

# Building Estonia's Gut Bacteria Genome Database

Kateryna Pantiukh

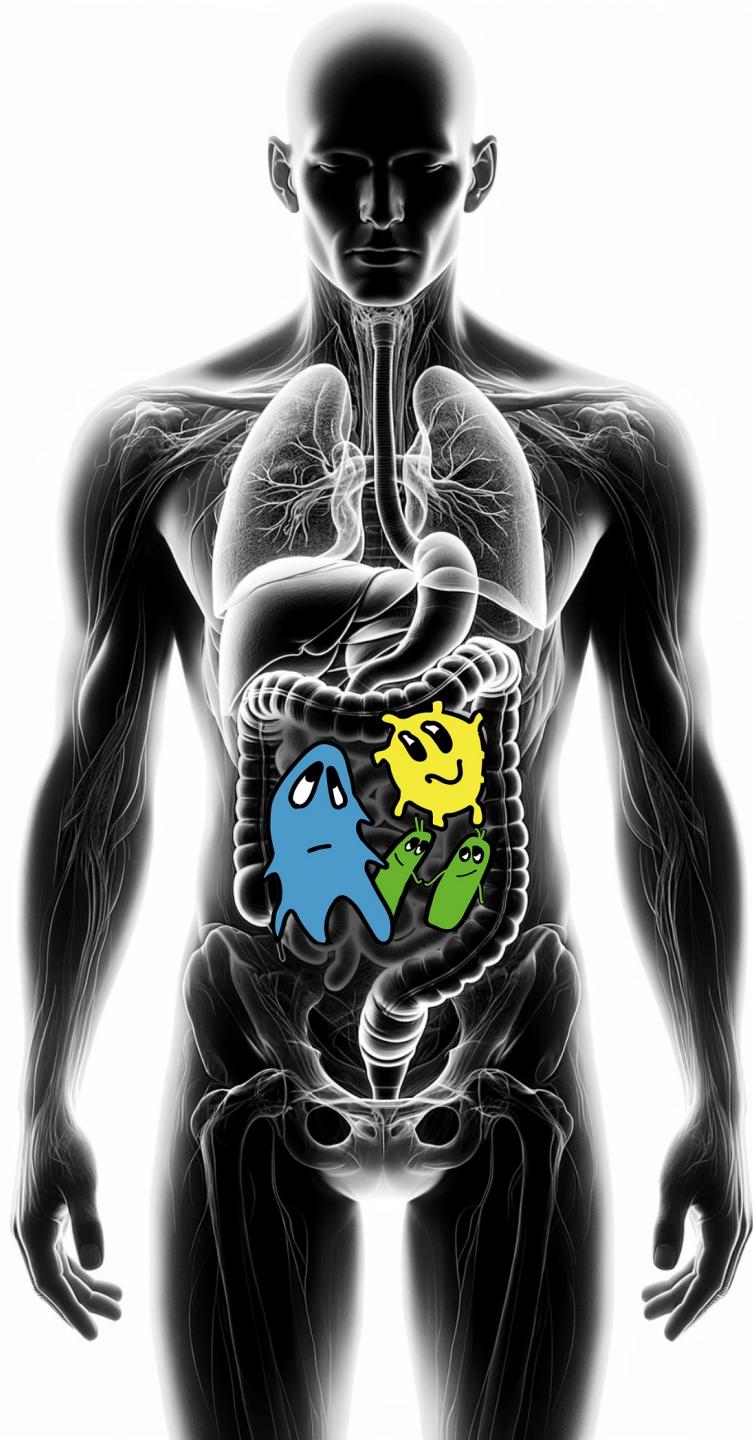


# Gut Bacteria MetaGenome Database

Almost complete genomes of  
microorganisms reconstructed  
from sequencing data from  
stool samples

