

Катерина Пантюх

Tartu microbiome group



Ukrainian bioinformatician
Born in Zaporizhzhia



estonian genome center
university of tartu

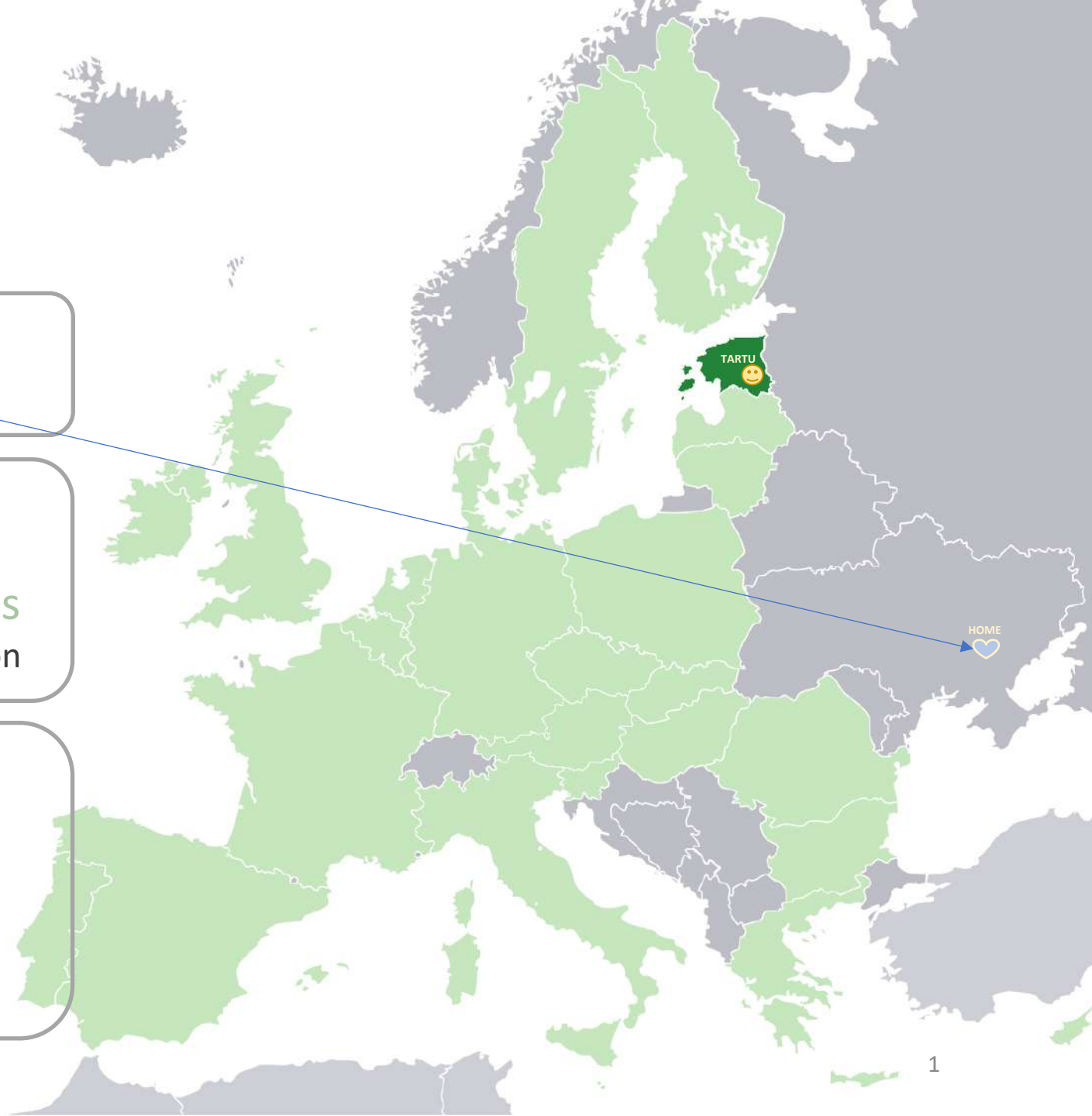
211 187

biobank participants
~20% of Estonian population



2 540

Metagenomes
+ electronic health record
+ host genotypes
+ metabolomic data



Естонський Біобанк

> 250 000 людей

Зразки біологічного матеріалу



Медичні та інші дані донорів

EHRs– electronic health records



Біобанк

Зразки біологічного матеріалу

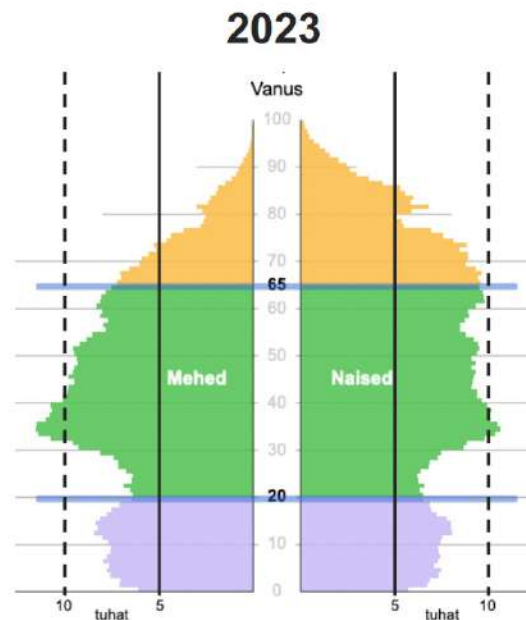


Медичні та інші дані донорів

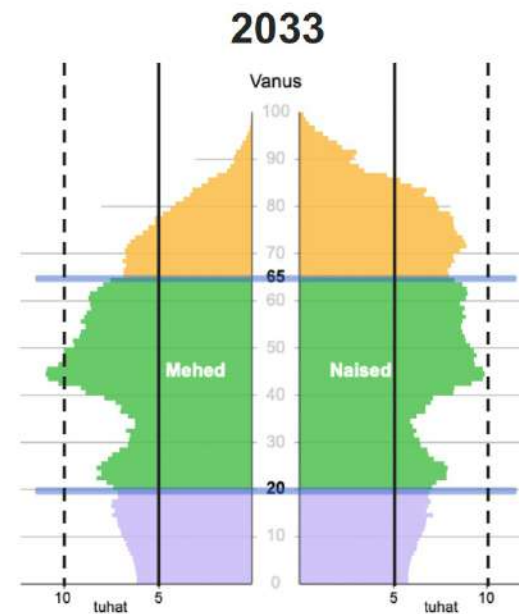
- *Геноміка (поногеномний сиквенс + генотипи)*
- *Метаболоміка (NMR)*
- *Метагеноміка (deep sequencing)*
- *Вік*
- *Стать*
- *BMI*
- *Статус по захворюванням*
- *Статус прийому ліків (антибіотиків та ін.)*
- *Дієта*
- *Частота заняття спортом та ін.*

Біобанк. Нащо?

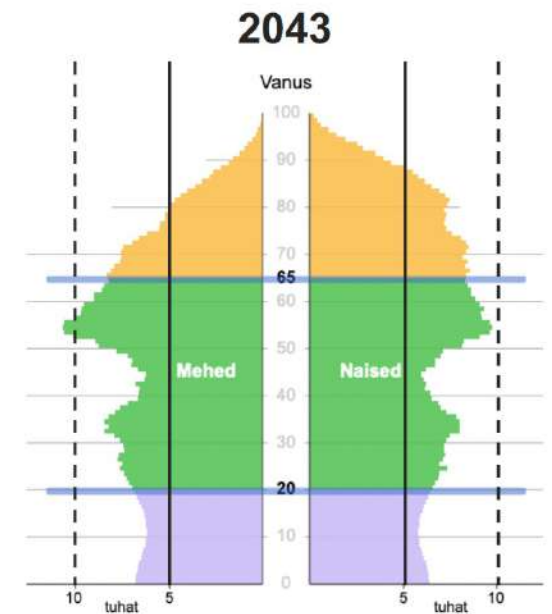
In the coming decades, the population will age and the proportion of people of working age will fall rapidly



