

Microbial metagenomics

MMB-901

Course information

- Practicals on Mon, Wed, Fri 10-16 (Mon Info, others Bio1)
- **No practicals on Fri 6.12, but on Thursday 5.12.**
- Start time? (<https://presemio.helsinki.fi/mmb901>)
- Practical work in groups of 4
- Instructions can be found from:
<https://github.com/karkman/MMB-901> Metagenomics
- **Self-learning materials for days in between** (in Moodle)

Learning outcomes

By completing this course, you will:

- Have a basic understanding of metagenomic sequencing technologies and bioinformatic approaches to analyse metagenomic data
- Be able to plan and execute a metagenomic sequencing project depending on the research questions.
- Have an up-to-date knowledge on the bioinformatic tools and best practices for the analysis of metagenomes.
- Be able to choose and critically evaluate new tools and approaches for specific research question
- Have confidence to learn and implement new bioinformatic methods using available documentation

Course completion and assessment

Completion

- Participation in teaching and practicals
- Weekly self-assessments during the course
- Group exam

Assessment

- Activity during the course (30 %)
- Weekly assessments (30 %)
- Group exam (20 %)
- Self evaluation at the end (20 %)

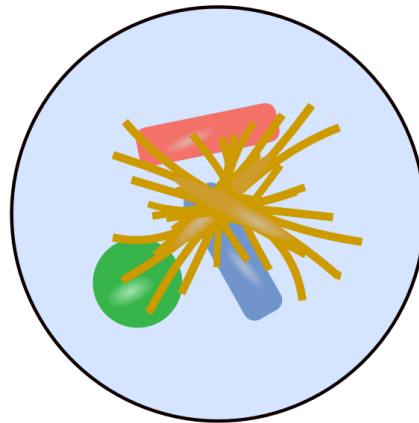
Metagenomics

Short introduction

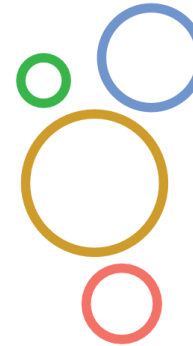
Metagenomics

- Study of all genetic material in an environment by sequencing

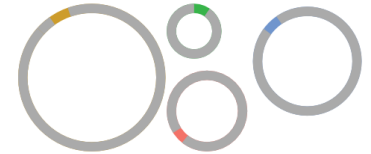
Mixed microbial community



DNA
Extraction



Amplicon sequencing



Multiple copies of fragments
from 1 target gene

Metagenomics sequencing



Short sequence
fragments from "all" DNA

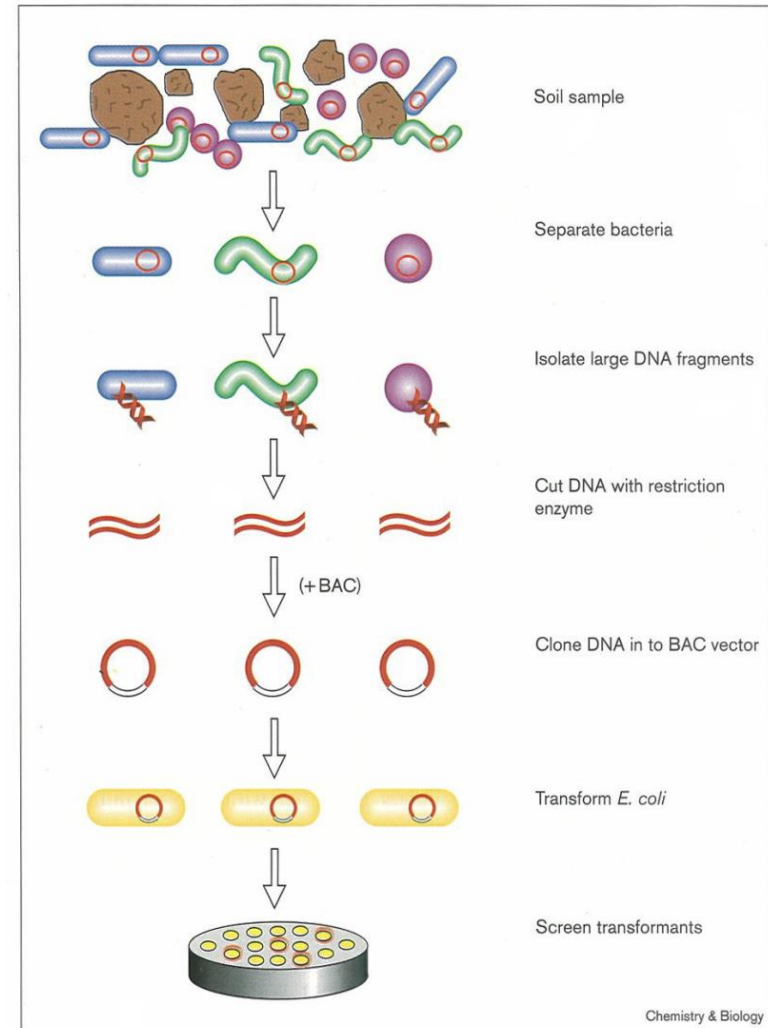
<https://doi.org/10.21105/jose.00053>

Metagenomics

- Jo Handelsman et al. 1998

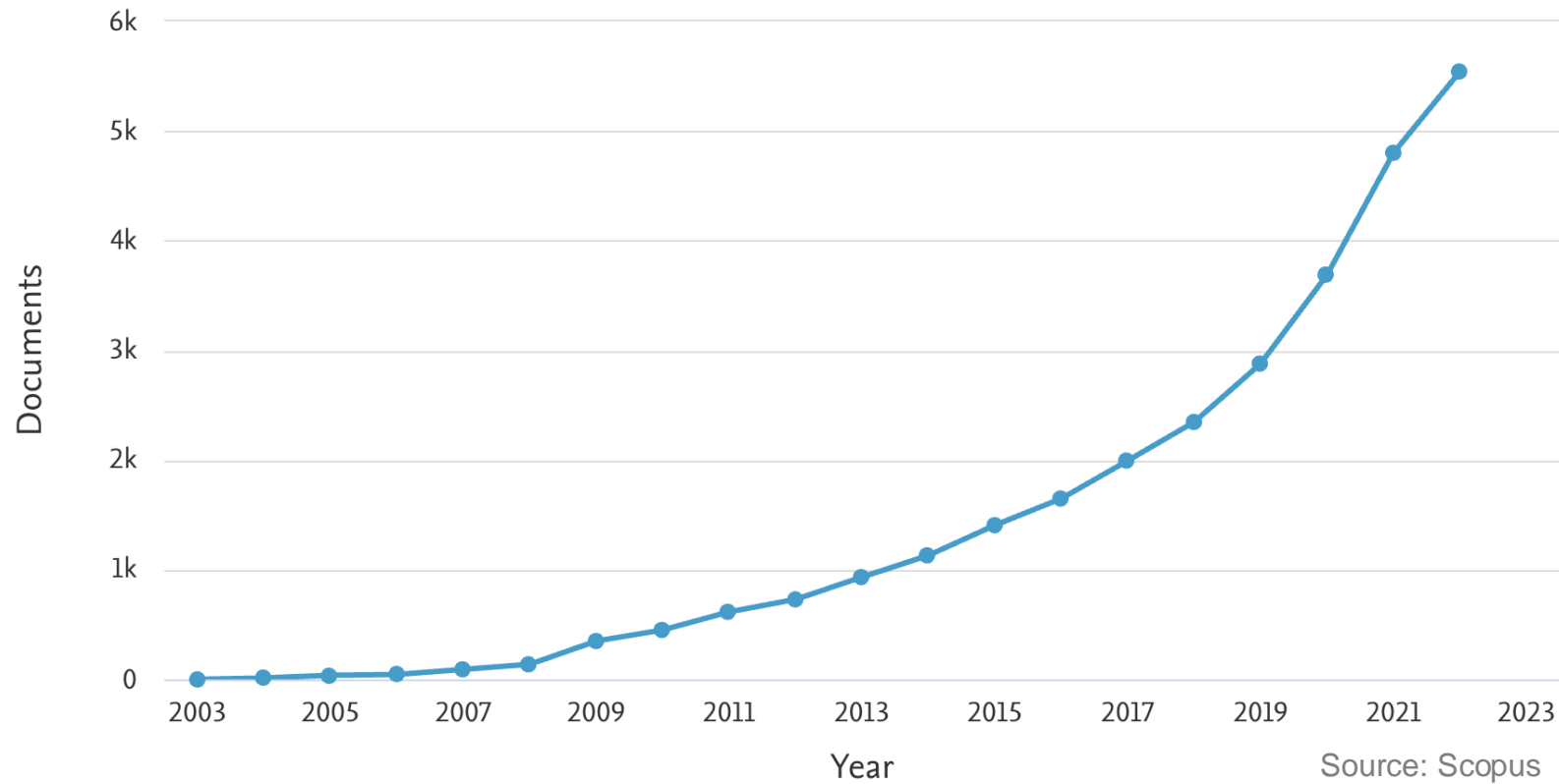
Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products

