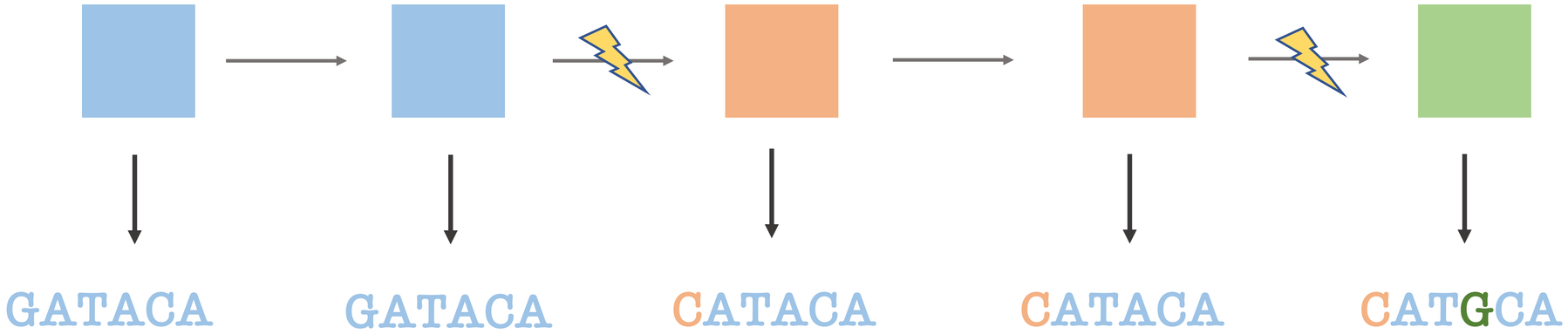
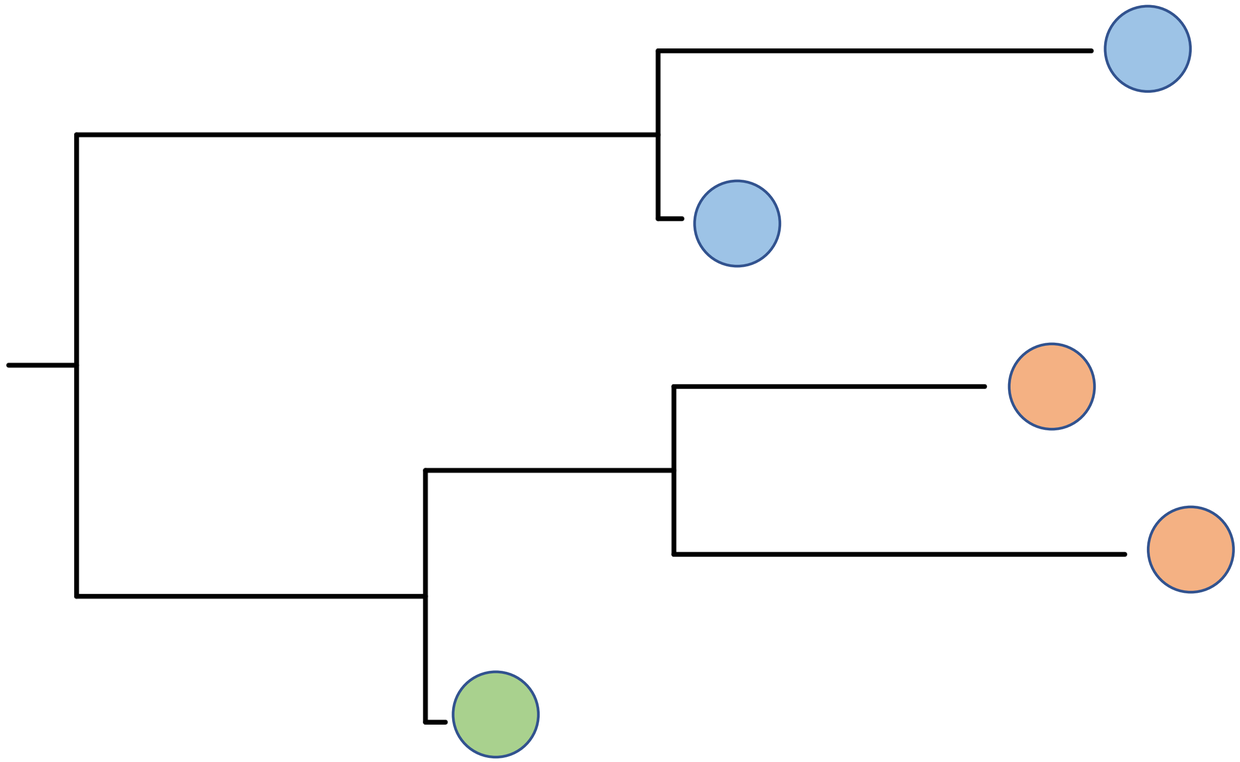
NEXTSTRAIN

