

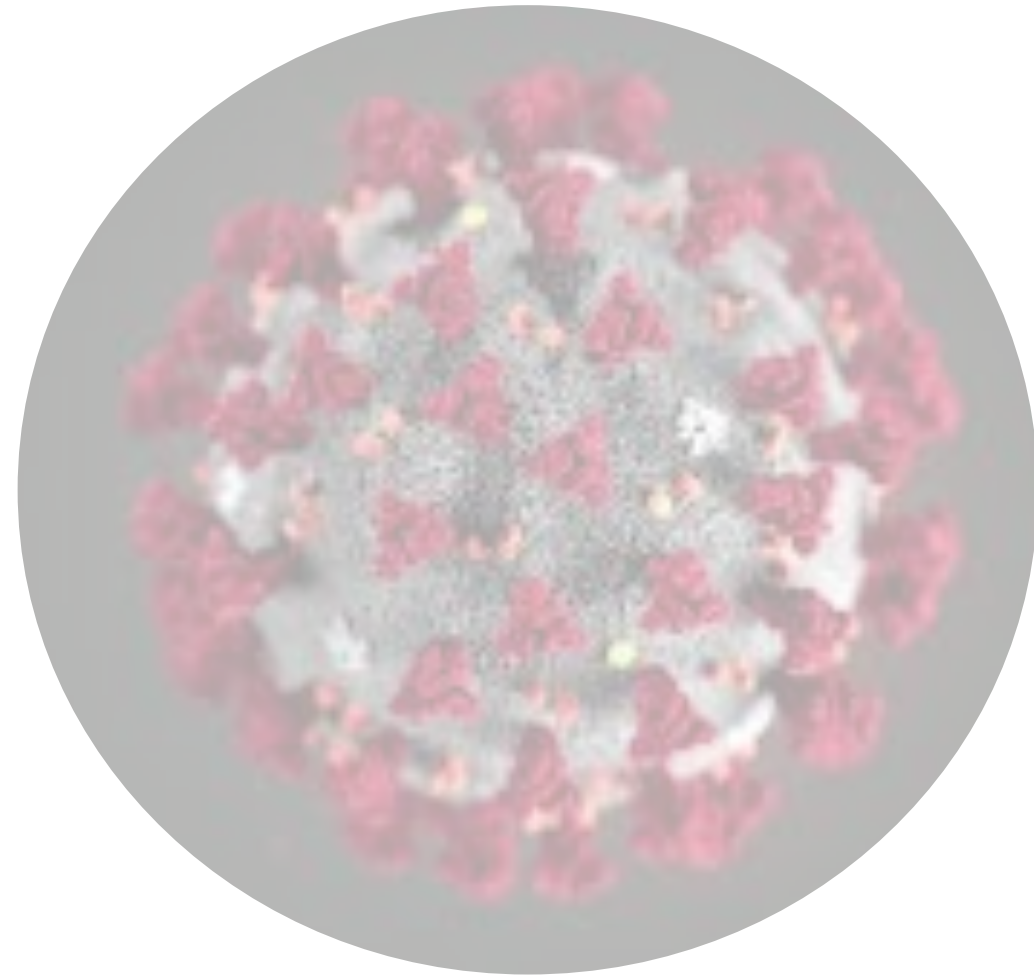
Understanding Modern Genomic Epidemiology Lecture 3

JOSEPH FAUVER, PH.D.

ASSISTANT PROFESSOR

UNMC CPH DEPARTMENT OF EPIDEMIOLOGY

5/7/2025



UNIVERSITY OF NEBRASKA MEDICAL CENTER™

COLLEGE OF PUBLIC HEALTH

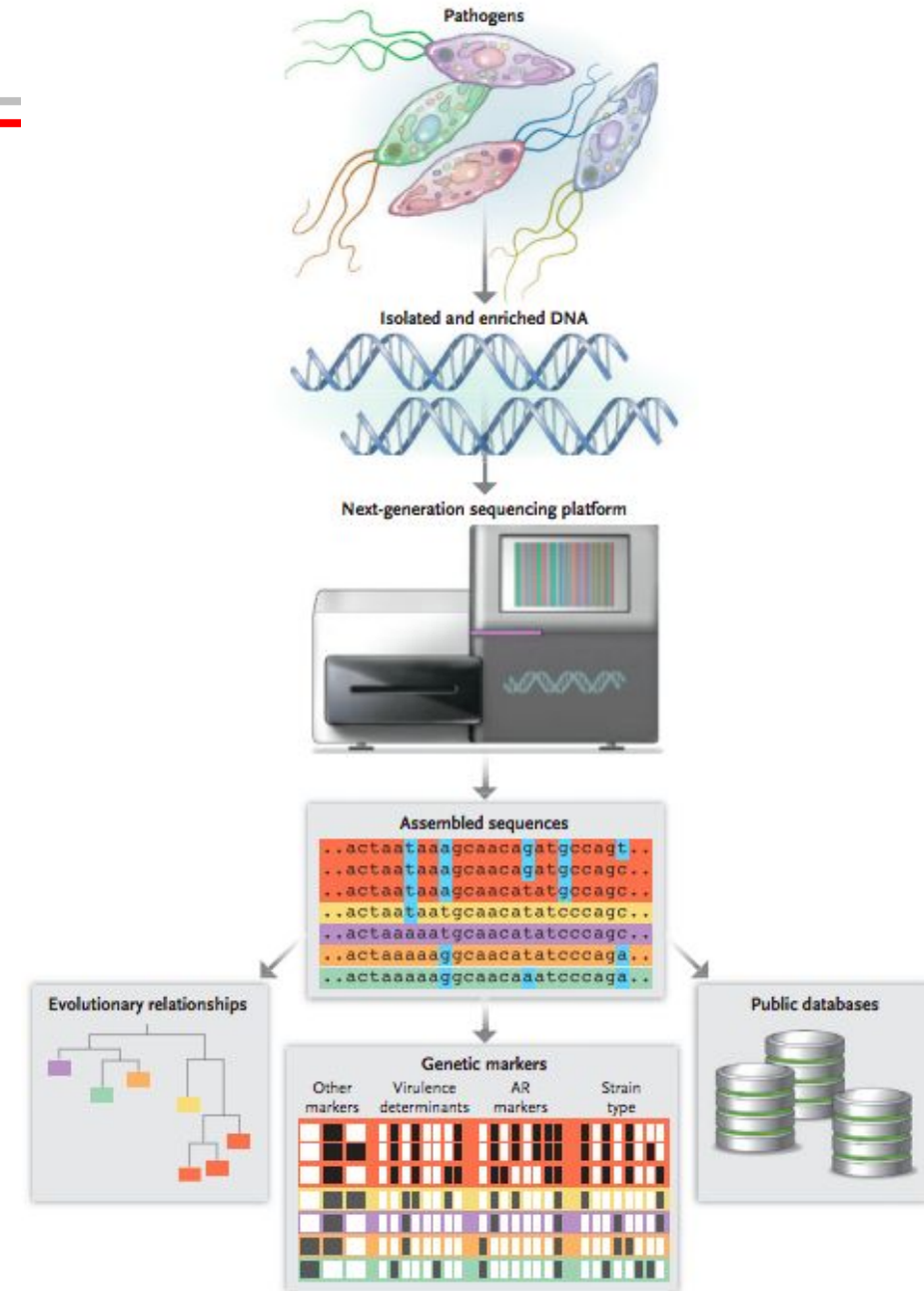
What does this look like?