Jo Handelsman¹, Michelle R Rondon¹, Sean F Brady², Jon Clardy² and Robert M Goodman¹



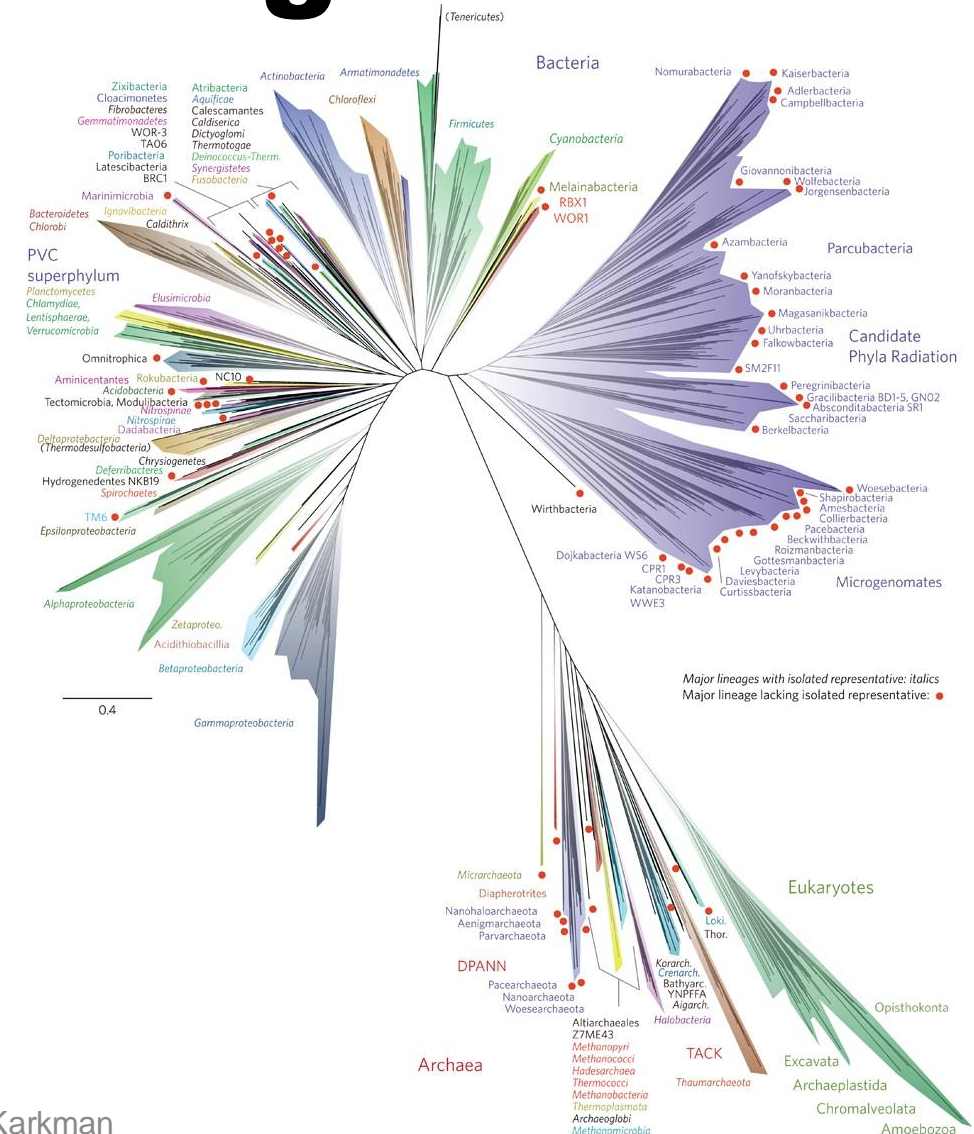
Articles published in metagenomics

Documents by year



Why do we need metagenomics?

