Create Influenza Trees with Nextstrain — seasonal-flu Workflow

# 1) Scope & outcomes

This SOP explains how to build influenza phylogenies using the official Nextstrain seasonal-flu workflow. You will start from sequences + metadata, run the nextstrain pipeline, and view the interactive tree in Auspice.

# 2) Prerequisites

* Familiarity with the command line (Linux/macOS/Ubuntu recommended).
* Access to influenza sequences + metadata on GISAID.
* Installed software (via conda/mamba is simplest): Nextstrain CLI.

# 3) Get the seasonal-flu workflow (We have done this step in the workshop already)

Clone the official repository containing rules, configs, and example builds.

git clone https://github.com/nextstrain/seasonal-flu.git  
cd seasonal-flu

* Directory overview:

seasonal-flu/  
├─ config/ # build-specific configuration (filters, regions, date ranges)  
├─ rules/ # Snakemake rules for alignment, tree, refinement, export  
├─ data/ # (you will place inputs here or set paths in config)  
└─ builds/ # output figures and auspice JSON will be written here

# 4) Prepare input data

You need aligned or raw nucleotide sequences (FASTA) plus a tabular metadata file (TSV/CSV) with sample names that match fasta headers.

* Navigate to GISAID. Select the "EpiFlu" link in the top navigation bar and then select "Search" from the EpiFlu navigation bar. From the search interface, select A/H3N2 human samples of your choice.
* Also, under the "Required Segments" section at the bottom of the page, select "HA". Then select the "Search" button. Select the checkbox in the top-left corner of the search results (the same row with the column headings), to select all matching records as shown below.
* Select the "Download" button. From the "Download" window that appears, select "Isolates as XLS (virus metadata only)" and then select the "Download" button.
* Create a new directory for these data in the seasonal-flu (from github above) working directory.

mkdir -p data/h3n2/

* Save the XLS file you downloaded (e.g., gisaid\_epiflu\_isolates.xls) as data/h3n2/metadata.xls. Return to the GISAID "Download" window, and select "Sequences (DNA) as FASTA". In the "DNA" section, select the checkbox for "HA". In the "FASTA Header" section, enter only Isolate name. Leave all other sections at the default values.
* Select the "Download" button. Save the FASTA file you downloaded (e.g., gisaid\_epiflu\_sequences.fasta) as data/h3n2/raw\_sequences\_ha.fasta.
* Ensure headers and metadata IDs match exactly (case/whitespace sensitive).
* Include useful columns in metadata (date, location, host, subtype).

# 5) Configure your build

Create/edit a config file (YAML) specifying subtype, date range, geographic focus, and paths to your inputs.

You can start from existing configs in the repository and modify fields for your study.

/config/h3n2/auspice\_config.json (*This is where you edit the visualization parameters; how the tree will look*)

/profiles/gisaid/builds.yaml (*make sure you check if the fields in your metadata is present and that the filters/subsampling number is more that your total number of sequences in your alignment*)

# 6) Run the seasonal-flu workflow

Activate your environment, then execute Snakemake through the Nextstrain CLI (preferred) or directly.

## Option A: Nextstrain CLI (recommended)

conda activate nextstrain  
nextstrain build . --configfile profiles/gisaid/builds.yaml

The pipeline will:

* Align with Nextalign and apply recommended masking.
* Infer a maximum-likelihood tree (IQ-TREE).
* Refine, root, and date the tree (Augur refine).
* Infer ancestors, translate, and annotate mutations (Augur).
* Export Auspice JSON for interactive visualization.

# 7) Outputs you care about

* Auspice JSON: in the path given by 'auspice\_prefix' (e.g., builds/h3n2.json).
* Tree files and logs: under 'builds/' or a 'results/' subfolder, depending on config.
* QC reports: filtered counts, alignment stats, masking summaries (see logs).

# 8) Visualize locally (Auspice)

Serve the JSON interactively in your browser:

# From the seasonal-flu directory (or where your JSON lives)  
nextstrain view auspice

* Use the sidebar to color by clade, location, or custom metadata columns.
* Export PNG/SVG for figures or screenshots for slides.

# 9) Share or deploy your build

* Share the JSON file (plus a small README with data sources and dates).
* Deploy to a static web host with Auspice’s 'export' mode, or use Nextstrain’s community builds (if applicable).
* For GISAID-based builds, follow attribution rules and do not redistribute raw sequences.

# 10) Good practices & QC

* Record the dataset time window, sequence counts, and filtering criteria in your README.
* Remove sequences failing QC (excess Ns, inconsistent dates) and re-run.
* Check tree for long branches (possible assembly or date errors).
* Document software versions:

nextstrain --version  
augur --version  
auspice --version  
iqtree2 -version

# 11) Troubleshooting

* Names mismatch: FASTA headers must match 'strain'/'name' in metadata.
* Date parsing errors: ensure ISO format (YYYY-MM-DD) or provide 'date' resolvable by Augur.
* Memory/timeouts: reduce 'max\_sequences' or subset by region/date; increase --cores / use HPC.
* Masked sites: verify your reference and mask BED align with your subtype.