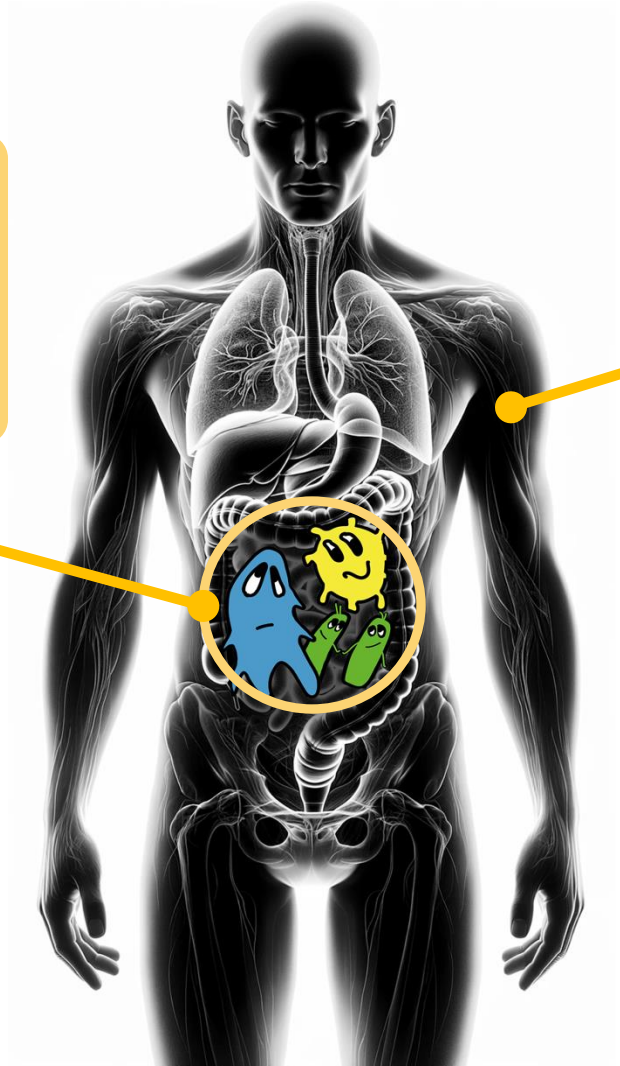
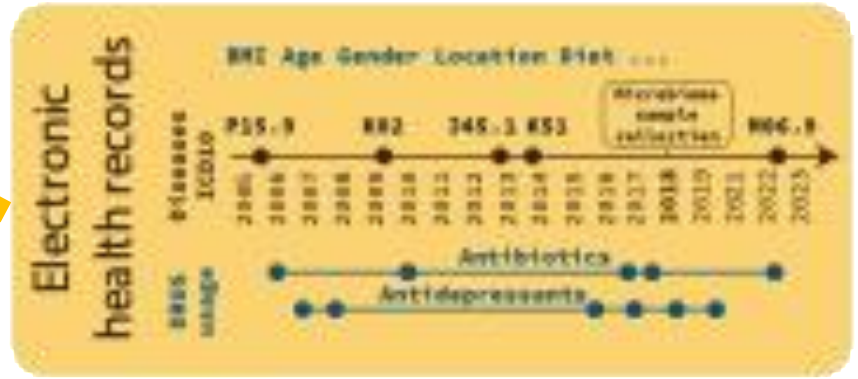
Metagenome-assembled genomes from a population-based cohort uncover novel gut species and strain diversity, revealing prevalent disease associations

Kateryna Pantiukh, Kertu Liis Krigul, Oliver Aasmets, Elin Org
Institute of Genomics, Estonian Genome Centre, University of Tartu, Tartu, Estonia

Estonian Microbiome Cohort

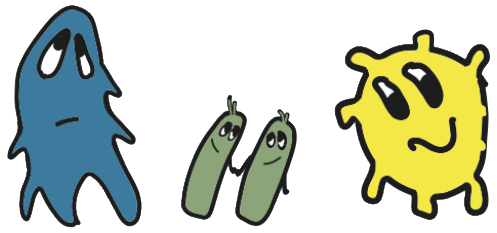
- 2,500 human gut microbiome samples
- 2nd TP for 332 samples