“Wet” lab:

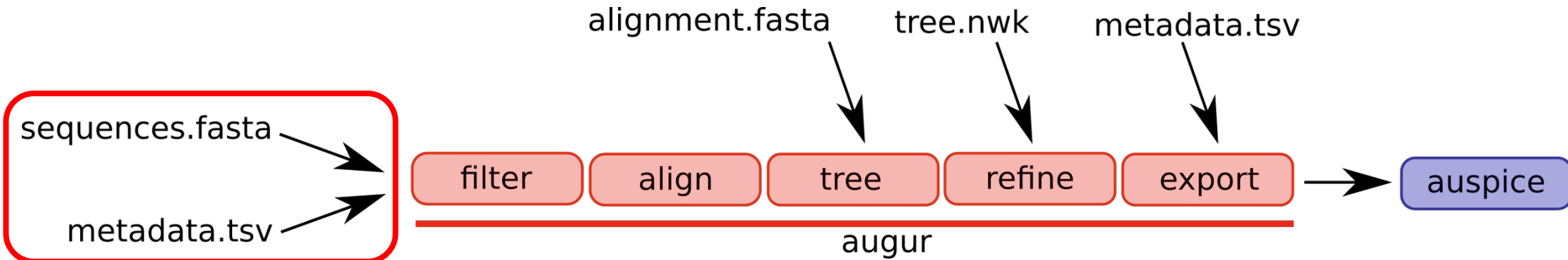
- Clinical samples/Microbial isolation
- RNA or DNA extraction
- Library preparation
- Next-Generation Sequencing

“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



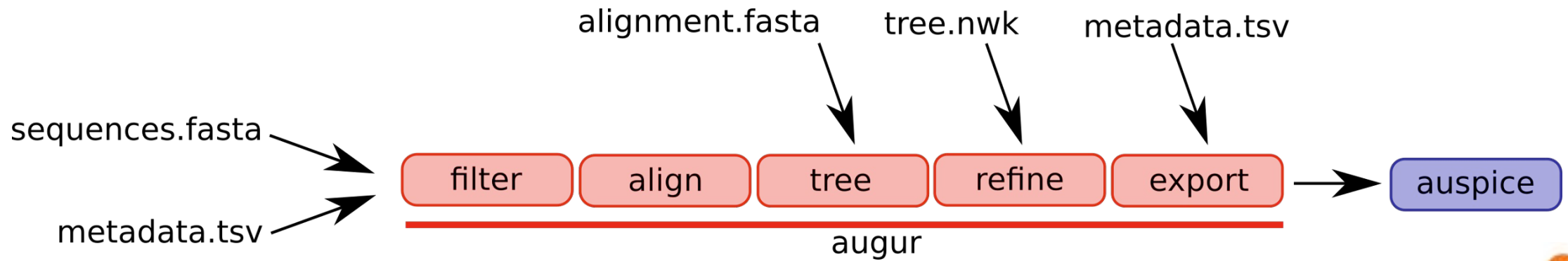
Genomes + Metadata = Start of Nextstrain Build

	A	B	C	D	E	F	G	H	K	R	S
1	strain	virus	accession	date	region	country	state	division	segment	latitude	longitude
2	AF196835	wnv	AF196835	1999-XX-X	North Ame	USA	NY	NY/Bronx	genome	42.1657	-74.9481
3	AF206518	wnv	AF206518	1999-XX-X	North Ame	USA	CT	CT/Fairfiel	genome	41.5978	-72.7554
4	DQ211652	wnv	DQ211652	1999-XX-X	North Ame	USA			genome		
5	EF530047	wnv	EF530047	2000-XX-X	North Ame	USA			genome		
6	EF571854	wnv	EF571854	1999-XX-X	North Ame	USA			genome		
7	EF657887	wnv	EF657887	2000-XX-X	North Ame	USA			genome		
8	FJ527738	wnv	FJ527738	2001-XX-X	North Ame	USA	LA	LA	genome	31.1695	-91.8678
9	GQ379156	wnv	GQ379156	2001-07-X	North Ame	USA			genome		
10	GQ379157	wnv	GQ379157	2008-08-X	North Ame	USA	CA	CA	genome	36.1162	-119.682
11	GQ379158	wnv	GQ379158	2007-08-X	North Ame	USA	CA	CA	genome	36.1162	-119.682
12	GQ379159	wnv	GQ379159	2008-08-X	North Ame	USA	CA	CA	genome	36.1162	-119.682
13	GQ379161	wnv	GQ379161	2006-02-X	North Ame	USA			genome		
14	HM538578	wnv	HM538578	2001-XX-X	North Ame	USA	CT	CT	genome	41.5978	-72.7554
15	HM538579	wnv	HM538579	2003-XX-X	North Ame	USA	CT	CT	genome	41.5978	-72.7554
16	HM538580	wnv	HM538580	2005-XX-X	North Ame	USA	CT	CT	genome	41.5978	-72.7554
17	HM538581	wnv	HM538581	2007-XX-X	North Ame	USA	CT	CT	genome	41.5978	-72.7554
18	HM538582	wnv	HM538582	2004-XX-X	North Ame	USA	IL	IL	genome	40.3495	-88.9861
19	HM538583	wnv	HM538583	2004-XX-X	North Ame	USA	IL	IL	genome	40.3495	-88.9861
20	HM756655	wnv	HM756655	2003-XX-X	North Ame	USA	CT	CT	genome	41.5978	-72.7554
21	HM756674	wnv	HM756674	2004-XX-X	North Ame	USA	NY	NY	genome	42.1657	-74.9481
22	HQ596519	wnv	HQ596519	1999-XX-X	North Ame	USA			genome		
23	HQ671736	wnv	HQ671736	2001-XX-X	North Ame	USA	NY	NY	genome	42.1657	-74.9481
24	HQ671737	wnv	HQ671737	2002-XX-X	North Ame	USA	NY	NY	genome	42.1657	-74.9481
25	HQ671738	wnv	HQ671738	2004-XX-X	North Ame	USA	NY	NY	genome	42.1657	-74.9481
26	HQ671739	wnv	HQ671739	2004-XX-X	North Ame	USA	NY	NY	genome	42.1657	-74.9481
27	HQ671740	wnv	HQ671740	2005-XX-X	North Ame	USA	NY	NY	genome	42.1657	-74.9481
28	HQ671741	wnv	HQ671741	2006-XX-X	North Ame	USA	NY	NY	genome	42.1657	-74.9481
29	HQ671743	wnv	HQ671743	2003-XX-X	North Ame	USA	IL	IL	genome	40.3495	-88.9861
30	HQ671744	wnv	HQ671744	2005-XX-X	North Ame	USA	IL	IL	genome	40.3495	-88.9861
31	HQ671745	wnv	HQ671745	2006-XX-X	North Ame	USA	IL	IL	genome	40.3495	-88.9861