Organisms acquire mutations

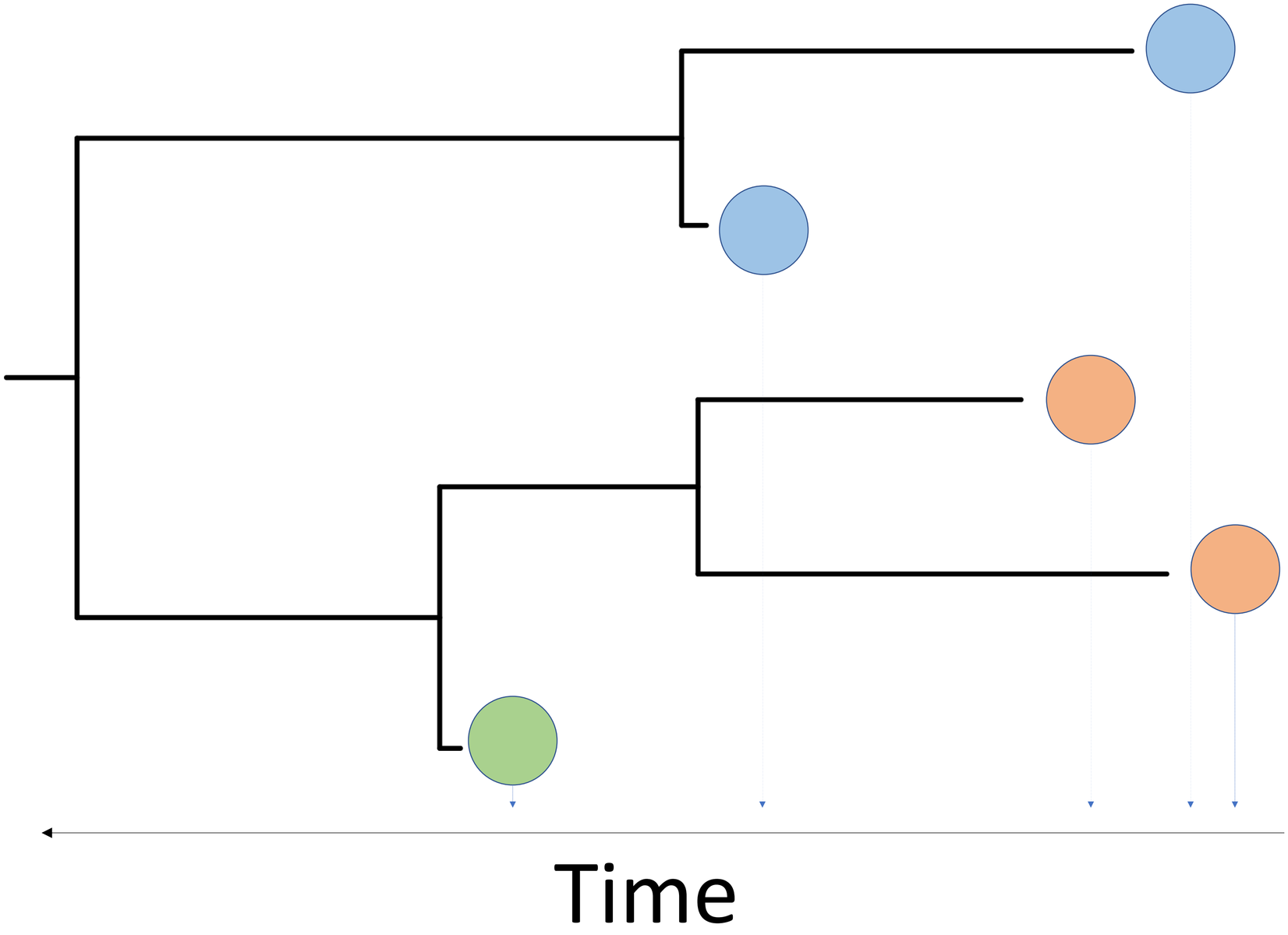


Phylogenetic trees reveal relationships



Genetic similarity

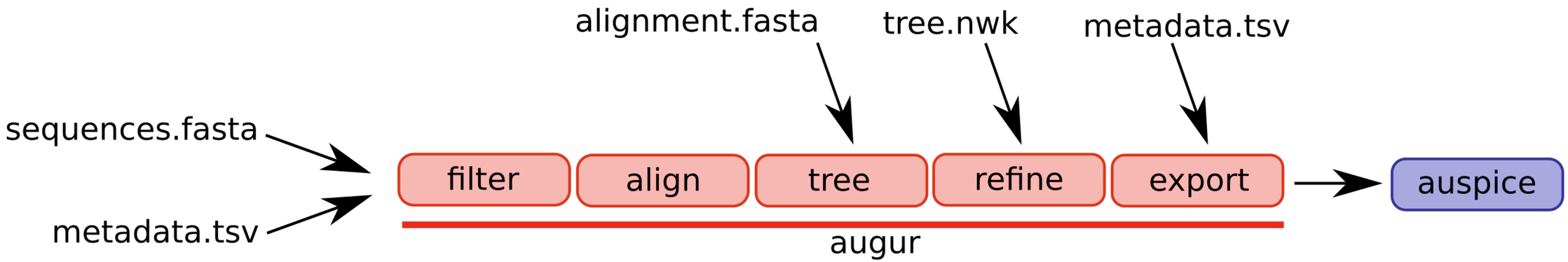
Trees reveal timing



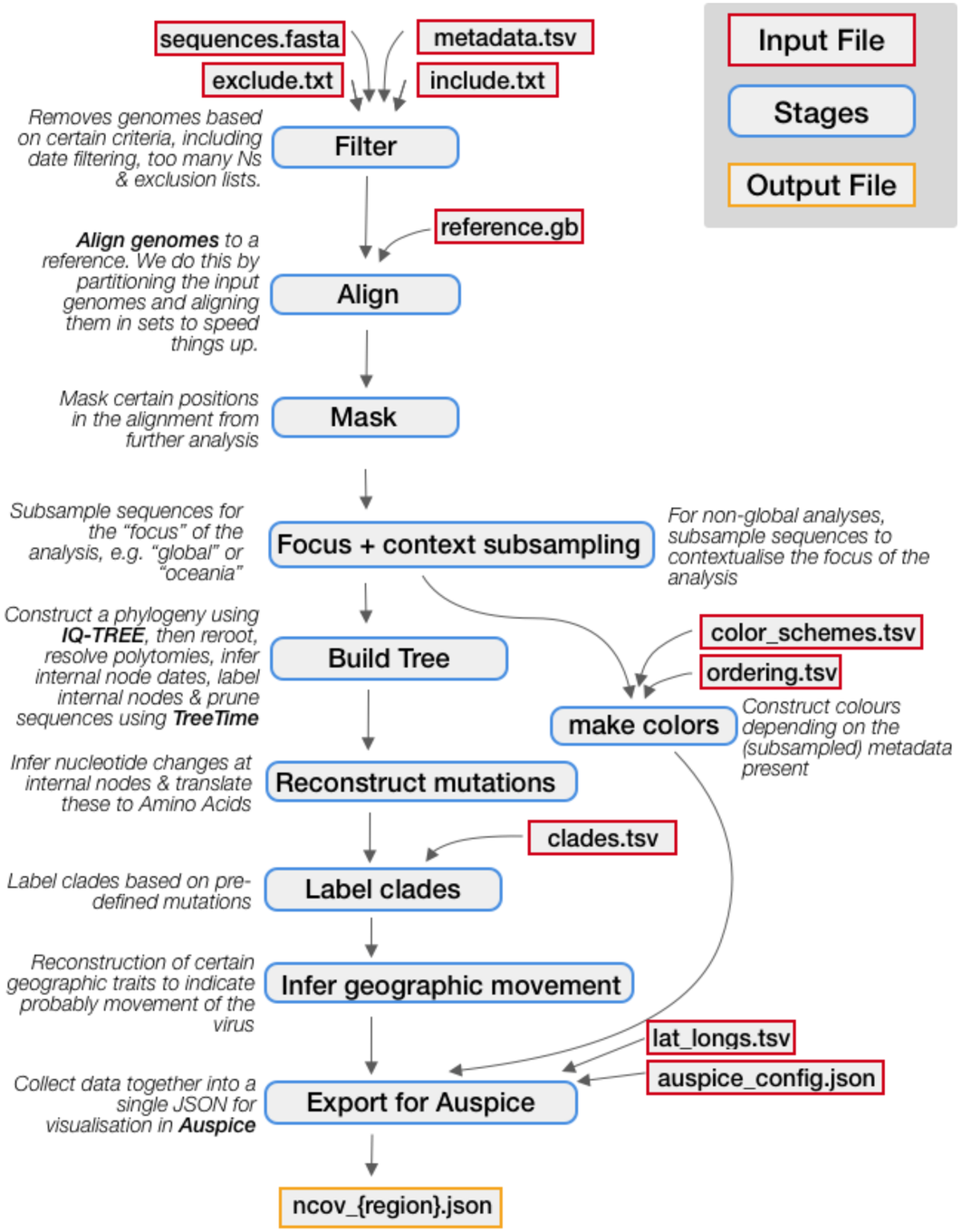
Nextstrain

- Ecosystem for performing sampling, alignment, and tree building
- Sort of a “quick and fast” way of performing genomic epidemiology
- Allows for the contextualization of sequencing data
- Easily updatable and customizable

