GPSC assignment

Install PopPUNK 2.4 as per instructions at [PopPUNK documentation](#) and download the GPS reference database <download [button](#)> and the GPS designation <download [button](#)>

Files required to run GPSC assignment using PopPUNK 2.4:
1. A 2-column tab-delimited file containing sample name and path to the corresponding assembly (no header)
2. GPS reference database <GPS_v5>
3. GPS designation <GPS_v5_external_clusters.csv>

output directory name is assigned using *--output*
number of threads can be changed using *–threads*

Run GPSC assignment:
poppunk_assign --db GPS_v5 --distances GPS_v5/GPS_v5.dists --query <2-column path to assembly> --output <GPSC_assignment> --external-clustering GPS_v5_external_clusters.csv

Outputs:
_clusters.csv: popPUNK clusters with dataset specific nomenclature
_external_clusters.csv: GPSC v5 scheme designations
*Novel Clusters: Will be assigned NA in the _external_clusters.csv as they have not been defined in the v5 dataset used to designate the GPSCs. Please email: **globalpneumoseq@gmail.com** to have novel clusters added to the database and a GPSC cluster name assigned after you have checked for low level contamination which may contribute to biased accessory distances.*

Merged clusters: Unsampled diversity may represent missing variation linking two clusters. GPSCs are then merged. For example if GPSC23 and GPSC362 merged, the GPSC would be then reported as GPSC23, with a merge history of GPSC23;362.

History

The instruction for PopPUNK v2.4 are available to download <[download button](#)>

The instruction for PopPUNK v1 are available to download <[download button](#)>