- The great plate count anomaly
- Taxonomy \neq function



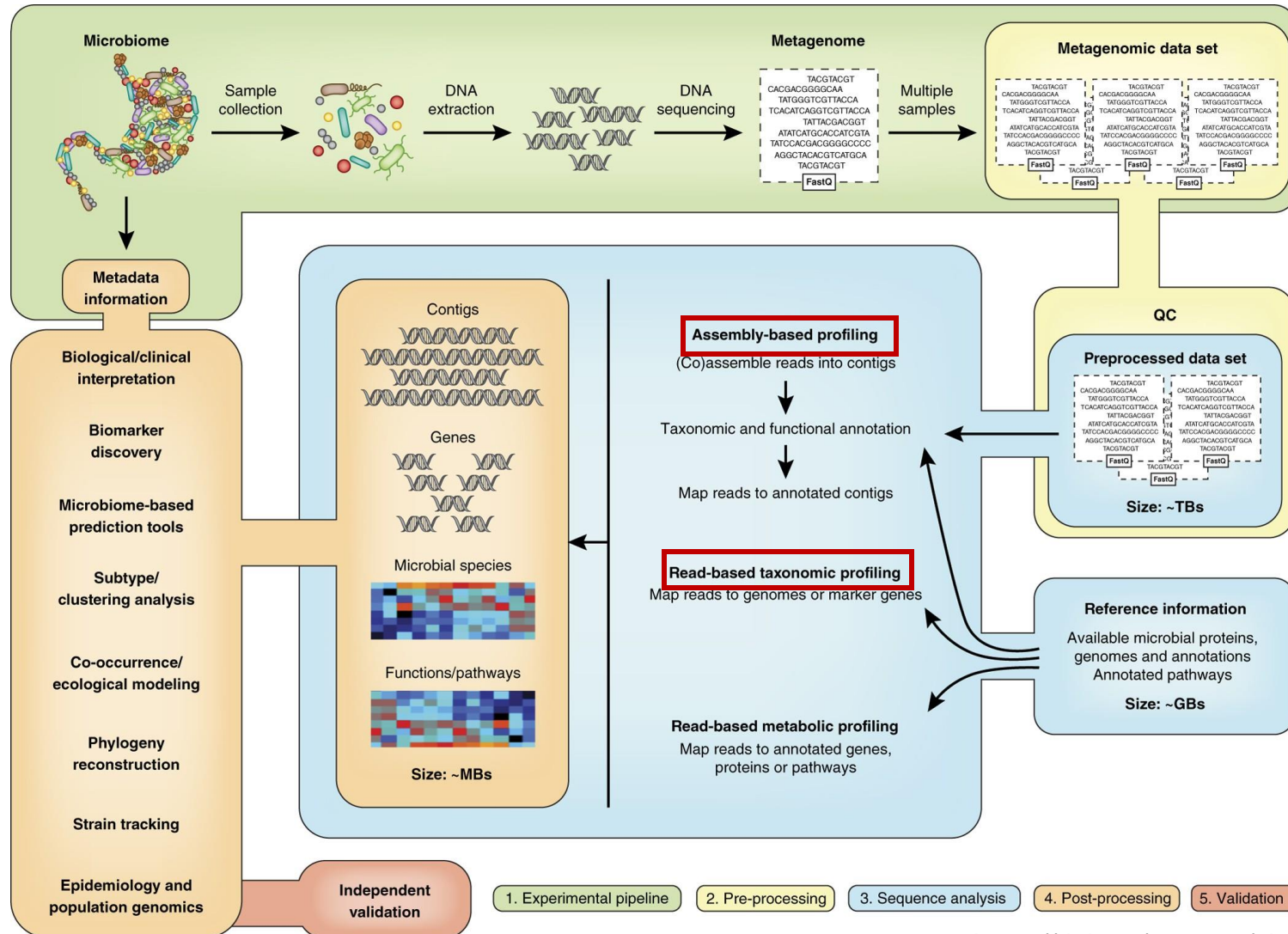
Metagenomic sequencing



Group work:

