

Tools for taxonomic assignment of microbial SAGs

Jacob H. Munson-McGee, Greg Gavelis

Bioinformatics of Microbial Single Cells

Tuesday April 5th, 2022



Taxonomic Assignment

- After isolation and sequencing of SAGs one of the first tasks is the identification of the species the cell belonged to.
- Helps in comparing cells both within a study/sample as well as between studies.



Small Ribosomal Subunit classification

Pros

- Traditional method
- All species have SSU sequences
- very well characterized sequence and multiple tools for identification
- Robust methods for creation of phylogenetic trees

Cons

- Many SAGs do not contain 16S sequences
 - In AG-910 only 67/226 have SSU sequences identified (30%)
-
- Can supplement 16S with PCR and targeted sequencing of SSU genes. However this relies on good primer design, it also increases the cost of identification

Whole Genome Classification Methods

- Typically rely on a combination of:
 - Identification of marker genes
 - Whole genome alignment (e.g. Average Nucleotide Identity)
 - 16S identification



CheckM (2015)

- Uses marker genes specific to lineages to estimate completeness and identify the lineage
- Flawed model as a result of the overrepresentation of cultured species and medically relevant species in NCBI database.
- Poor estimation of contamination
- Poor agreement with 16S sequences in SAGs
- We have stopped using it for SAGs



Genome Taxonomy Database

- Uses a combination of reference genes and ANI
- Extensive database with the best representation of environmental species that we have identified.
-



Hands on Activity

- At this point open up jupyter_hub/terminal and navigate to

`/mnt/storage/lessons/Day2AM_taxonomic_classification/`

and follow the instructions in the Day2AM_setup notebook

