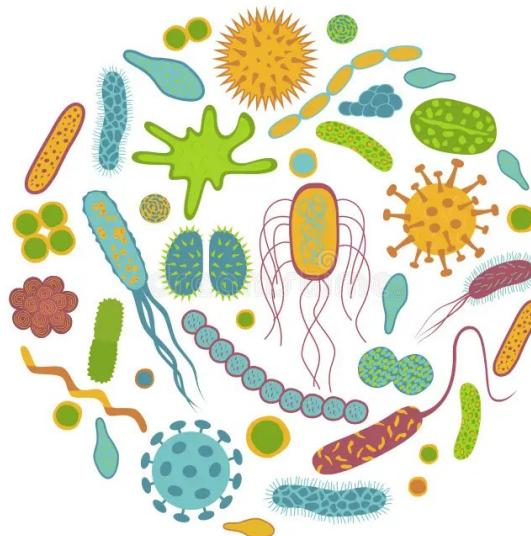
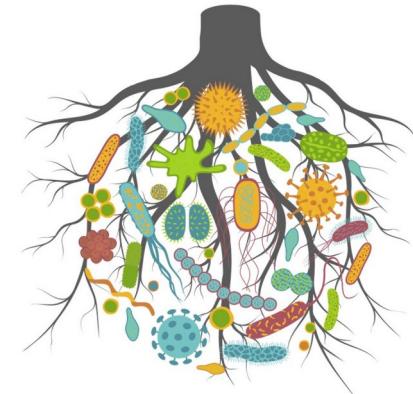
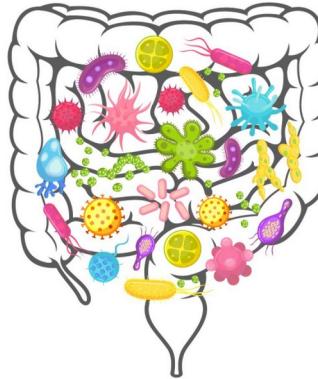


Introduction to Metagenomics



What is a microbiome?

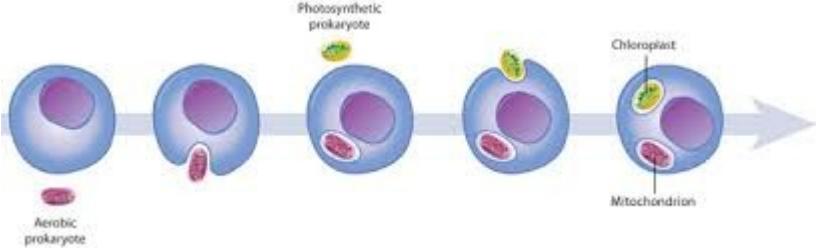


The totality of microbes in a defined environment, especially their genomes and interactions with each other and surrounding environment.

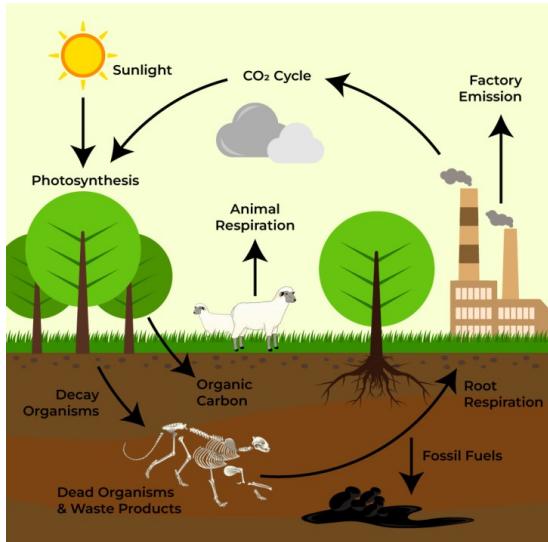
- A population of a single species/strain is a culture, extremely rare outside of lab, some infections
- A microbiome is a mixed population of different microbial species (microbial ecosystem)

A mixed community is the norm!