**MAG** *Metagenome Assembled Genome*





# Microbiome

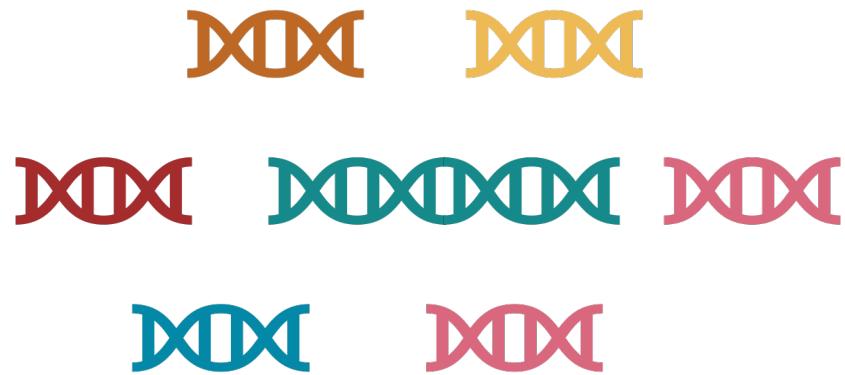
*Community of microorganisms*



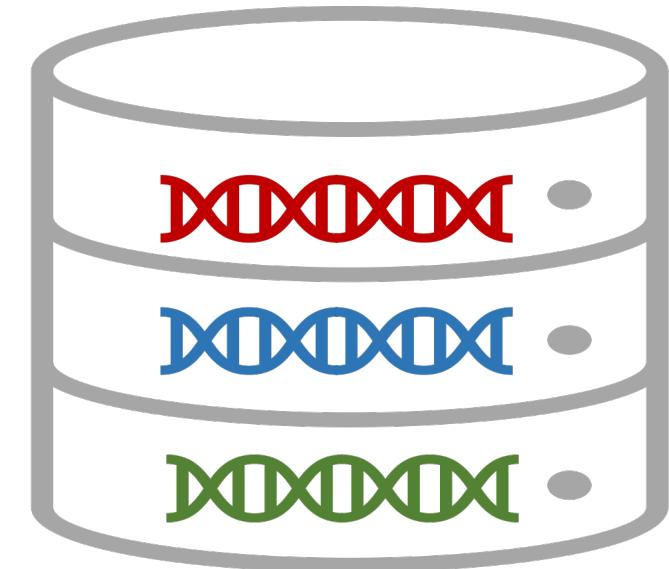
# Metagenome

*pool of DNA*

# Metagenome profiling

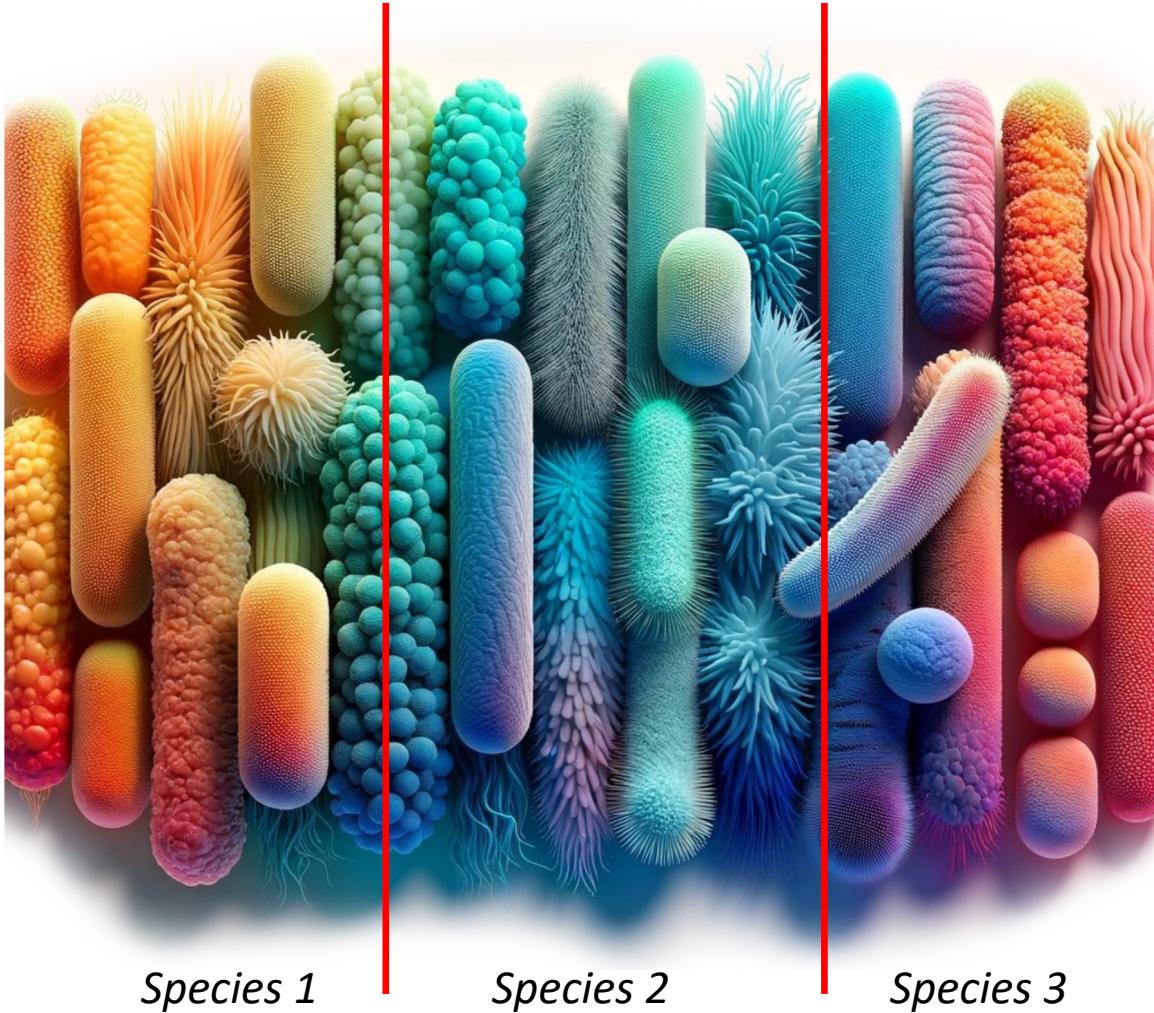


*Metagenome*



*Public DB  
of bacterial genomes*

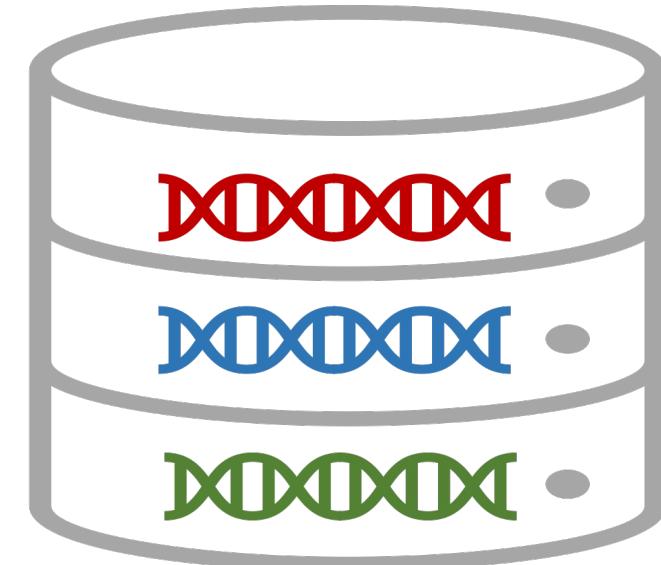
# The Problem. Where is a species border?



