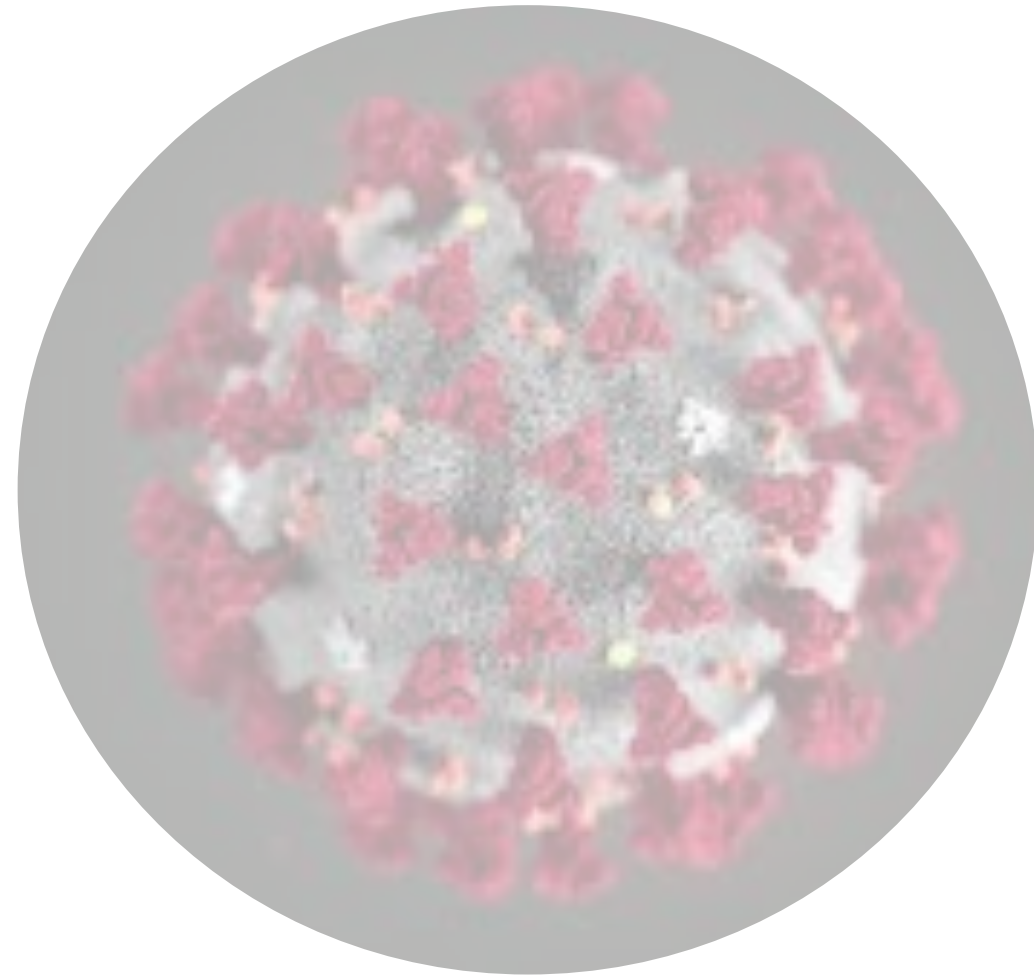# Understanding Modern Genomic Epidemiology Lecture 3

JOSEPH FAUVER, PH.D.

ASSISTANT PROFESSOR

UNMC COPH DEPARTMENT OF EPIDEMIOLOGY

5/7/2025

UNIVERSITY OF NEBRASKA MEDICAL CENTER℠
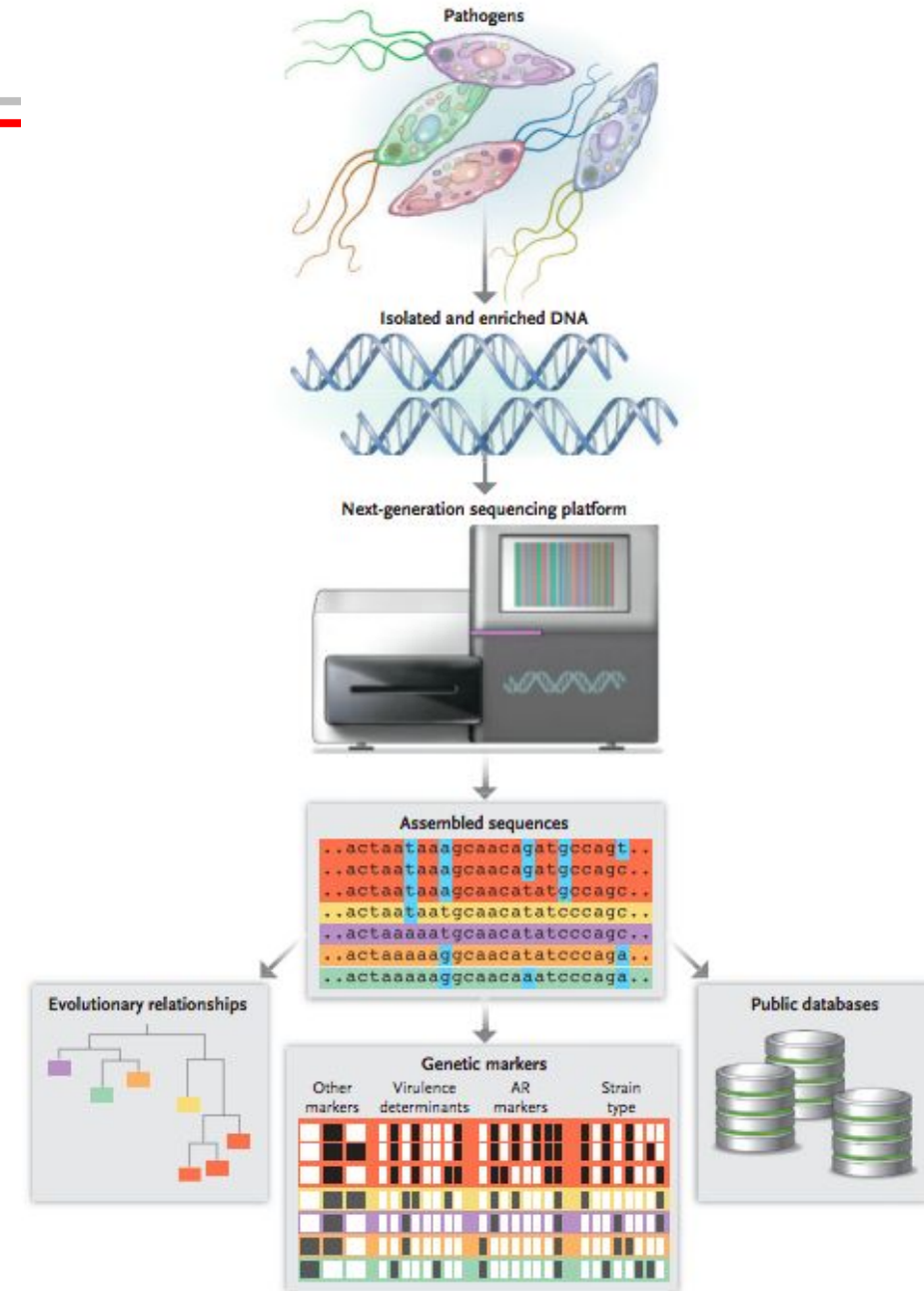**COLLEGE OF PUBLIC HEALTH**
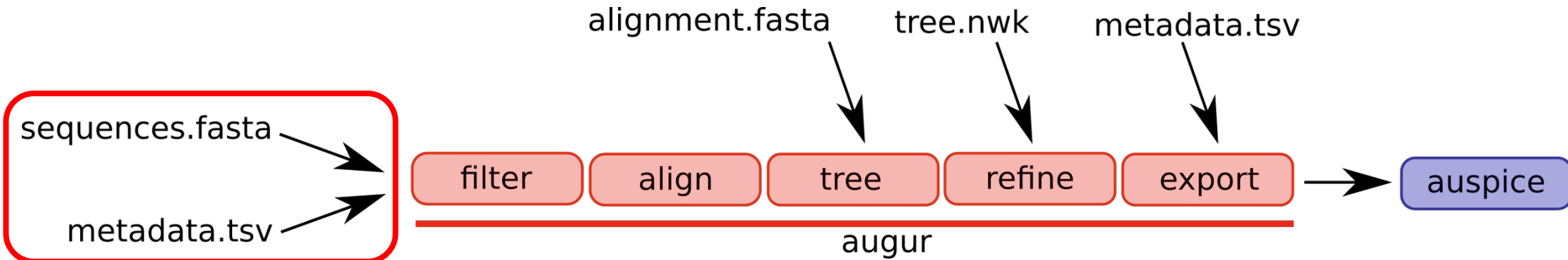
What does this look like?