- 1-3 specific features for each technology
- Which one(s) is/are suitable for metagenomics

Metagenomic data analysis



Read-based vs. Assembly-based



Read-based

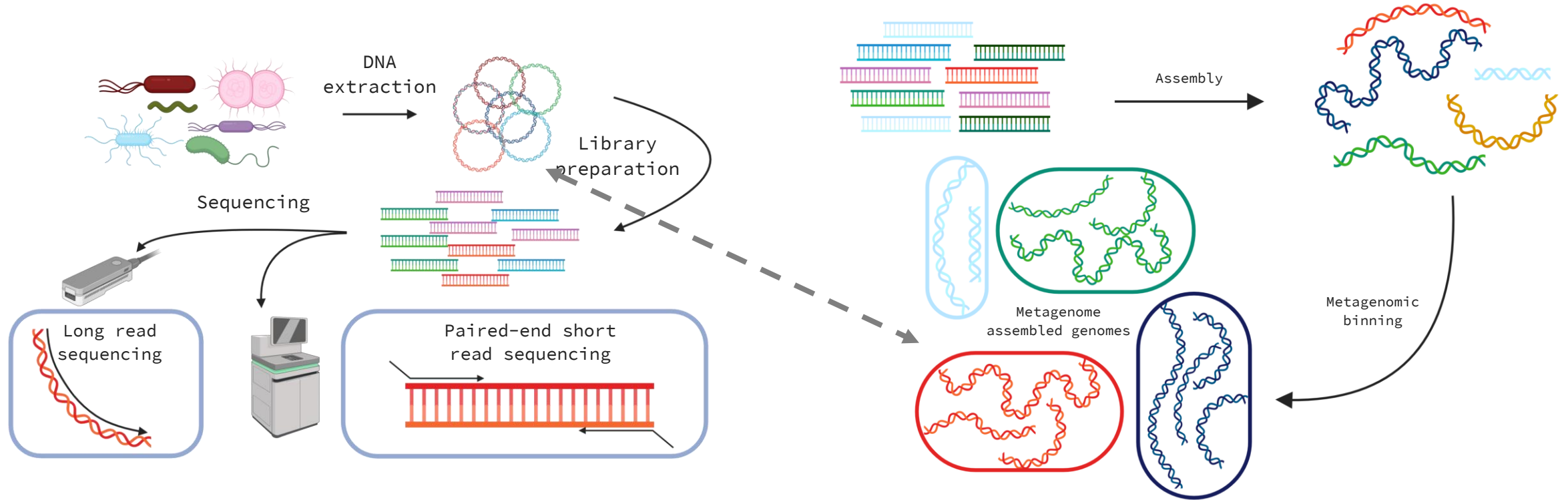


Assembly-based



Genome-resolved metagenomics

Reconstructing genomes from metagenomic sequencing data



Databases for (meta)genomics

Sequence Read Archive (SRA)

European Nucleotide Archive (ENA)

- Publicly available repositories for high-throughput sequencing data
- All sequencing data should be deposited to repositories upon publication
- Sequencing data (Runs) organised under projects (BioProject) and linked to samples (BioSample)
- Various levels of metadata depending on the project
- Web and (several) command line access options

<https://www.ncbi.nlm.nih.gov/sra> <https://www.ebi.ac.uk/ena>

MGnify

- Website to browse, analyse, discover and compare microbiome data
- Data from ENA/SRA
- Includes analyses, assemblies and MAG collections

<https://www.ebi.ac.uk/metagenomics>

Branchwater & Sandpiper

- Annotation of sequencing experiments in SRA/ENA
- Web interface to search for sequences/taxonomy

<https://sandpiper.qut.edu.au/> <https://branchwater.sourmash.bio/>

Database exercises

Database exercise 1

Search for a publication that has metagenomic sequencing and look for the repository accession. Then go to SRA or ENA and find the study.

1. What data can you find for the project? Number of sequencing runs? Sample biome information? Ways to download the data?
2. Then go to the other and look for the same project and the same things.
3. Which repository do you find easier to use/more informative?

Database exercise 2

Go to Mgnify (<https://www.ebi.ac.uk/metagenomics>) and try to find the same project from there.