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



Nextstrain

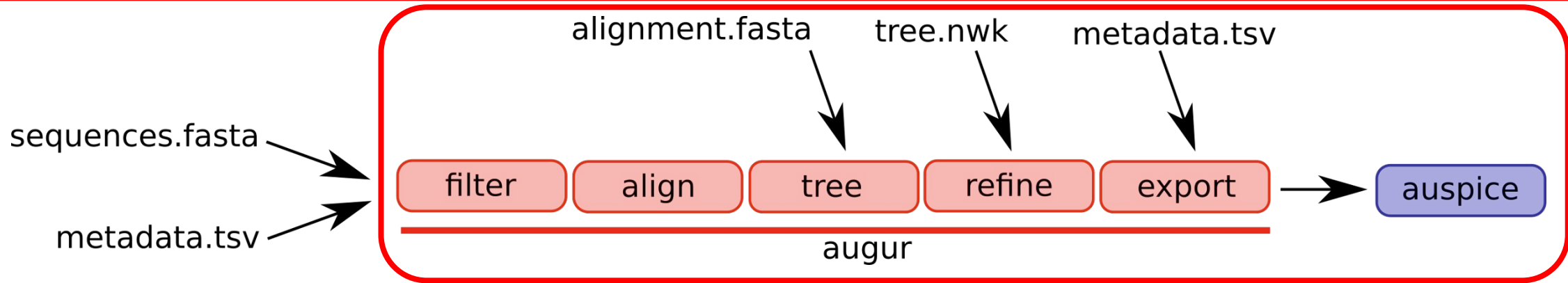
Real-time tracking of pathogen evolution

geneious
prime



UNIVERSITY OF NEBRASKA MEDICAL CENTER™
COLLEGE OF PUBLIC HEALTH

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



Nextstrain

Real-time tracking of pathogen evolution

Hadfield et al., [Nextstrain: real-time tracking of pathogen evolution](#), *Bioinformatics* (2018)

The core Nextstrain team is



Please see the [team page](#) for more details.

- 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 epidemiology

Nextstrain- Landing Page



About us

An open-source project to harness the scientific and public health potential of pathogen genome data

Core pathogens

Continually updated views of a range of pathogens maintained by the Nextstrain team

SARS-CoV-2

Up-to-date analyses and a range of resources for SARS-CoV-2, the virus responsible for COVID-19 disease

Open source tooling

Bioinformatic workflows, analysis tools and visualization apps for use by the community

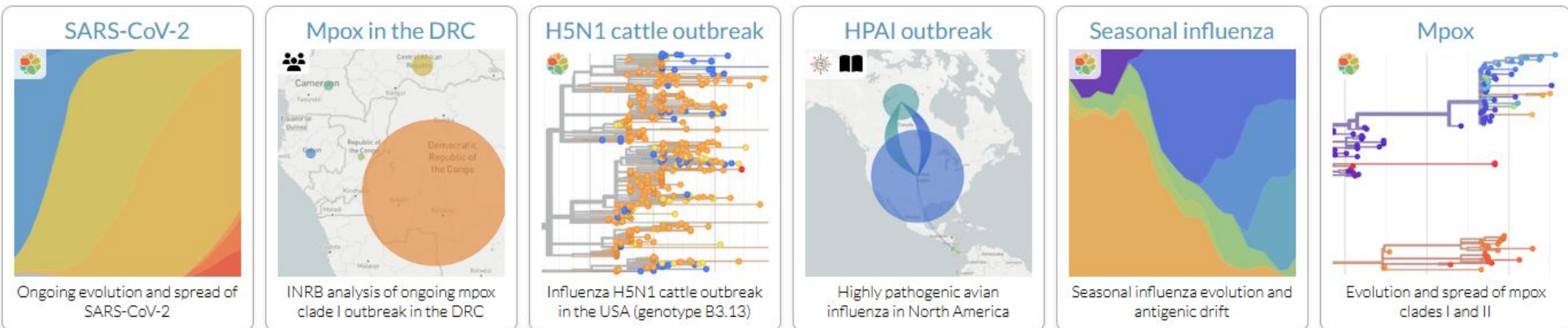
Nextclade

In-browser phylogenetic placement, clade assignment, mutation calling and sequence quality checks

Nextstrain Groups

Datasets and narratives shared by research labs, public health entities and others

Featured analyses



Nextstrain- Core Build

DOCS CONTACT LOGIN

Dataset

avian-flu

h5n1

ha

2y

Date Range

2006-03-28 2025-02-02

PLAY RESET

Color By

Region

Filter Data

Type filter query here...

Tree

Layout

RECTANGULAR

RADIAL

UNROOTED

CLOCK

SCATTER

Branch Length

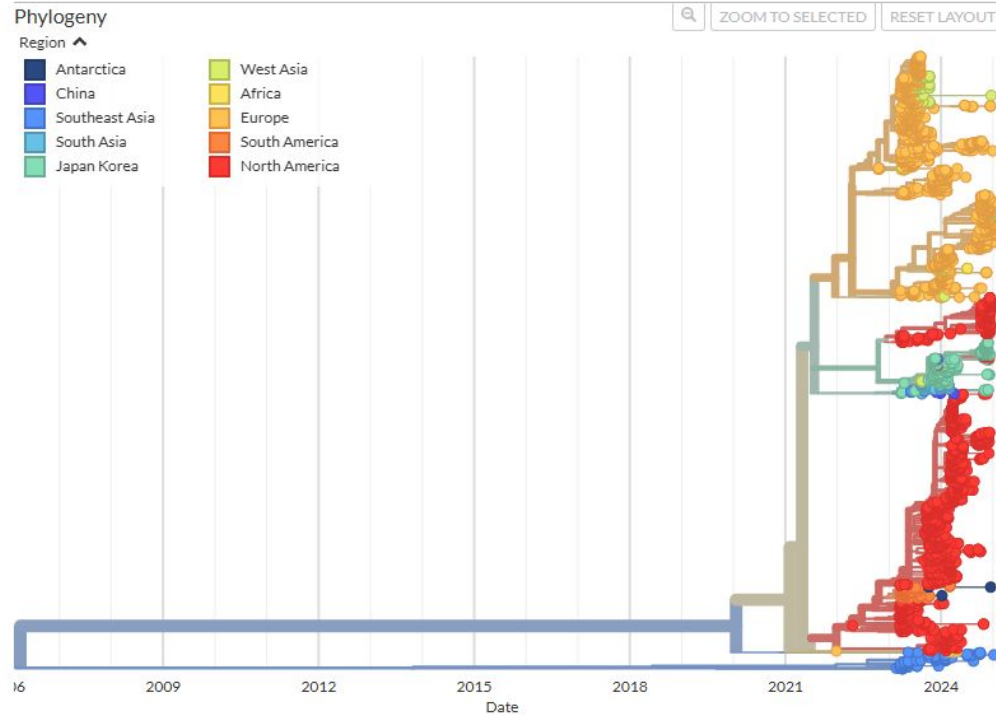
TIME DIVERGENCE

Show confidence intervals

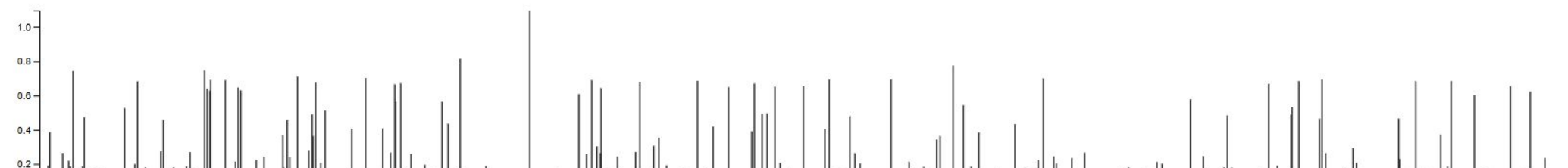
Real-time tracking of influenza A/H5N1 virus evolution

Built with nextstrain/avian-flu. Maintained by Louise Moncla and the Nextstrain team. Data updated 2025-03-20. Enabled by data from USDA and GISAID.

