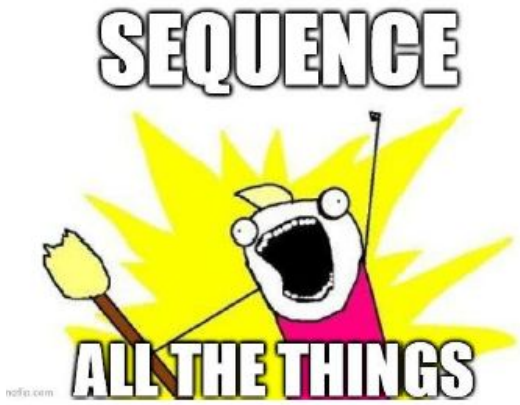

Enrichment



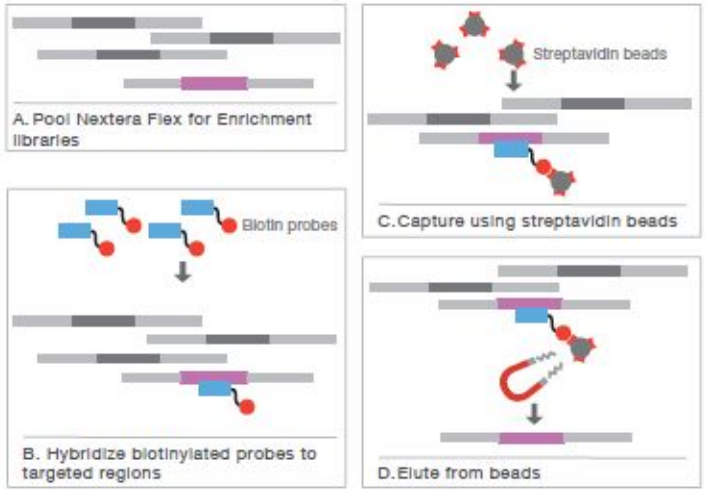

**Tiled Amplicon**

- **ARTIC Protocol**
- **Tailed Amplicon**

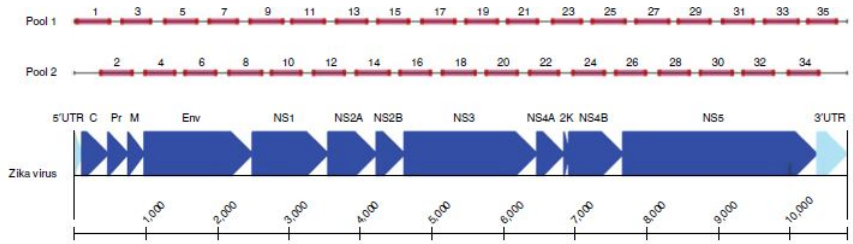
# Sequencing - Four Methods and Counting




**Metagenomics**



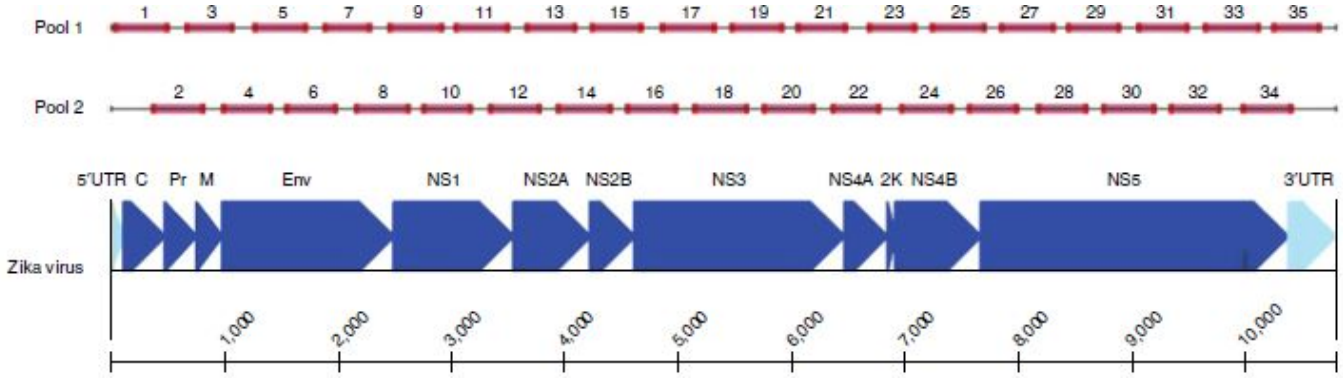

**Enrichment**




**Tiled Amplicon**

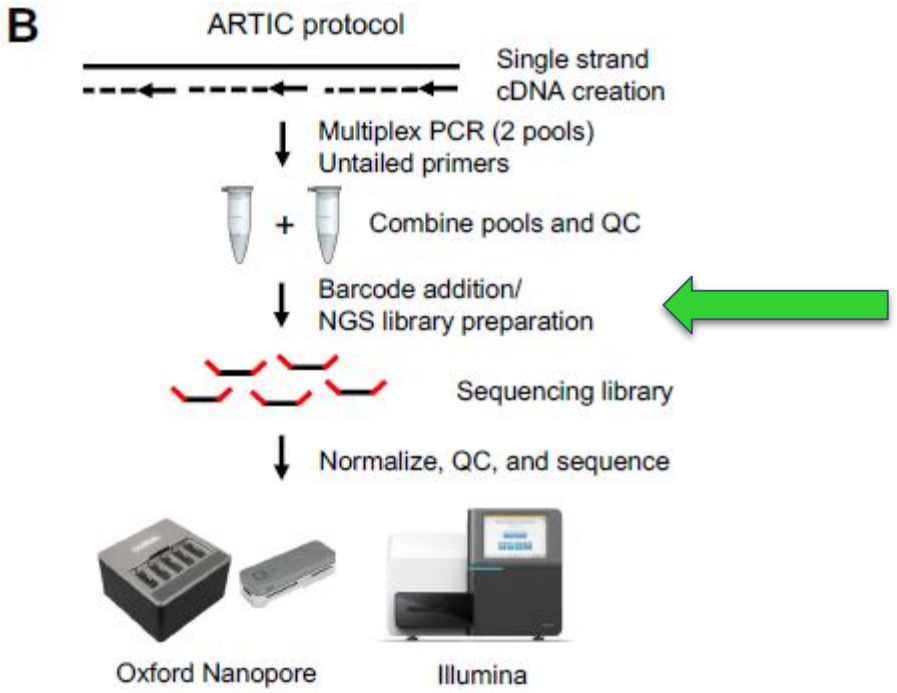
- **ARTIC Protocol**
- **Tailed Amplicon**

# ARTIC Protocol




- Set of PCR primers that covers the SARS-CoV-2 genome
- ~200 primers for ~100 overlapping ~400 bp amplicons
- Primers split into 2 pools to prevent overlapping amplicons in the same reaction.
- https://artic.network/ncov-2019

# ARTIC Protocol




Key step is deciding whether to use barcodes or DNA Flex.