"Wet" lab:

- Clinical samples/Microbial isolation
- RNA or DNA extraction
- Library preparation
- Next-Generation Sequencing

"Dry" lab:

- Data processing
- Primary genomic analyses
- Data interpretation
- Submission to repositories

Preparing to run a Nextstrain build

- **Next-Generation Sequencing Data -> Consensus Sequence Generation**
  - Data types (.fastq, .fasta, .bam)
  - Reference mapping
  - Variant calling
- **Identifying relevant data sources/repositories for contextual data**
  - Pathoplexus
  - NCBI GenBank (SRA, Virus, Microbe)
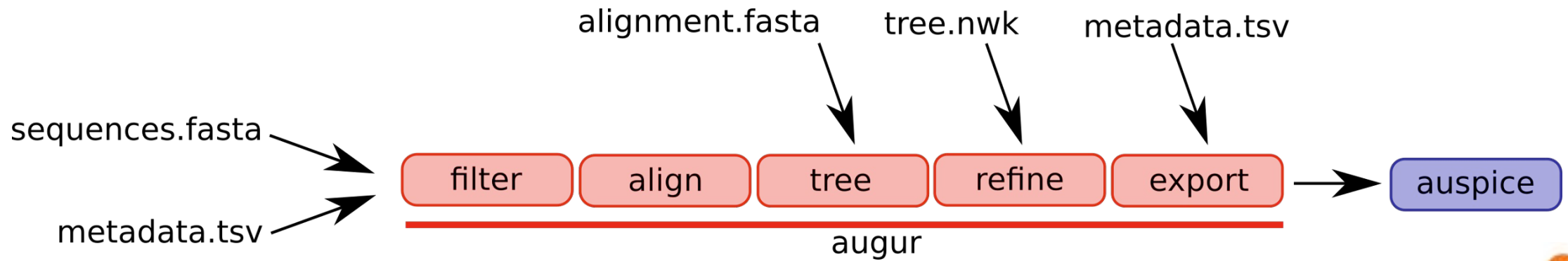 - GISAID
  - PlsmoDB
  - BacWGSTdb

UNIVERSITY OF NEBRASKA MEDICAL CENTER℠
COLLEGE OF PUBLIC HEALTH

Genomes + Metadata = Start of Nextstrain Build

What is next?

- NGS data, consensus sequence generation, data repositories
- **Description of Nextstrain (what it is, what it isn't, how it works, etc.)**
- Basics of phylogenetic tree interpretations

Preparing to run a Nextstrain build
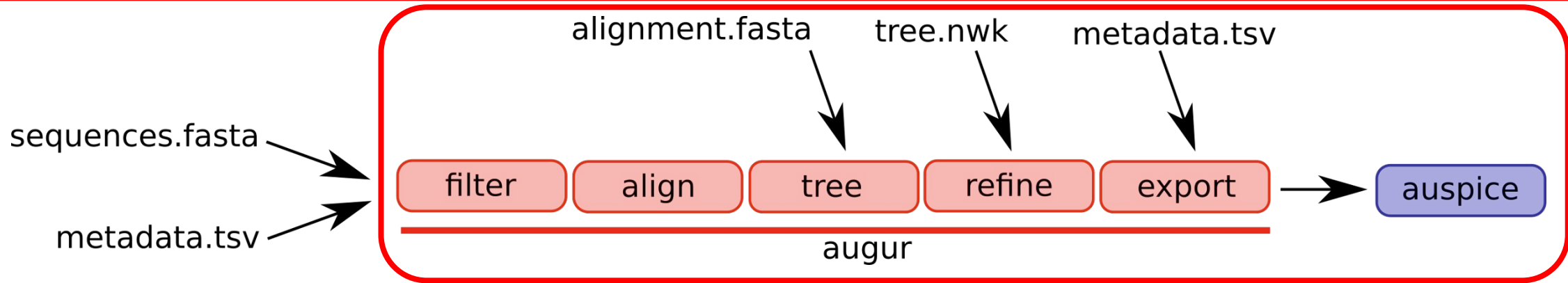

- **Overview of Nextstrain**
  - What is Nextstrain
  - How is it used

- **Augur**
  - Data filtration
  - Multi-sequence alignment from sequences.fasta file
  - Phylogenetic analysis
  - Refinement/annotation