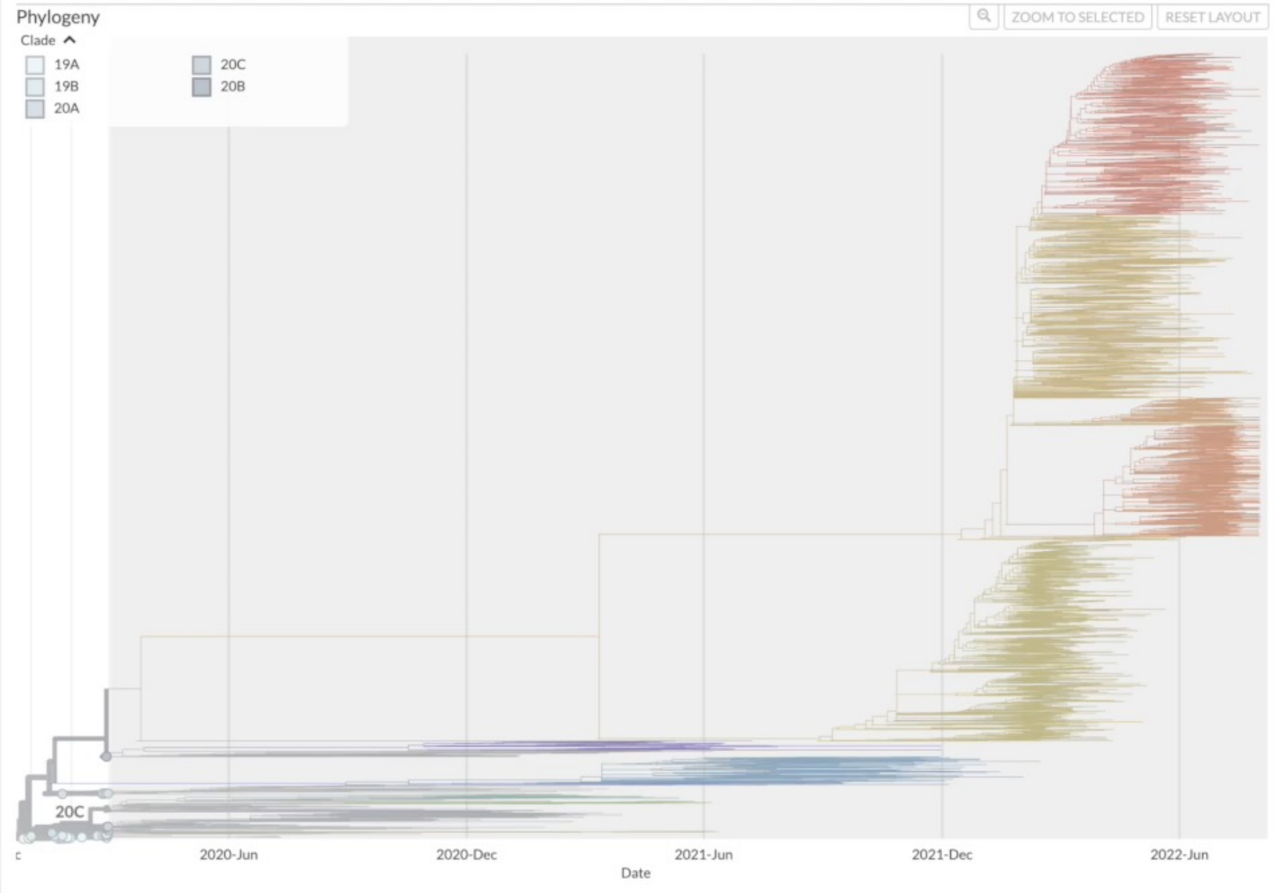
Nextstrain

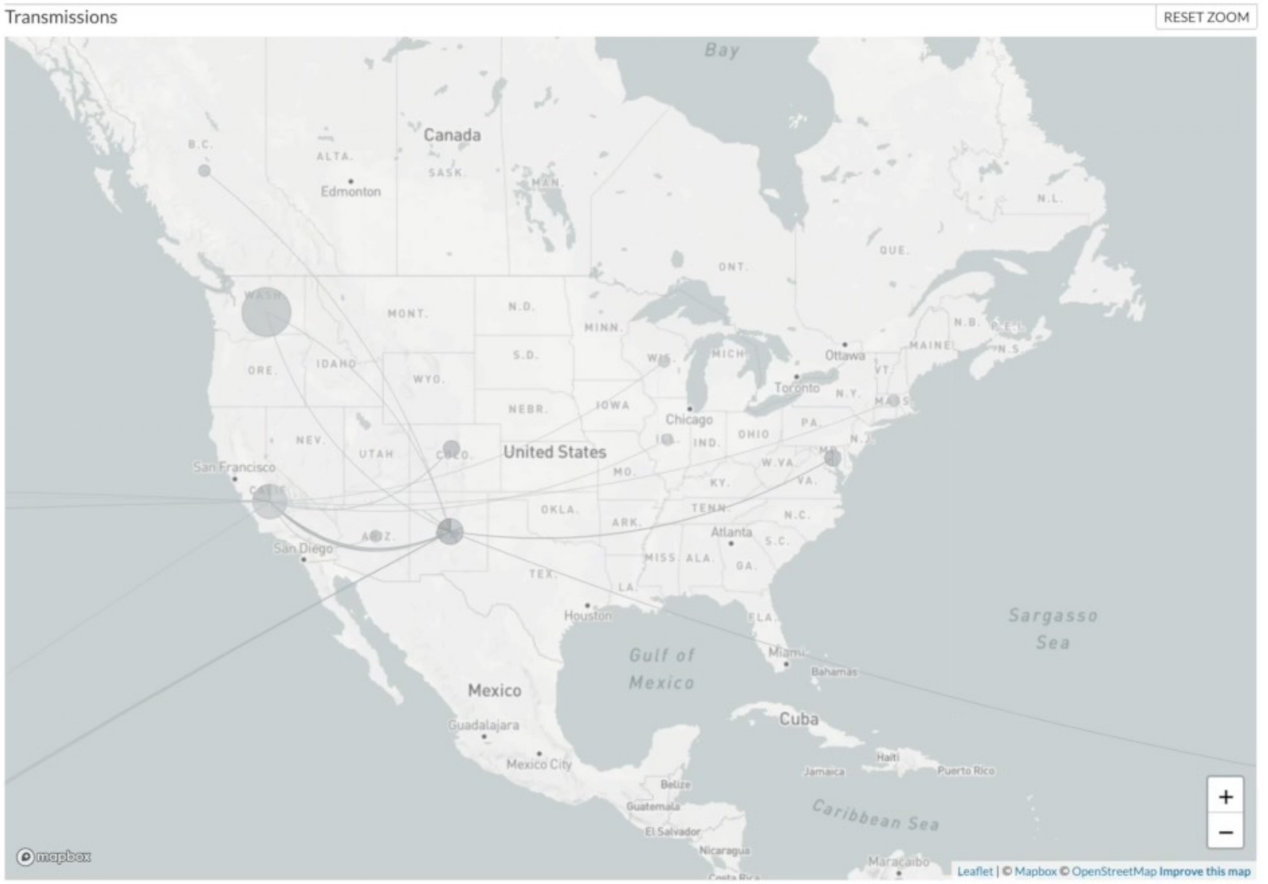

alignment.fasta
tree.nwk
metadata.tsv
sequences.fasta
metadata.tsv
filter
align
tree
refine
export
auspice
augur