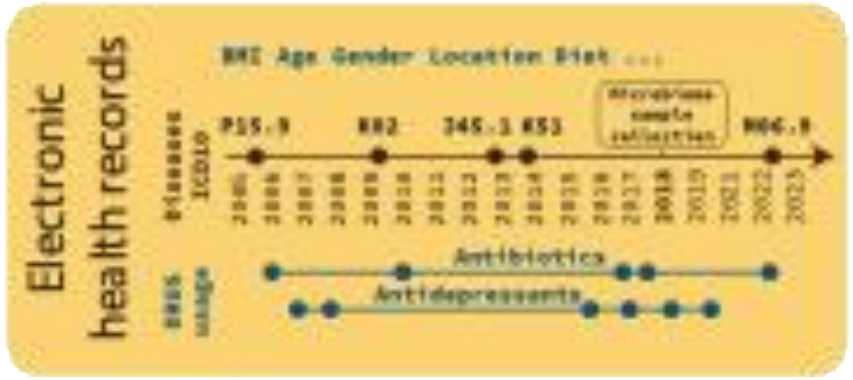


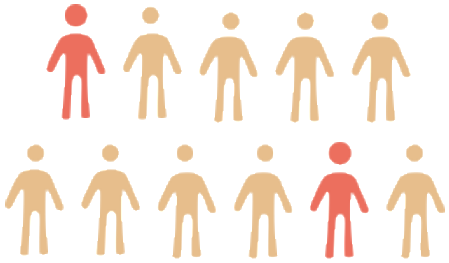


Estonian Microbiome Cohort

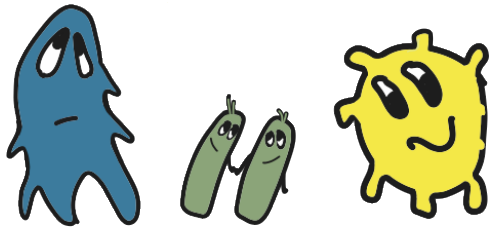
- 2,500 human gut microbiome samples
- 2nd TP for 332 samples



