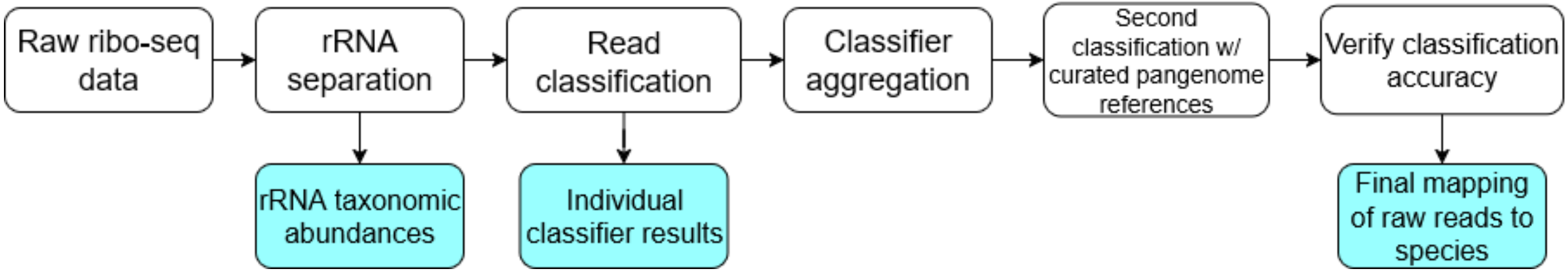
Pore Water Microbiome Characterization using Ribosome Profiling

Yochen Zhong

Objectives

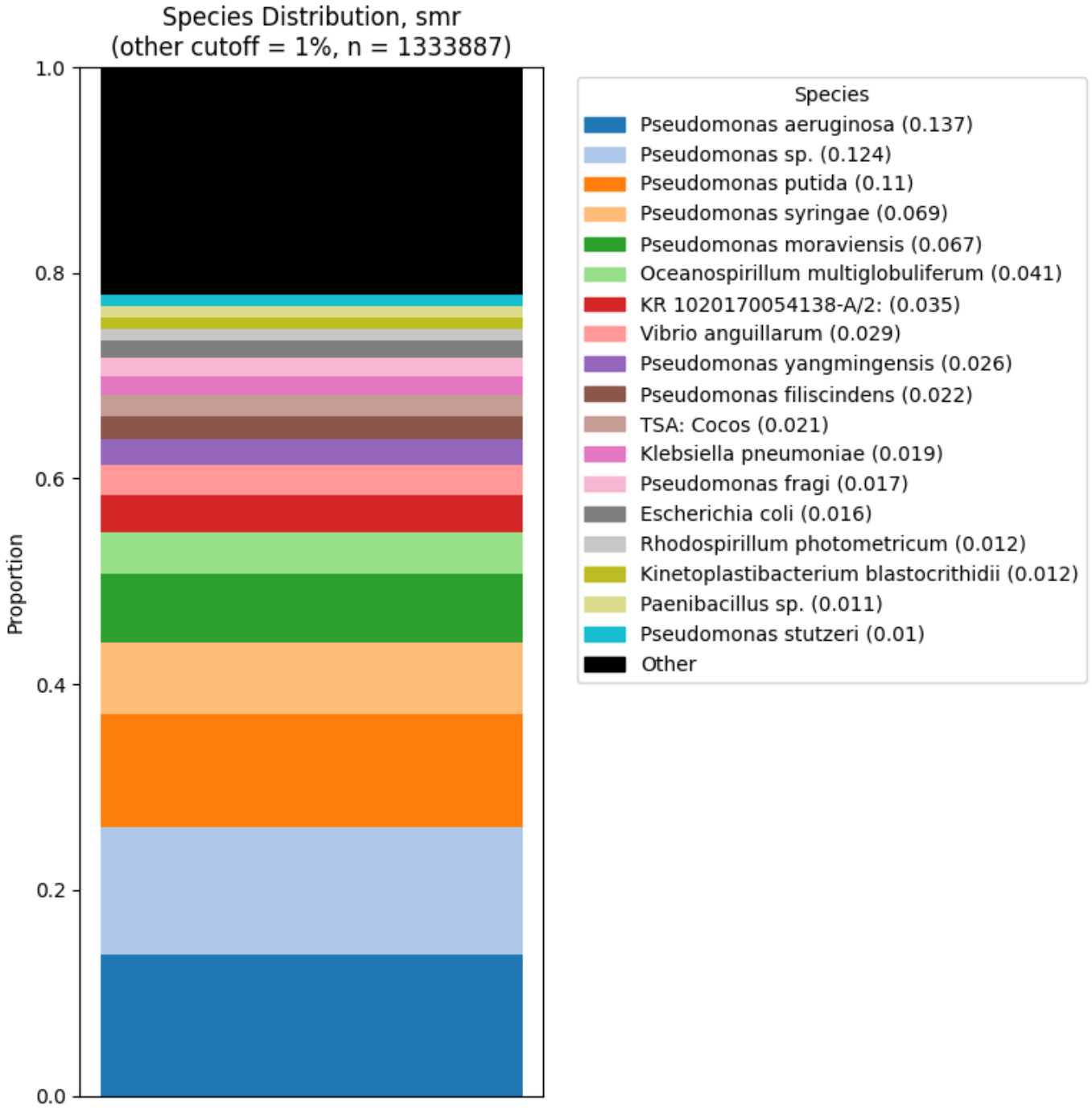
1) Assign and validate taxonomic identity of individual Ribo-ITP reads from pore water sample
2) Identify functional capabilities of the microbiome