Nextstrain

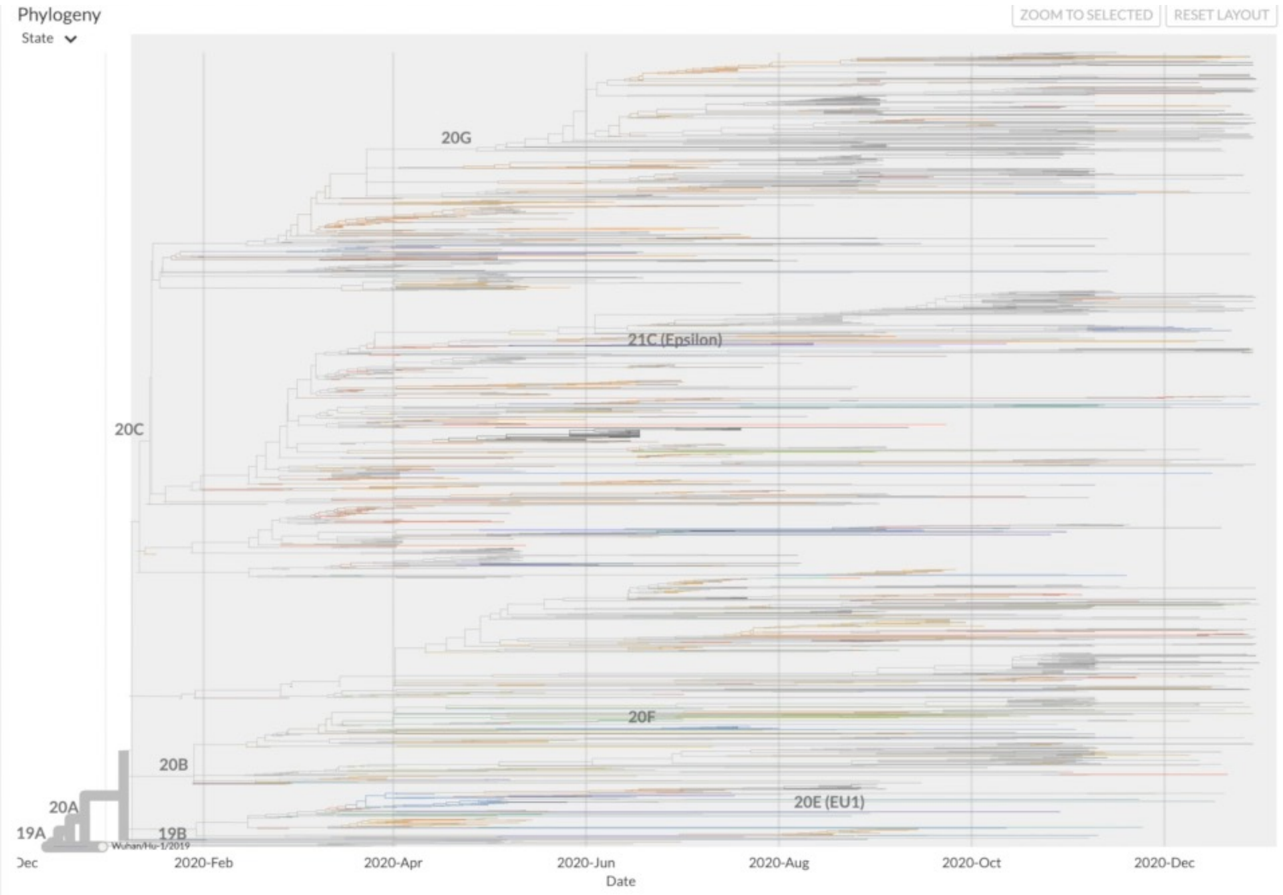

Input File
Stages
Output File
sequences.fasta
metadata.tsv
exclude.txt
include.txt
Removes genomes based on certain criteria, including date filtering, too many Ns & exclusion lists.
Filter
reference.gb
Align genomes to a reference. We do this by partitioning the input genomes and aligning them in sets to speed things up.
Align
Mask certain positions in the alignment from further analysis
Mask
Subsample sequences for the "focus" of the analysis, e.g. "global" or "oceania"
Focus + context subsampling
For non-global analyses, subsample sequences to contextualise the focus of the analysis
Construct a phylogeny using IQ-TREE, then reroot, resolve polytomies, infer internal node dates, label internal nodes & prune sequences using TreeTime