- Barcodes for ONT and Illumina
- DNA Flex only for Illumina

DNA Flex allows for PulseNet like protocol after PCR reactions.

https://www.biorxiv.org/content/10.1101/2020.05.11.088724v1.full.pdf

# Sequencing ARTIC Amplicons

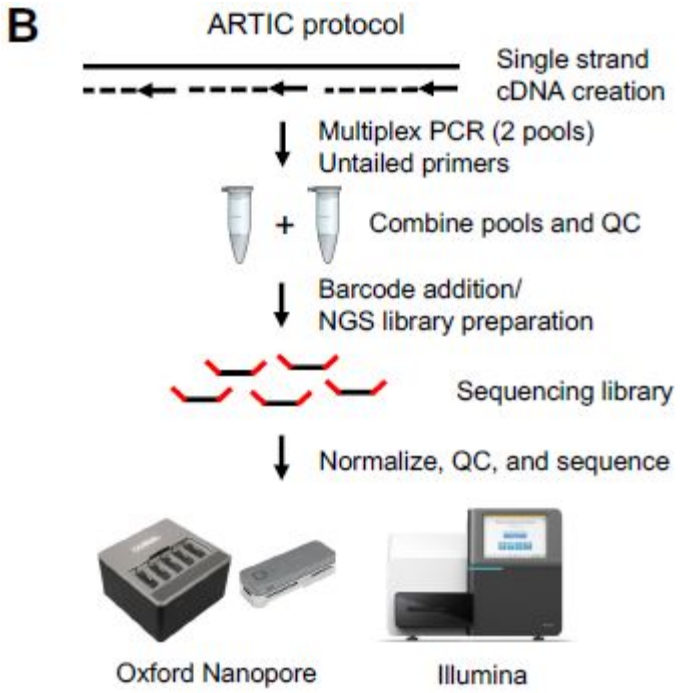
- Both Illumina and Oxford Nanopore sequencers can sequence ARTIC amplicons.
- Illumina has experienced rapid adoption in public health labs due to the availability of Illumina instruments and the ability to process ARTIC amplicons with the Illumina DNA Flex library kit in order to barcode.

# ARTIC V3 Illumina Protocol

Protocols.io for labs with PulseNet expertise:
https://www.protocols.io/groups/staphb

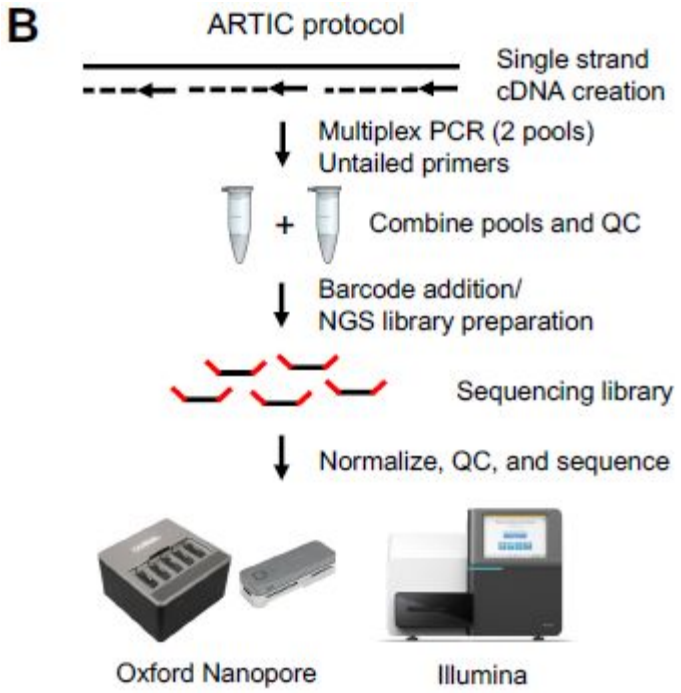
- Does not go into details on primer design.
- Does not yet have bioinformatics protocols.