Part 1: Taxonomic Classification



rRNA separation

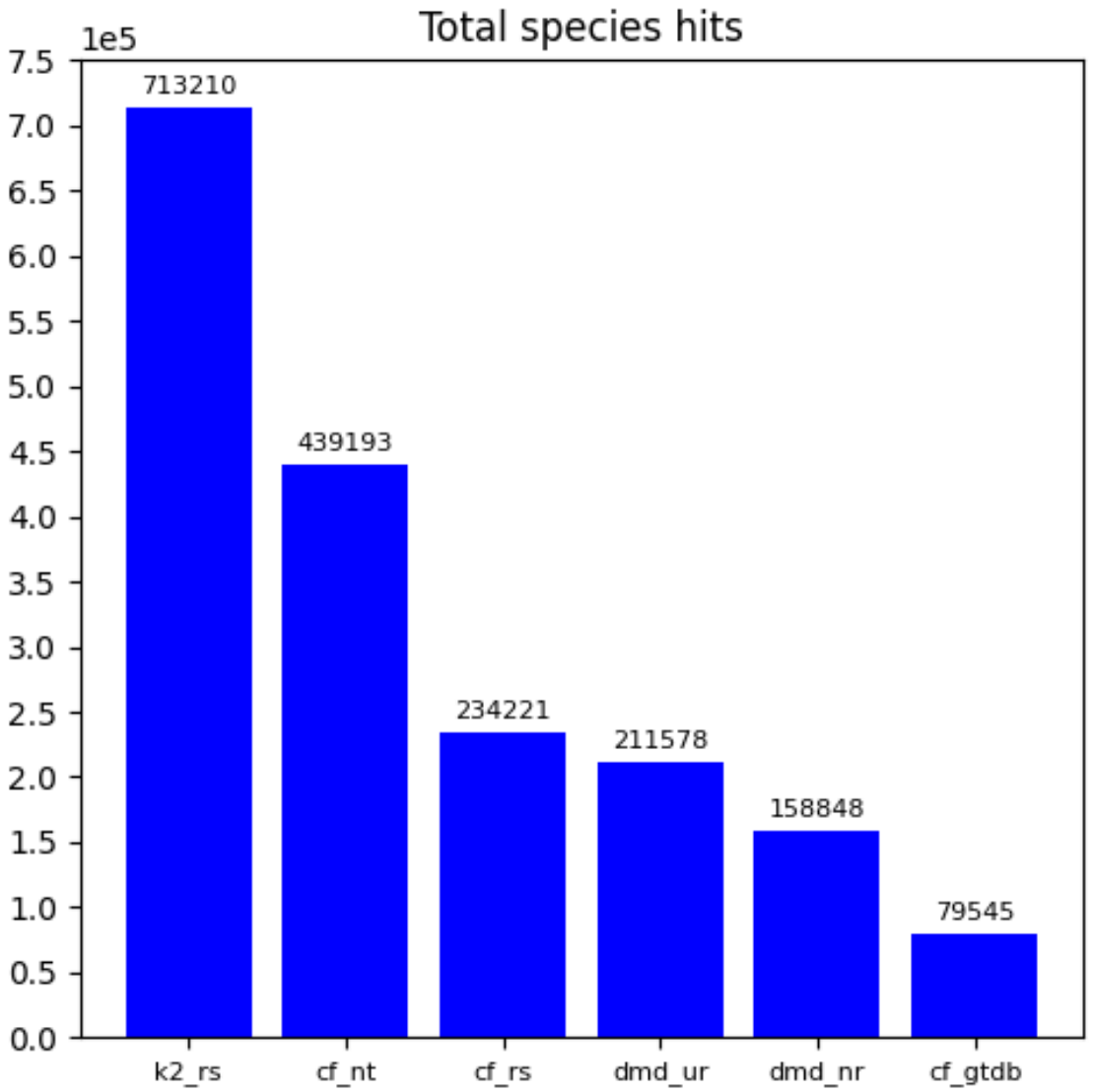
- SortMeRNA (Kopylova et al., 2012)
- 7219457 raw reads:
  - 1333887 (18.48%) classified as rRNA
  - 5885570 (81.52%) retained for downstream classification