Build Tree
color_schemes.tsv
ordering.tsv
make colors
Construct colours depending on the (subsampled) metadata present
Infer nucleotide changes at internal nodes & translate these to Amino Acids
Reconstruct mutations
clades.tsv
Label clades based on pre-defined mutations
Label clades
Reconstruction of certain geographic traits to indicate probably movement of the virus
Infer geographic movement
lat_longs.tsv
auspice_config.json
Collect data together into a single JSON for visualisation in Auspice
Export for Auspice
ncov_{region}.json

Viral diversity over time





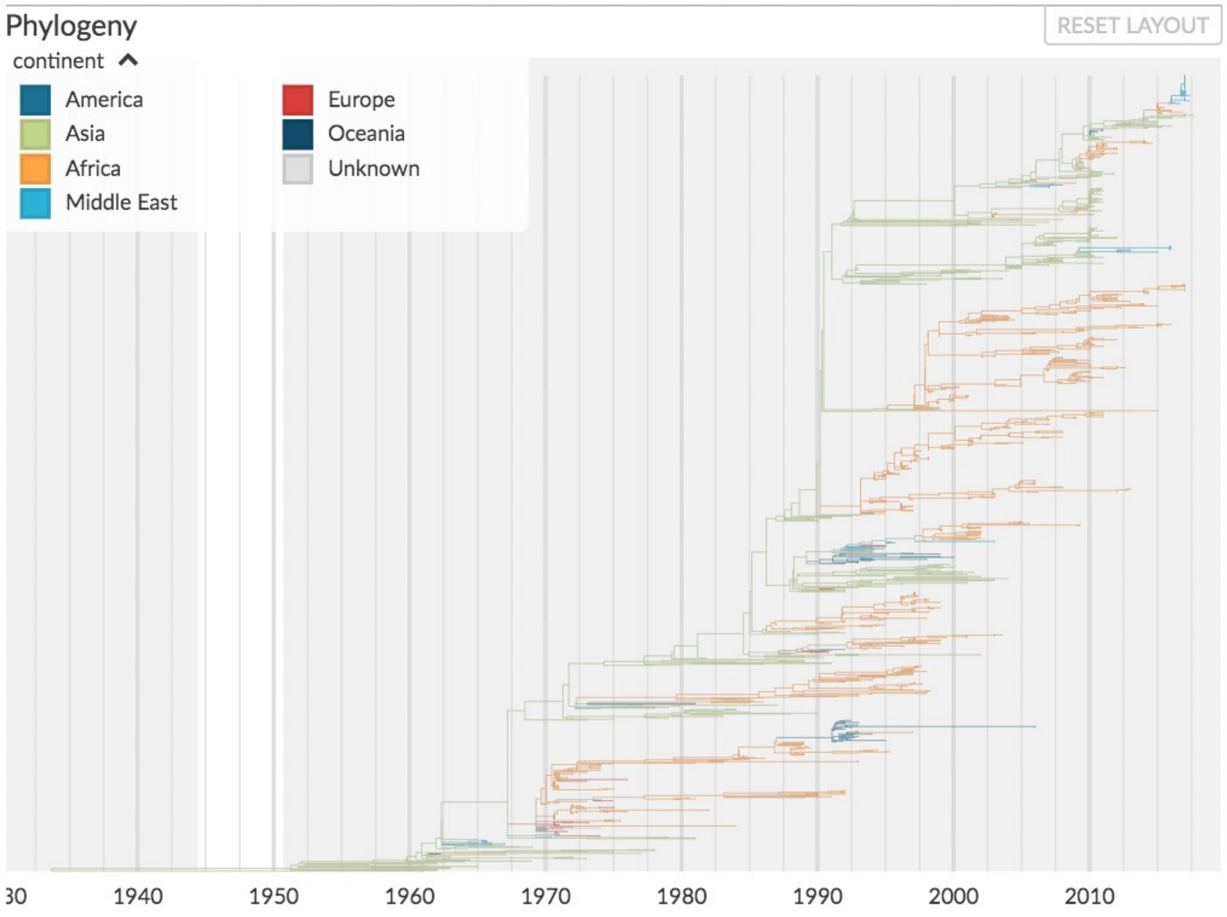
Define more fine grained geographic designations