Showing 3318 of 3318 genomes sampled between Dec 2021 and Jan 2025.



Nucleotide diversity of genome



RESET LAYOUT ENTROPY EVENTS

Nextstrain- Custom Build

auspice.us

Date Range ⓘ

1992-10-16 2025-05-25

PLAY RESET

Color By ⓘ

Nebraska Region

Filter Data ⓘ

Type filter query here...

Tree ⓘ

Layout

RECTANGULAR

RADIAL

UNROOTED

CLOCK

SCATTER

Branch Length

TIME DIVERGENCE

Show confidence intervals

Focus on Selected ⓘ

Branch Labels

none

Tip Labels

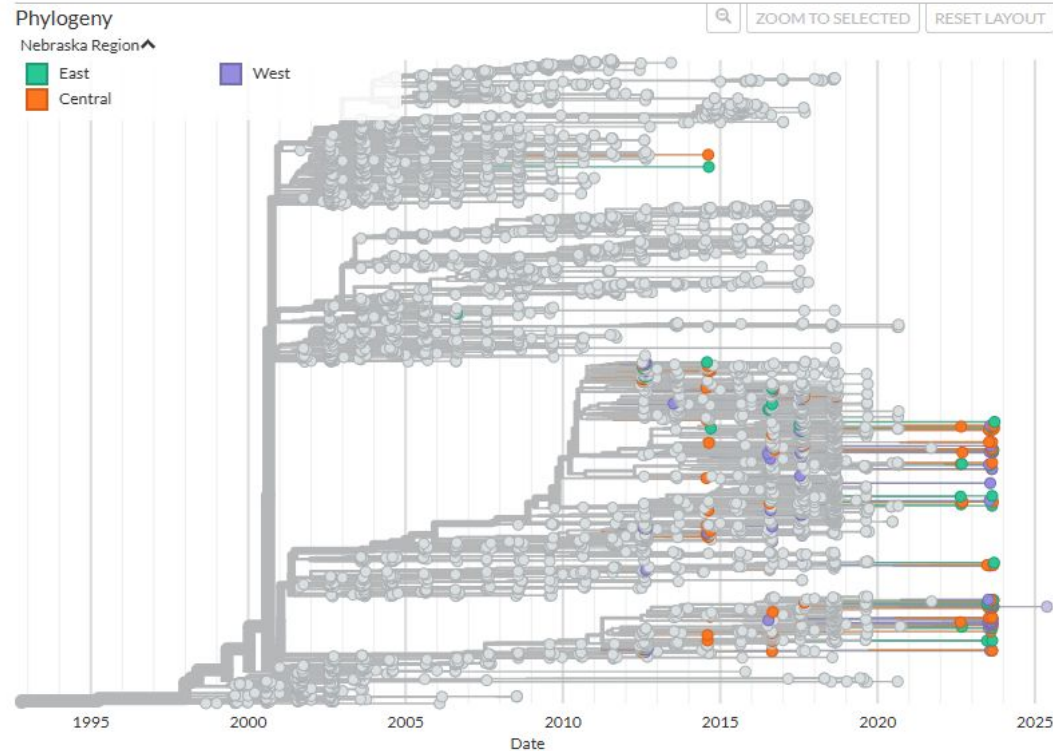
Sample Name

Explode Tree By ⓘ

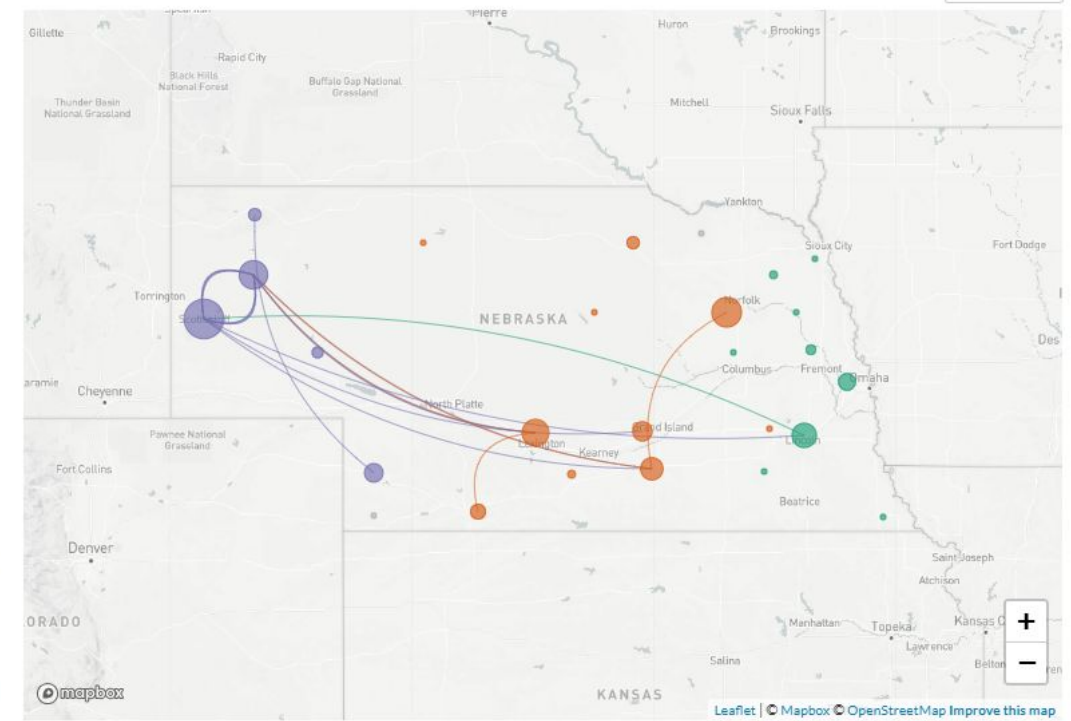
Nextstrain WNV

Built with [nextstrain/zika-tutorial](#). Maintained by [Joseph Fauver](#). Data updated 2025-03-12.