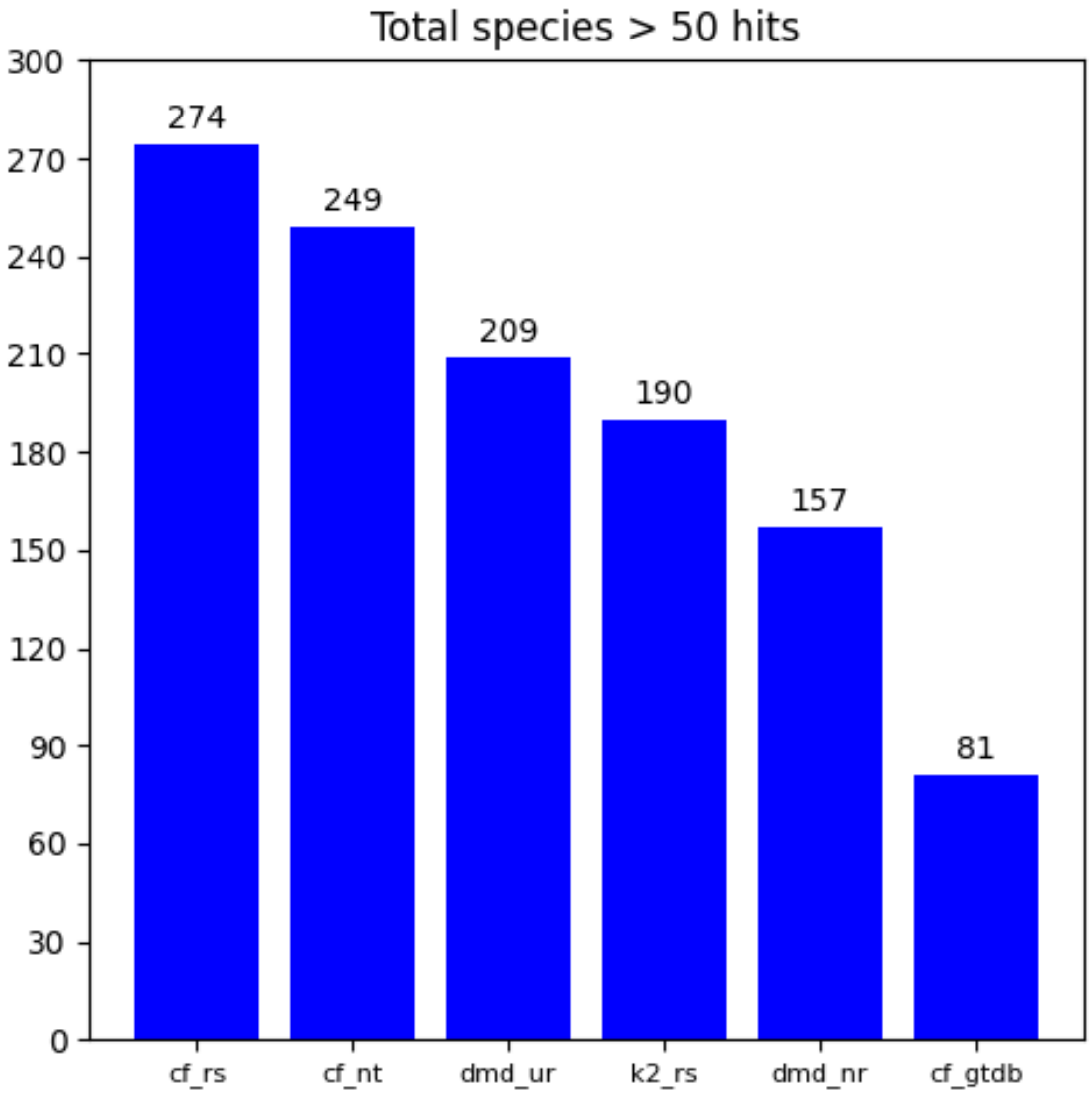


Classifiers and databases used for initial read assignments

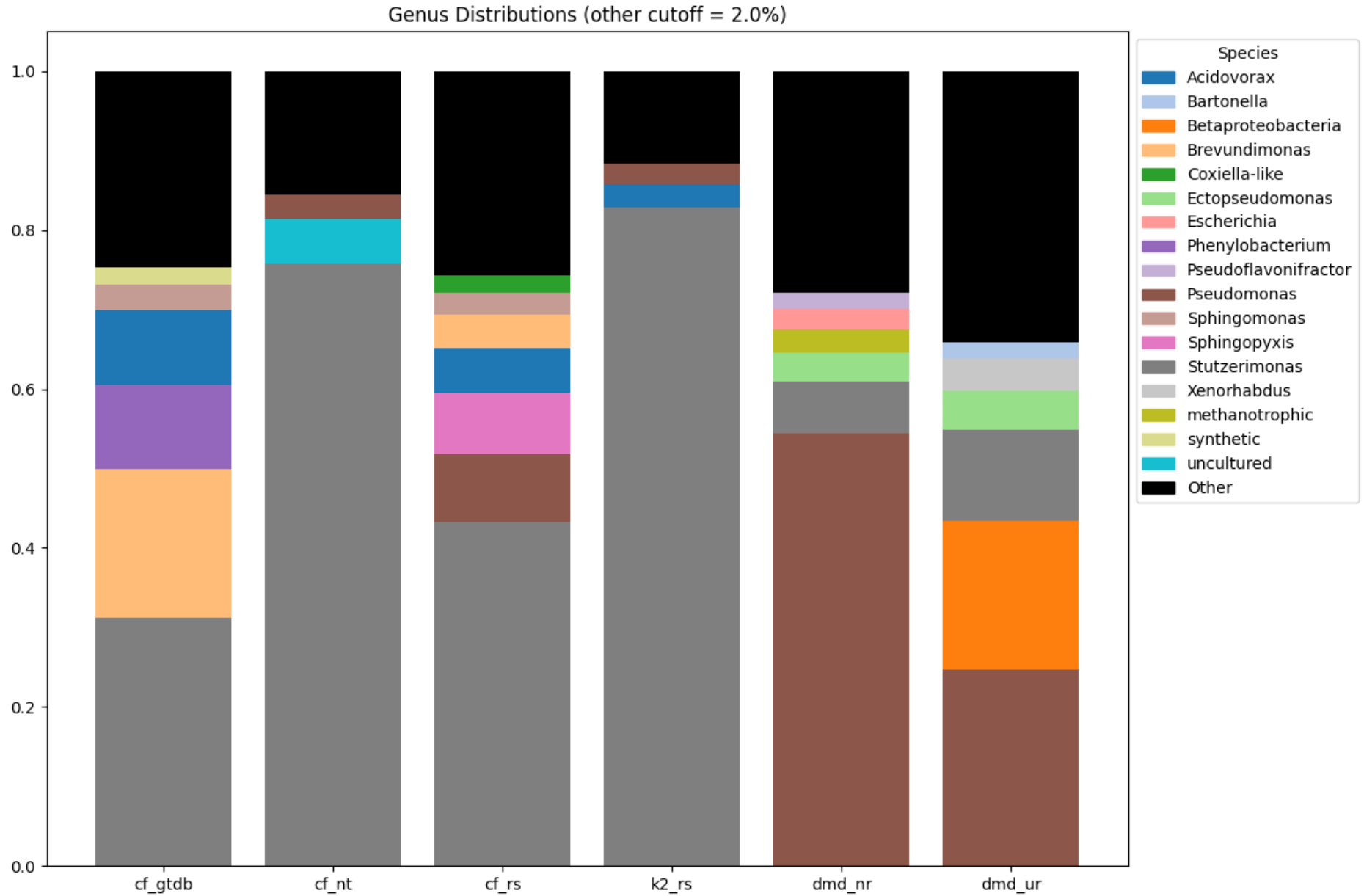
- Nucleotide matching
  - Kraken2 (Wood et al., 2014)
    - NCBI RefSeq Bacteria/Human complete genomes (O'Leary et al., 2016)
  - Centrifuger (Song & Langmead, 2024)
    - NCBI RefSeq Bacteria/Human/Virus/Archea, GenBank SARS-CoV2
    - GTDB r226 + RefSeq Human/Virus/Fungi/Contaminants
    - NCBI nt
- Protein matching
  - DIAMOND (Buchfink et al. 2021)
    - NCBI nr
    - UniRef100