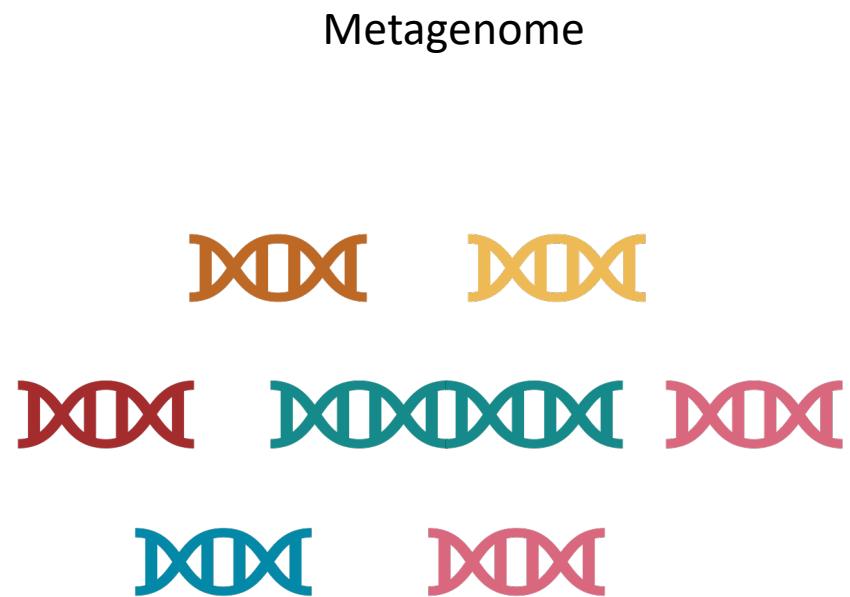
Bacteria with  
genome similarity  $> 95\%$   
are different species

# Metagenome profiling

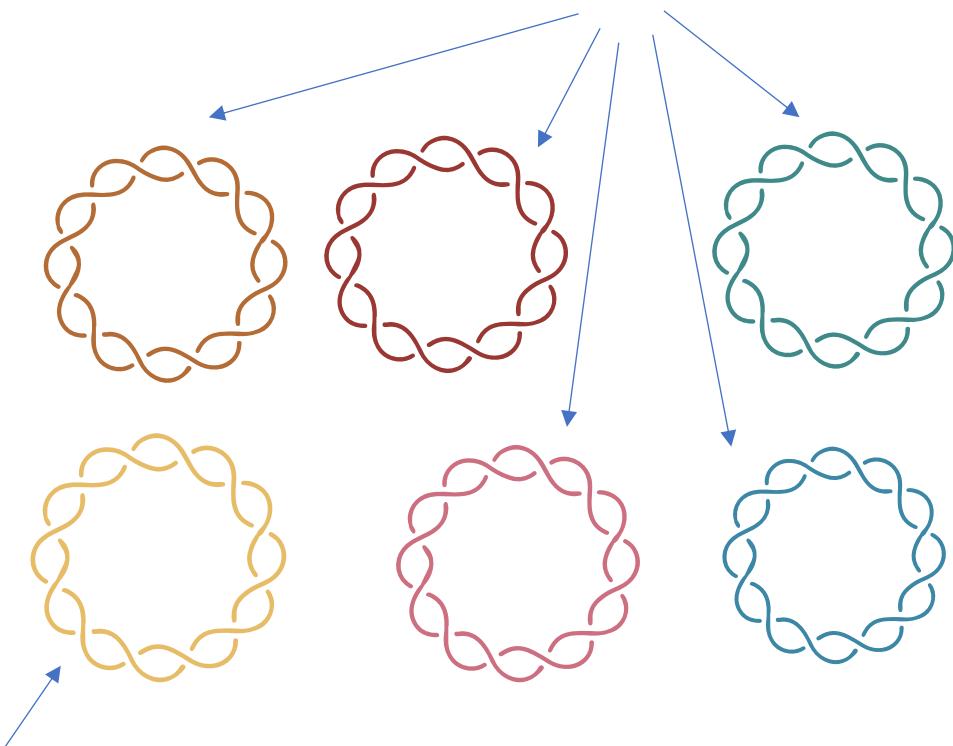
New ???