- **Auspice**

Nextstrain

The core Nextstrain team is

Trevor Bedford, Richard Neher, Ivan Aksamentov, John SJ Anderson,

Kim Andrews, Jennifer Chang, James Hadfield, Emma Hodcroft,

John Huddleston, Jover Lee, Victor Lin, Cornelius Roemer,

Thomas Sibley

Please see the team page for more details.

Real-time tracking of pathogen evolution

- Software package that uses interaction phylogenetic analysis to inform pathogen surveillance for many important infectious diseases (SARS-CoV-2, Influenza, mPox, Lassa, etc.)

- Open source software- depends on community members to maintain and update builds that provide "real-time" snapshot of pathogen genomics

- Quick (once you know how to use it), adaptable, accurate, open-sourced, program that has become the standard for routine phylogenetic analysis for genomic epidemiolgy

Nextstrain- Landing Page

Nextstrain- Core Build

Nextstrain- Custom Build
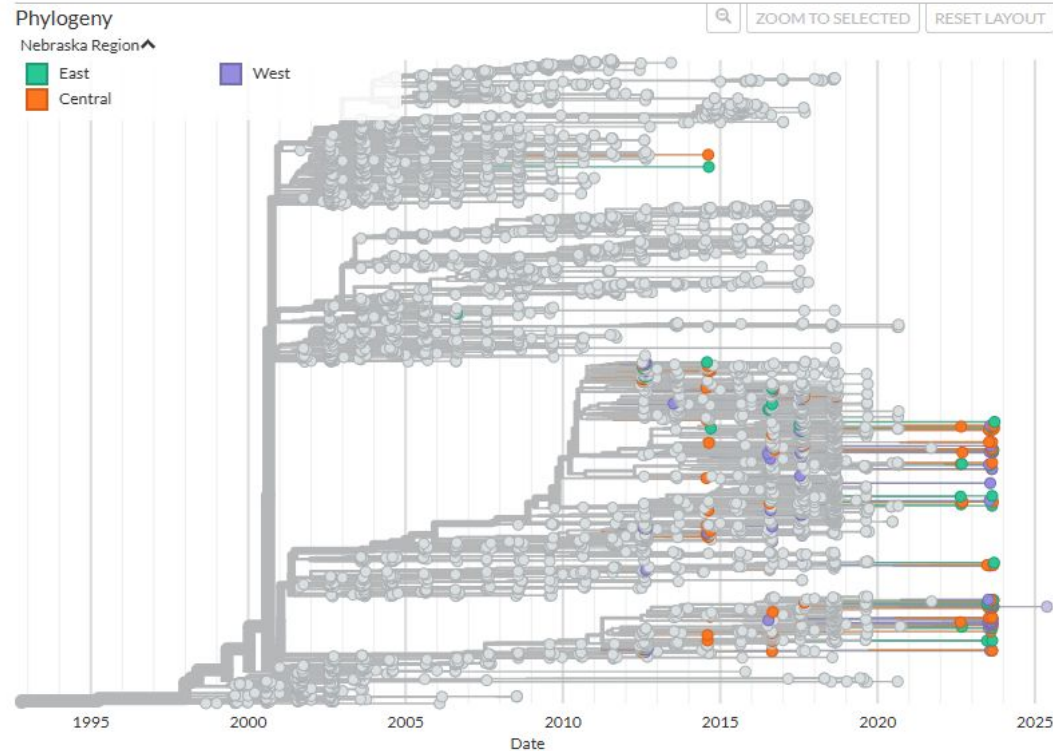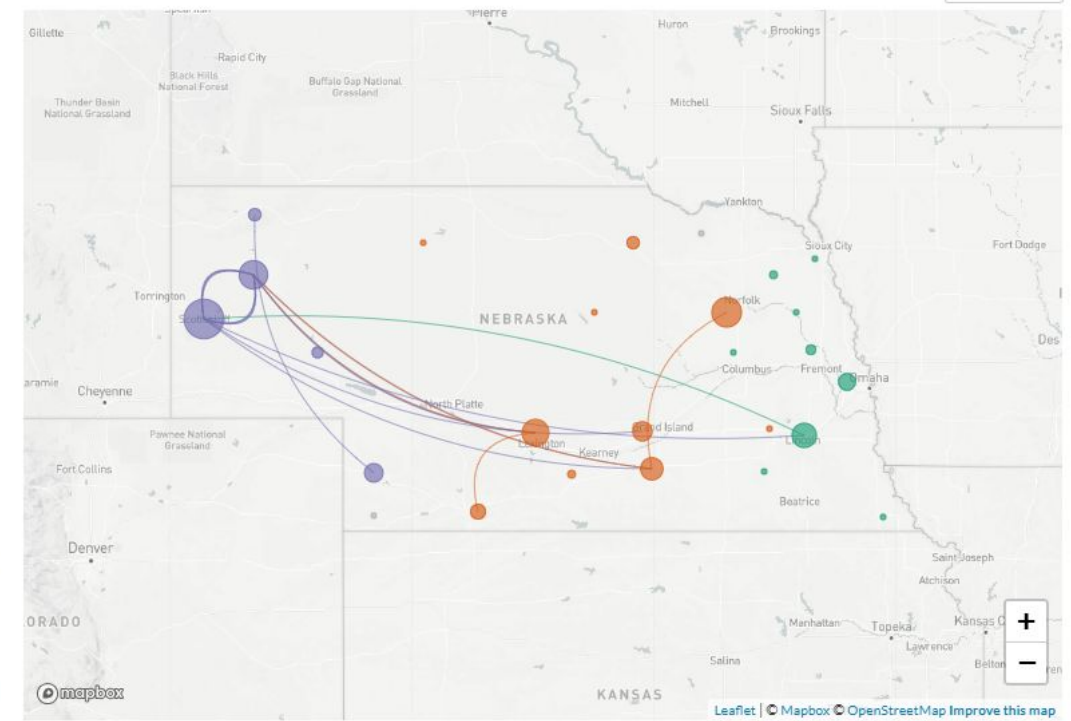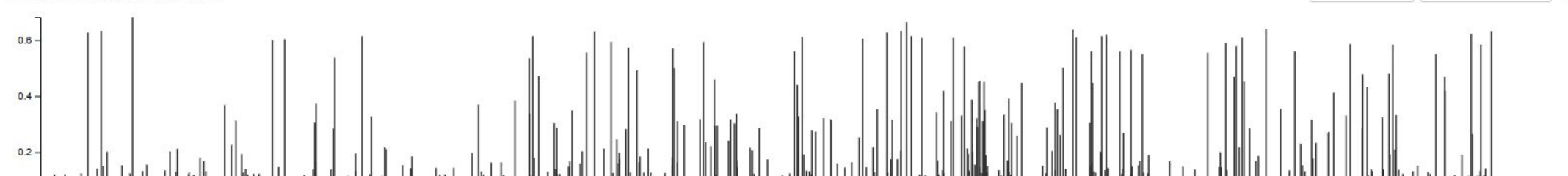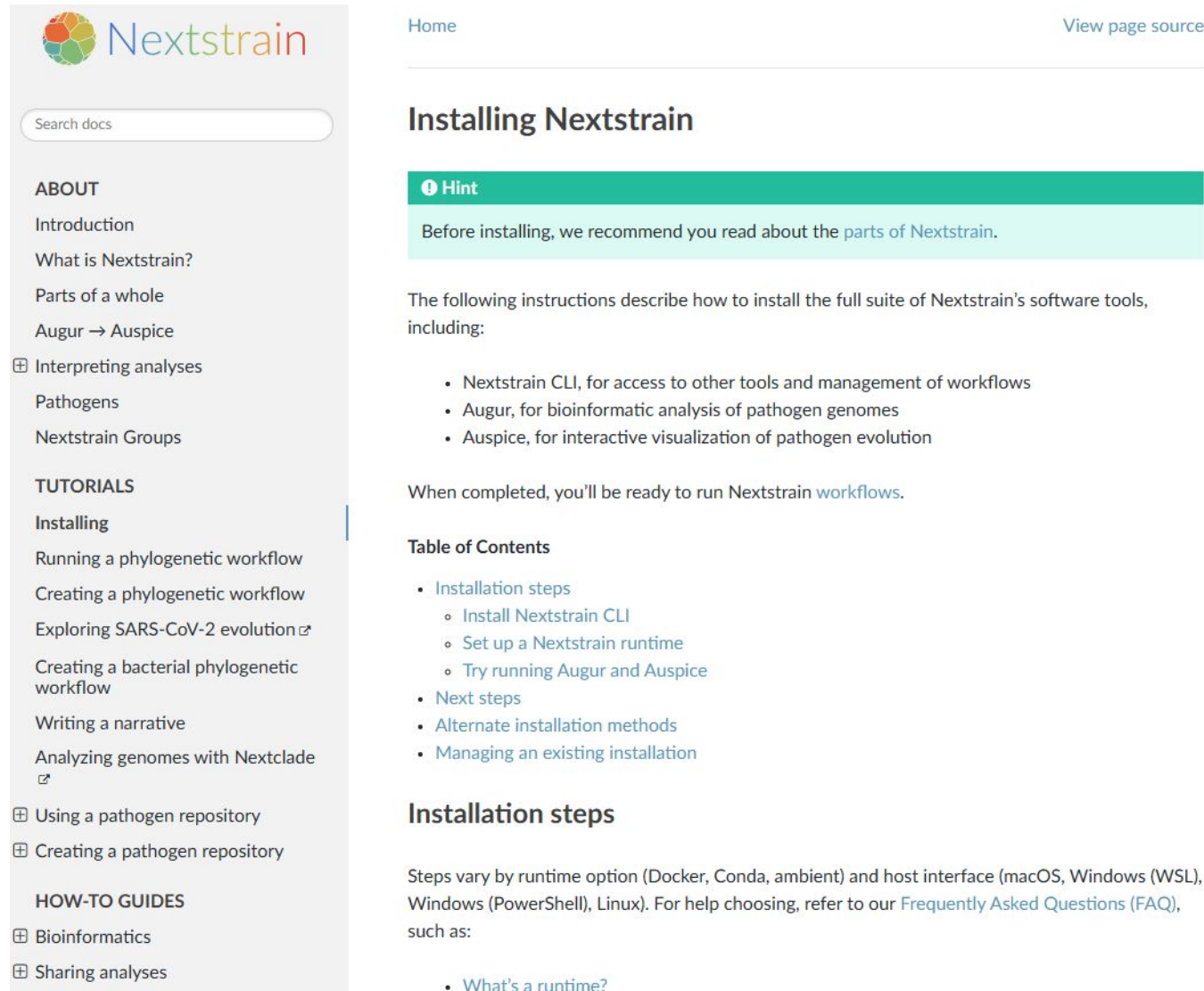
Nextstrain- An Abundance of Information

- Didactic tutorials with standard dataset to assist in learning the basic usage, language, and approach of Nextstrain

- ^ Dr. Shaun Cross will provide in-depth video on how to work through the Nextstrain tutorial

- Additional documentation for building custom Nextstrain builds (in-person training)