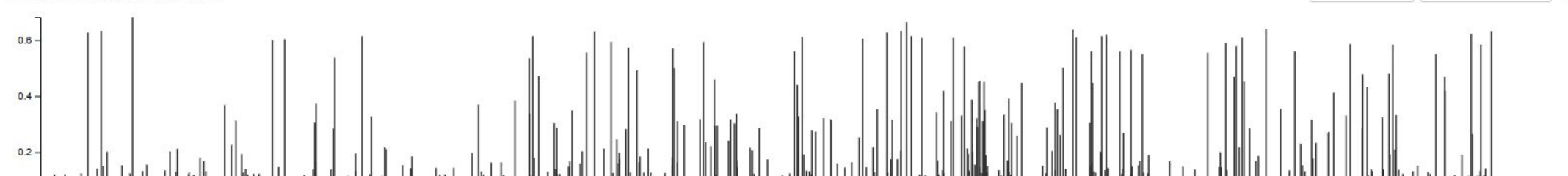
Showing 3938 of 3938 genomes sampled between Aug 1998 and May 2025.



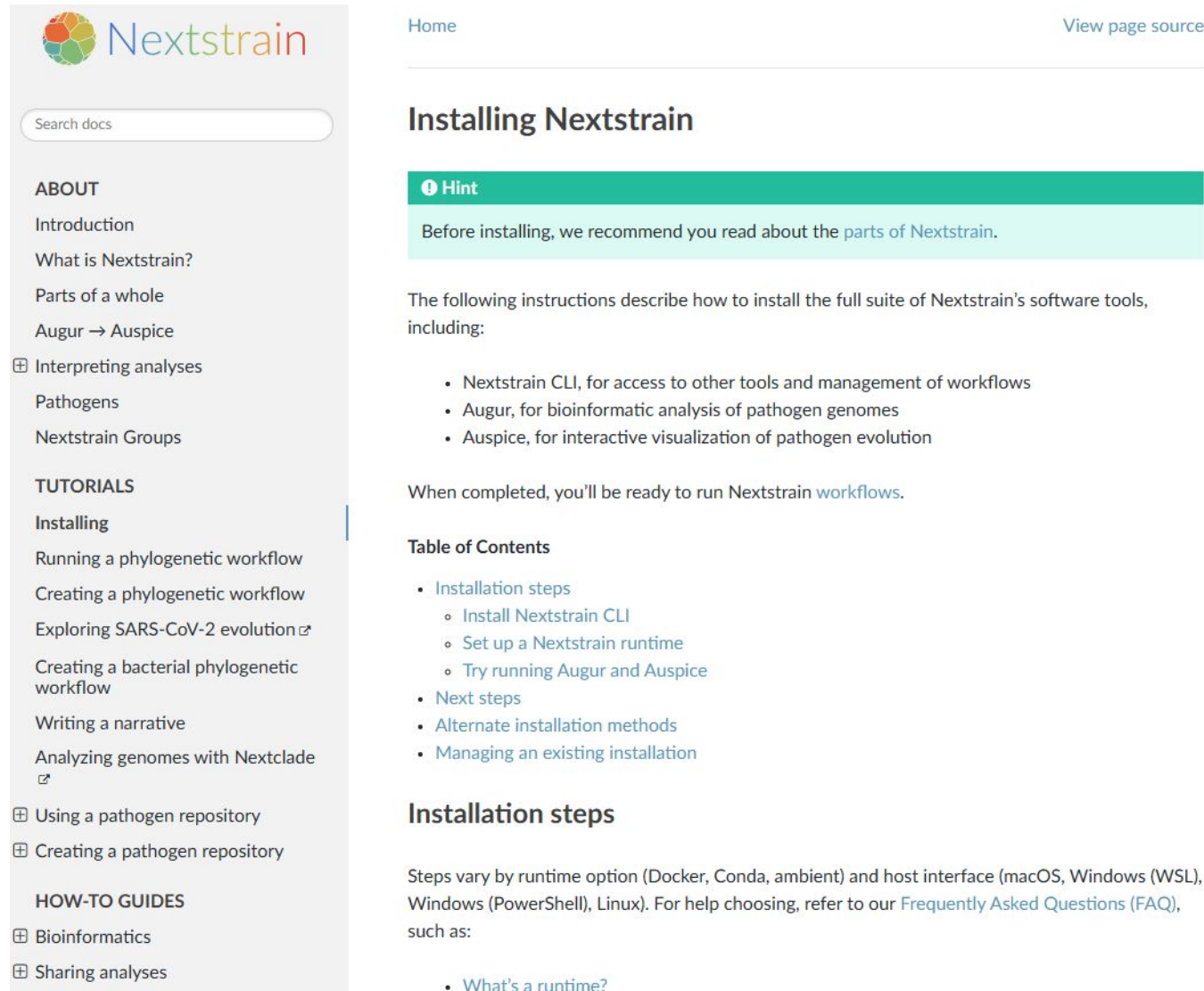
Transmissions



Nucleotide diversity of genome



Nextstrain- An Abundance of Information



The screenshot shows the Nextstrain website's 'Installing Nextstrain' page. The left sidebar contains navigation links under 'ABOUT', 'TUTORIALS', and 'HOW-TO GUIDES'. The main content area has a 'Hint' box, a paragraph about installation instructions, a 'Table of Contents' list, and 'Installation steps'.

Nextstrain

Search docs

ABOUT

- Introduction
- What is Nextstrain?
- Parts of a whole
- Augur → Auspice

Interpreting analyses

- Pathogens
- Nextstrain Groups

TUTORIALS

- Installing
- Running a phylogenetic workflow
- Creating a phylogenetic workflow
- Exploring SARS-CoV-2 evolution
- Creating a bacterial phylogenetic workflow
- Writing a narrative
- Analyzing genomes with Nextclade

Using a pathogen repository

- Creating a pathogen repository

HOW-TO GUIDES

- Bioinformatics
- Sharing analyses

Home [View page source](#)

Installing Nextstrain

Hint

Before installing, we recommend you read about the [parts of Nextstrain](#).

The following instructions describe how to install the full suite of Nextstrain's software tools, including:

- Nextstrain CLI, for access to other tools and management of workflows
- Augur, for bioinformatic analysis of pathogen genomes
- Auspice, for interactive visualization of pathogen evolution

When completed, you'll be ready to run Nextstrain [workflows](#).