Why study microbiomes?



- Environmental Science
 - Critical elemental cycles (carbon, nitrogen, sulfur, iron, ...)
 - Pollution control, cleanup
 - Ecology / Evolution (chloroplasts, mitochondria, symbiosis, competition, ...)



Why study microbiomes?

Environmental Science

- Critical elemental cycles (carbon, nitrogen, sulfur, iron, ...)
- Pollution control, cleanup
- Ecology / Evolution (chloroplasts, mitochondria, symbiosis, competition, ...)

Industrial Applications

- Wastewater treatment (*V. cholera*, algal blooms, etc.)
- Bioprospecting (novel enzymes, compounds)
- Novel biosynthesis
- Fermentations: Consortia (yogurt) / wild (kombucha, Belgian ales)



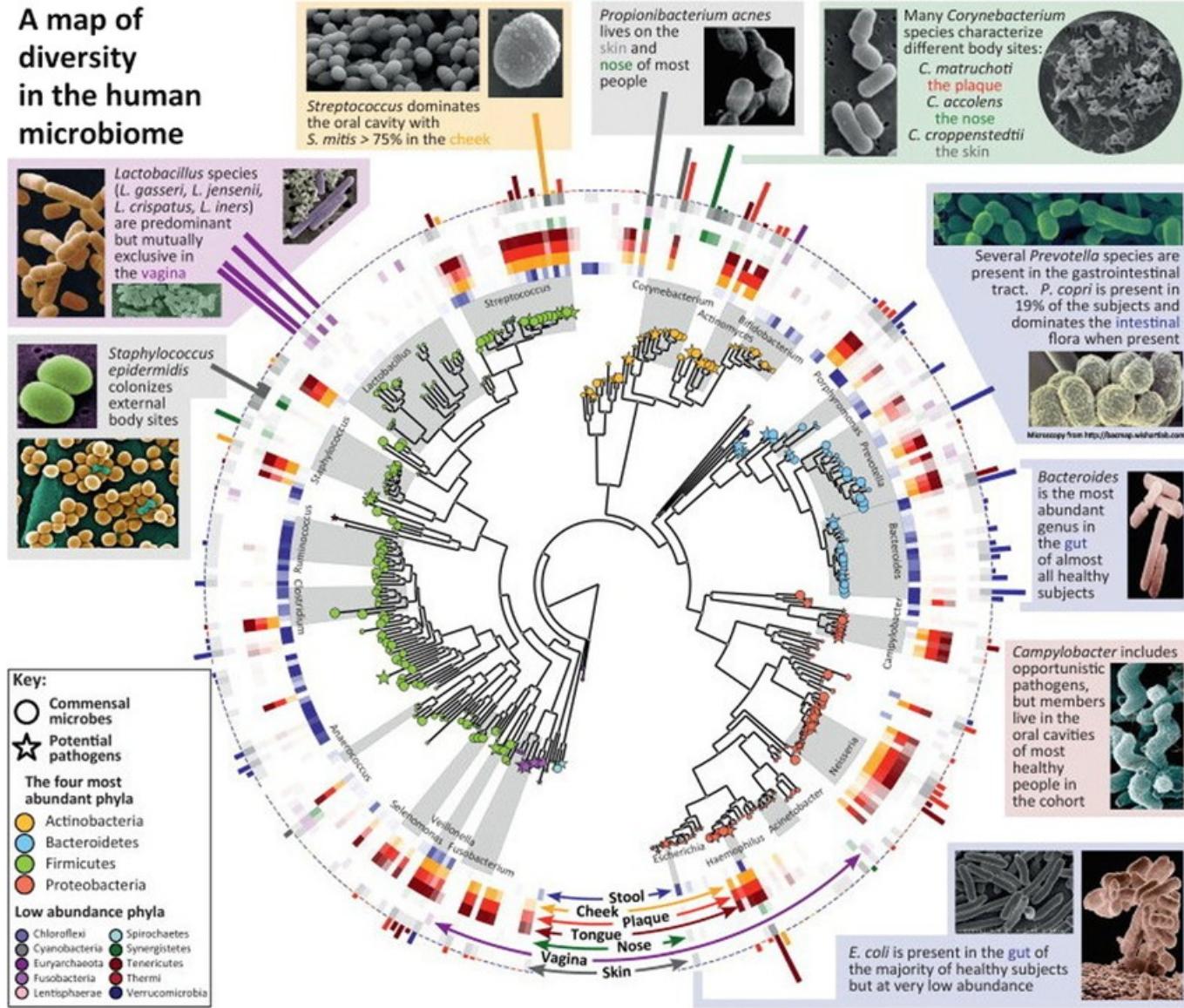
Why study microbiomes?