Augur- the engine behind Nextstrain
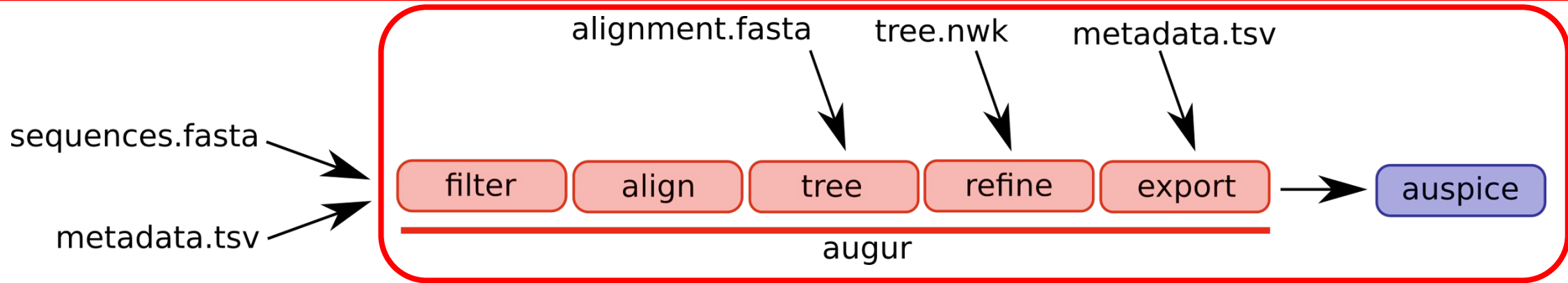
- Modular phylogenetic workflow that is run through Snakemake
- Inputs are sequences.fasta file (genomes) and metadata.tsv (metadata)
- Custom builds will also need to edit standard "config" files Augur will use:
  - Colors.tsv
  - Lat_longs.tsv
  - Reference_genome.gb
  - auspice_config.json
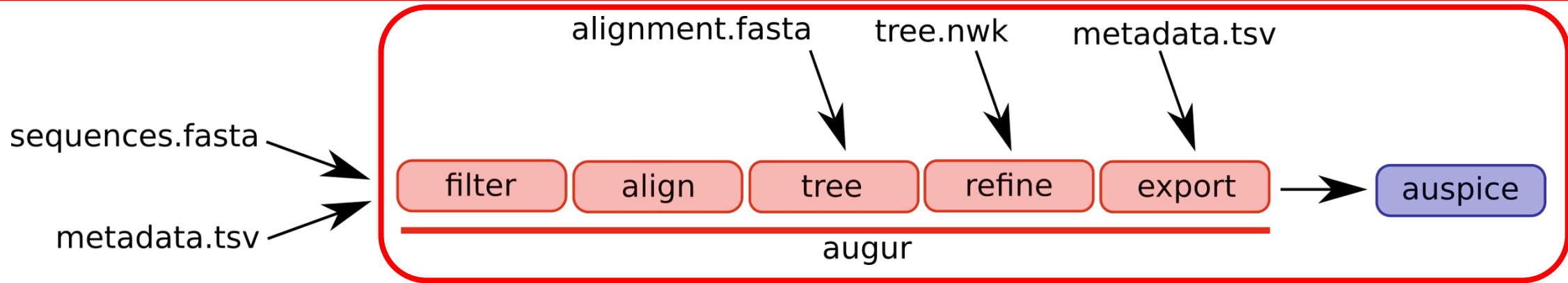
UNIVERSITY OF NEBRASKA MEDICAL CENTER℠
COLLEGE OF PUBLIC HEALTH

Augur- the engine behind Nextstrain

- What augur does
  - Prepare pathogen sequences and metadata
  - Align sequences
  - Construct a phylogeny from aligned sequences
  - Annotate the phylogeny with inferred ancestral pathogen dates, sequences, and traits
  - Export the annotated phylogeny and corresponding metadata into auspice-readable format (JSON)

Augur- the engine behind Nextstrain

- What augur does
  - **Prepare pathogen sequences and metadata**
  - Align sequences
  - Construct a phylogeny from aligned sequences
  - Annotate the phylogeny with inferred ancestral pathogen dates, sequences, and traits
  - Export the annotated phylogeny and corresponding metadata into auspice-readable format (JSON)

UNIVERSITY OF NEBRASKA MEDICAL CENTER℠
COLLEGE OF PUBLIC HEALTH

Augur- the engine behind Nextstrain
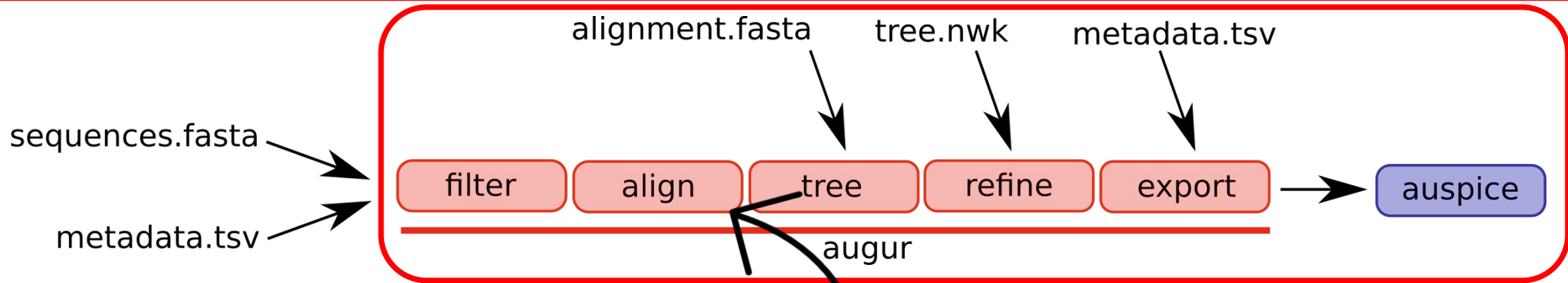
- What augur does
  - Prepare pathogen sequences and metadata
  - **Align sequences** (Nextstrain uses MAFFT)
  - Construct a phylogeny from aligned sequences
  - Annotate the phylogeny with inferred ancestral pathogen dates, sequences, and traits
  - Export the annotated phylogeny and corresponding metadata into auspice-readable format (JSON)