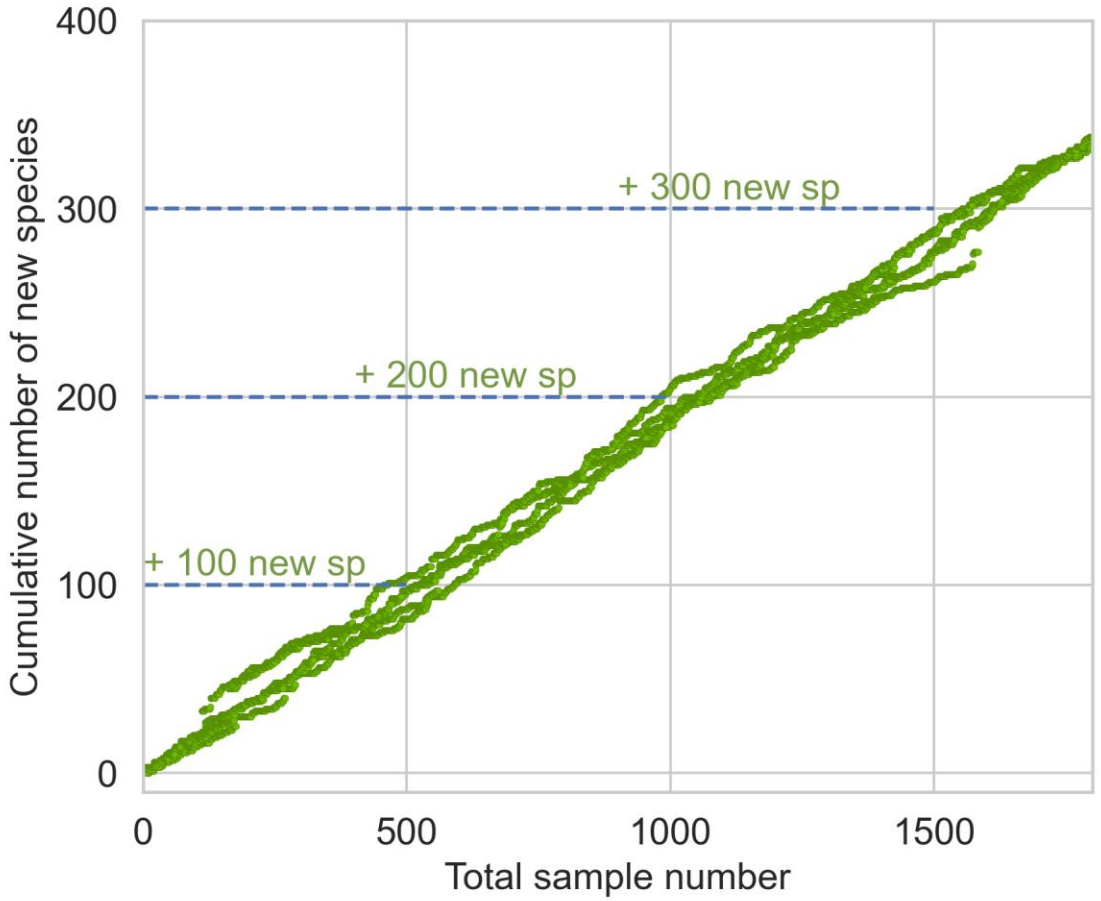




Estonian Microbiome Cohort



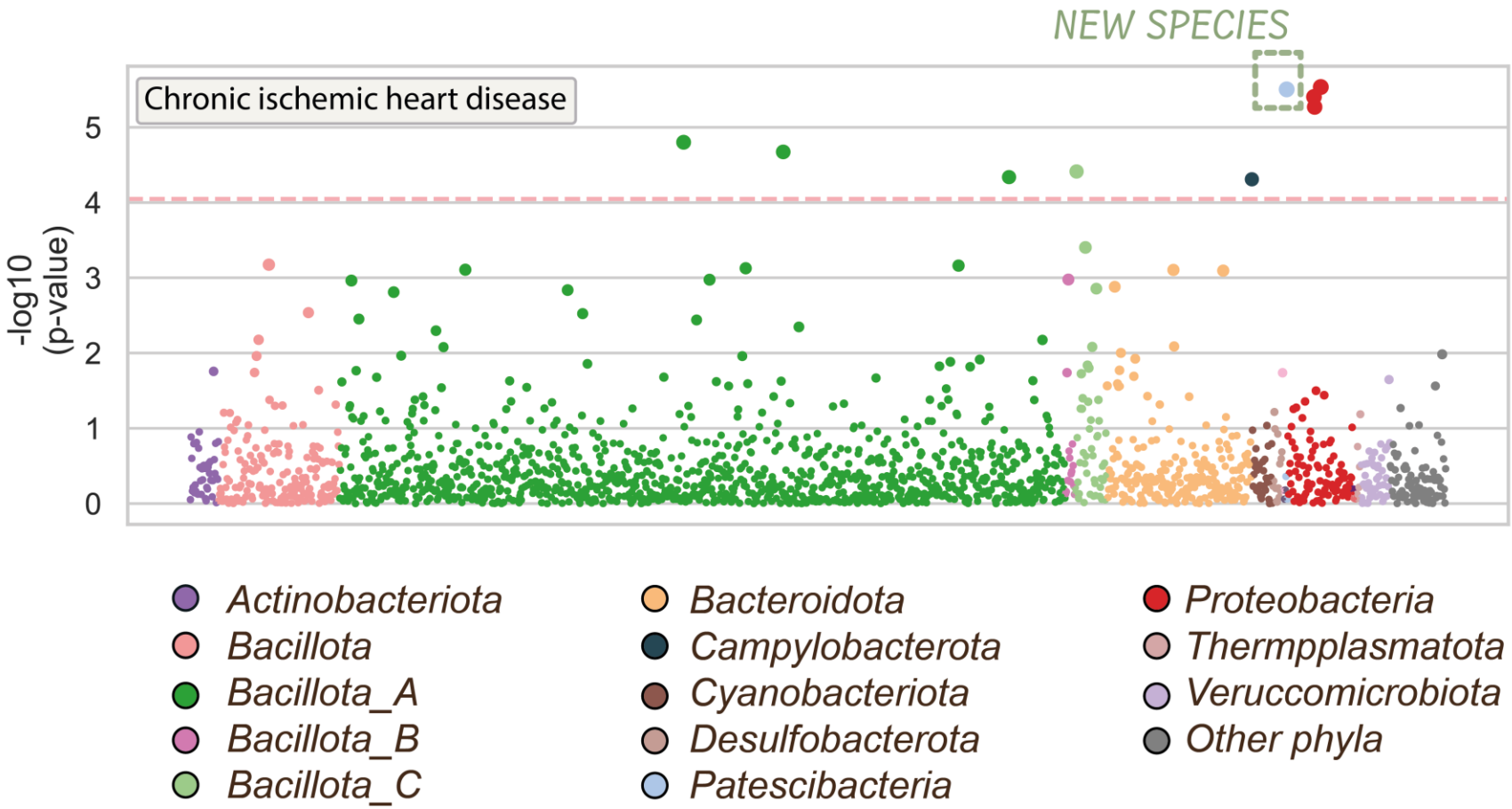
- Many gut bacteria are missing from databases
- Bacterial species form a continuum, not strict groups