Human Health

- Protection from pathogens (e.g. *Clostridium difficile*)
- Cancer
- Absorption/Production of nutrients in the gut (obesity, T2D)
- Development/regulation of immune system, e.g. chronic diseases (T1D, RA, IBD, other autoimmune, UTIs, periodontitis, ...)



A map of diversity in the human microbiome



Who's there?

(Taxonomic profiling)

What are they doing?

(Functional profiling)

What does it all mean?

(Statistical analysis)

A summary of metagenomics technique



Amplicon Sequencing:
PCR amplify a single marker
gene, e.g. 16S rRNA

Lyse all cells
Extract Total DNA (and/or RNA)



Shotgun Sequencing:
Direct sequencing of (fragments
of) total DNA/RNA



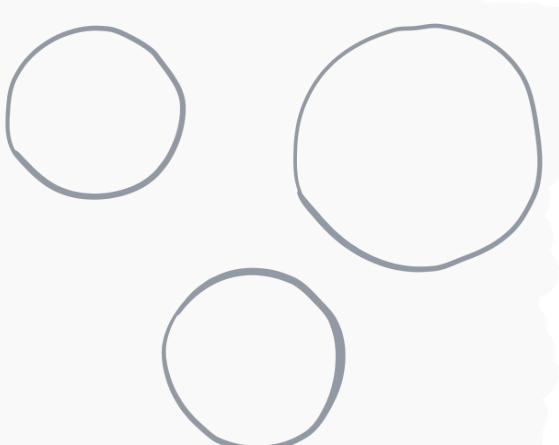
DNA
sequencer

Hello
my name is

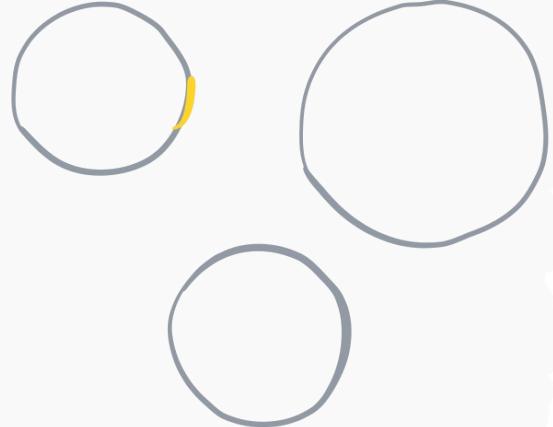
Amplicon
analysis

Relative abundances,
Genomes,
Genes,
Metabolic profiling,
Genomic structure,
Genetic variants...