MSA- Multi-Sequence Alignment

- Phylogenetic methods are used to determine the evolutionary relationship of organisms based on variations at sequence level (polymorphisms)
  - Due to polymorphisms, sequences derived from common ancestors (homologs) have distinct levels (%) of similarity
  - Only homologous sequences can be compared, and their variations can be used to determine how closely/distantly related the organisms are to each other
  - Goal is to get homologous sites arranged in columns.

```
DENV1   AAAAGTCAGGTCAAACGCAGCTATTGGAGCAGTGTTCGTTGATGAAAATCA
DENV2   AAAGGTGAGAAGCAATGCAGCCTTGGGGGCCATATTCACTGATGAGAACAA
DENV3   AAAGGTCAGAACTAACGCAGCCATGGGCGCCGTTTTCACAGAGGAGAACCA
DENV4   AAAAGTTAGATCAAACGCAGCCATAGGCGCAGTCTTTCAGGAAGAACAGGG
ZIKV    CAAGGTGCGCAGCAATGCAGCACTGGGAGCAATATTTGAAGAGGAAAAAGA
WNV     AAAAGTCAACAGTAATGCCGCCCTAGGAGCGATGTTTGAAGAACAGAACCA
YFV     AAAAGTCCGAAGTCATGCAGCCATTGGAGCTTACCTGGAAGAACAAGAACA
POWV    GAAGGTGAGGTCCAACGCTGCTCTAGGTGCATGGTCGGATGAACAGAATAA
```

Alignments of homologous sequences are the input into building a phylogenetic tree!

```
*AA*GT*******A*GC*GC**T*GG*GC**********GA**A**A***
```
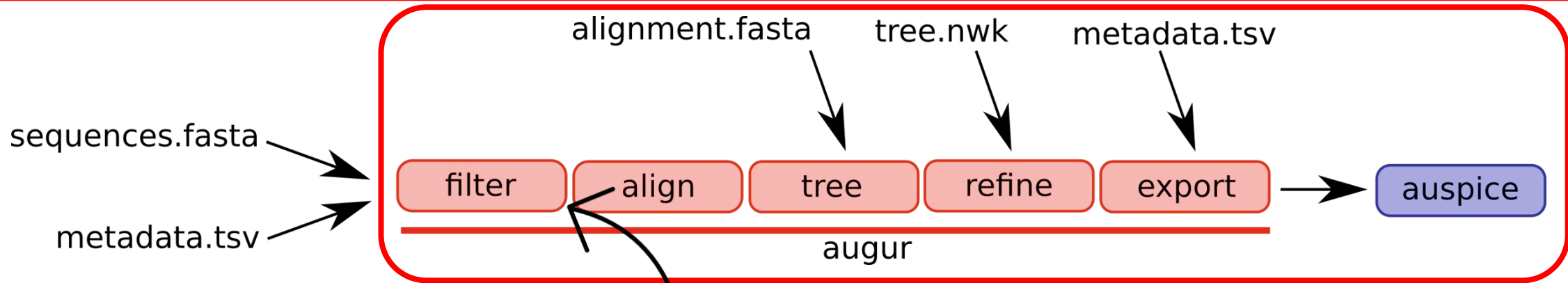
Augur- the engine behind Nextstrain
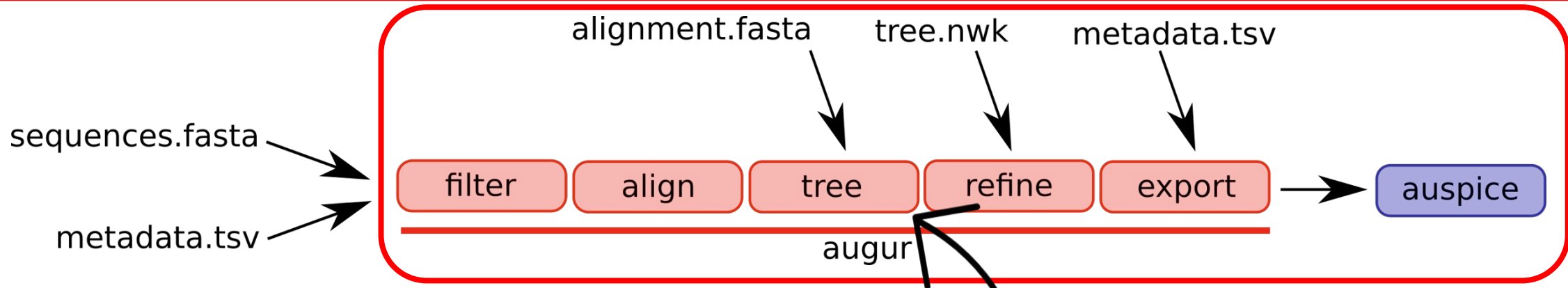
- What augur does
  - Prepare pathogen sequences and metadata
  - Align sequences
  - **Construct a phylogeny from aligned sequences**
  - Annotate the phylogeny with inferred ancestral pathogen dates, sequences, and traits
  - Export the annotated phylogeny and corresponding metadata into auspice-readable format (JSON)

Building Phylogenies in Nextstrain

- IQ-TREE
  - Maximum-likelihood (ML)***; Fast
  - ***There are A LOT of ways to make phylogenetic trees


```
>SWE-2016   NNAGCAGTA---GGTAGCAATAACNNN
>ESP-2014   GTAGCAGTA---GGTAGCAGTNNNNNN
>USA-2016   NNNNNAGTGAATGGTAGTAGTAATAAT
>CAN-2015   NNNGCAGTGAATAGTAGTGGTNNNNNN
```
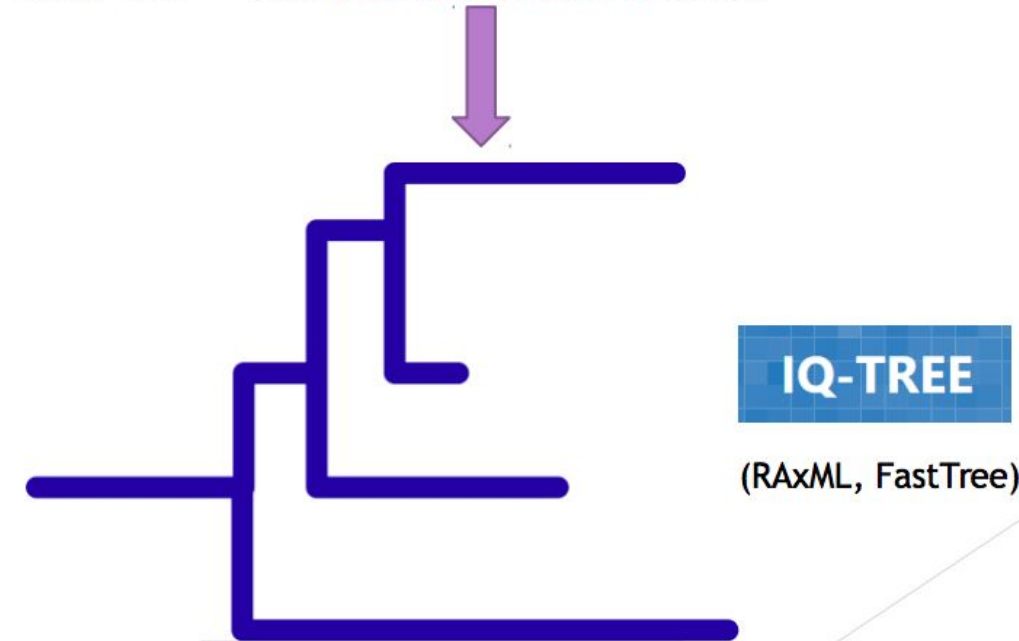
IQ-TREE