# Metagenome assembly



Bacteria carry specific & unique genome features



New Bacteria Species  
(Estonian population)

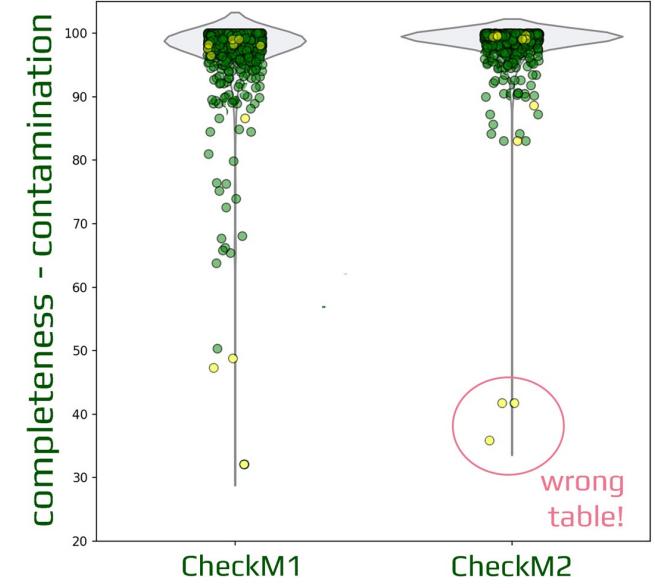
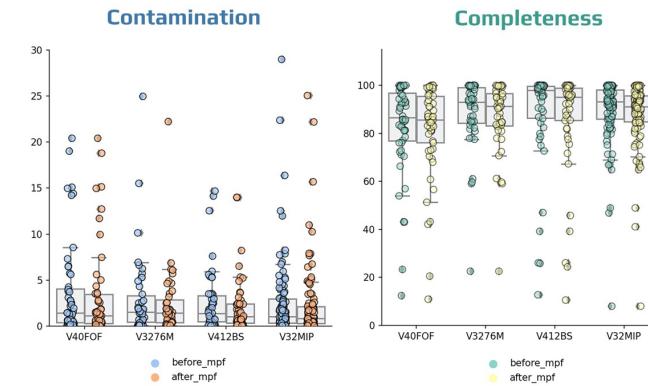
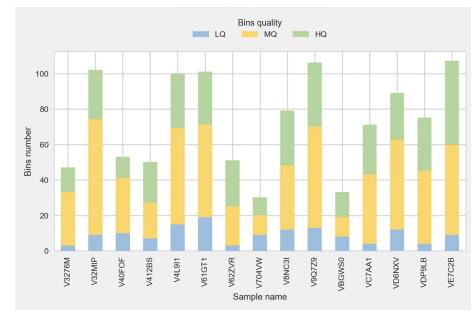
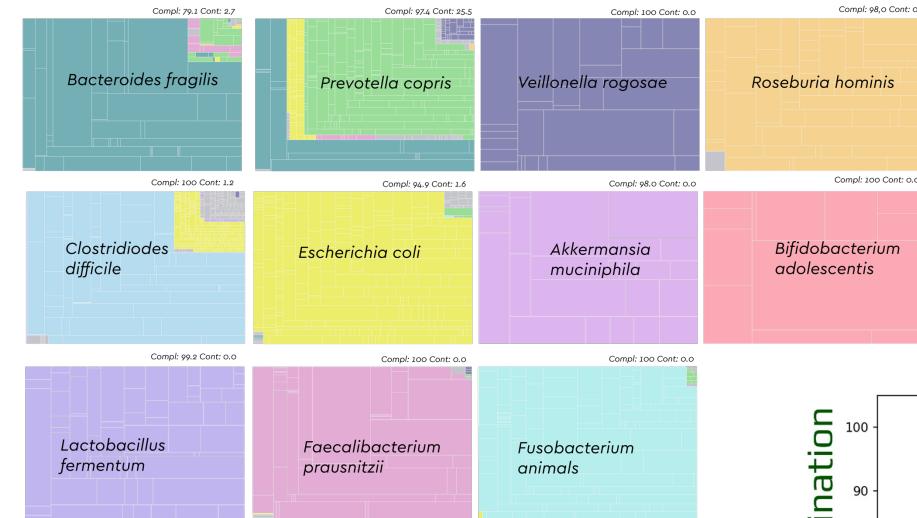
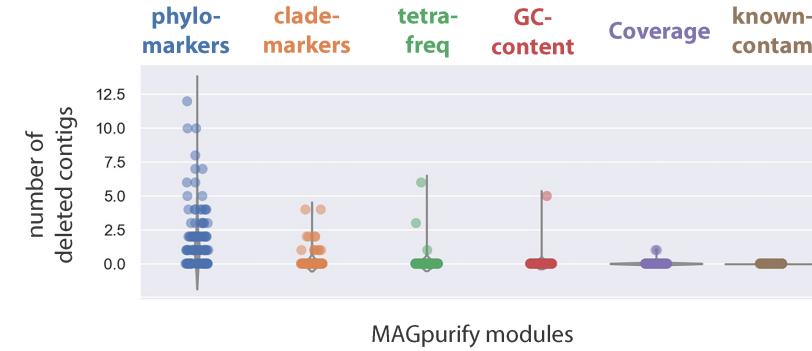
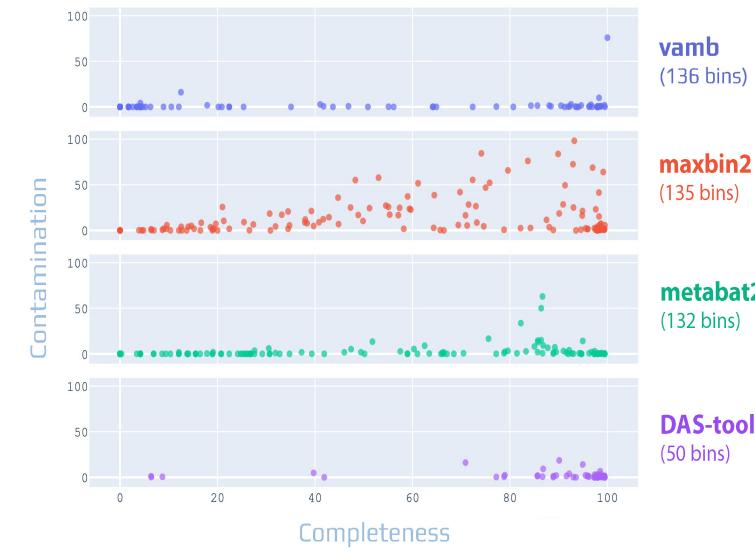
# Gut Bacteria Genome Database