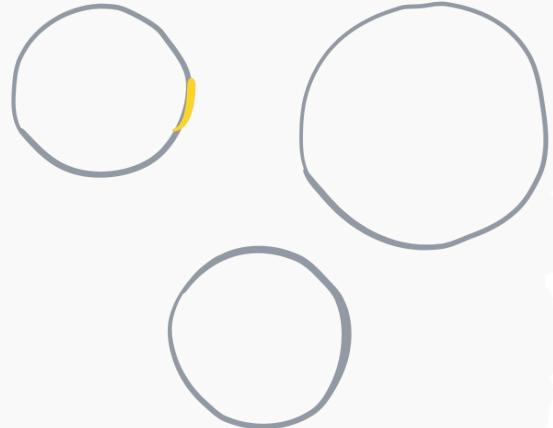
AMPLICON SEQUENCING



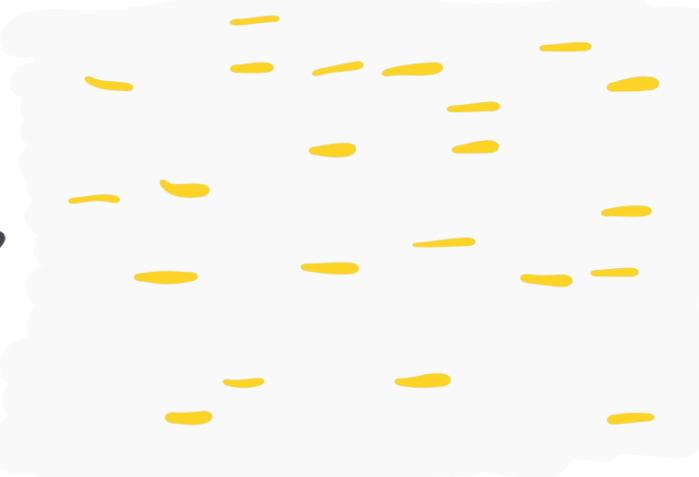
AMPLICON SEQUENCING

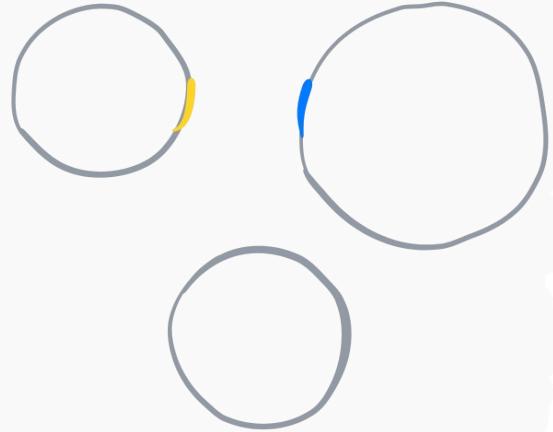