35



42

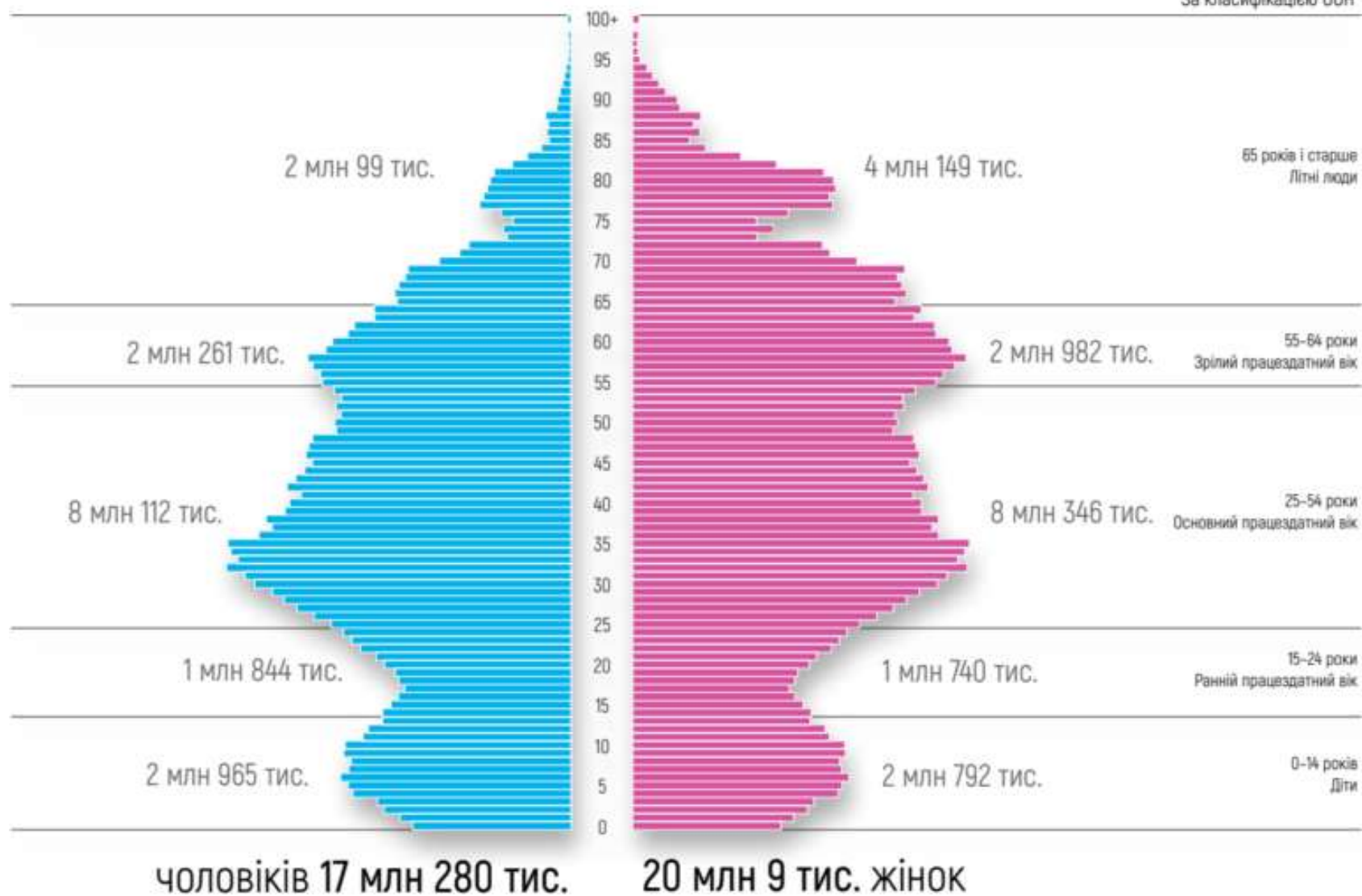


48

Кількість осіб
старшої вікової
групи на 100 осіб
середньої вікової
групи.

СТАТЕВО-ВІКОВА СТРУКТУРА НАЯВНОГО НАСЕЛЕННЯ

За класифікацією ООН



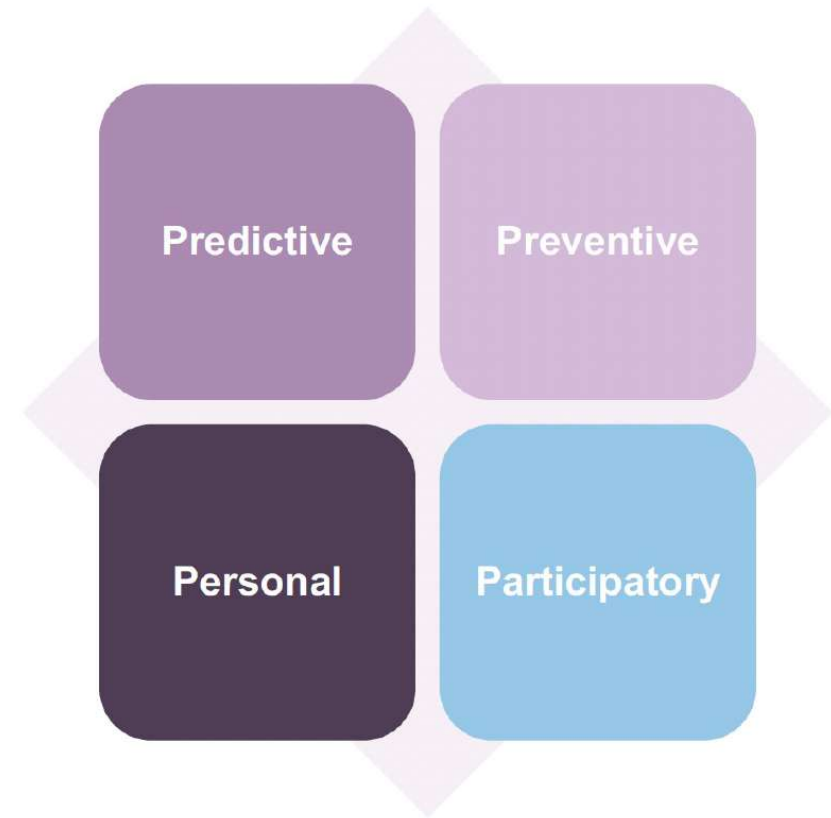
4П медицина

- прогностна
- профілактична
- персоналізована
- партисипативна



Біобанк. Мета

Знайти **біомаркери**, які можуть допомогти спрогнозувати захворювання з урахуванням **індивідуальних особливостей** людей. Виявити захворювання до появи симптомів, **попередити** його розвиток і **залучити** до цього процесу якомога більше людей, що допоможе **знижити витрати на медичну систему**.



Біомаркери

- SNP – однонуклеотидні поліморфізми геному (аналіз GWAS). Популяційний ризик виникнення захворювання обумовлений вашим геномом.