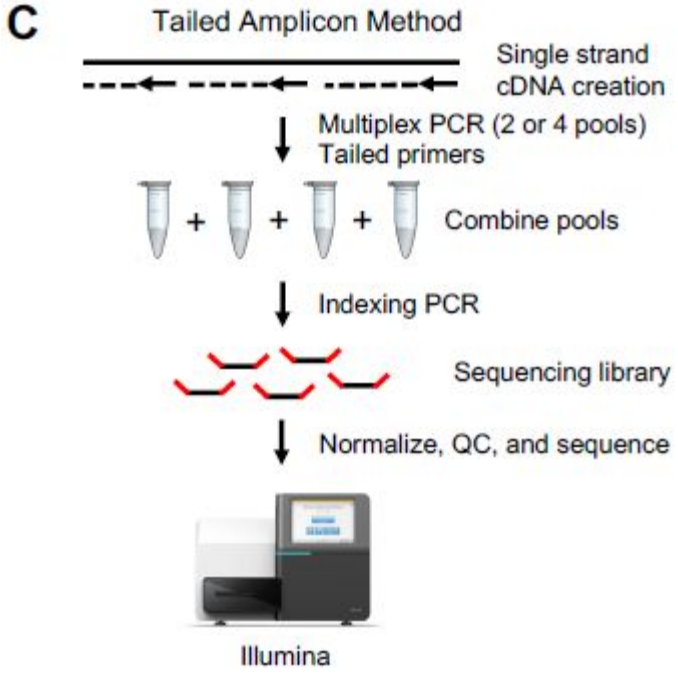
# ARTIC vs Tailed Amplicon Sequencing




https://www.biorxiv.org/content/10.1101/2020.05.11.088724v1.full.pdf

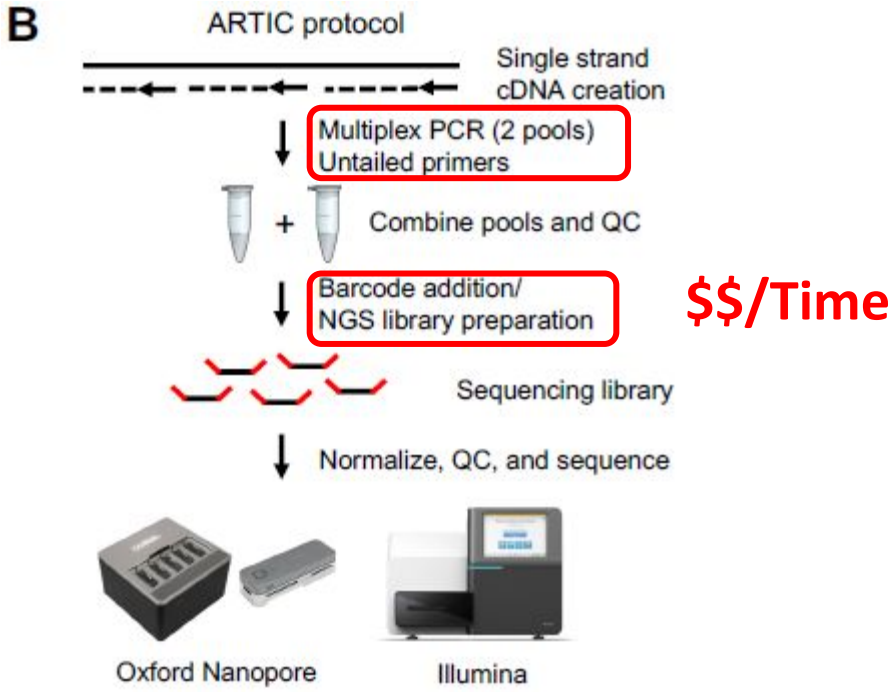
# ARTIC vs Tailed Amplicon Sequencing



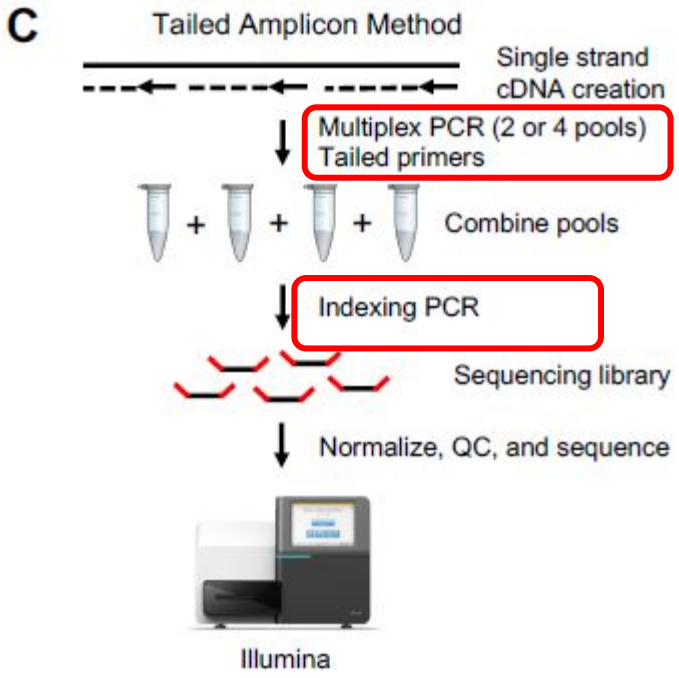




https://www.biorxiv.org/content/10.1101/2020.05.11.088724v1.full.pdf

# ARTIC vs Tailed Amplicon Sequencing

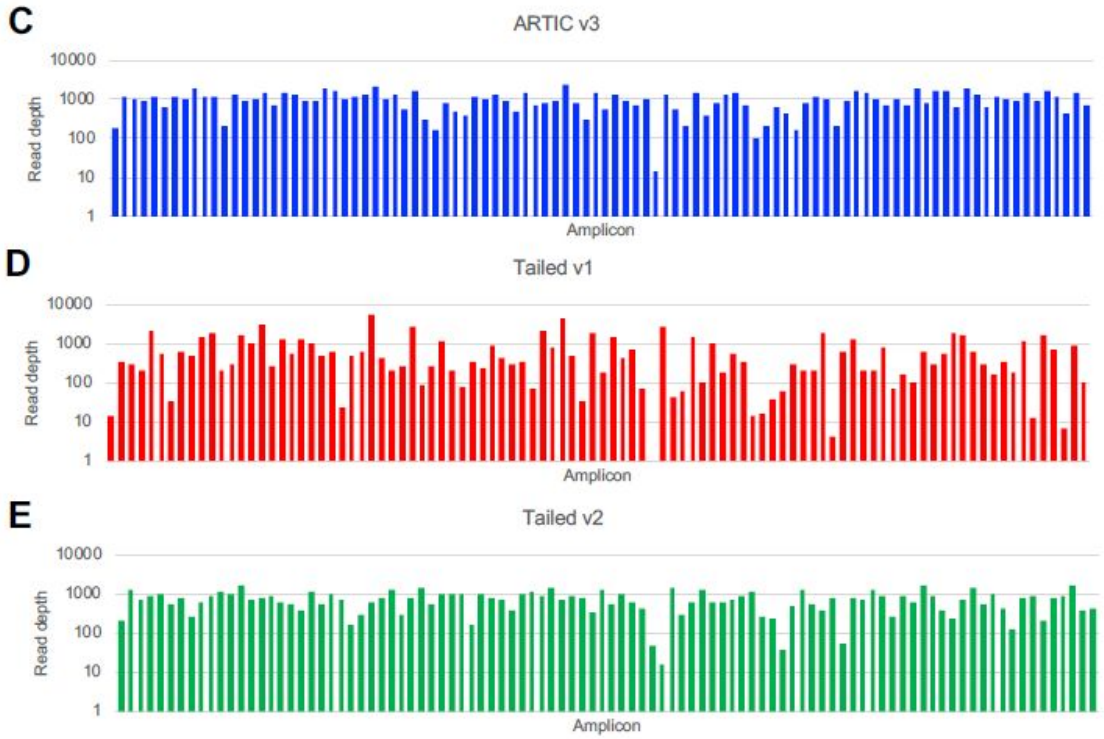


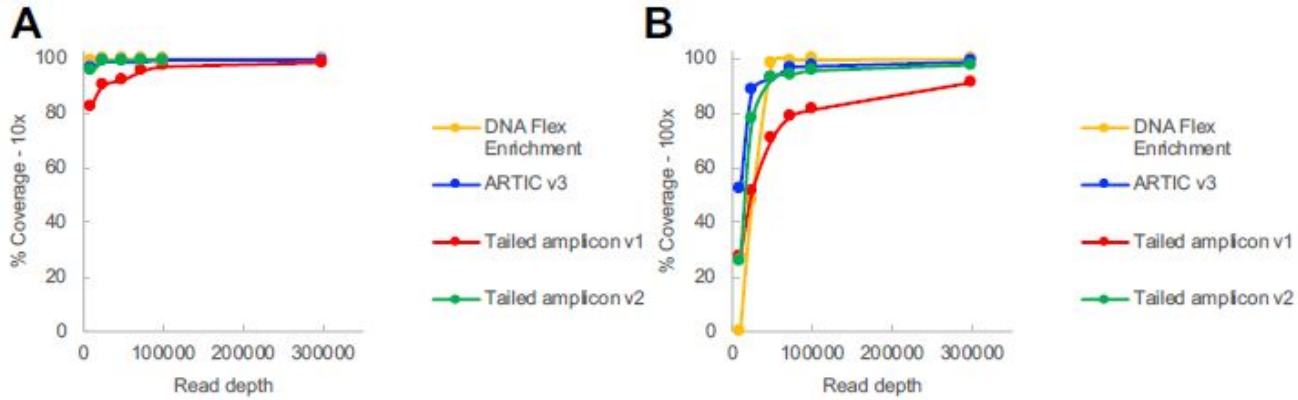

$$/Time




https://www.biorxiv.org/content/10.1101/2020.05.11.088724v1.full.pdf