Almost complete genomes of  
microorganisms reconstructed  
from sequencing data from  
stool samples

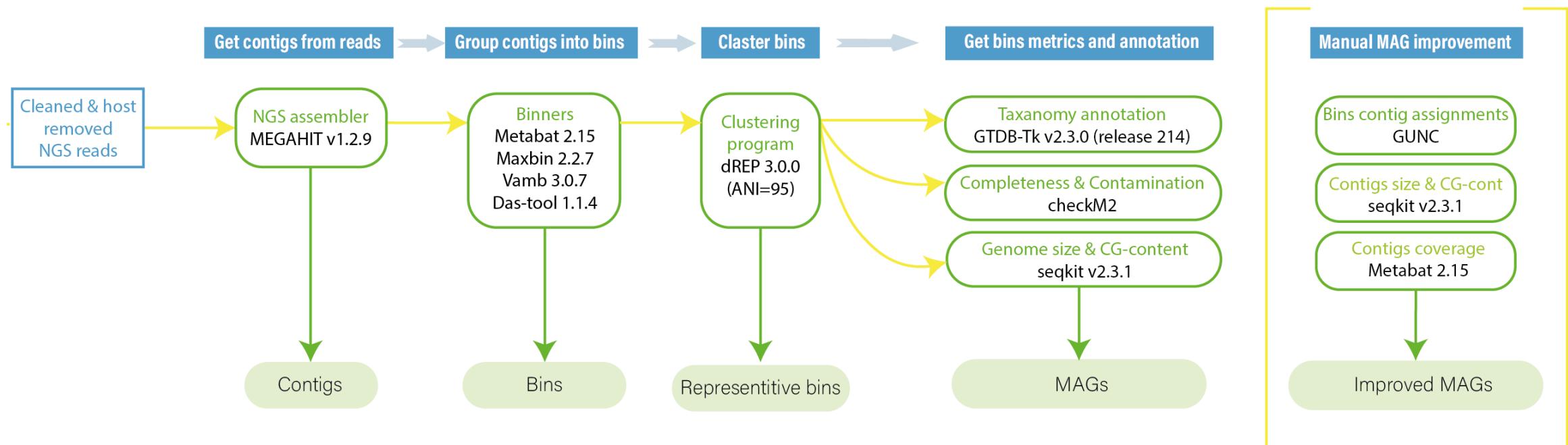
**MAG** *Metagenome Assembled Genome*



# MAGs extraction pipeline tests

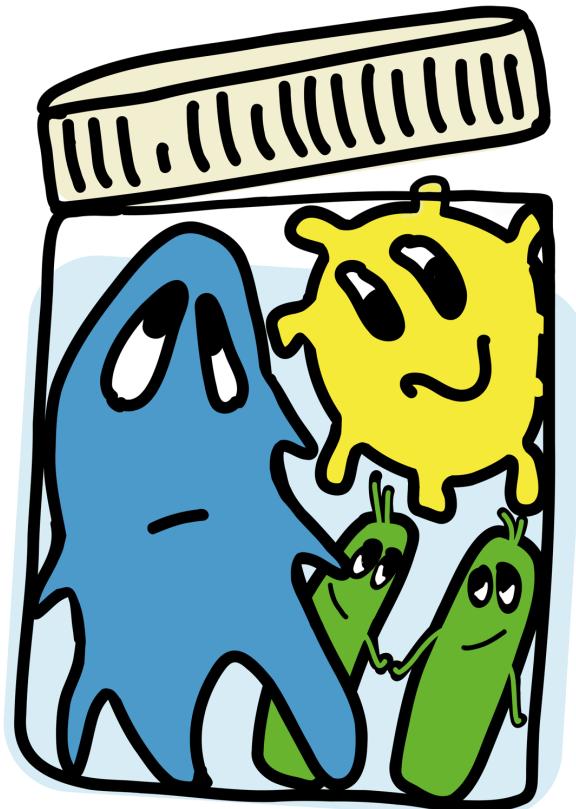


# Final Pipeline



# Metagenome-Assembled Genome database

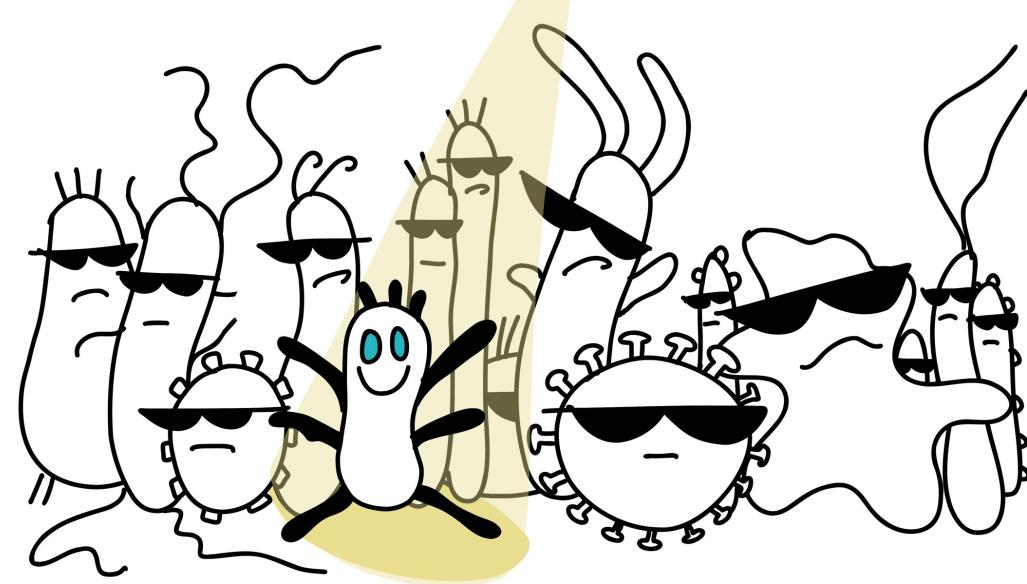
1881 human gut samples



- + **84 793 MAGs**