AMPLICON SEQUENCING

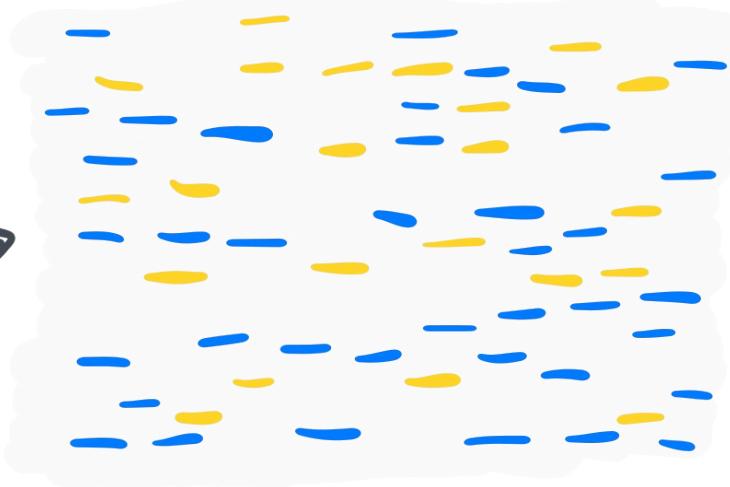


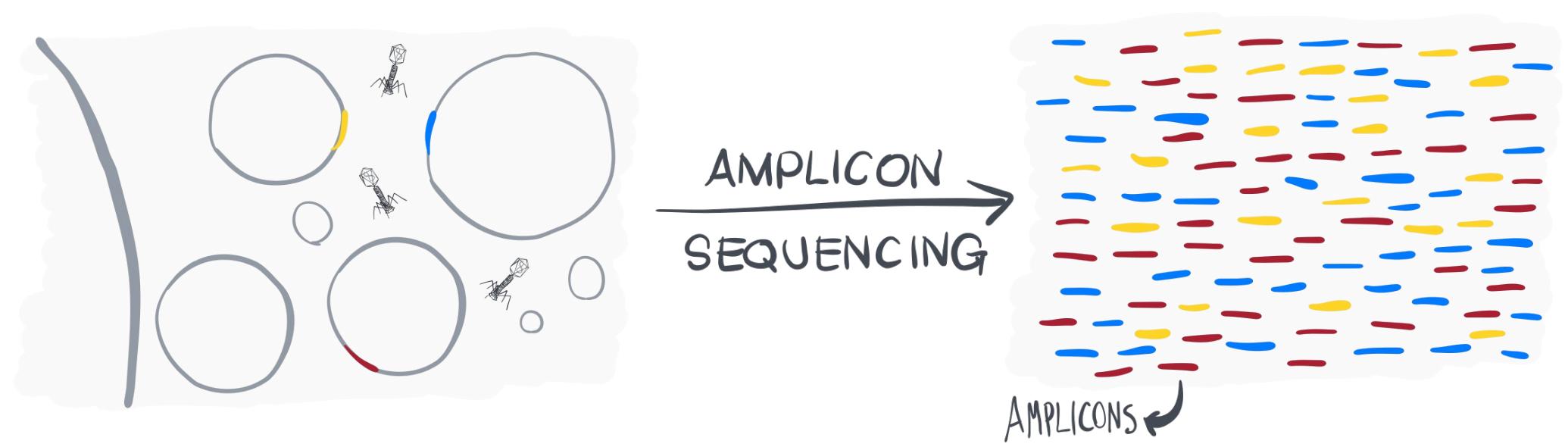
AMPLICON
SEQUENCING

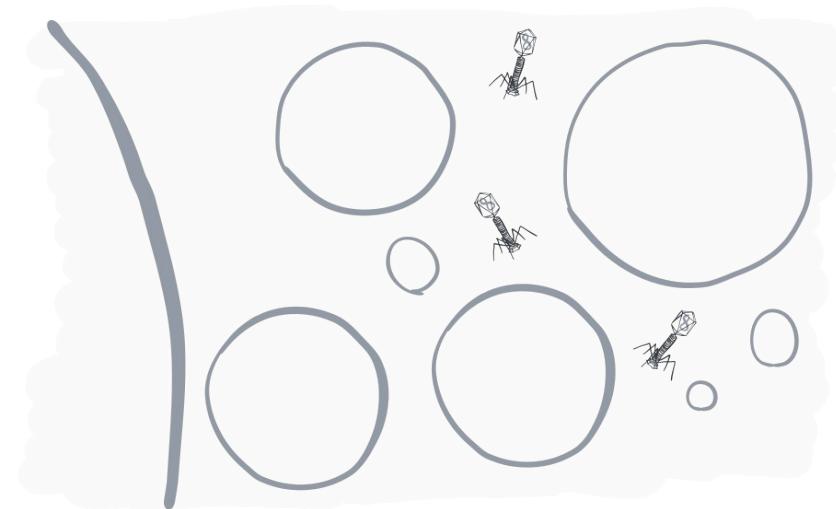