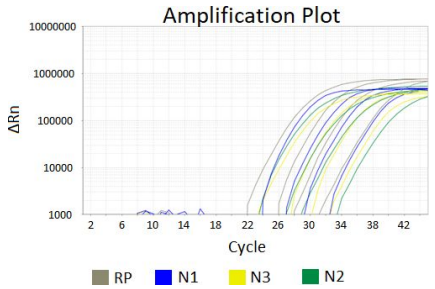
# Tailed Amplicon Sequencing



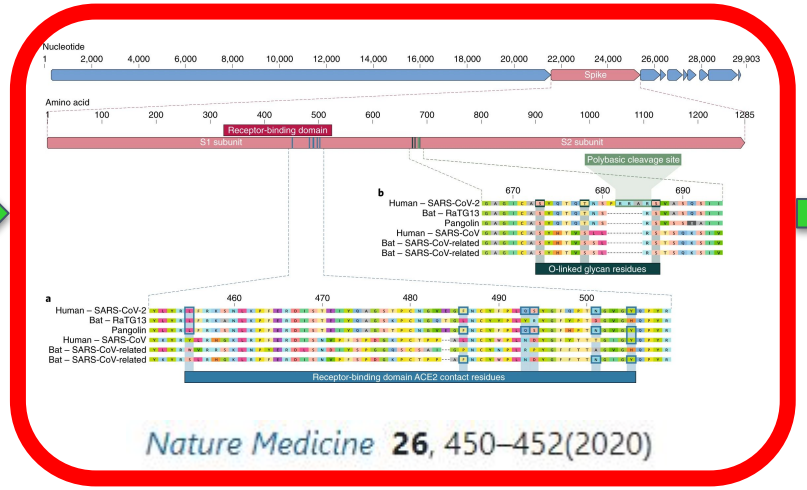




https://www.biorxiv.org/content/10.1101/2020.05.11.088724v1.full.pdf

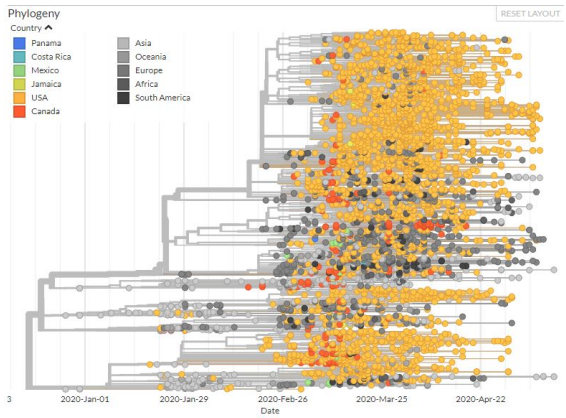
# SARS-CoV-2 Sequencing Workflow




If positive, proceed to sequencing workflow.




nextstrain.org



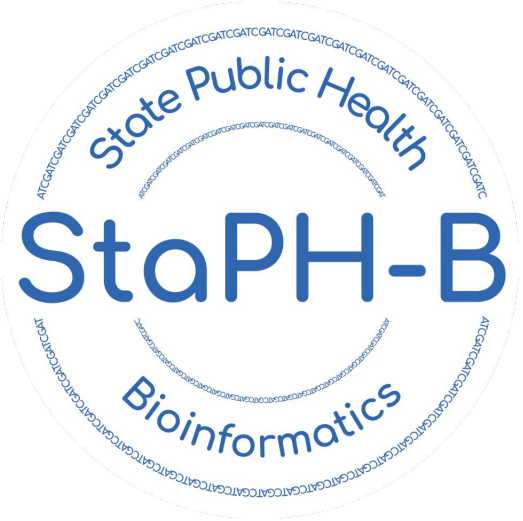
# Bioinformatics

At this point in time, most bioinformatics approaches are focused on reference based assembly rather that *de novo*.