(RAxML, FastTree)

Building Phylogenies in Nextstrain

- IQ-TREE
  - Maximum-likelihood (ML); Fast
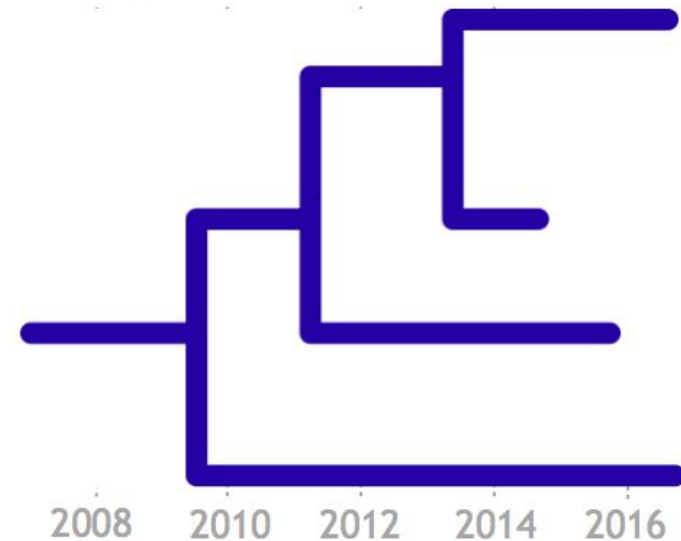- TreeTime
  - Also ML – much faster than Bayesian!
  - What it does:
    - estimate discrete ancestral states
    - infer evolution models
    - reroot trees to maximize temporal signals
    - estimate molecular clock phylogenies
    - estimate population size histories
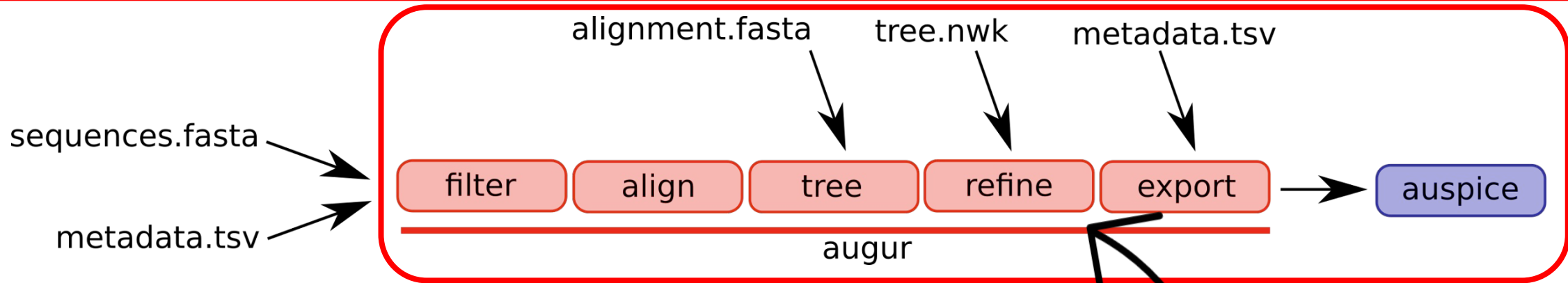
- *You can use TreeTime independently*
  - https://github.com/neherlab/treetime
  - https://treetime.biozentrum.unibas.ch/

Augur- the engine behind Nextstrain

- What augur does
  - Prepare pathogen sequences and metadata
  - Align sequences
  - Construct a phylogeny from aligned sequences
  - **Annotate the phylogeny with inferred ancestral pathogen dates, sequences, and traits**
  - Export the annotated phylogeny and corresponding metadata into auspice-readable format (JSON)