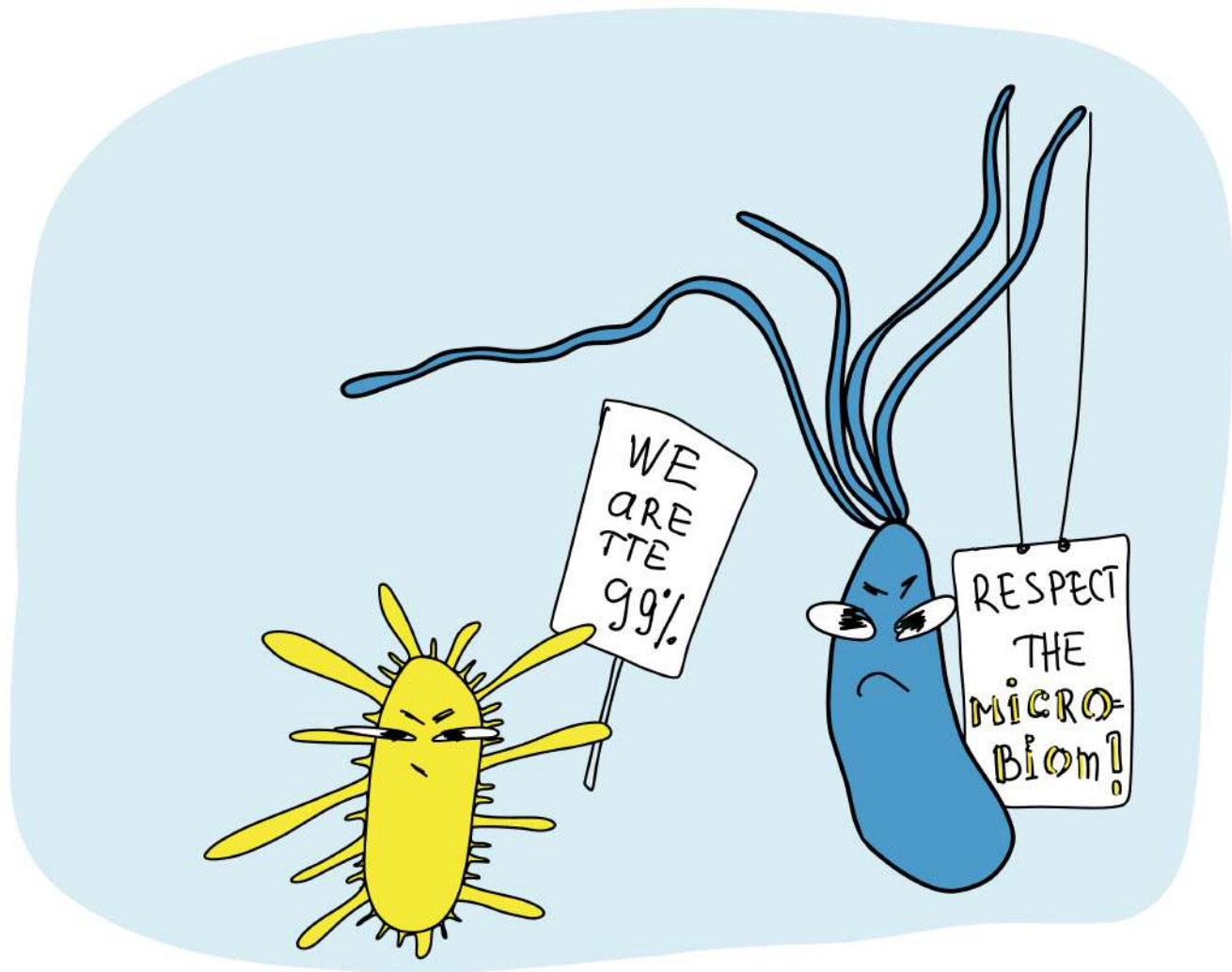
Біомаркери

Не тільки геном.

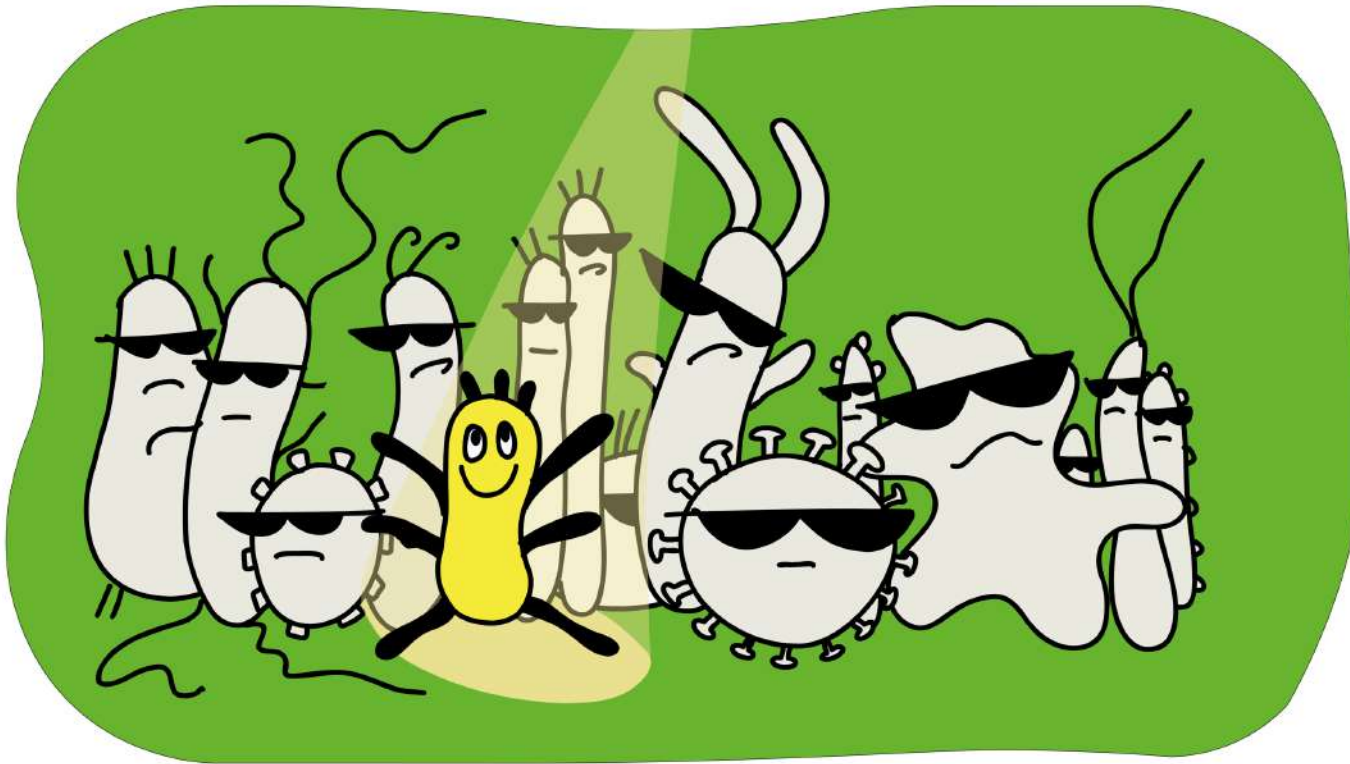
протеом,

метаболом,

... МІКРОБІОМЕ!



МІКРОБІОМ – спільнота мікроорганізмів



- Бактерії
- Археї
- Віруси
- Мікроскопічні гриби
- Одноклітинні еукаріоти

Аналіз мікробіому кишечнику набуває популярності. Чому?



- Мікробіом пов'язаний з багатьма захворюваннями

- Мікробіом на відміну від геному можна змінити

Компенсація непереносимості лактози

Ссавці не можуть перетравлювати лактозу в дорослому віці

Корисна мутація дозволила зберігати активність лактази в дорослому віці. Ця мутація розповсюдилась в популяції.

Більшість людей можуть пити молоко, але деякі ні ...

... якщо вони не мають специфічних бактерій, які перетравлюють лактозу за них.

Компенсація непереносимості лактози



Lactobacillus acidophilus та ***Lactobacillus bulgaricus***

та деякі інші може виробляти бета-галактозидазу, фермент, схожий на людську лактазу, який розщеплює лактозу на глюкозу та галактозу.

Ці прості цукру легше засвоюються і не викликають симптомів, пов'язаних з непереносимістю лактози.

Синтез вітаміну B12



Особи з мутаціями чи люди в групі ризику (люди похилого віку, вегетаріанці) можуть мати дефіцит вітаміну B12.

Певні кишкові бактерії, наприклад види з родів ***Lactobacillus*** і ***Bifidobacterium***, можуть синтезувати вітамін B12.

МІКРОБІОМ

Бактерії - більше, ніж лише загроза



Гіпотеза "зникаючого мікробіому"

Передбачає, що зменшення різноманітності мікробіому людини через

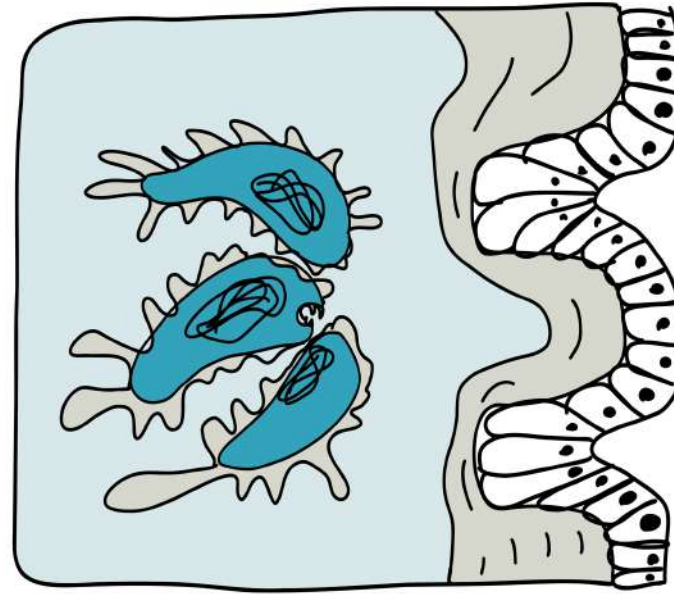
- зміни в харчуванні,
- способі життя,
- надмірне використання антибіотиків та
- зростаючу гігієну

сприяє **збільшенню частоти хронічних захворювань**, таких як алергії, астма, аутоімунні розлади та запальні захворювання кишечника.