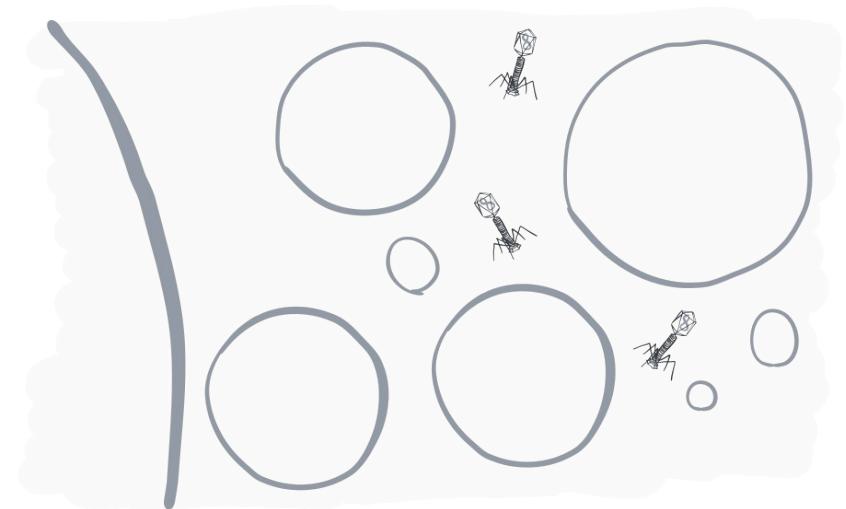


AMPLICON
SEQUENCING

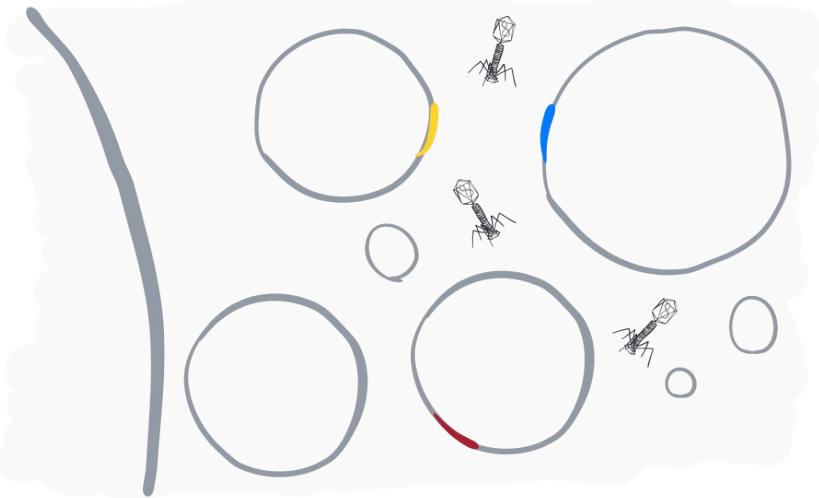




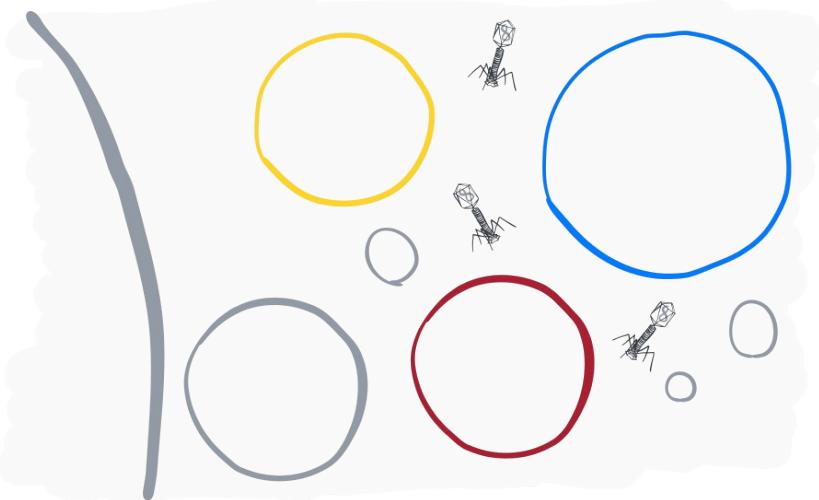
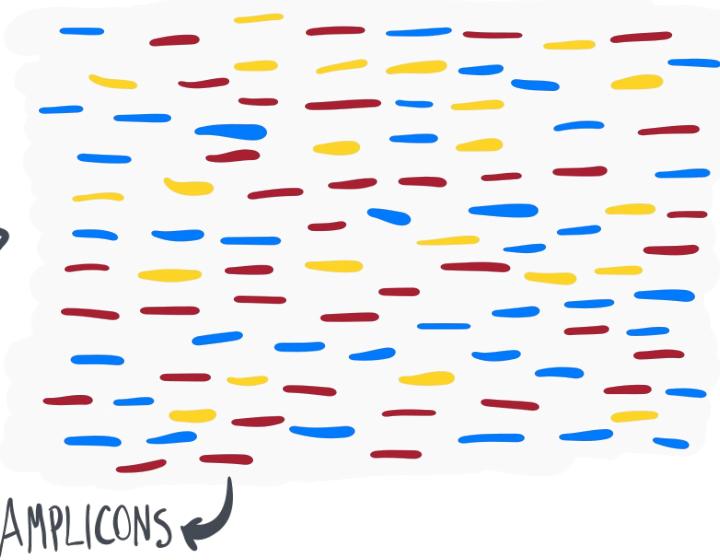