- + **2 257 species-level clusters** with **references (represenitives) genomes** for each cluster
- + Species-level **Abundancy table**  
**Strain-level Abundancy tables**

# Known species New species

From 2,257 representatives genomes

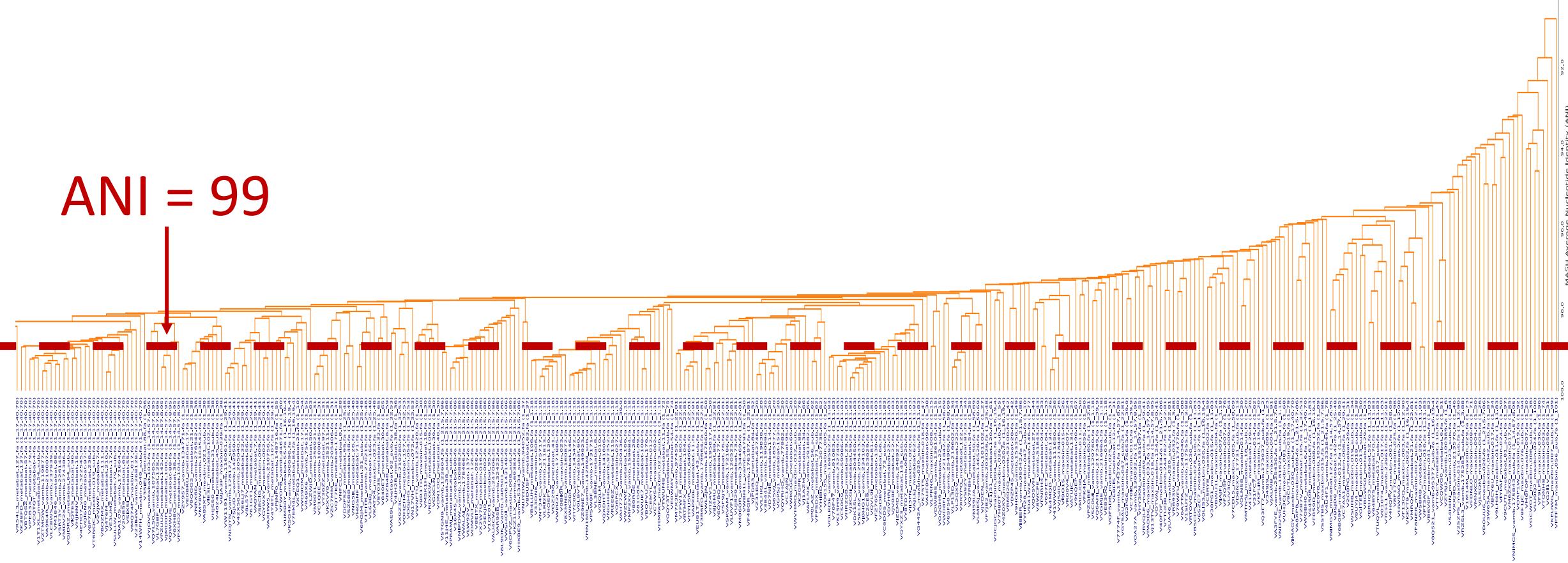


- 1 belonged to a potentially new FAMILY
- 25 belonged to a potentially new GENUS
- 333 belonged to a potentially new SPECIES