Table of Contents

- [Installation steps](#)
 - [Install Nextstrain CLI](#)
 - [Set up a Nextstrain runtime](#)
 - [Try running Augur and Auspice](#)
- [Next steps](#)
- [Alternate installation methods](#)
- [Managing an existing installation](#)

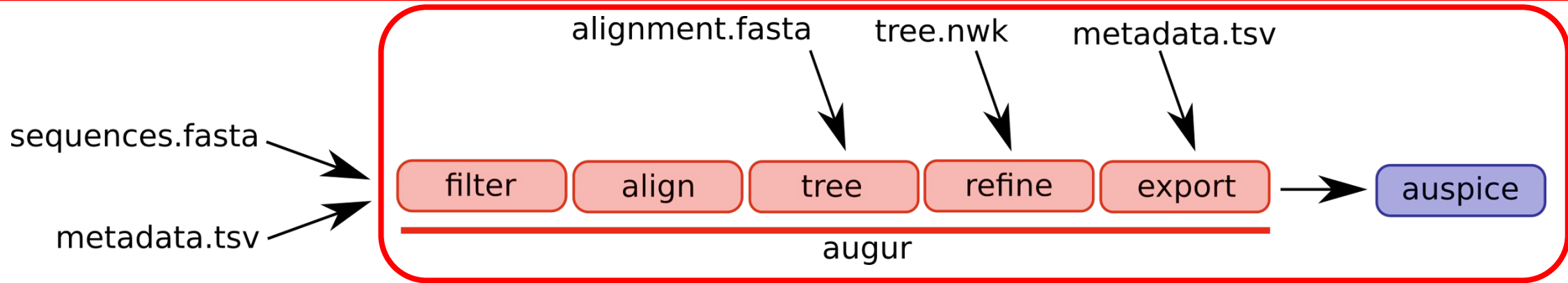
Installation steps

Steps vary by runtime option (Docker, Conda, ambient) and host interface (macOS, Windows (WSL), Windows (PowerShell), Linux). For help choosing, refer to our [Frequently Asked Questions \(FAQ\)](#), such as:

- [What's a runtime?](#)

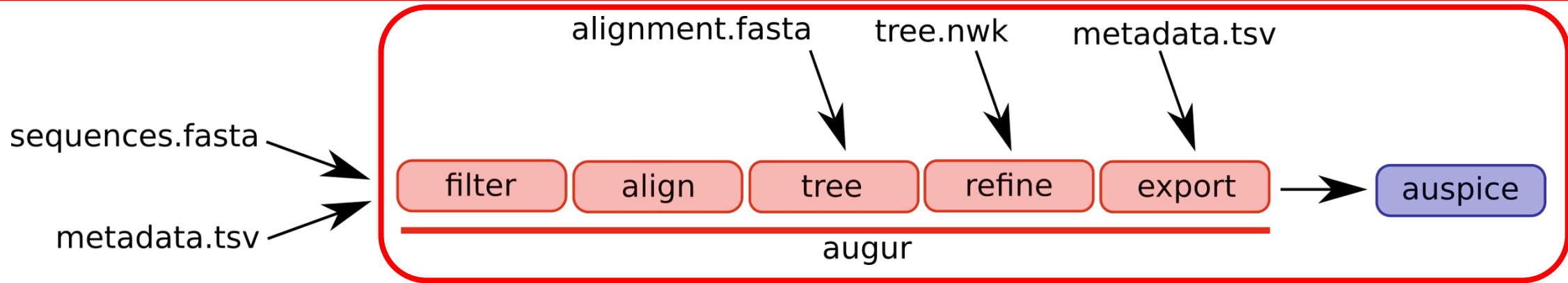
- Didactic tutorials with standard dataset to assist in learning the basic usage, language, and approach of Nextstrain
- ^ Dr. Shaun Cross will provide in-depth video on how to work through the Nextstrain tutorial
- Additional documentation for building custom Nextstrain builds (in-person training)

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`

Augur- the engine behind Nextstrain

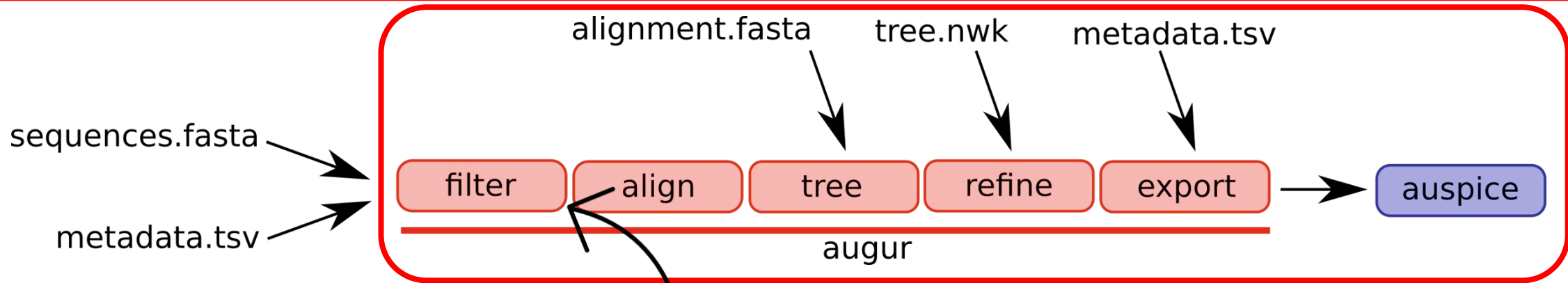


- What augur does

- 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)



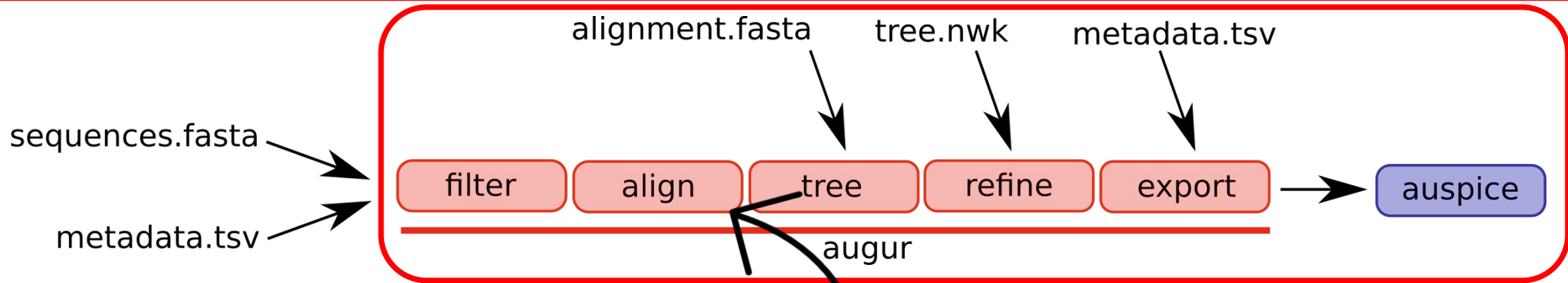
Augur- the engine behind Nextstrain



- 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)



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.