Втрата корисних мікроорганізмів може порушувати баланс між мікроорганізмами і людським організмом, що призводить до **зниженої резистентності до патогенів**

Бактерія

- Одна клітина
- Тисячі видів
- Вони живуть всюди
- Малий кільцевий геном (3-5 Мб)
- + плазмід (гени стійкості до антибіотиків)
- Вони виділяють метаболіти в навколишнє середовище.



Аналіз чистих
культур в лабораторії

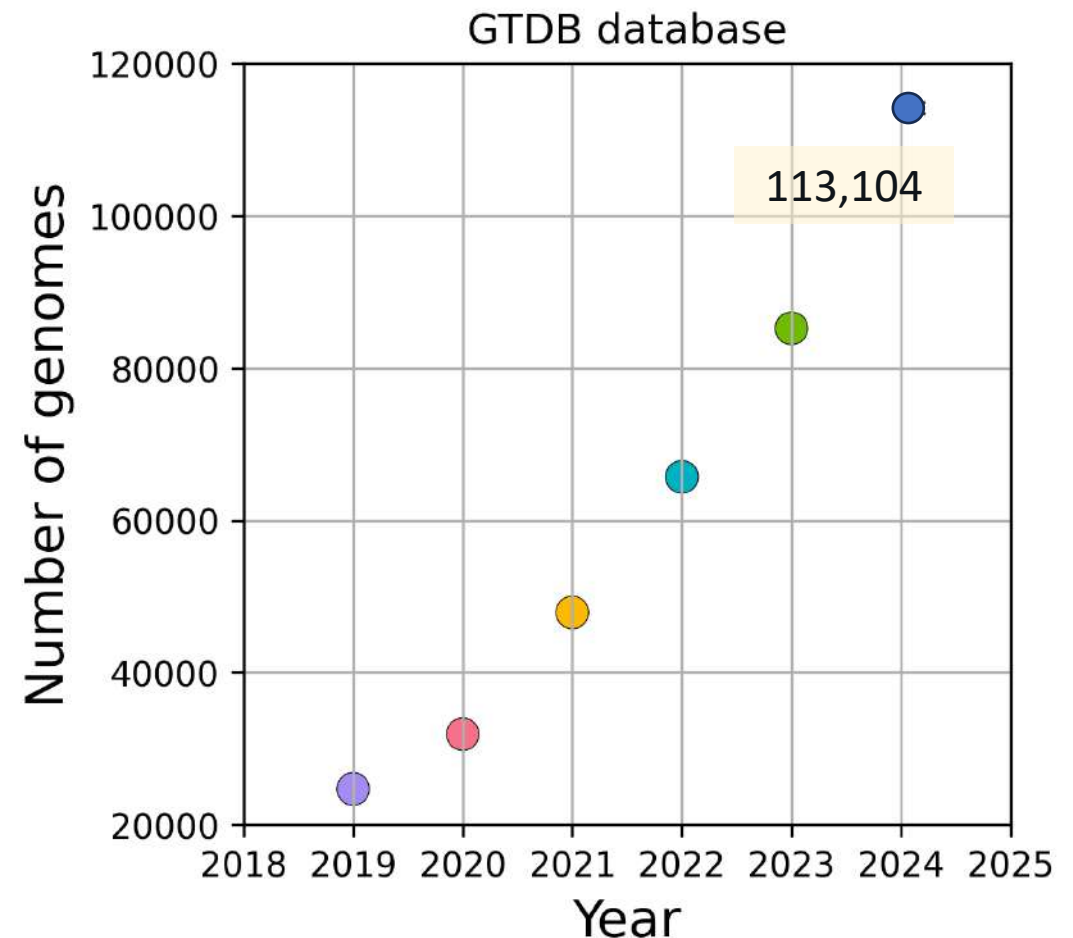
до 2000 року



NGS сиквенування
та метагеноміка

після 2000 року

- описано менше 5000 видів бактерій
- фокус на патогенах
- кожен новий вид – подія!



Метагеном

суміш ДНК від різних бактерій



CTGTCGGTAC

TGTCGGT

GTTCGGT

ATTGTCGGTTCT

Метагеном

суміш ДНК від різних бактерій

Як проаналізувати метегеном?