Individual classifiers identify a moderate amount of reads



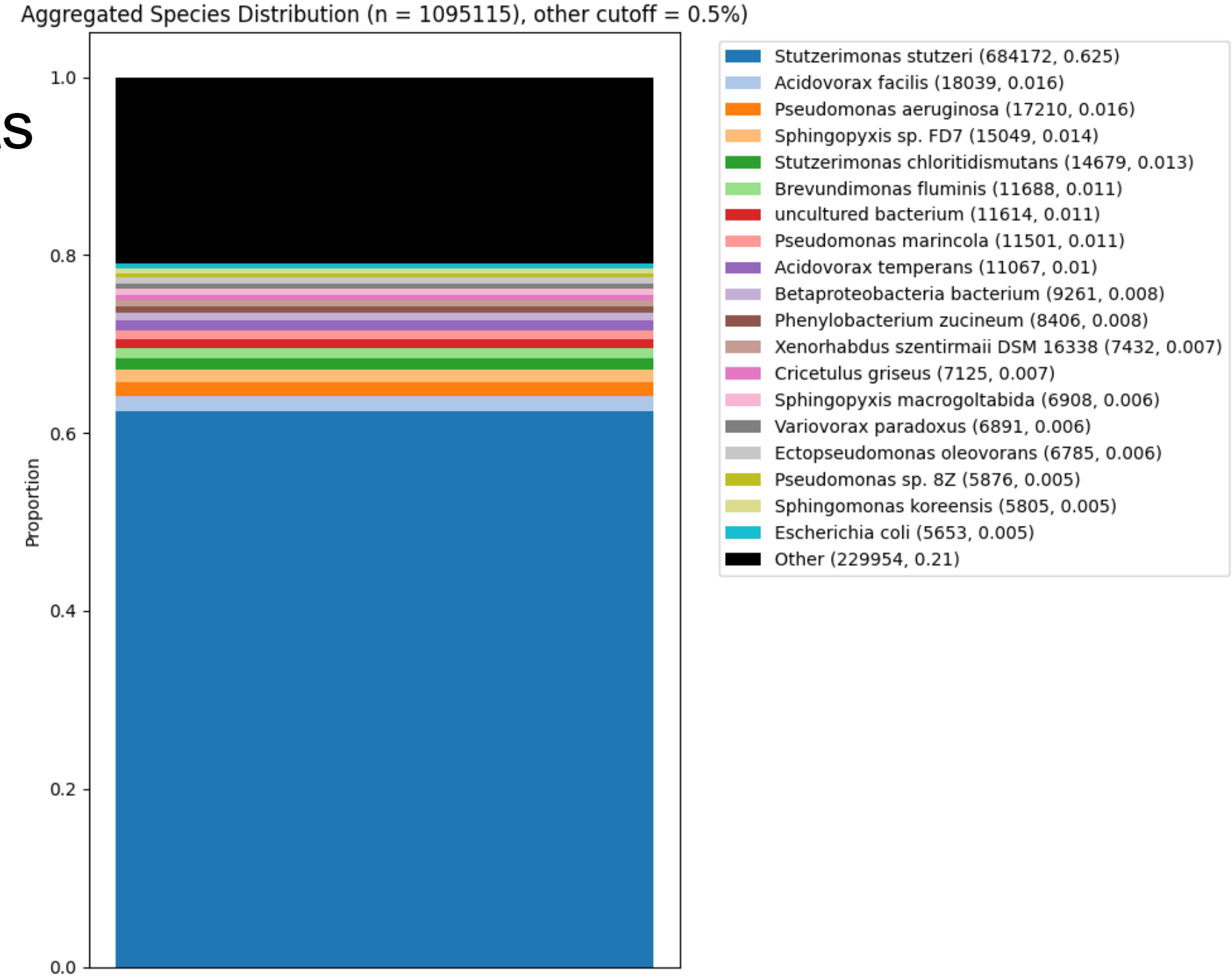


Initial classifier results, genera

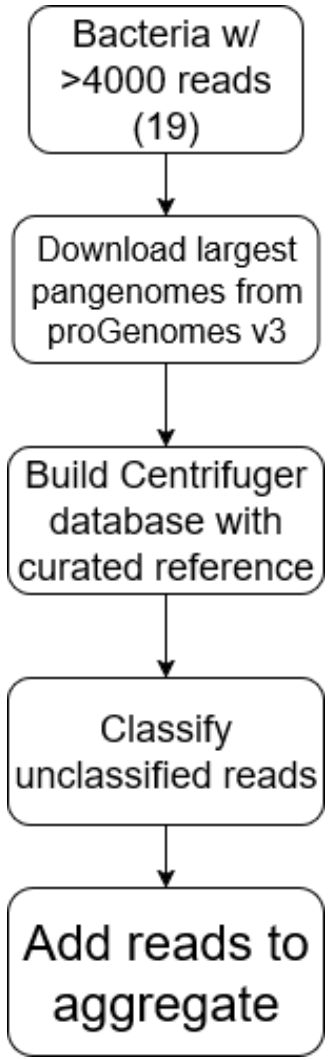