~ 16% are NEW

# New strains *Bacteroides uniformis* based on ANI index

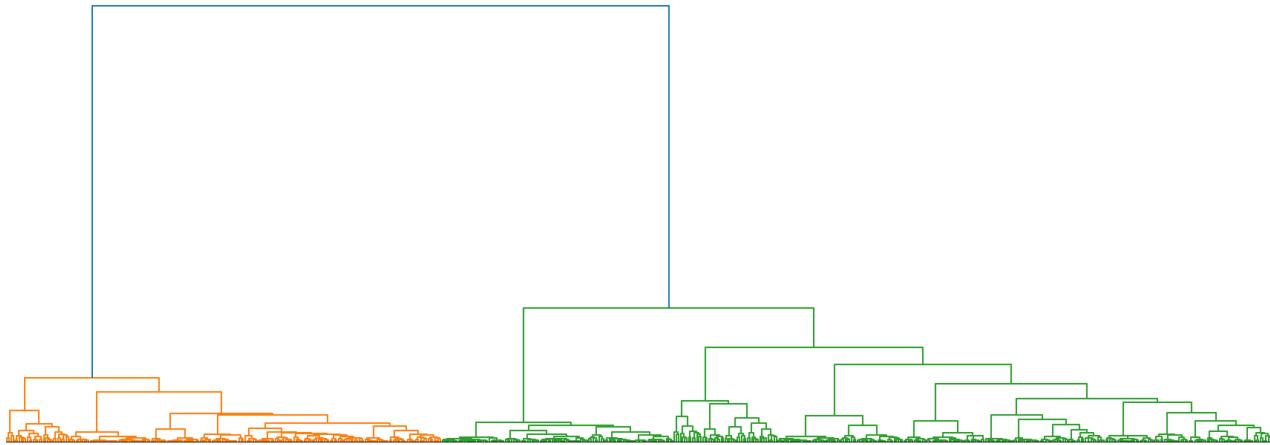
ANI = 99



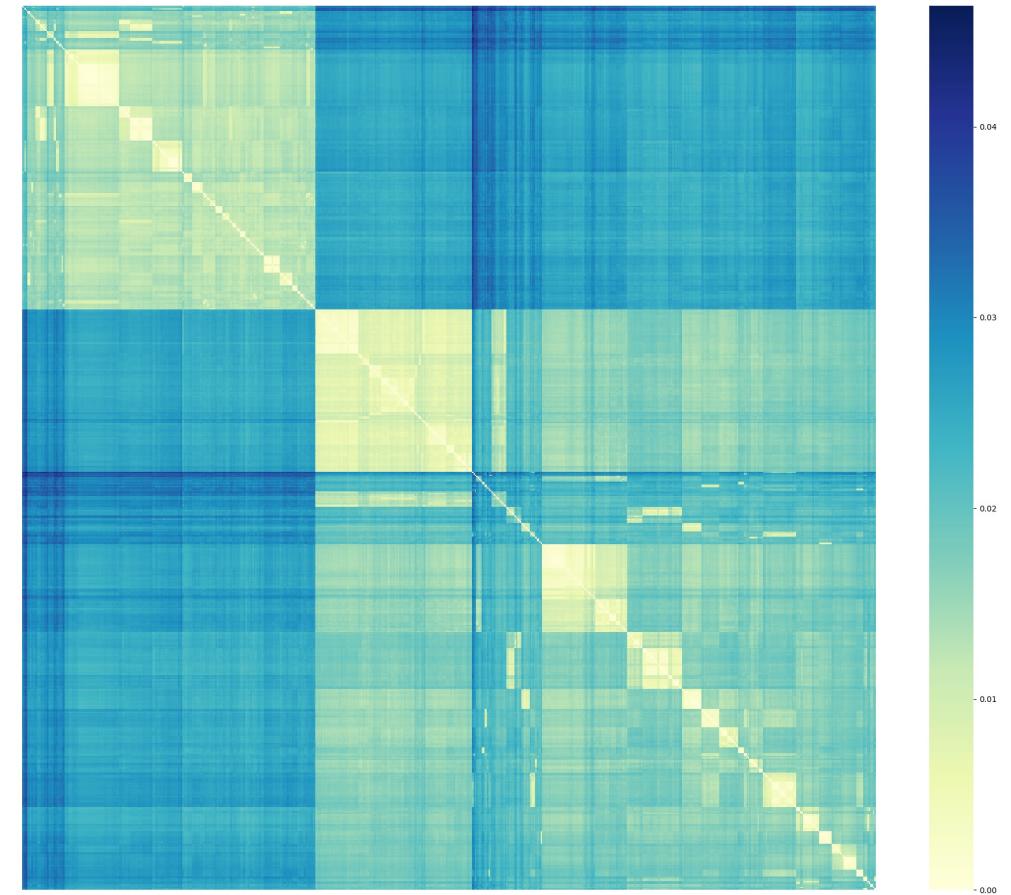
1 species → 357 genomes (MAGs) → 81 strain clusters