- 16S amplicon sequencing
- Metagenome profiling
- Metagenome de-novo assembly

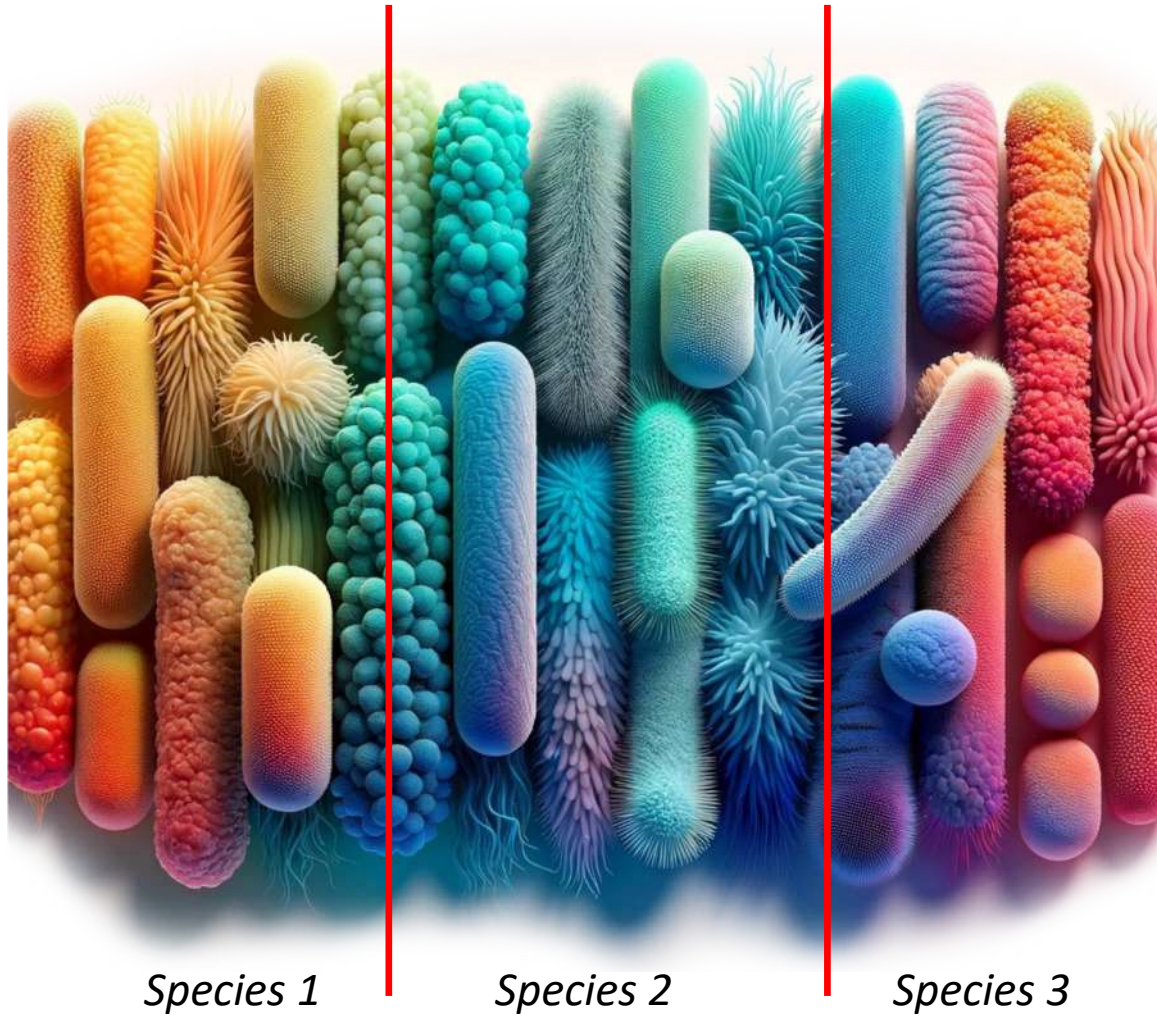
CTGTCGGTAC

TGTCGGT

GTTCGGT

ATTGTCGGTTCT

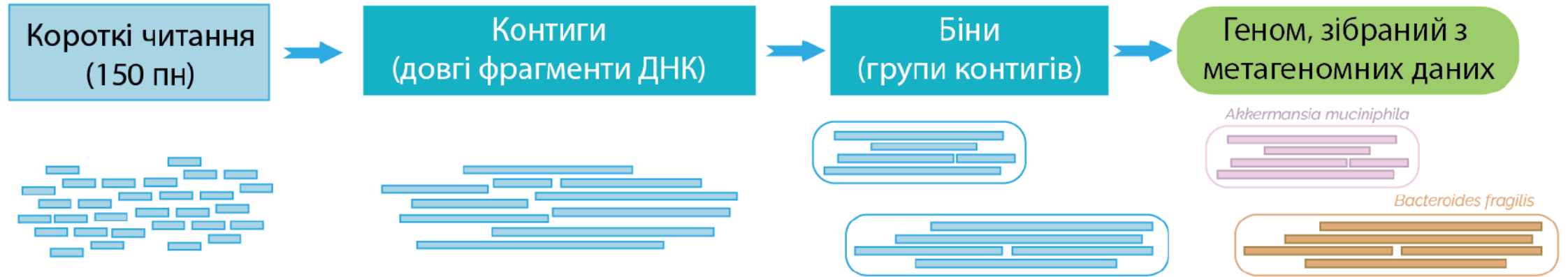
Проблема. Що таке вид?



Бактерії зі схожістю
геному $> 95\%$
належать до одного
виду

Збір геномів з Метагеномних даних

MAG – metagenome assembled genomes



Збір геномів з Метагеномних даних

MAG – metagenome assembled genomes

Короткі читання
(150 пн)



ААТСТГТСТГТСТ

ТГТСТГТСТГТСТ



АТТТГТСТГТСТГТСТ

Контиги
(довгі фрагменти ДНК)



Збір геномів з Метагеномних даних

MAG – metagenome assembled genomes

Контиги
(довгі фрагменти ДНК)



Покриття (Coverage): Параметр, що відображає, як часто певний континг представлений у даному наборі даних.

GC-зміст (GC-content): Співвідношення гуаніну та цитозину в ДНК-контингу.

Таксономічні мітки (Taxonomic markers): Наявність специфічних генів або генетичних маркерів, що асоціюються з певними таксономічними групами.

Тетра-нуклеотидні частоти (Tetranucleotide frequencies): Частота появи всіх можливих тетрануклеотидних (чотири-базових) послідовностей.

Біни
(групи контигів)



Збір геномів з Метагеномних даних

MAG – metagenome assembled genomes

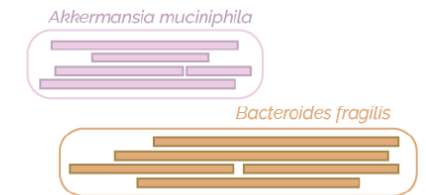


Таксономія. Порівняння щойно зібраного геному з усіма відомими геномами та визначення його таксономії на основі подібності.

Оцінка якості: анотація генів і порівняння списку знайдених генів з коровими генами. Оцінка повноти геному (Completeness) і забруднення геному фрагментами інших геномів (Contamination)

Розмір геному, фрагментація: оцінка розміру геному та рівня фрагментації.

Геном, зібраний з метабеномних даних



Що в результаті?



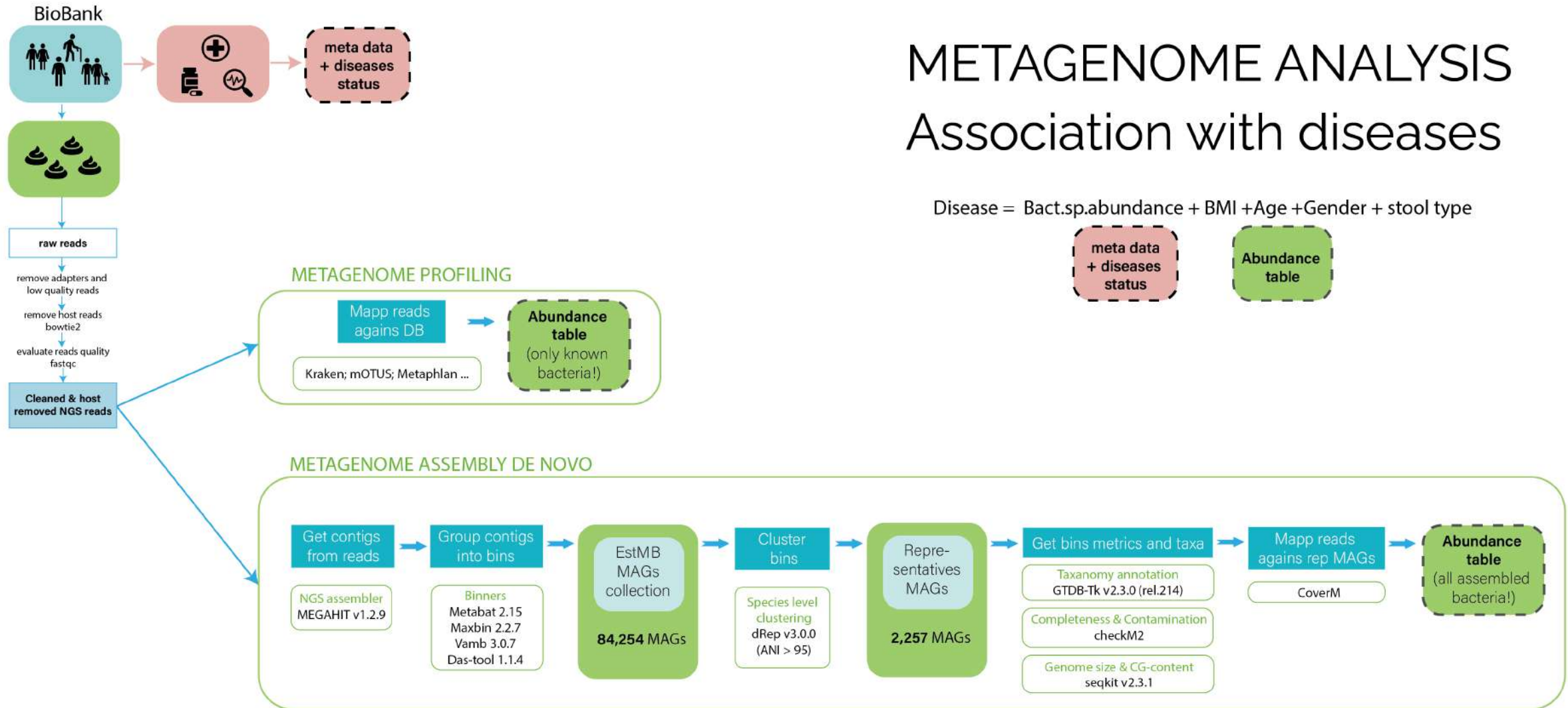
Колекція геномів бактерій які присутні в даному зразку біологічного матеріалу.

В подальшому можна використовувати як референс для досліджень.