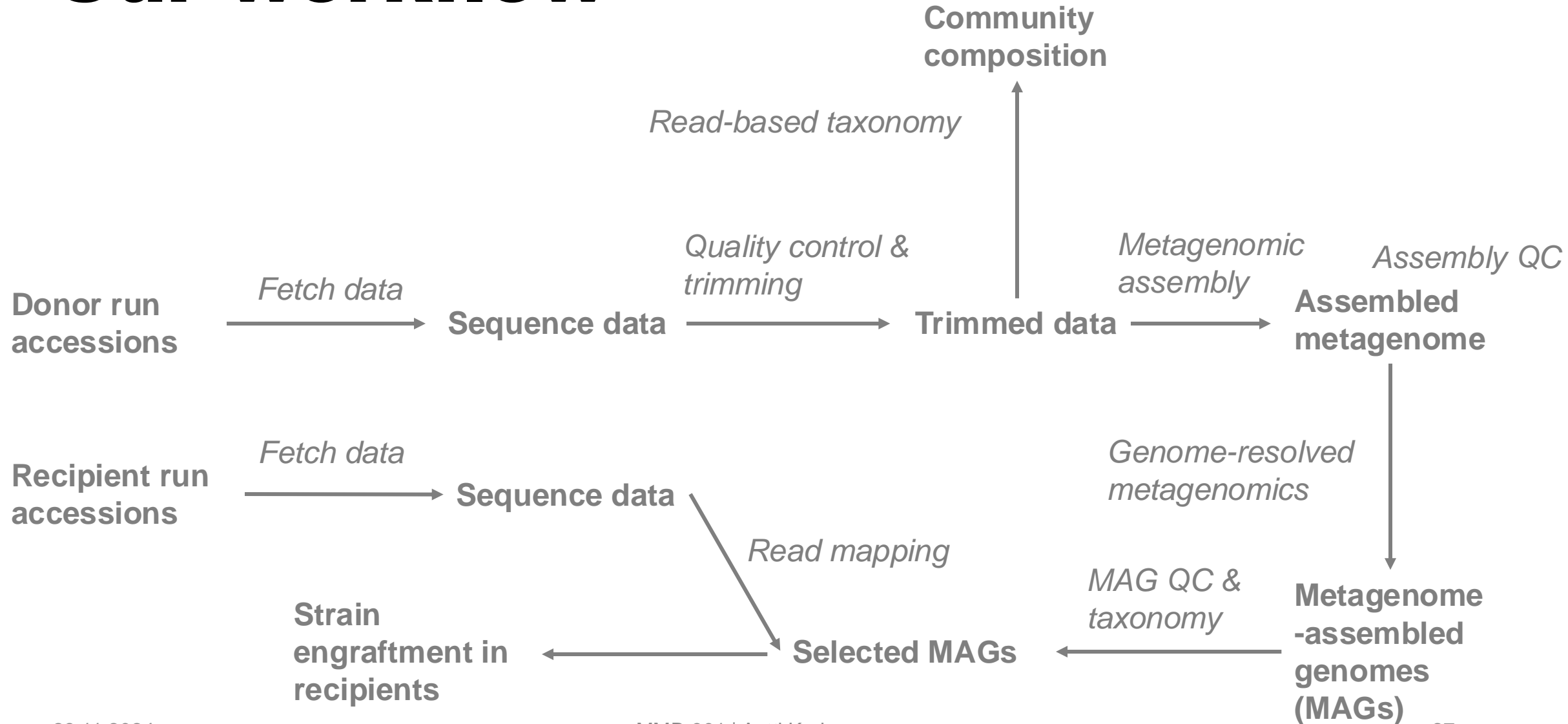
1. Can you find any other information or analyses for the project?
2. How many metagenomes can you find from the same biome?

Database exercise 3

Go to Sandpiper (<https://sandpiper.qut.edu.au/>) and look for *Megamonas funiformis*

1. How many samples had this bacterium?
2. How is the geographic distribution of the runs where the bacterium is found? Is it found globally?
3. In which environment the bacterium is most often found?
4. What information can you find for the sample where it is most abundant?

Our workflow



Let's get to work

https://github.com/karkman/MMB-901_Metagenomics