

Assigning your own data to GPSCs

Install popPUNK as per instructions at

<https://poppunk.readthedocs.io/en/latest/installation.html> and downloaded the GPS reference database “GPS_query.tar.bz2” from <https://www.pneumogen.net/gps/assigningGPSCs>

Files required to run GPSC assignment using popPUNK:

queries.txt: a list of paths to assemblies you wish to assign GPSCs to

GPS_query: GPS reference database uncompress

GPS_query.tar.bz2

output directory name is assigned using --output

number of threads can be changed using --threads

Run GPSC assignment:

```
poppunk --assign-query --ref-db GPS_query --distances  
GPS_query/GPS_query.dists --model-dir GPS_query --q-files  
queries.txt --output GPSC_assignment --threads 8 --no-stream  
--full-db --external-clustering GPS_query/gpsc_definitive.csv
```

Output files:

_clusters.csv: popPUNK clusters with dataset specific nomenclature

_external_clusters.csv: GPSC scheme designations

Novel Clusters: Will be assigned NA in the

_external_clusters.csv as they have not been seen in the dataset used to designate the GPSCs. The popPUNK _clusters.csv file can be used to determine if NA isolates are the same cluster or not. Please email

globalpneumoseq@gmail.com to have novel clusters added to the database and a GPSC cluster name assigned.

Merged clusters: Unsampling diversity may represent variation between two clusters, GPSCs are the merged, for example if GPSC23 and GPSC362 merged the GPSC would be then reported to as GPSC23;362