# New strains *Akkermansia muciniphila*

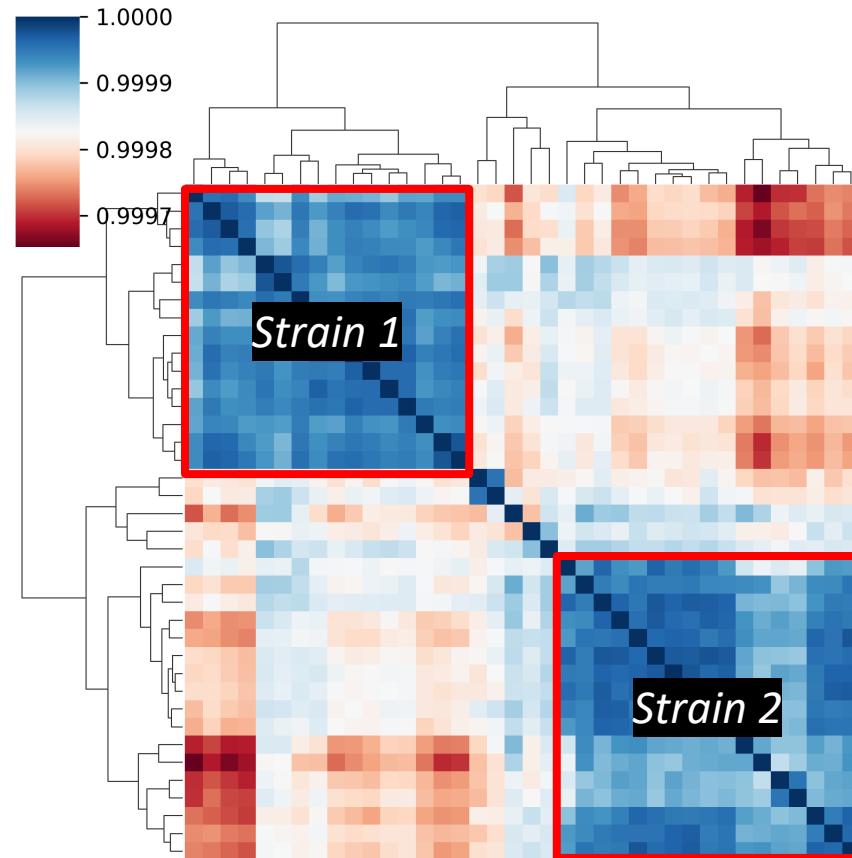
623 bins  $\longrightarrow$  2 strains



Mash-distance heatmap



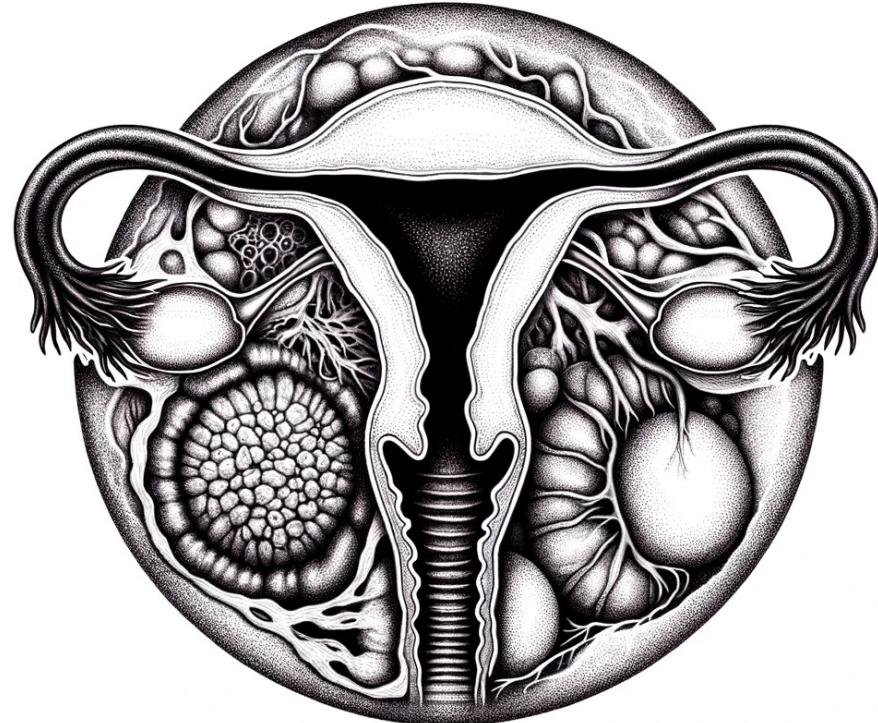
# New strains *UBA-71* *based on functional annotation*





# Find new correlation

\* at collaboration with Granada University



**Endometriosis phenotype**  
(136 cases, 864 controls)

vkood	H0001_Odoribacter_splanchnicus.fa	H0002_Barnesiella_intestinihominis.fa
V00H9L	1	1
V010T4	1	1
V03HFT	1	1
V03ZU2	0	1
V0404P	1	1
...	...	...

**Species level Abundance table**  
(1293 bacteria after filter out  
bacteria presented at <20 samples)

# Find new correlation

\* at collaboration with Granada University

<i>H0008_Ruminococcus_D_bicirculans</i>
<i>H0045_51-20_sp001917175</i>
<i>H0074_Coprobacter_fastidiosus</i>
<i>H0148_Holdemanella_biformis</i>
<i>H0241_UBA10281_sp900767815</i>
<i>H0273_CAG-603_sp900066105</i>
<i>H0286_Mesosutterella_multiformis</i>
<i>H0318_Scatenecus_faecavium</i>
<i>H0332_Flemingibacterium_sp900546315</i>
<i>H0473_Alistipes_A_pullcola</i>
<i>H0547_QALR01_sp944381075</i>

21 new significant correlation\*  
Included 11 species not presented at public DB

\*Bonferroni corrected p value>0.01

<i>H0653_Aphodousia_sp900544345</i>
<i>H1471_Gallalistipes_aquisgranensis</i>
<i>H1483_Butyricicoccus_A_intestinisimiae</i>
<i>H1592_HGM12789_sp900766585</i>
<i>H1652_Acutalibacteraceae_undG</i>
<i>H1683_Ruminococcus_D_sp900539095</i>
<i>H1755_UBA737_sp905193585</i>
<i>H1820_RGIG3732_sp022768945</i>
<i>H2245_Blautia_A_caecimuris</i>
<i>H2249_V9D3004_sp900760345</i>

# This is the Way

Develop a  
pipeline



Find new  
correlations

Create  
MAGs DB

Explore  
genomic features  
behind correlations