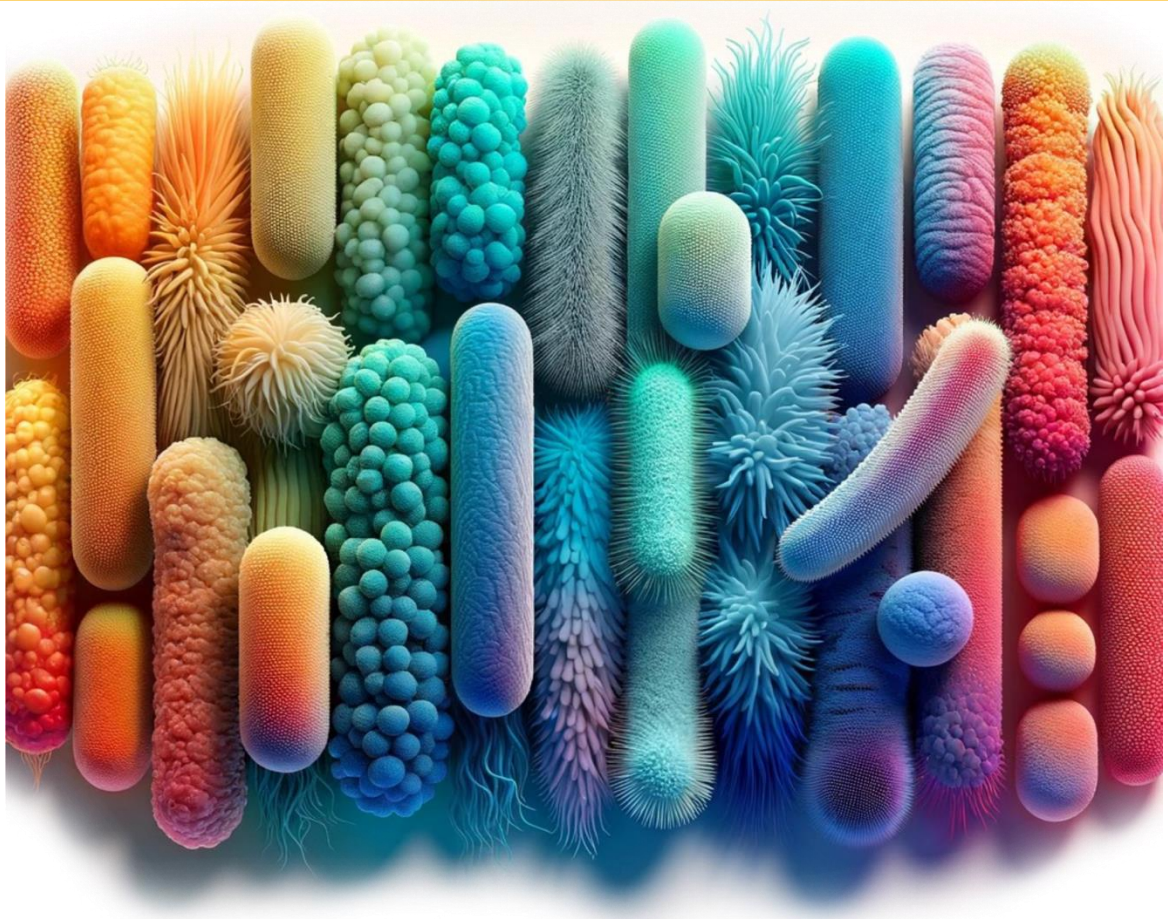
Association study on species level

- 84,254 MAGs
- 2,257 species
- 353 new species

- **33** diseases analysed
- Species-level associations identified in **25** diseases
- **9** diseases showed links to **newly** discovered species



