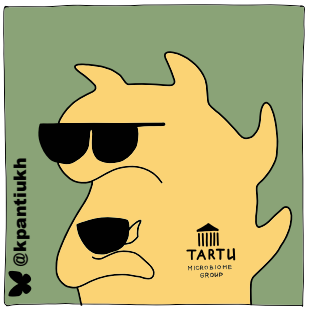


Biobank Cohorts as a Platform for Gut Microbiome Research

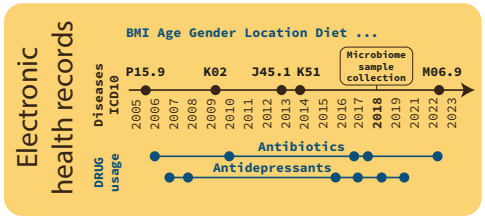
Kateryna Pantiukh, Kertu Liis Krigul, Oliver Aasmets, Elin Org
Institure of Genomics, University of Tartu, Tartu Microbiome

INTRODUCTION

ESTONIAN BIOBANK

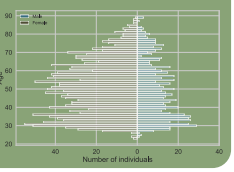
211,187
biobank participants
~20% of the Estonian population

211,187 genotyped
3,000 sequenced
2509 microbiome
202,000 NMR



DATA ESTONIAN MICROBIOME COHORT

- 2,509 individuals sampled
- Longitudinal data for 332 individuals
- 56 million paired-end reads per sample
- Participants from diverse regions across Estonia



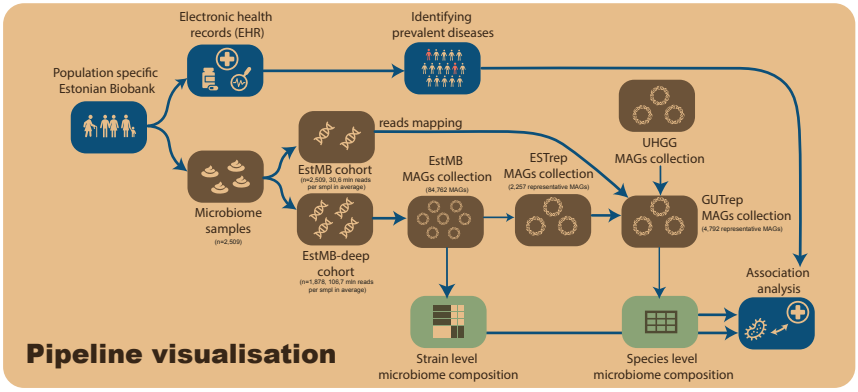
METHODS

Sequencing: shotgun paired-end 150 bp reads with deep coverage (56M paired reads/sample on average).

MAGs Assembly: Multi-binning metagenome assembly (VAMB, MaxBin, MetaBAT), refined with DAS-Tools. MAG quality was assessed using CheckM2, GUNC, and SeqKit. MAGs were cluster with dRep. Taxonomic annotation was performed with GTDB-Tk. Abundance & prevalence estimated with CoverM.

MWAS: Association analysis was performed using a linear model between CLR-transformed relative abundances of species (prevalence >1%) and disease, or metadata variables. Significance levels were adjusted using Bonferroni correction.

Pan genome: Anvio pangenome pipeline.



RESULTS

- A population-specific MAG database was constructed for the Estonian population, comprising 86,254 metagenome-assembled genomes (MAGs) representing 2,257 gut bacterial species.
- The analysis showed that prior MAG assembly remains essential for capturing novel species in population-specific microbiome studies - the number of new species continued to rise with increasing sample size, showing no sign of plateau (Figure 1).
- Gut microbiome composition in the Estonian population was estimated, followed by a microbiome-wide association study (MWAS) across 33 common diseases, revealing multiple associations, including links to newly reconstructed species. (Figure 2).
- Analysis of strain structure across human gut species revealed substantial variation in genetic diversity. To quantify this, the Strain Richness Index (SRI) was developed and used to estimate strain-level diversity among 376 gut species (Figure 3).
- Using SRI, species were selected for strain-level association studies, demonstrating that such analyses uncover novel phenotype associations not detectable at the species level (Figure 4-6).