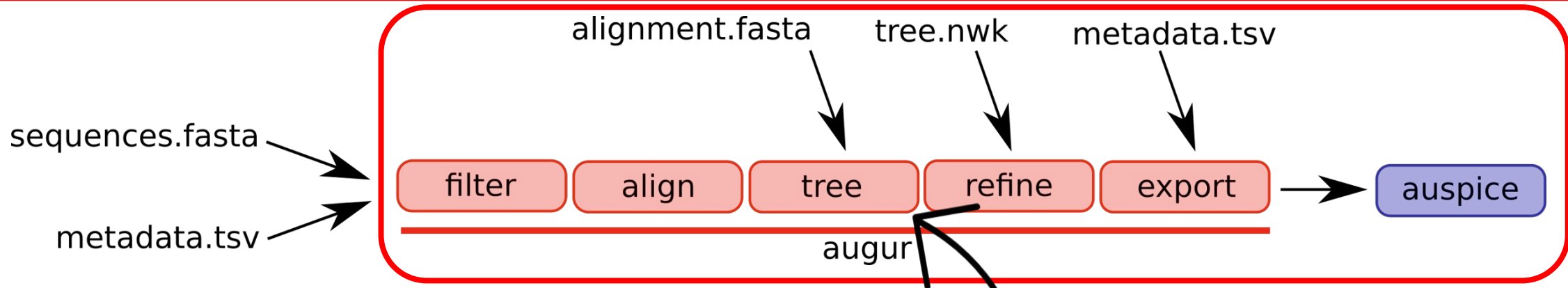
```
DENV1 AAAAGTCAAGGTCAAACGCAGCTATTGGAGCAGTGTTCTGTTGATGAAAATCA
DENV2 AAAGGTGAGAAAGCAATGCAGCCTTGGGGGCCATATTCAGTATGAGAACAA
DENV3 AAAGGTCAAGAACTAACGCAGCCATGGGCGCCGTTTTTCACAGAGGAGAACCA
DENV4 AAAAGTTAGATCAAACGCAGCCATAGGCGCAGTCTTTTCAGGAAGAACAGGG
ZIKV CAAGGTGCGCAGCAATGCAGCACTGGGAGCAATATTTGAAGAGGAAAAAGA
WNV AAAAGTCAACAGTAATGCCGCCCTAGGAGCGATGTTTGAAGAACAGAACCA
YFV AAAAGTCCGAAGTCATGCAGCCATTGGAGCTTACCTGGAAGAACAAAGACA
POWV GAAGGTGAGGTCCAACGCTGCTCTAGGTGCATGGTCTGGATGAACAGAAATAA
```

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)