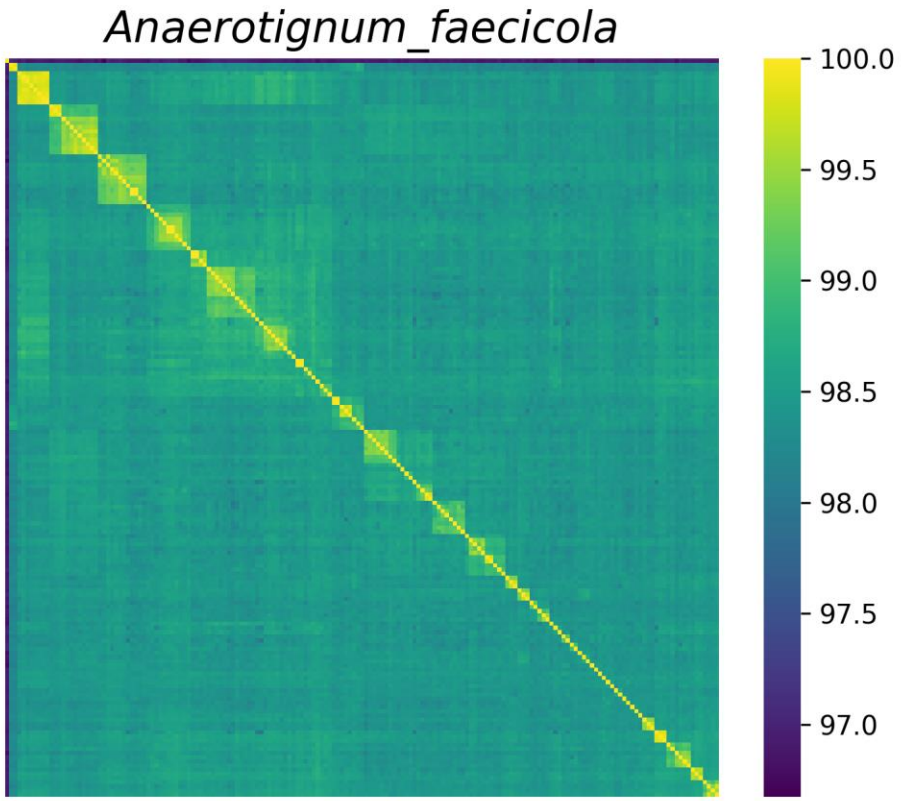
When species is just not enough

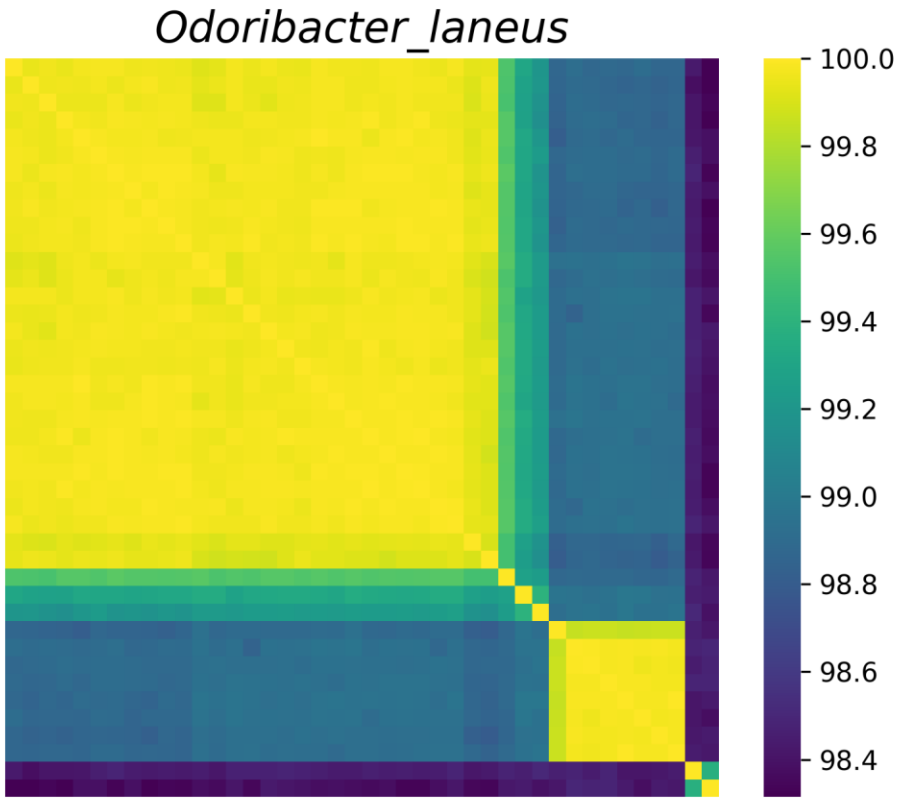


Species 1 *Species 2* *Species 3*

Bacteria with
genome similarity > 95%
are different species

When species is just not enough





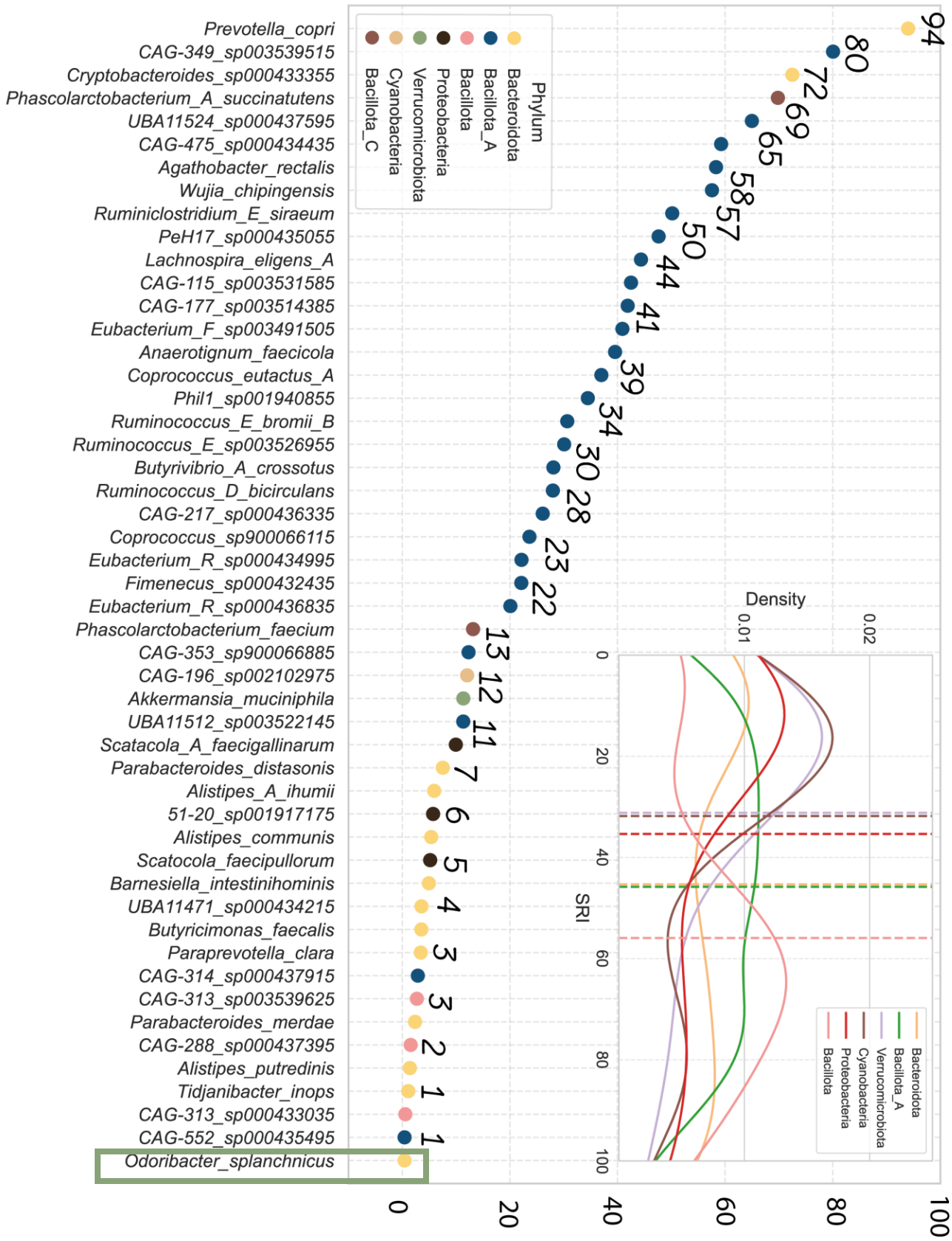
The Strain Richness Index - SRI

$$SRI = \frac{Number\ of\ strain\ cluster}{Total\ number\ of\ MAGs} * 100\%$$

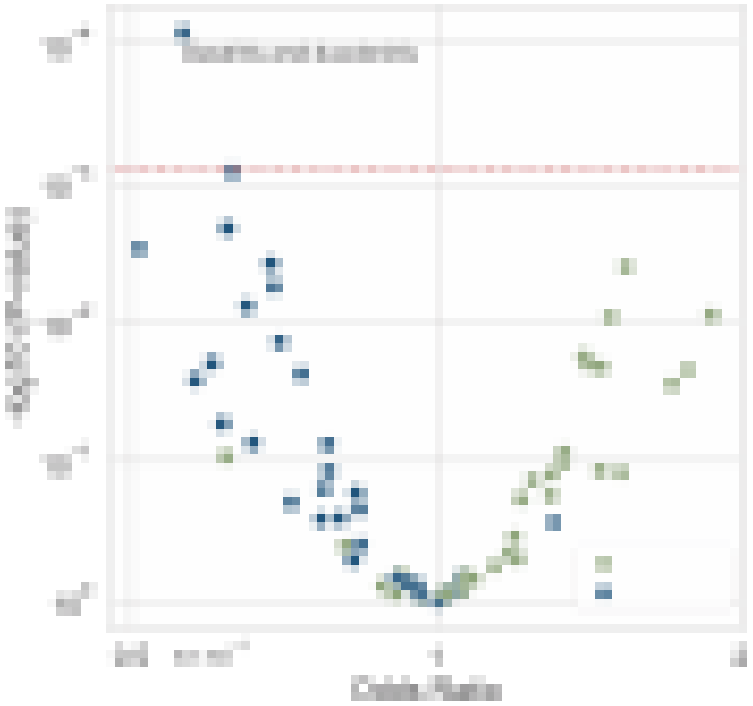