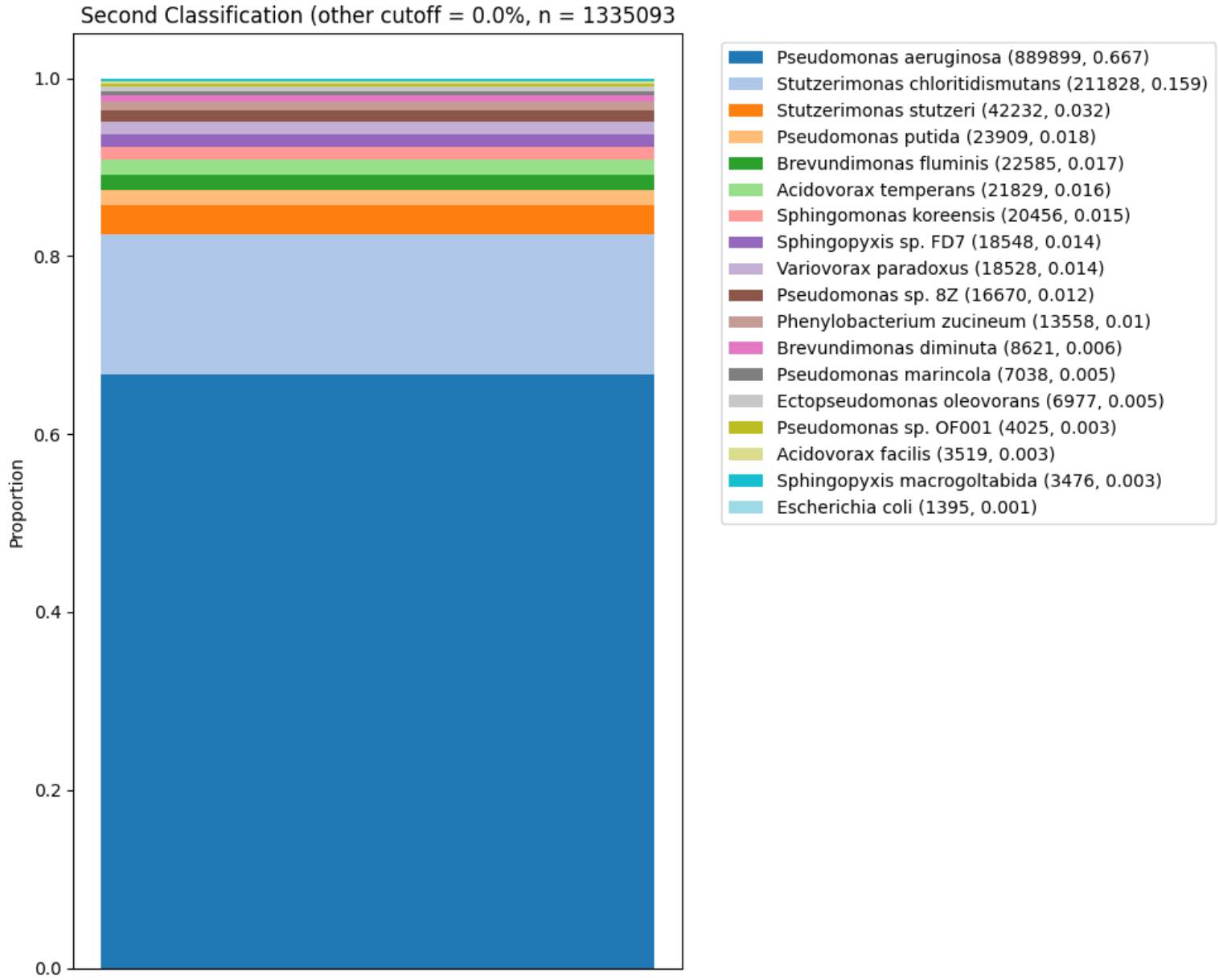
Aggregated classifier results

- Aggregated classifier information by combining single hits and unanimous hits
- 1095115/5885570 (18.6%) of raw reads classified