Most pipelines for reference alignment, primer trimming, and variant calling include a step with iVar from the Andersen lab at the Scripps Research Institute.

Several pipelines that utilize iVar for ARTIC V3 amplicons are currently available.

# Bioinformatics - StaPH-B Toolkit

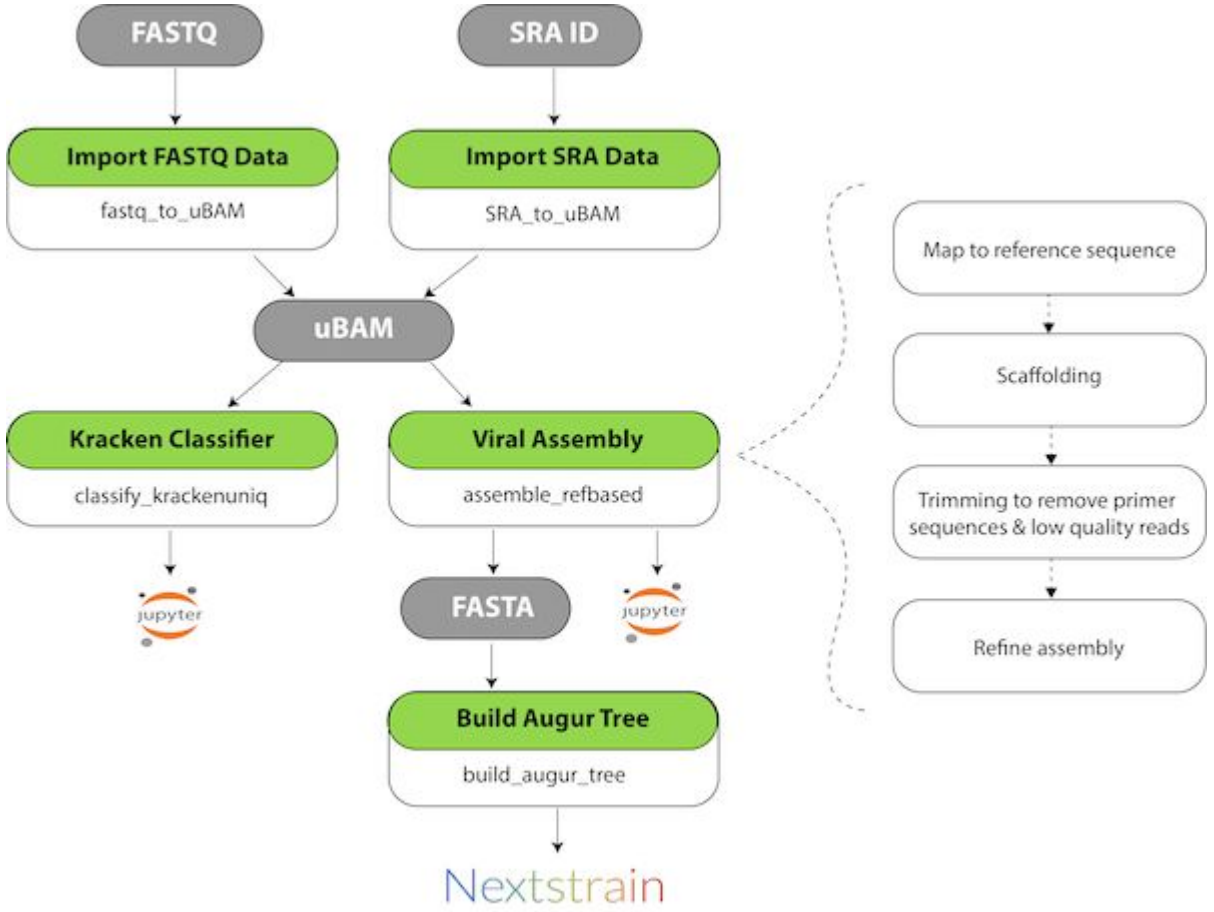



Overview - https://github.com/StaPH-B/staphb_toolkit

Installation - https://staph-b.github.io/staphb_toolkit/install/

StaPH-B toolkit contains commands for both specified docker containers as well as NextFlow workflows. The Monroe workflow (staphb-wf monroe) can accept both ONT and Illumina ARTIC protocol data.

# Bioinformatics - Terra.bio




Example pipeline from Terra.bio

- Google account required
- Cloud based but in *your* cloud, not anyone else's.
- Lot's of workflows available.

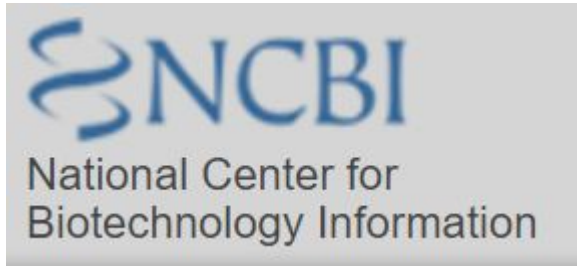
# Bioinformatics - Many Others

- Commercial
  - CLC Genomics Workbench
  - Bionumerics
- Other open source
  - ARTIC Network
  - UTPHL
  - Connor Lab

# Sharing Data

The GISAID Initiative promotes the rapid sharing of data from all influenza viruses and the coronavirus causing COVID-19.

GISAID Bulk Upload




SRA

- Sequence Read Archive, stores raw sequencing data and alignment information to enhance reproducibility and facilitate new discoveries through data analysis.

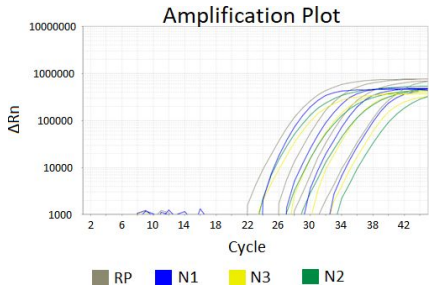
Genbank

- the NIH genetic sequence database, an annotated collection of all publicly available DNA sequences.