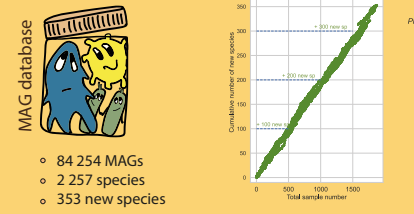


FIGURE 1. The relationship between the number of samples analyzed and the cumulative number of new species identified.

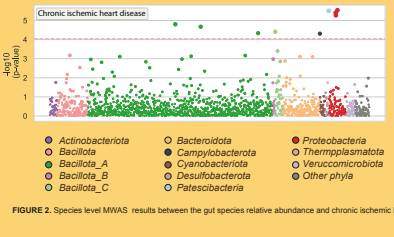


FIGURE 2. Species level MWAS results between the gut species relative abundance and chronic ischemic heart disease.

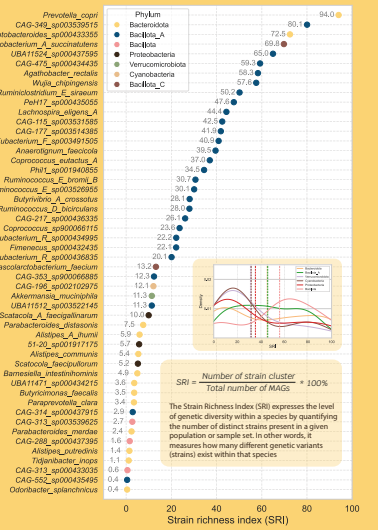


FIGURE 3. Strain richness index (SRI) for the top 50 species with the highest number of metagenome-assembled genomes.

PROJECT GOAL

To demonstrate the use of a biobank cohort platform for in-depth microbiome research, aiming to uncover health-related microbiome associations and insights into microbial community structure.

CONCLUSIONS

We expanded the human gut bacterial reference by adding new MAGs, providing a scalable framework for cohort-based microbiome wide association studies (MWAS). Using this framework, we identified associations with 25 common diseases and introduced an SRI metric to evaluate within species genetic diversity. As a proof of concept, we conducted a strain-level MWAS that revealed a novel disease association undetectable at the species level.

PREPRINT

FIGURE 4. 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.

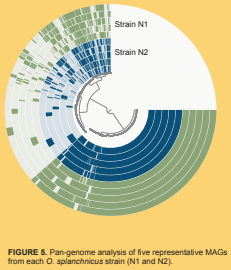


FIGURE 5. Pan-genome analysis of five representative MAGs from each *O. splanchnicus* strain (N1 and N2).

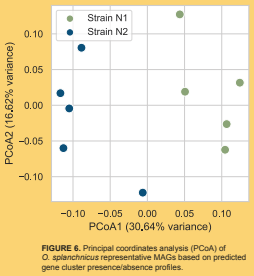


FIGURE 6. Principal coordinates analysis (PCoA) of *O. splanchnicus* representative MAGs based on predicted gene cluster presence/absence profiles.

This project was supported by:

European Union European Regional Development Fund
Investing in your future
EMBO
Estonian Research Council

NEXT steps (unpublished data):

In-depth analysis of species within the community, focusing on differing behaviors linked to distinct genomic features.

FIGURE 7. Human gut microbiome richness in the Estonian cohort

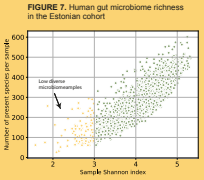


FIGURE 8. PCoA of Beta Diversity for samples with low diverse microbiome (Shannon <3.0)

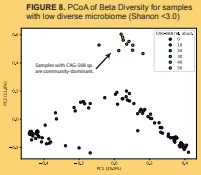
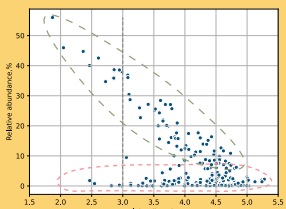


FIGURE 9. Relationship between the Relative Abundance of CAG-568 Genus Species and Sample Microbiome Richness Reveals Two Distinct Behavioral Groups.



CONTACTS