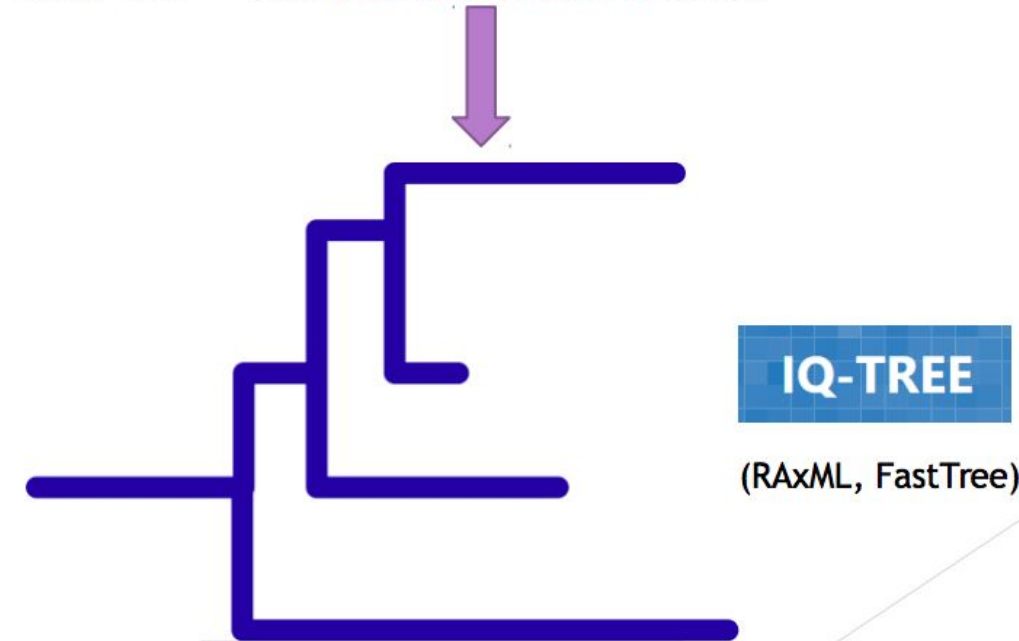


Building Phylogenies in Nextstrain

- 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
```



Building Phylogenies in Nextstrain

- IQ-TREE

- Maximum-likelihood (ML); Fast

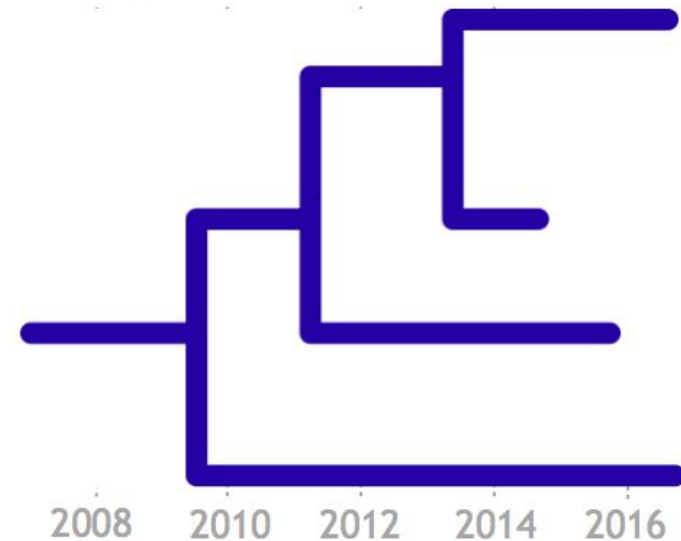
- TreeTime

- 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/>

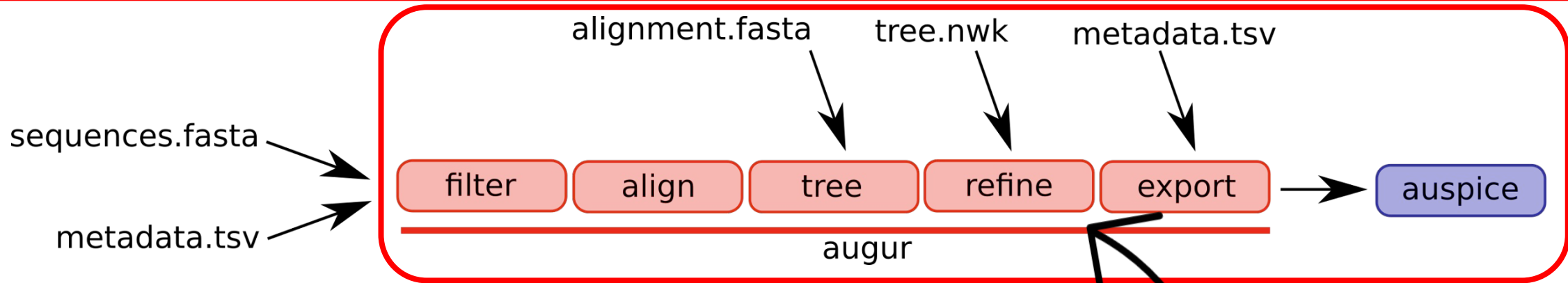
```
>SWE-2016 NNAGCAGTA---GGTAGCAATAACNNN
>ESP-2014 GTAGCAGTA---GGTAGCAGTNNNNNN
>USA-2016 NNNNNAGTGAATGGTAGTAGTAATAAT
>CAN-2015 NNNGCAGTGAATAGTAGTGGTNNNNNN
```



UNIVERSITY OF NEBRASKA MEDICAL CENTER™

COLLEGE OF PUBLIC HEALTH

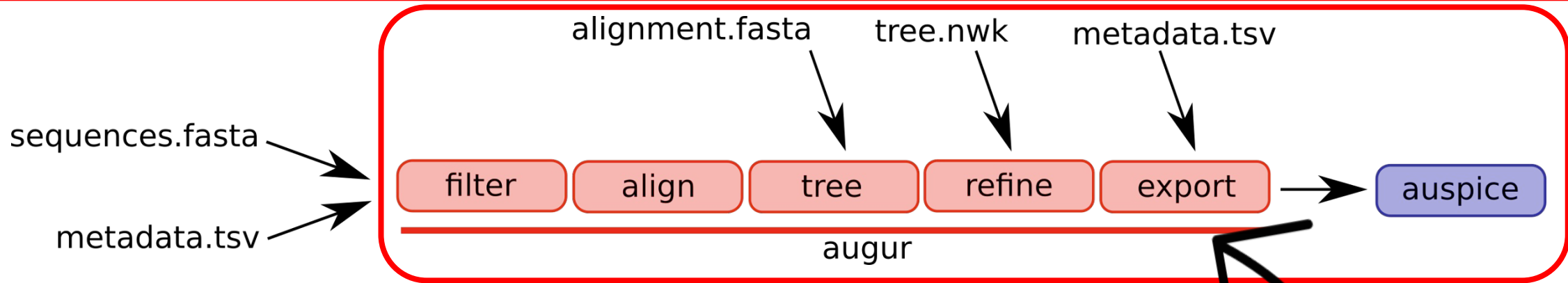
Augur- the engine behind Nextstrain



- 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



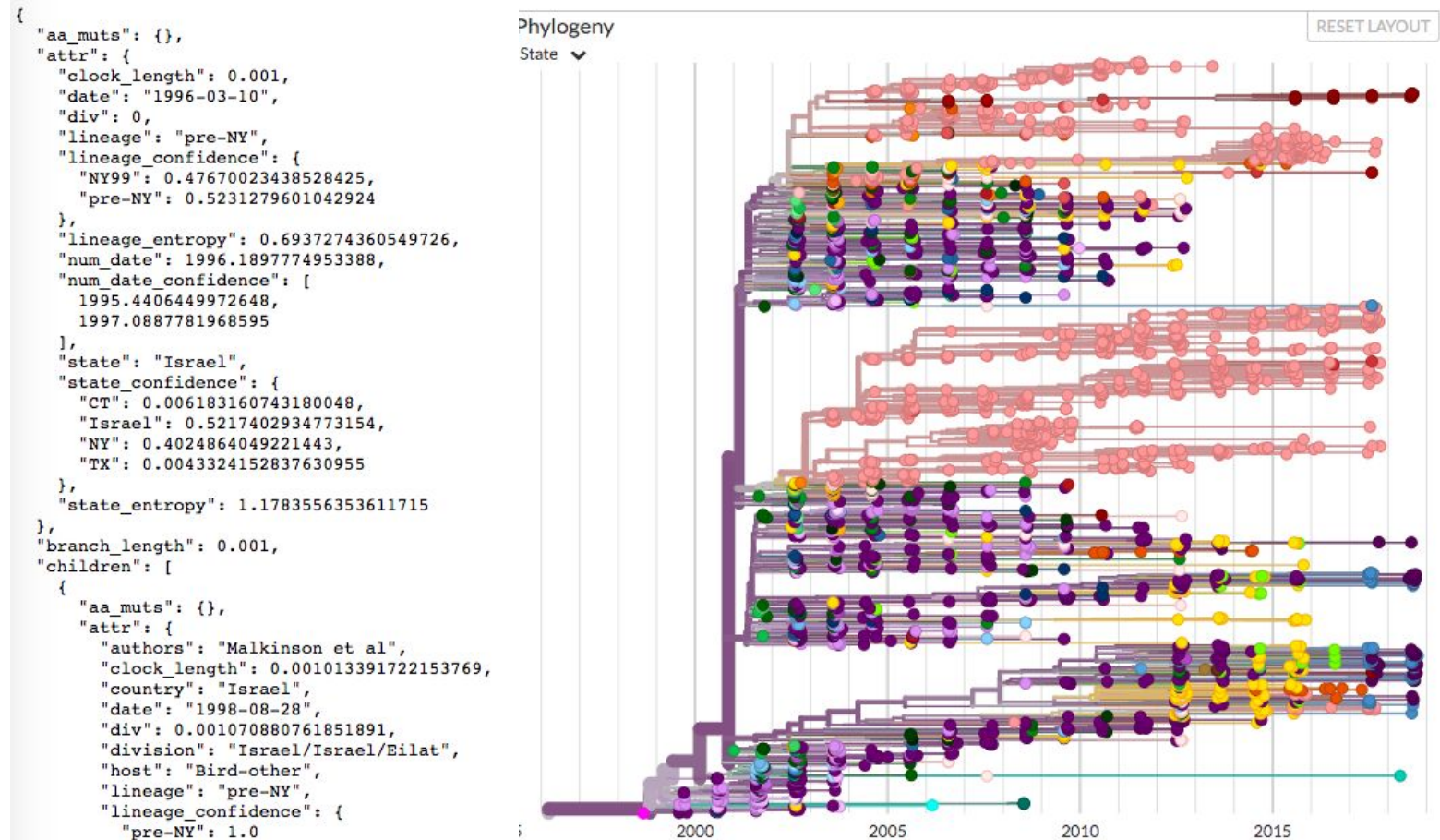
- What augur does

- 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)**



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>

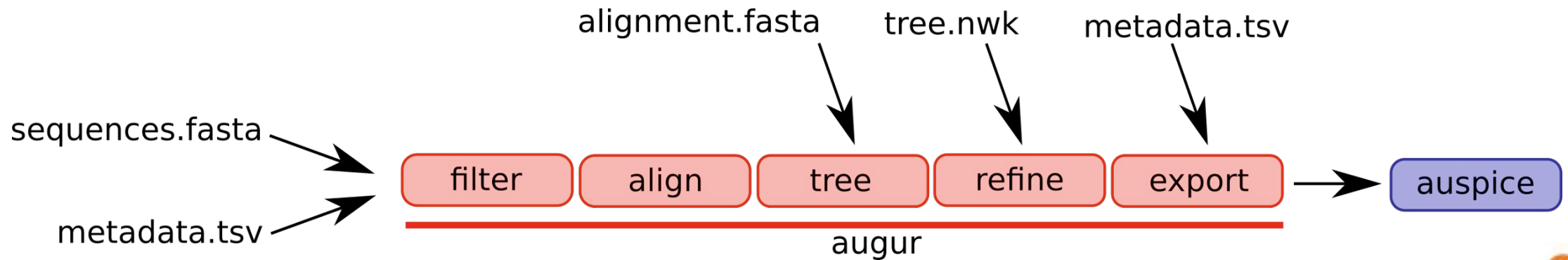
Auspice: <https://auspice.us/>

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**



Nextstrain

Real-time tracking of pathogen evolution

geneious
prime



UNIVERSITY OF NEBRASKA MEDICAL CENTER™
COLLEGE OF PUBLIC HEALTH

Questions?

Reach out via email: jfauver@unmc.edu