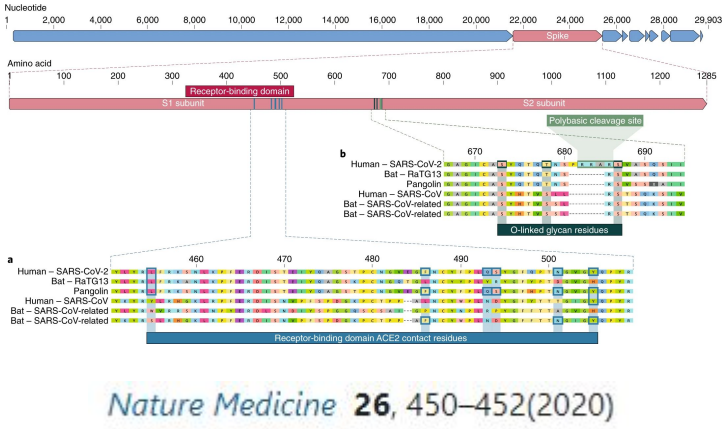
SRA and Genbank Upload

METAGENOTE: Home - https://youtu.be/TLjcASgsm_8

# SARS-CoV-2 Sequencing Workflow

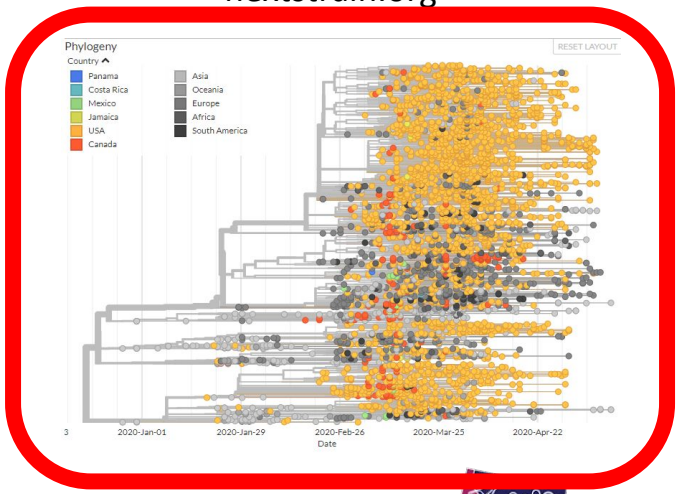



If positive, proceed to sequencing workflow.

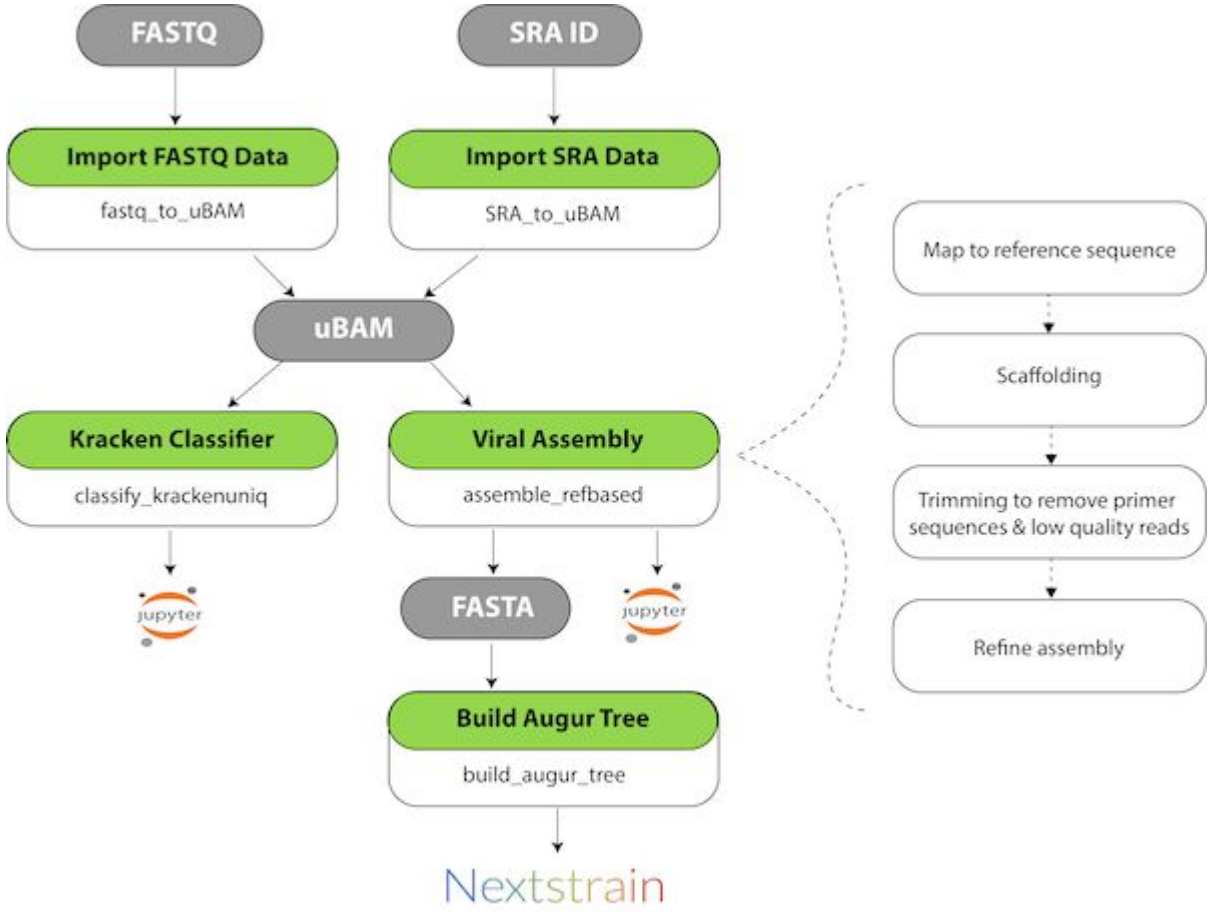



*Nature Medicine* **26**, 450–452(2020)

nextstrain.org



# Bioinformatics → Genomic Epidemiology




Example pipeline from [Terra.bio](Terra.bio)