METAGENOME ANALYSIS

Association with diseases

$$\text{Disease} = \text{Bact.sp.abundance} + \text{BMI} + \text{Age} + \text{Gender} + \text{stool type}$$



Bacterial Genome database

2504 human gut samples from Estonian Biobank



84 793 bacterial genomes

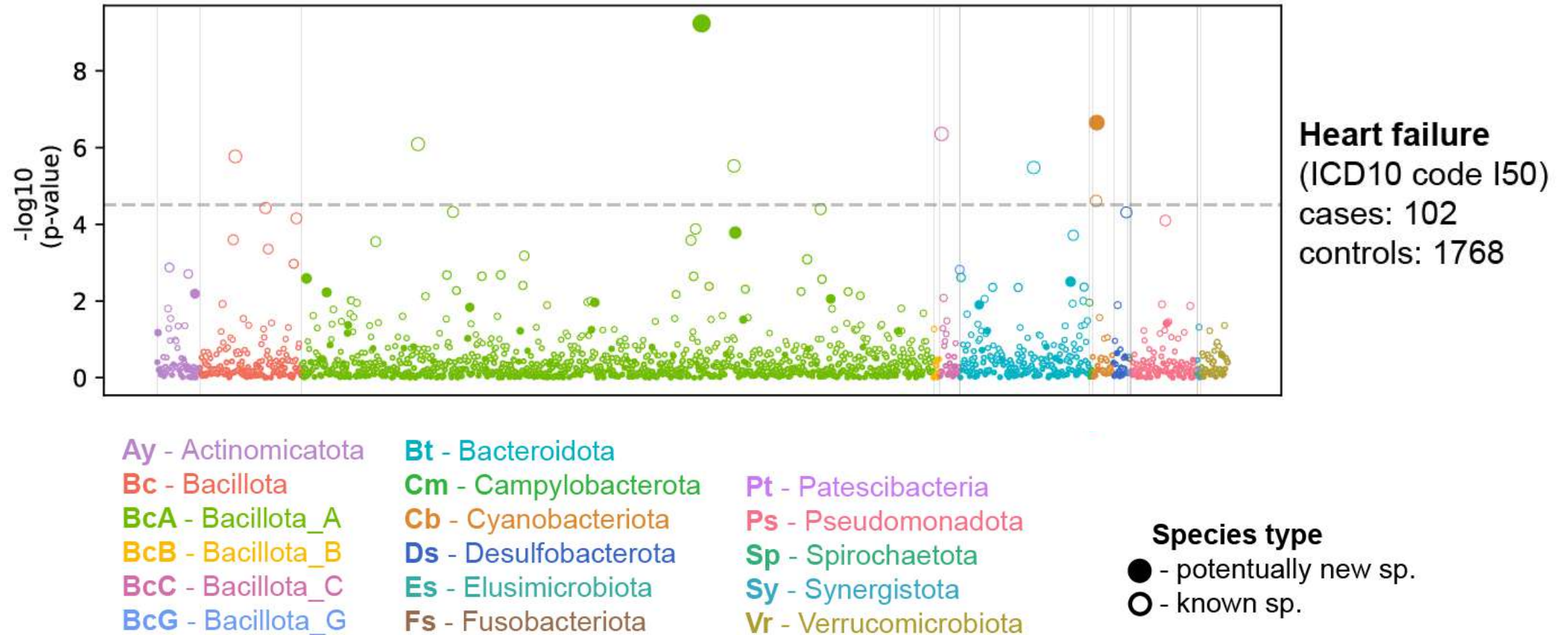
organised into

2 257 species clusters

including

353 NEW species

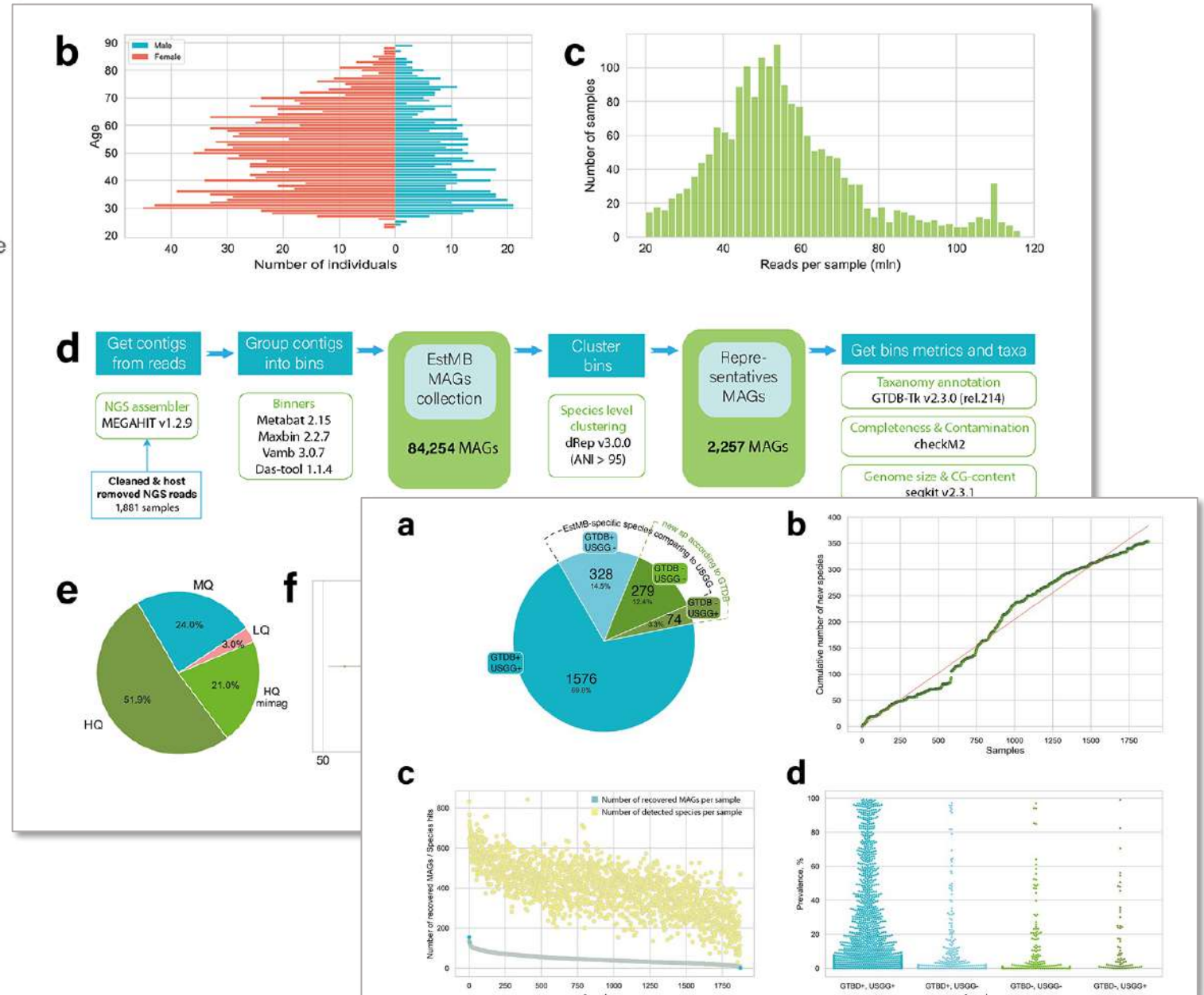
Кореляційний аналіз



Препринт!

QR Code for 10.1101/2024.07.06.602324

Metagenome-assembled genomes of Estonian Microbiome
links with prevalent diseases

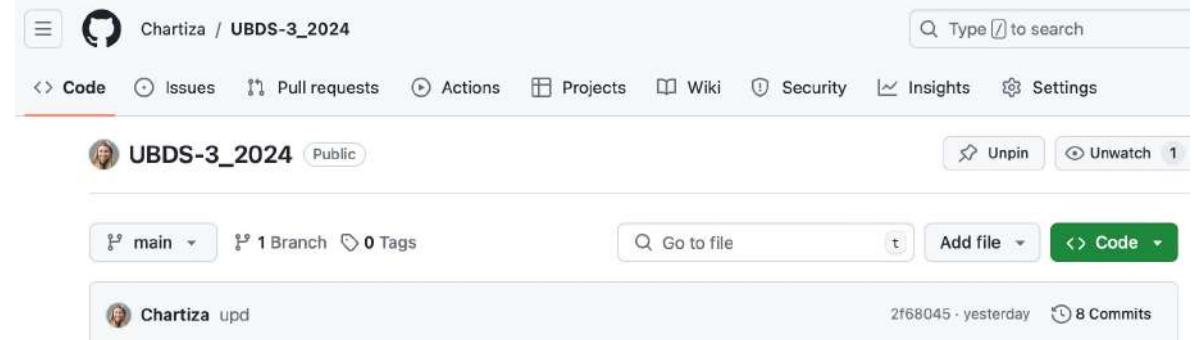


Correlation \neq Causation

Кореляція не означає причинно-
наслідковий зв'язок

Практичний курс

Анонс



step1_metadata.ipynb Аналіз метаданих людей, зразки яких доступні до аналізу. Структура популяції. histogram, piechart, swarmplot, violinplot.

step2_metagenome_profiling.ipynb Приклади результатів найпоширеніших програм для профілювання метагеному. Аналіз продуктивності програм за допомогою штучних даних. Вивчення barplot, swarmplot, scatterplot.

step3_mags_qual.ipynb Поняття MAG - Metagenome-Assembled Genomes чи геном, зібраний з метагеномних даних. Показники якості MAG. Чим відрізняється аналіз метагеному за допомогою профайлінгу та збірки геномів з метагеномних даних *de novo*?