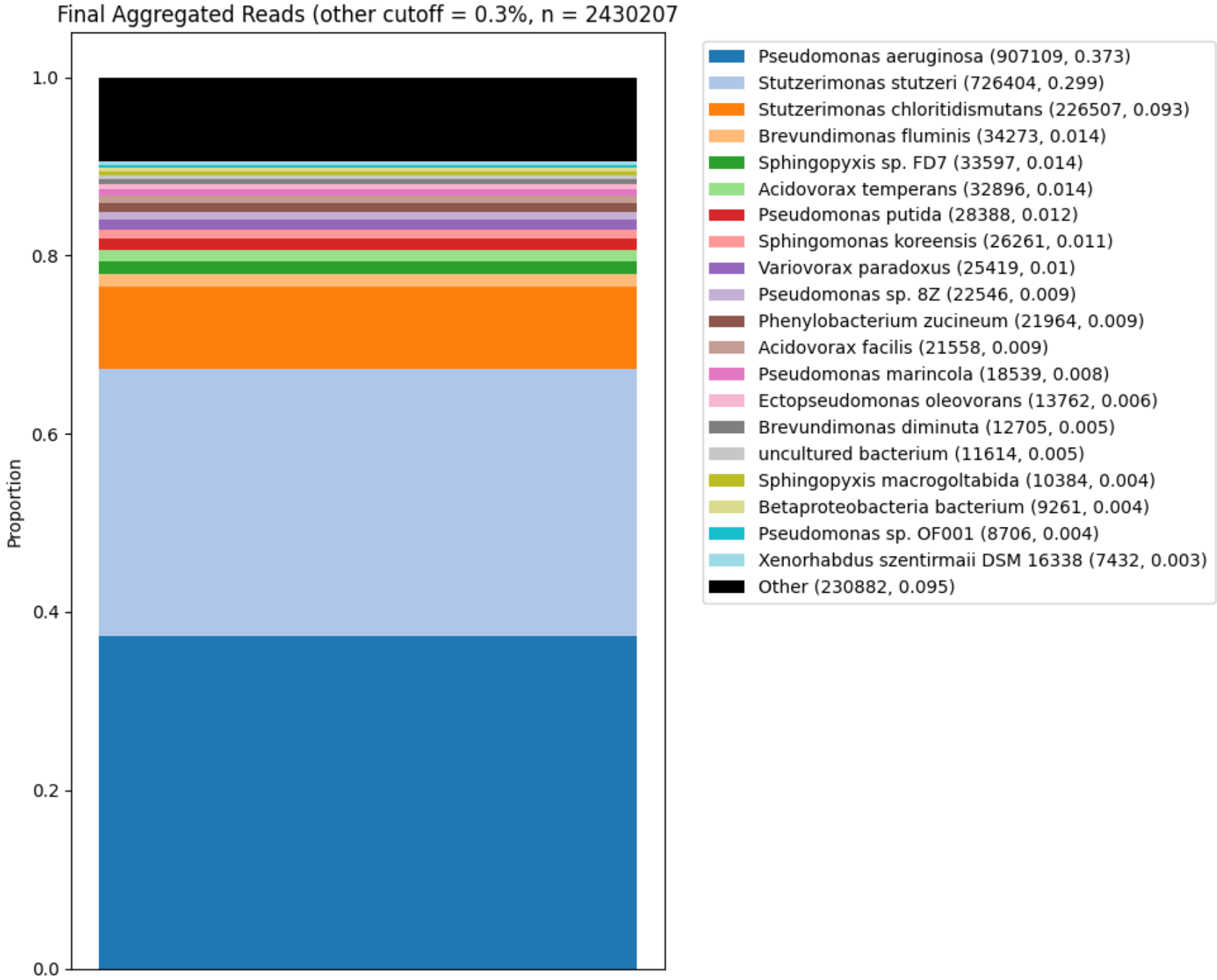
Second classification round



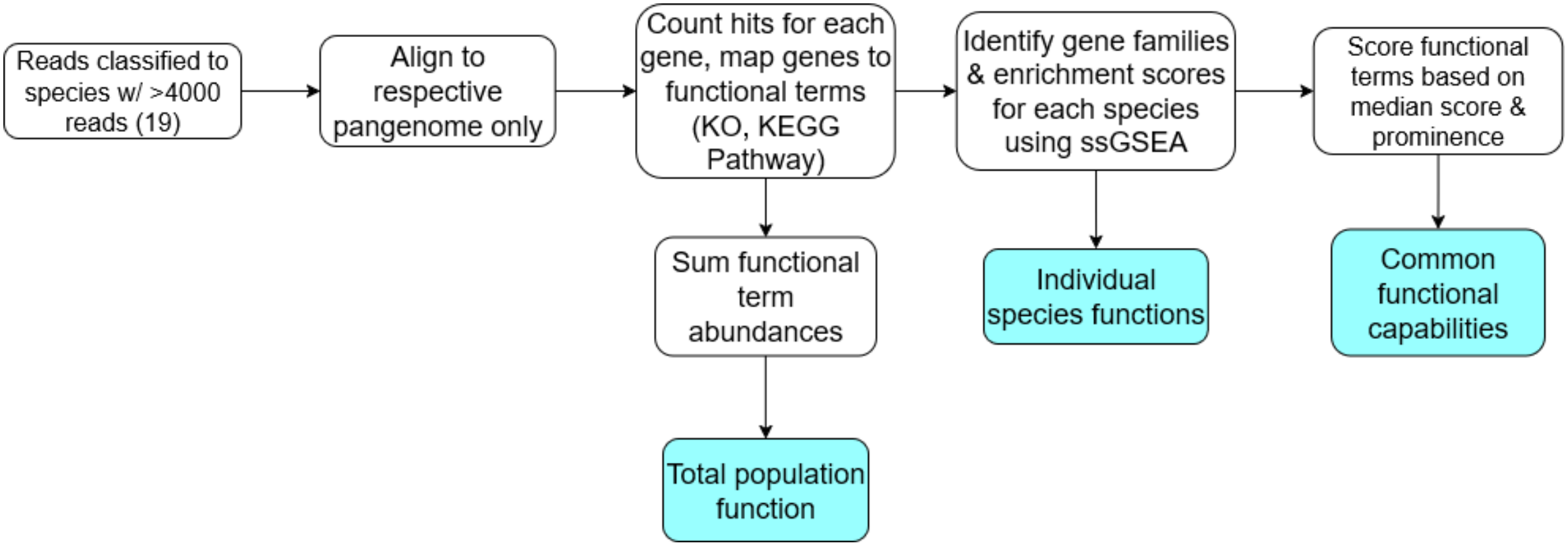


Final taxonomic results

- 2430207/5885570 (41.3%) of raw reads classified

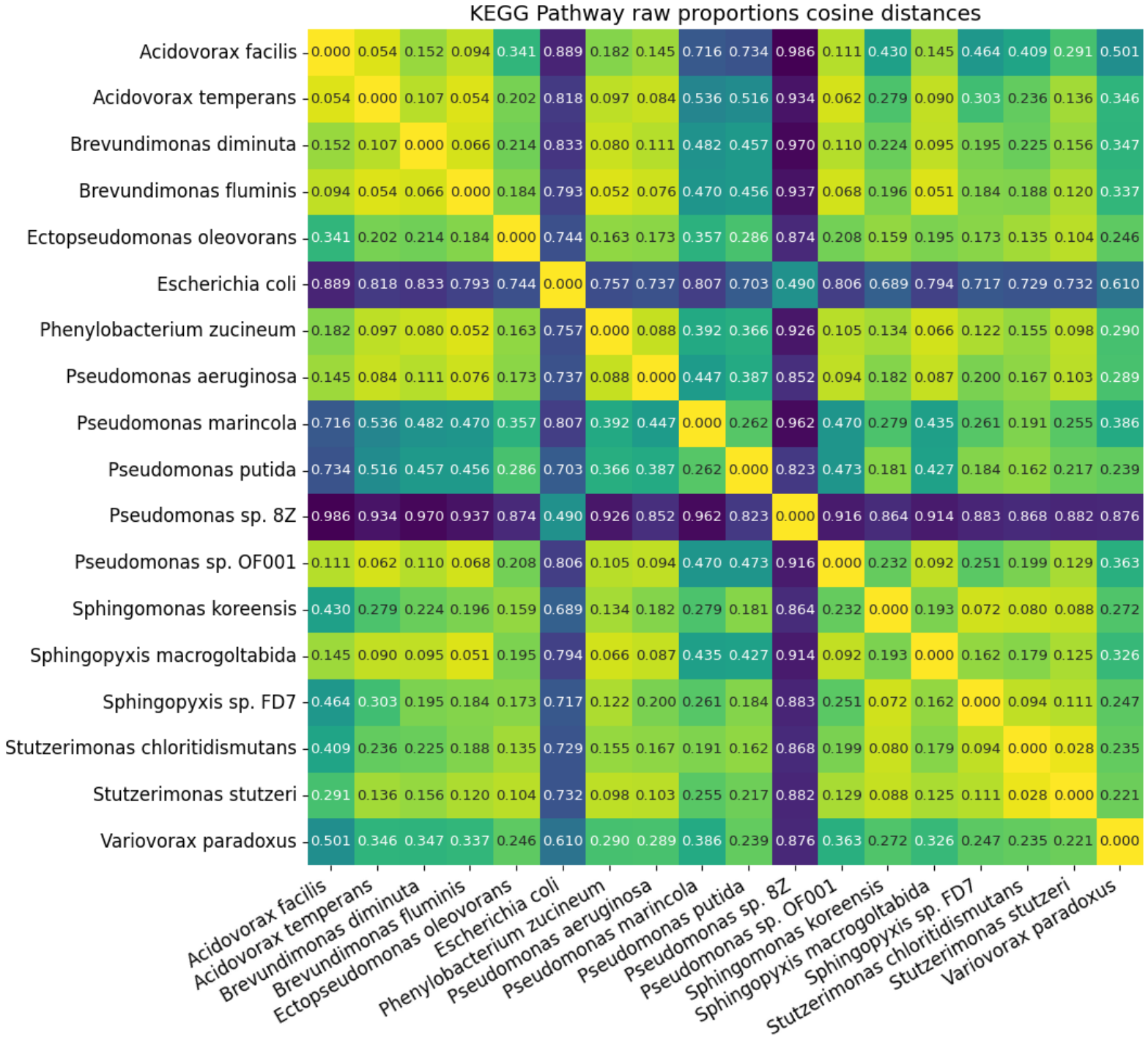


Part 2: Functional Analysis



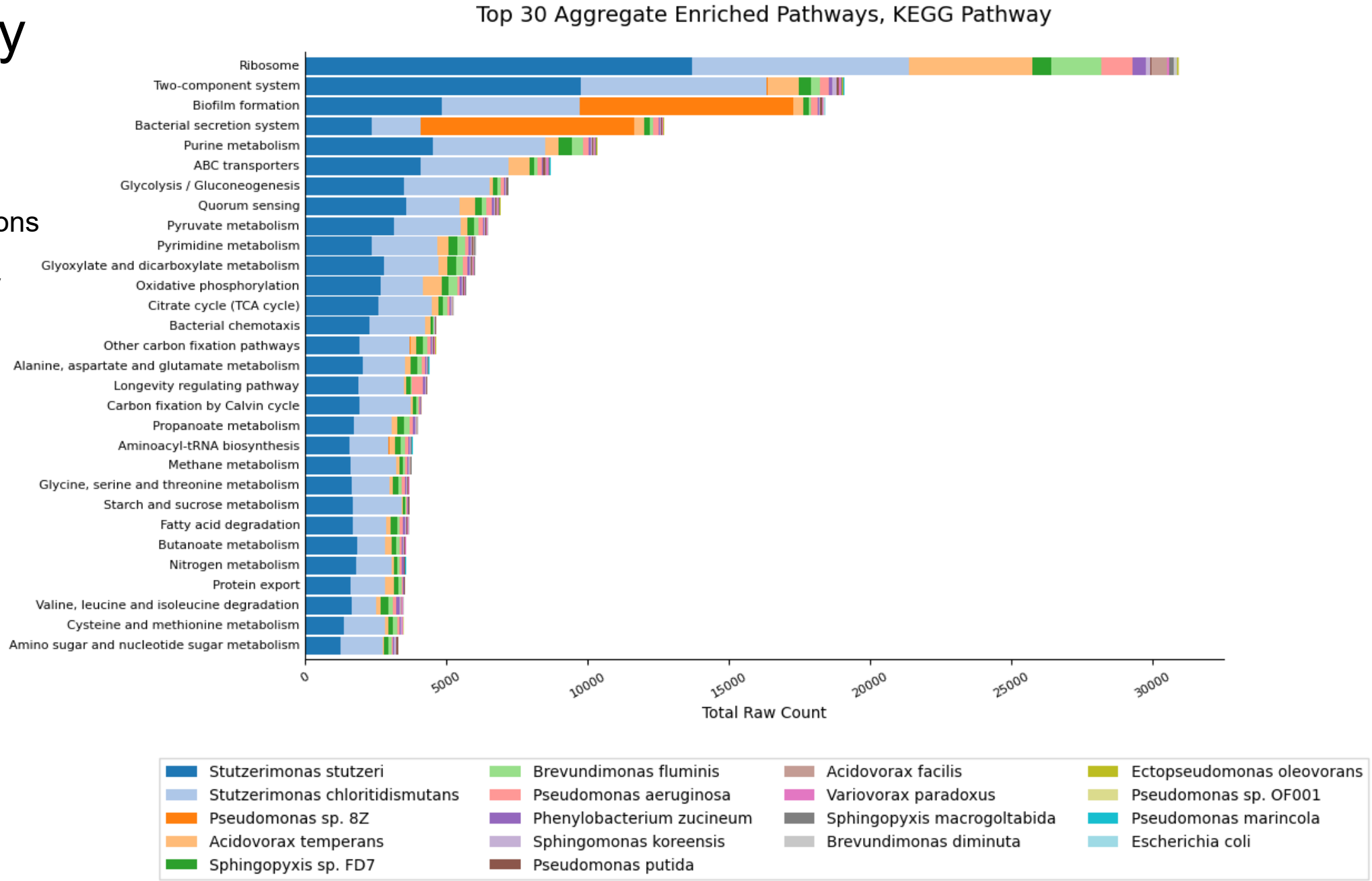
Functional Relatedness: KEGG Pathways

- Related species show similar functional roles
- *X. szentirmaii* reads returned zero functional terms; dropped from analysis



Aggregate community functions

- Essential functions populate top KEGG Pathway terms



Summary

- A multi-classifier approach is effective in characterizing environmental microbiome ribosome profiling data
- The microbial community translational activity is dominated by a few bacteria, particularly *S. stutzeri*, *P. aeruginosa*, and *S. chloriditismutans*
- Key community functions can be identified from ribosome profiling data

Limitations/Next Steps

- Ribo-seq is not suited for taxonomic abundance estimation
  - Integration with metagenomics data
- Ribo-seq only describes a portion of bacterial activity
  - Integration with metatranscriptomics data
- Many unmapped reads; only proceeded with taxa >4000 raw aggregate reads
- Pangenomes only partially describe a species
- Comparison is needed for significant conclusions
  - Differential analysis
  - Correlation with geochemical/environmental data

Questions

- Environmental sample results or expectations?
- Any specific bacterial processes of interest?