

**SupplementaryData1:** GIST logic, parameters and sampling numbers

EpiCoV - GISAID  
Data (15,271,032 genomes)

Augur

```
-min-date {min_date} --max-date {max_date} \  
--min-length 28400 \  
--query "(country == 'Brazil') &  
        (division == '{state-name}') &  
        (pango_lineage == {target_lineages}) &  
        (host == 'Human') " \  
--subsample-max-sequences {max_genomes} \  
--exclude-ambiguous-dates-by any \  
--group-by pango_lineage year month \  

```

Gamma = 3,085  
Delta = 2,916  
Omicron = 2,985

Subsampling

EpiCoV - GISAID  
Data without subsampled  
data

BLASTn

```
--task megablast --evaluate 0.00001 \  
qcovhsp >= 99.9
```

Subsampling with  
similar genomes

Gamma = 4,561  
Delta = 4,364  
Omicron = 4,654

MAFFT

```
--keeplength --compactoutput --anysymbol \  
--kimura 1 --6merpair
```

aligned subsampling data