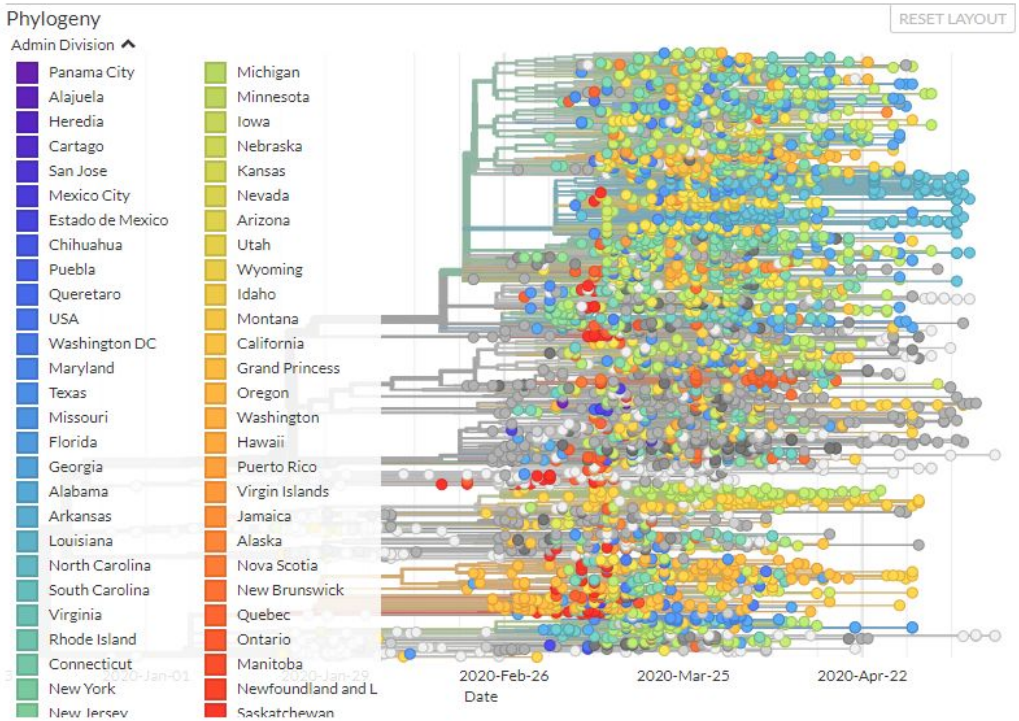
- Google account required

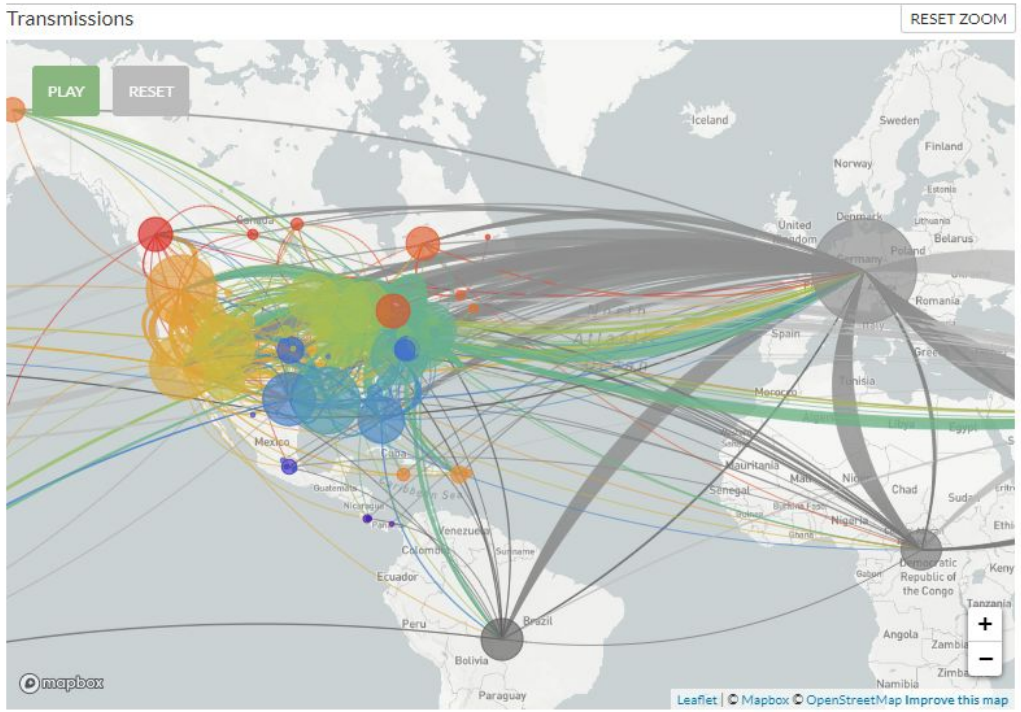
Pipeline incorporates tree building in augur for NextStrain.

What is NextStrain???