The Strain Richness Index (SRI) expresses the level of genetic diversity within a species by quantifying the number of distinct strains present in a given population or sample set.

SRI=94 means that per each 100 genomes of particular species we expected 94 different strains
-> Each strain cluster has 1-2 participant

Strain richness index (SRI)

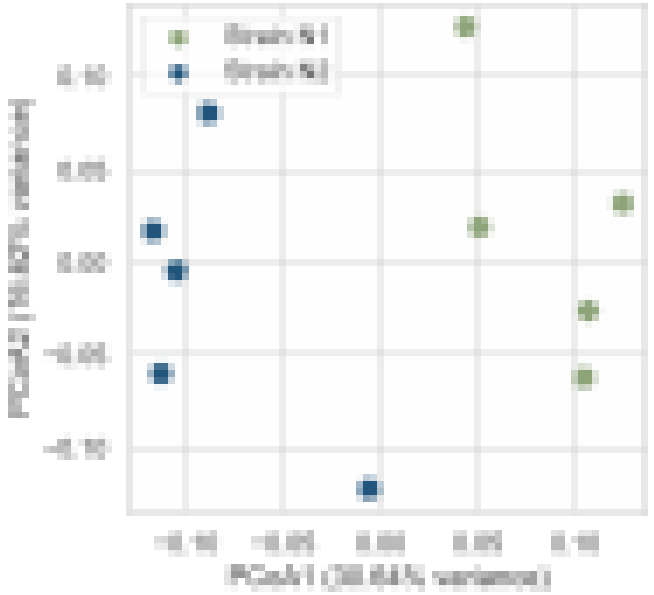


Strain level association study



Volcano plot showing the associations between the two major *O. splanchnicus* strains and 33 disease phenotypes. The red line indicates the Bonferroni-corrected significance threshold.

Pan-genome analysis of five representative MAGs from each *O. splanchnicus* strain (N1 and N2).



Principal coordinates analysis (PCoA) of *O. splanchnicus* representative MAGs based on predicted gene cluster presence/absence profiles.

Thoughts

- Not all species may be suitable for strain-level analysis
- Differences in strain richness need to be further explored using broader global microbiome data
- Why are some gut species highly diverse while others are not?

POSTER 20