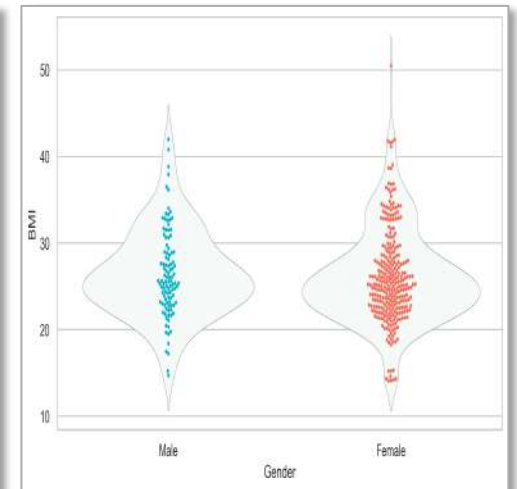
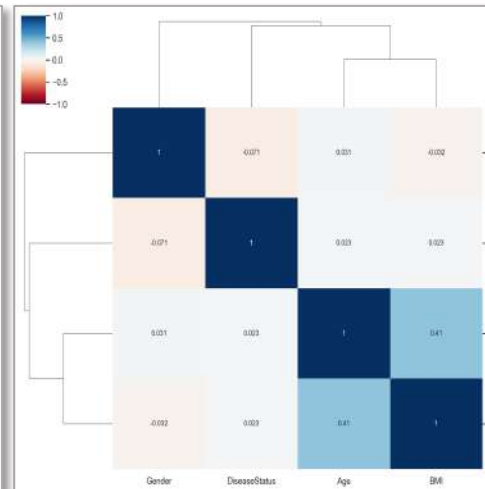
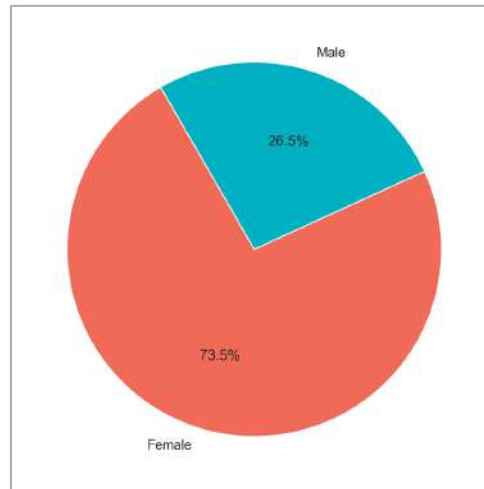
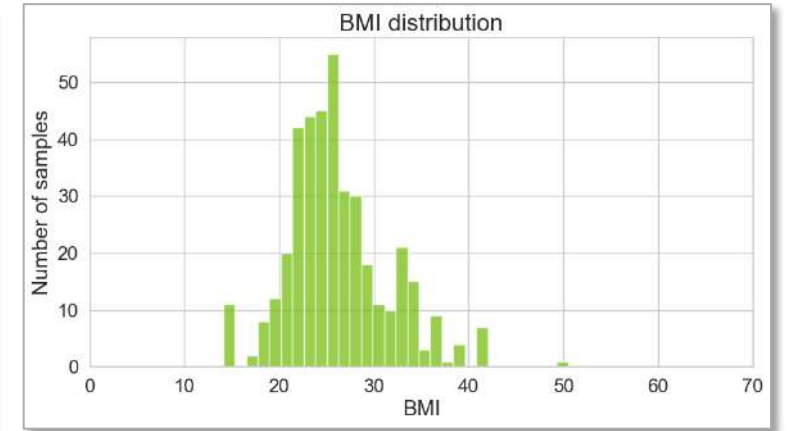
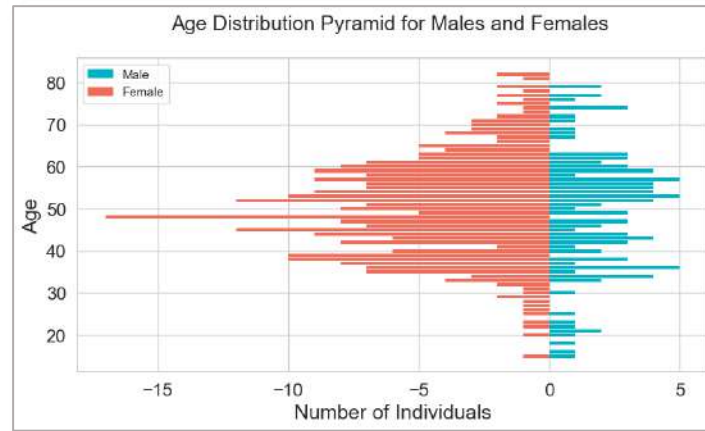
step4_taxa.ipynb Аналіз та візуалізація таксономії присутніх бактерій. Обговорення концепції GTDB та різниці між ncbi та gtdb таксономією (додатковий скрипт - пошук аналогів назв видів між цими двома системами).

step5_MWAS.ipynb Пошук асоціацій між бактеріями та хворобами (та іншими характеристиками).

step6_corr_plot.ipynb Візуалізація асоціацій (MWAS analysis, heatplot)

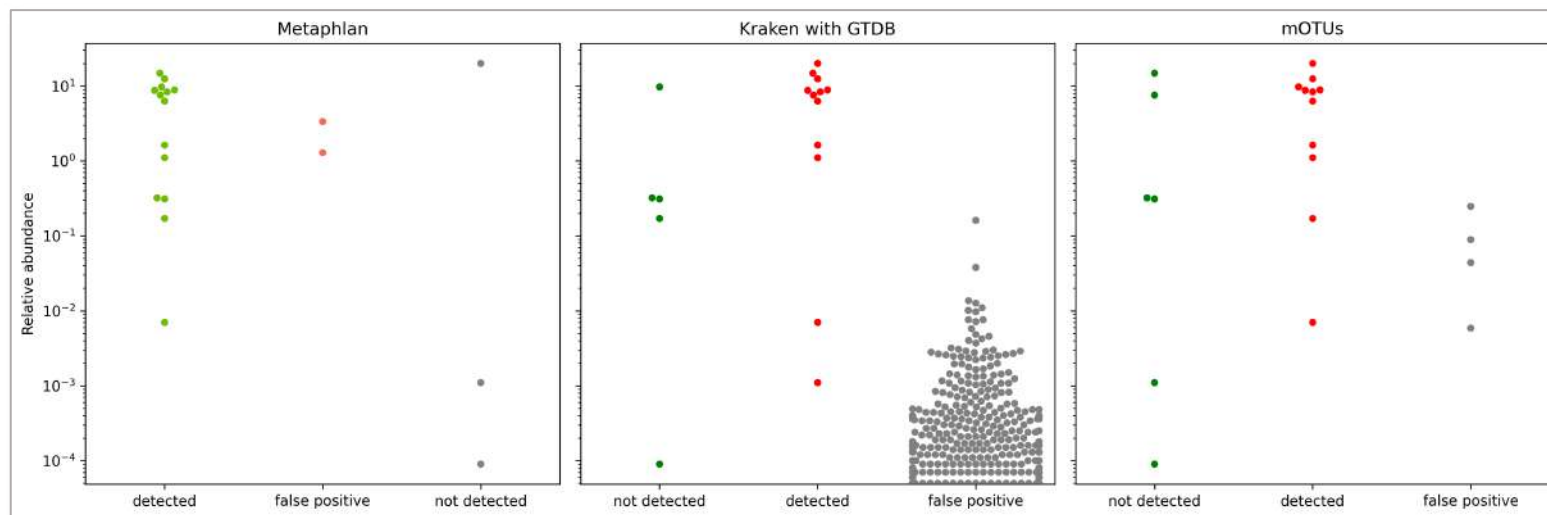
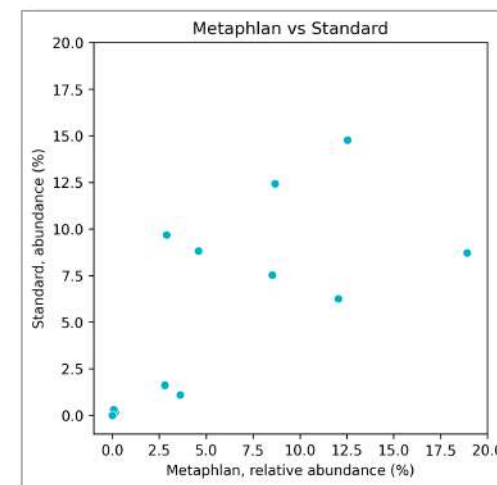
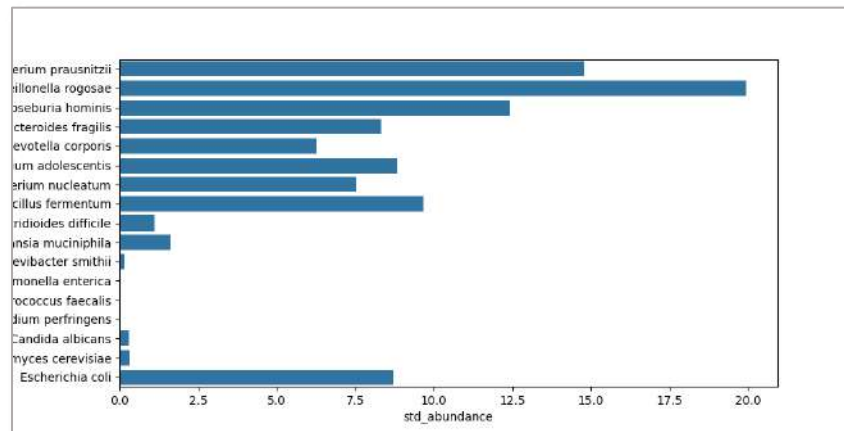
Практичний курс. Анонс