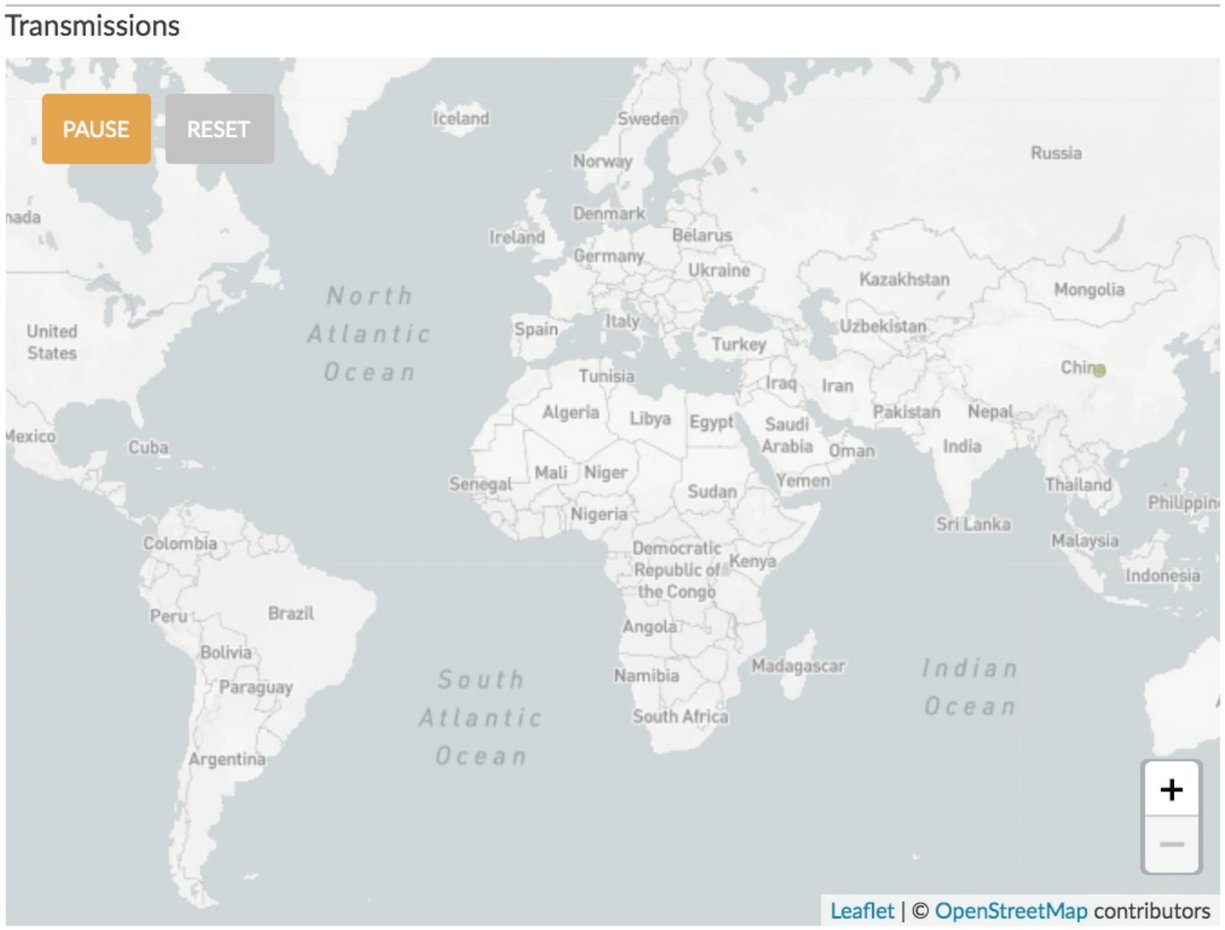


Global cholera dynamics

Cholera

Animation in progress.





nextstrain

Questions?