# Phylogenetic time tree construction

7,262 sequenced Alpha lineage genomes with less than 12,000 ambiguous bases were exported with the respective data from our local database. These were filtered to remove 609 non-Thuringian isolates, four duplicates, and 127 entries missing location information or isolation date. After filtering, the Thuringian dataset contained 6,522 isolates.

From the 1,091,655 public German genomes provided by the Robert Koch Institute, 146,347 Alpha lineage isolates were extracted. For these, more specific location data (country, state, district, city, latitude, longitude) were added based on the sending laboratory’s postal code. Subsequently, 9,600 Thuringian isolates were removed. The resulting 136,747 genomes were then analyzed with poreCov (version 1.7.2) to determine mutations/deletions and confirm the lineages.[1](https://www.zotero.org/google-docs/?A1Nxh8) The results were uploaded to a separate collection in our local database. poreCov identified a further 108 non-Alpha isolates, which were subsequently removed from the dataset. During the generation of a test phylogenetic time tree (see below), “Nextstrain” excluded 540 isolates due to quality issues.[2](https://www.zotero.org/google-docs/?wrZqqQ) After their removal, the final German Alpha dataset comprised 136,099 entries.

The prepared Thuringian and German Alpha lineage datasets were combined to build a phylogenetic time tree using the program “Nextstrain” (version 5.0.0).[2](https://www.zotero.org/google-docs/?kiYYXW) Following the program's instructions, two genomes of the original Wuhan lineage were added to the combined fasta-file, while their respective data were added to the combined data file (omitting this step would lead to a crash in the used Nextstrain version). It was ensured that the data file contained the necessary columns given in the manual. Further, in our version, the data file needed to contain at least 12 columns, or Nextstrain would have an error. The resulting fasta and metadata file for the nextstrain analysis are provided as Supplementary Files 3 and 4, respectively.

A custom “build.yaml”-file was used for the nextstrain analysis, which is provided as Supplementary File 5:

inputs:

- name: alpha-lin\_phylotree

metadata: data/alpha-lin\_phylotree\_metadata.tsv

sequences: data/alpha-lin\_phylotree.fasta

builds:

thuringia:

subsampling\_scheme: custom-division

region: Europe

country: Germany

division: Thuringia

subsampling:

custom-division:

focal:

group\_by: "year month"

seq\_per\_group: 10000000

query: --query "(country == '{country}') & (division == '{division}')"

related:

group\_by: "country year month"

seq\_per\_group: 16000

exclude: "--exclude-where 'division={division}'"

priorities:

type: "proximity"

focus: "focal"

files:

auspice\_config: "my\_profiles/alpha-lin\_analysis/my\_auspice\_config.json"

description: "my\_profiles/alpha-lin\_analysis/my\_description.md"

The analysis was started with the command:

“nextstrain build . --cores 32 --profile my\_profiles/path/to/build/dir”.

The resulting phylogenetic time tree comprises 64,131 German (non-Thuringian) and 6,298 Thuringian Alpha-lineage genomes. Additional 2,160 German (non-Thuringian) and 224 Thuringian genomes are included in the metadata but not visualized in the tree and, therefore, not considered for the following analysis. The resulting tree is shown in Figure 1-figure supplement 3. The subsampled genome-, tree-, and metadata file are provided as Supplementary Files 6, 7, and 8.

## References

[1 Brandt C, Krautwurst S, Spott R, *et al.* poreCov-An Easy to Use, Fast, and Robust Workflow for SARS-CoV-2 Genome Reconstruction via Nanopore Sequencing. *Front Genet* 2021; **12**: 711437.](https://www.zotero.org/google-docs/?9F4Uto)

[2 Hadfield J, Megill C, Bell SM, *et al.* Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* 2018; **34**: 4121–3.](https://www.zotero.org/google-docs/?9F4Uto)