UNIVERSITY OF NEBRASKA MEDICAL CENTER℠
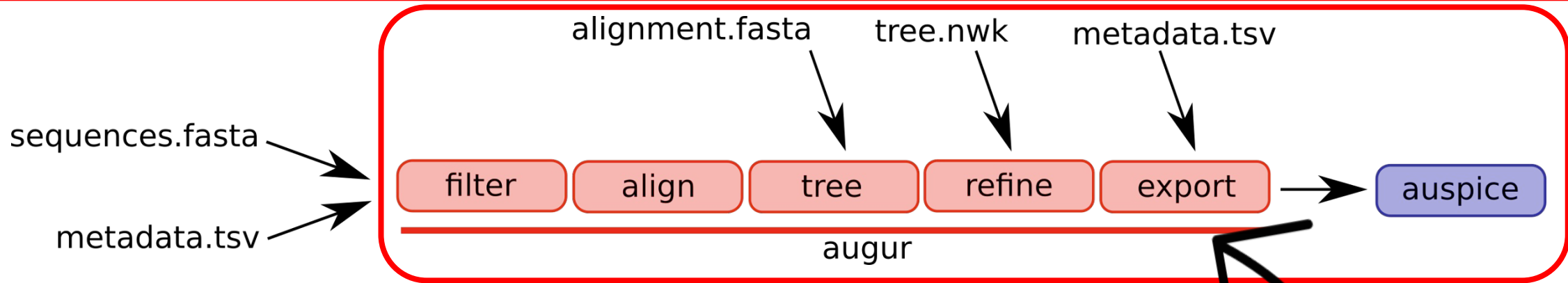**COLLEGE OF PUBLIC HEALTH**
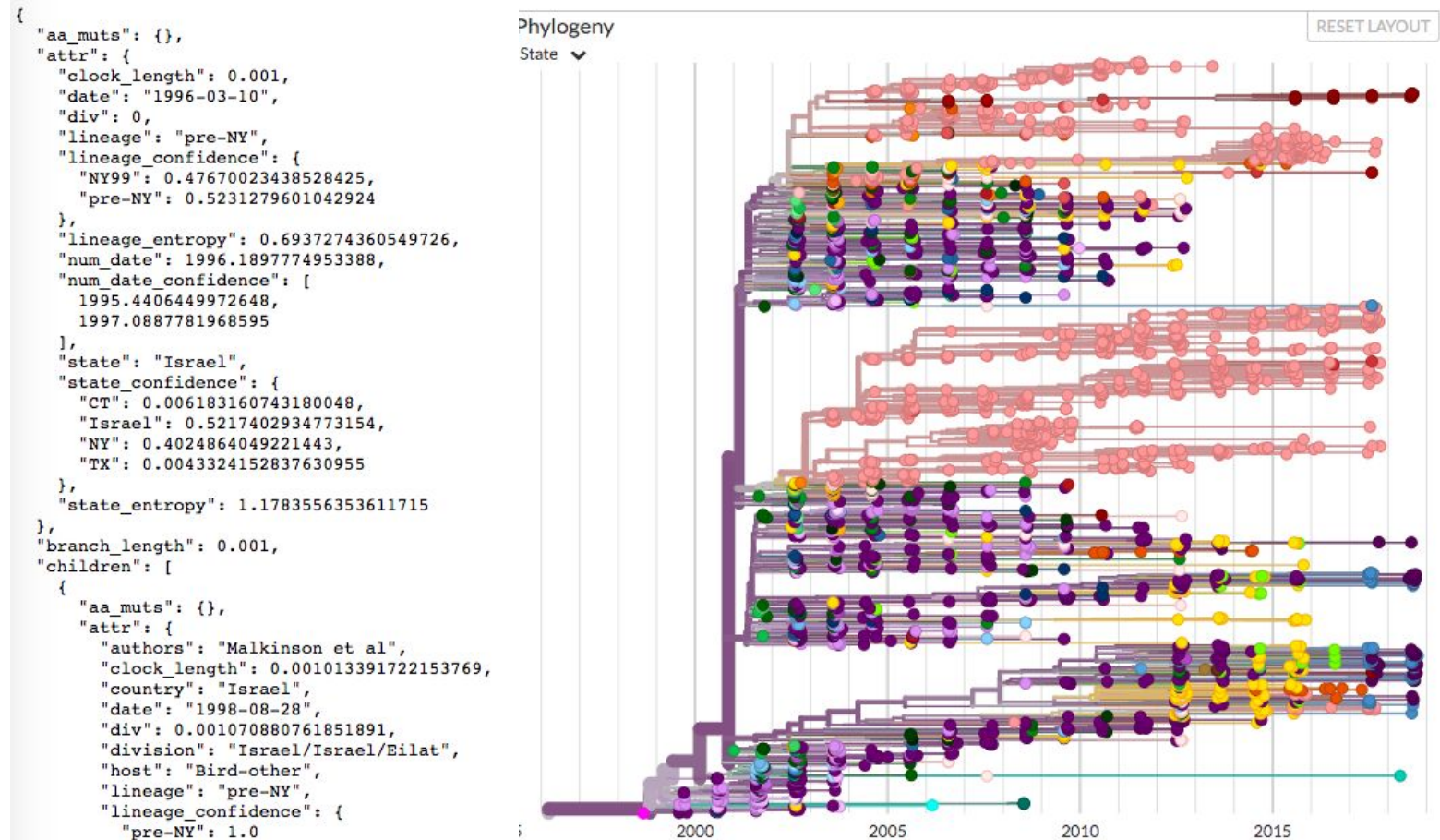
Augur- the engine behind Nextstrain

- What augur does
  - Prepare pathogen sequences and metadata
  - Align sequences
  - Construct a phylogeny from aligned sequences
  - Annotate the phylogeny with inferred ancestral pathogen dates, sequences, and traits
  - **Export the annotated phylogeny and corresponding metadata into auspice-readable format (JSON)**

Auspice- interactive phylogenetic visualization

- Inputs .json files exported from augor

Nextstrain- Useful Links

Glossary:
https://docs.nextstrain.org/en/latest/reference/glossary.html#term-Nextstrain-CLI

Website: https://nextstrain.org/

GitHub: https://github.com/nextstrain

Auspice: https://auspice.us/

Detailed Augur explanation:
https://docs.nextstrain.org/en/latest/learn/augur-to-auspice.html

What is next?

- NGS data, consensus sequence generation, data repositories

- Description of Nextstrain (what it is, what it isn't, how it works, etc.)

- **Basics of phylogenetic tree interpretations**
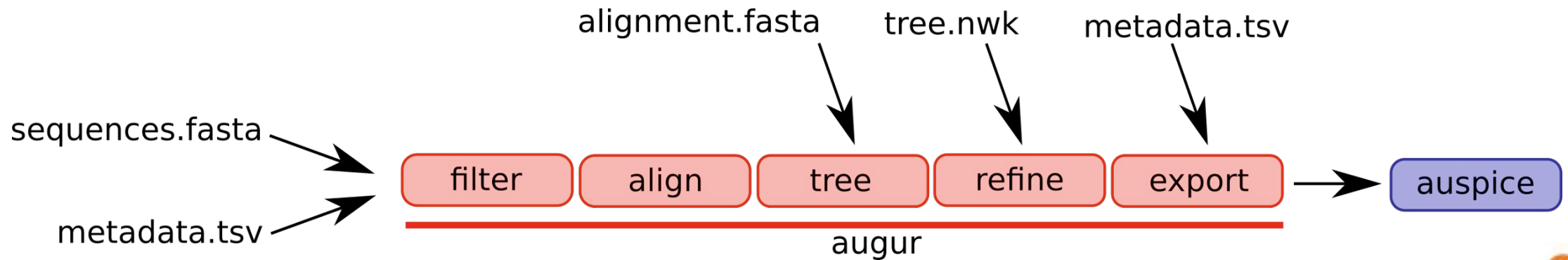
Questions?

Reach out via email: jfauver@unmc.edu