# Genomic Epidemiology

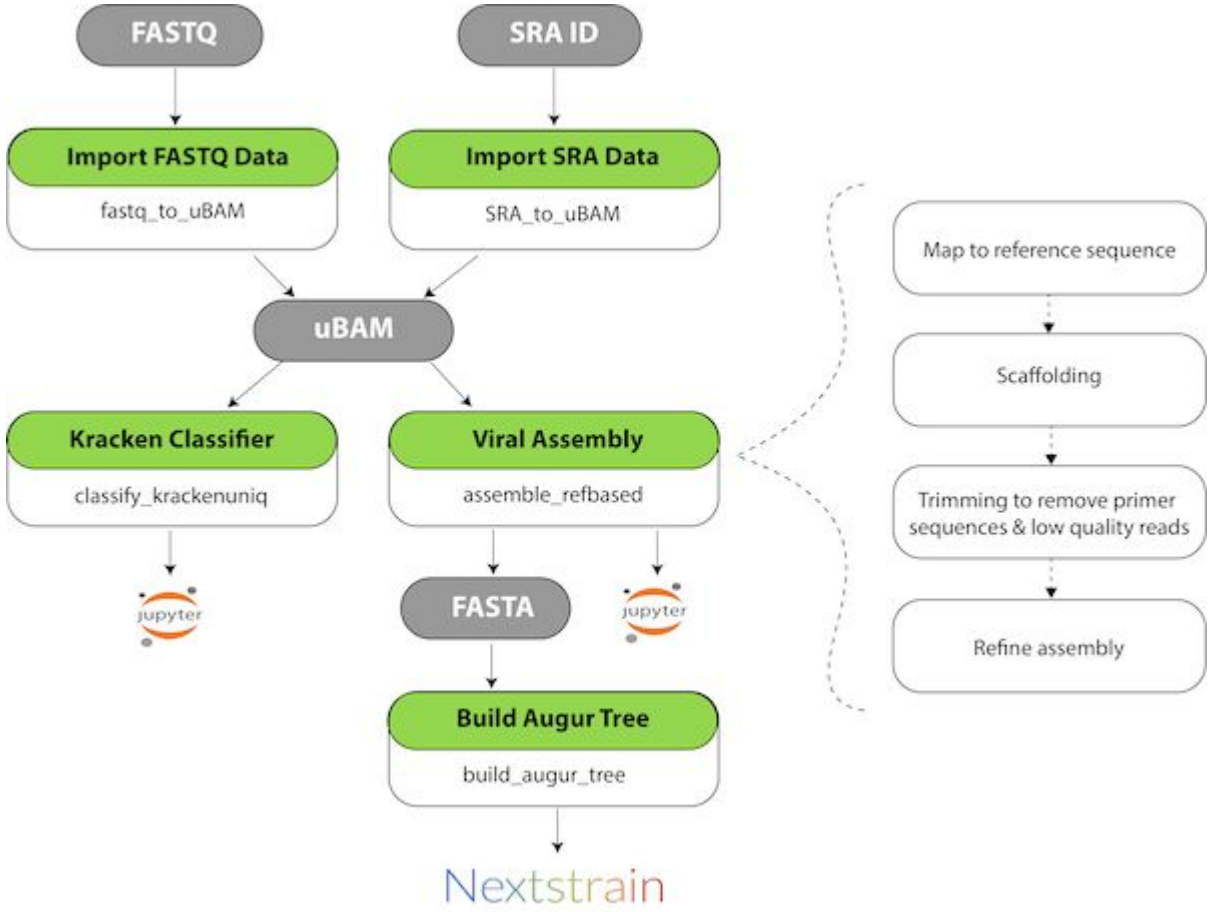
Nextstrain




https://nextstrain.org/ncov/north-america?c=division&r=division

# Bioinformatics → Genomic Epidemiology




Example pipeline from [Terra.bio](Terra.bio)

- Google account required

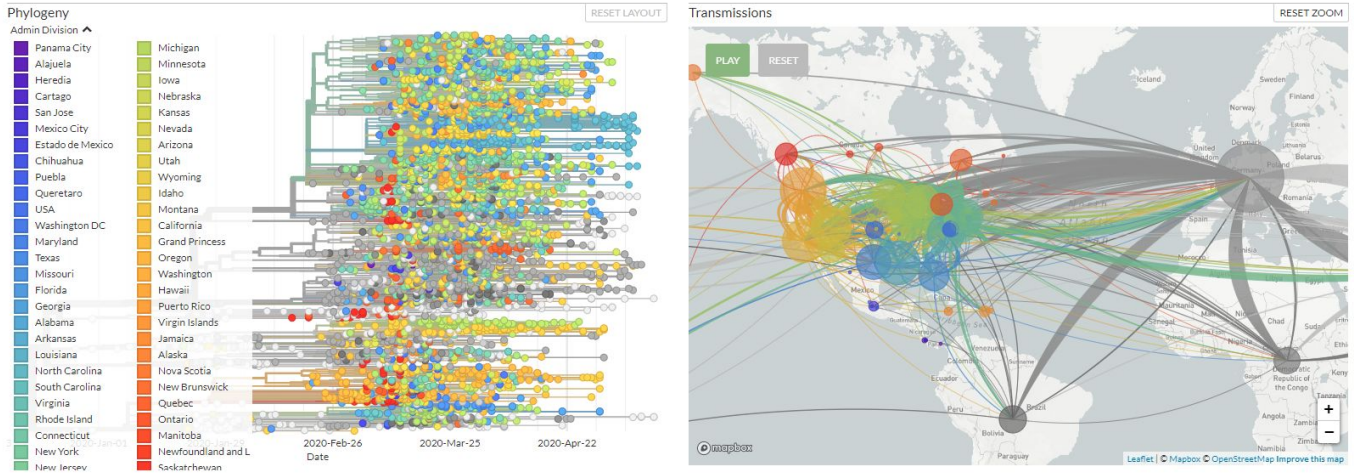
Pipeline incorporates tree building in augur for NextStrain.

What is NextStrain???

# But I can’t share my data with NextStrain

https://auspice-us.herokuapp.com/

Your data here



# It Takes a Village...

## StaPH-B Consortium

- http://www.staphb.org/

## CDC SPHERES (SARS-CoV-2 Sequencing for Public Health Emergency Response, Epidemiology, and Surveillance)

- https://www.cdc.gov/coronavirus/2019-ncov/cases-updates/spheres.html

# Resource List

Right now:

- https://github.com/CDCgov/SARS-CoV-2_Sequencing
- https://www.protocols.io/groups/staphb

Please go to (after Friday 5/29):

- http://www.staphb.org/resources/sars-cov-2/
- http://www.staphb.org/training/sars-cov-2/
- https://paperpile.com/shared/dU0ZnG

# Future Webinar Topics

Let me know if there are topics here that you would be interested in hearing about in more detail.

- ARTIC protocol
- Designing primers for other viral pathogens
- ONT sequencing
- StaPH-B toolkit
- Terra.bio
- Human read depletion
- Data sharing (GISAID, NCBI)
- NextStrain
- NextStrain local
- Tailed amplicon sequencing
- Enrichment protocol
- Metagenomics
- Others?