`step1_metadata.ipynb` Аналіз метаданих людей, зразки яких доступні до аналізу. Структура популяції.
histogram, piechart, swarmplot, violinplot.



Практичний курс. Анонс

`step2_metagenome_profiling.ipynb` Приклади результатів найпоширеніших програм для профілювання метабеному. Аналіз продуктивності програм за допомогою штучних даних. Вивчення `barplot`, `swarmplot`, `scatterplot`.

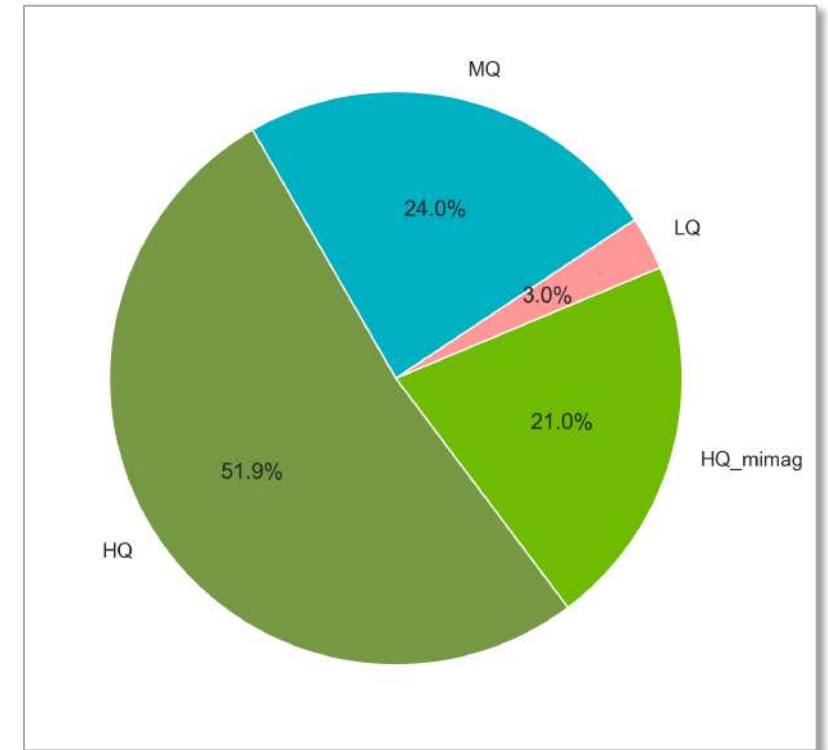


Практичний курс. Анонс

`step3_mags_qual.ipynb` Поняття MAG - Metagenome-Assembled Genomes чи геном, зібраний з метагеномних даних. Показники якості MAG. Чим відрізняється аналіз метагеному за допомогою профайлінгу та збірки геномів з метагеномних даних *de novo*?

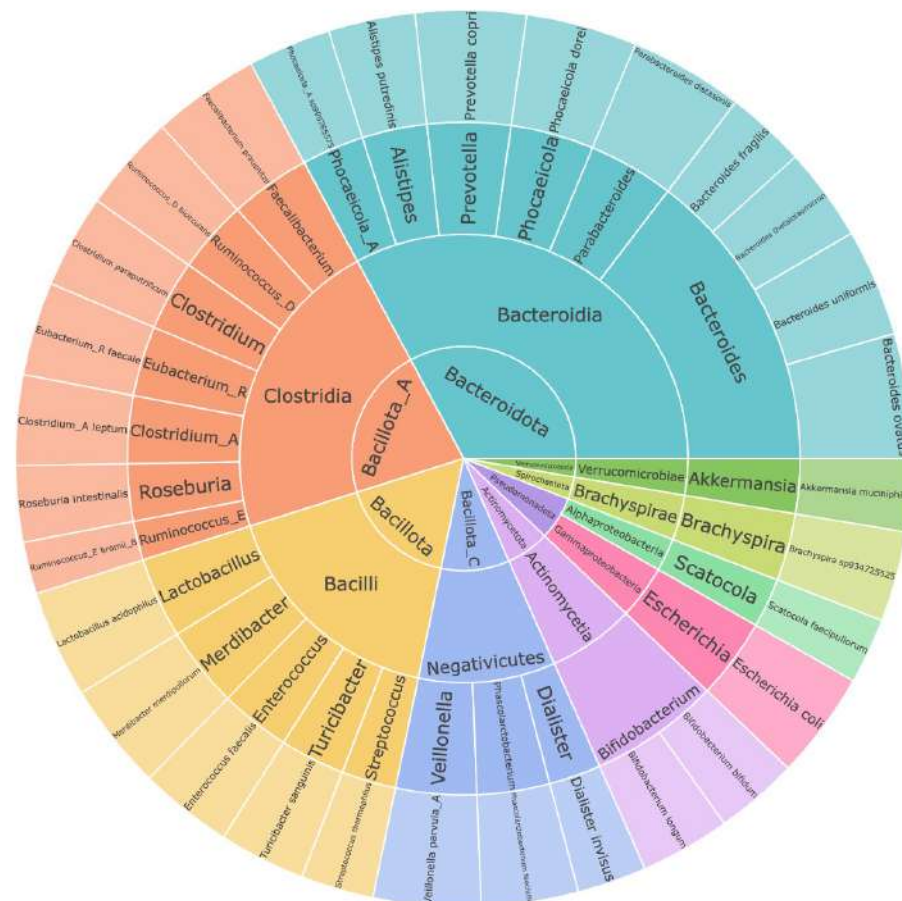


**MIMAG
standard**



Практичний курс. Анонс

step4_taxa.ipynb Аналіз та візуалізація таксономії присутніх бактерій. Обговорення концепції GTDB та різниці між ncbi та gtdb таксономією (додатковий скрипт - пошук аналогів назв видів між цими двома системами).



Практичний курс. Анонс

step5_MWAS.ipynb Пошук асоціацій між бактеріями та хворобами (та іншими характеристиками).



```
20
21 for b in bac:
22     # Create a design matrix by adding the covariates to the model
23     X = mrg[covariates + [ph_name]]
24     X = sm.add_constant(X)
25
26     y = mrg[b]
27     model = sm.OLS(y, X).fit()
28     if model.pvalues[ph_name] < alpha:
29         print(b)
30         print('pValue', model.pvalues[ph_name])
31         print('beta', model.params[ph_name])
32         print('-----')
33
```

[29]

```
.. Bifidobacterium longum
   pValue 0.0009882978398366952
   beta 0.10540482430154978
   -----
   Bacteroides ovatus
   pValue 0.008940609871037033
   beta 0.06904217231202944
   -----
```

Практичний курс. Анонс

step6_corr_plot.ipynb Візуалізація асоціацій (MWAS analysis, heatmap)