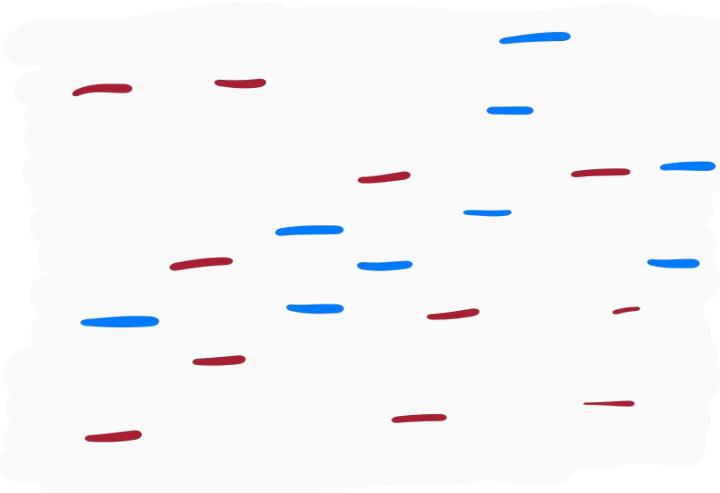
SHOTGUN
SEQUENCING

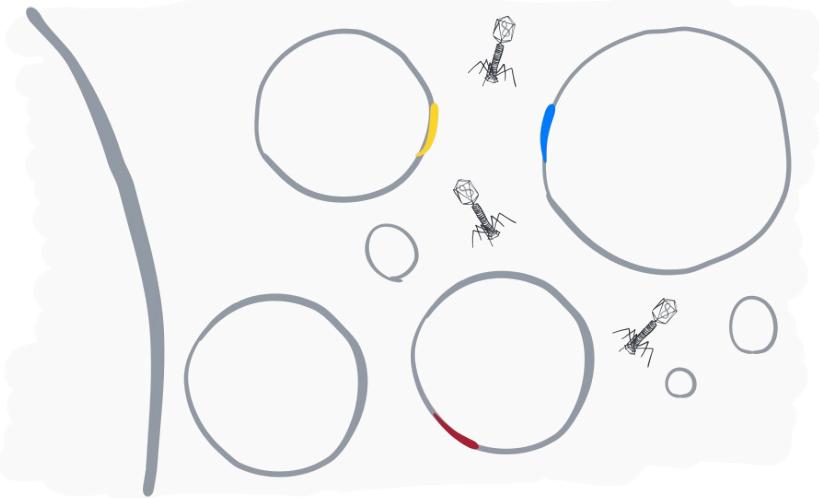


AMPLICON
SEQUENCING

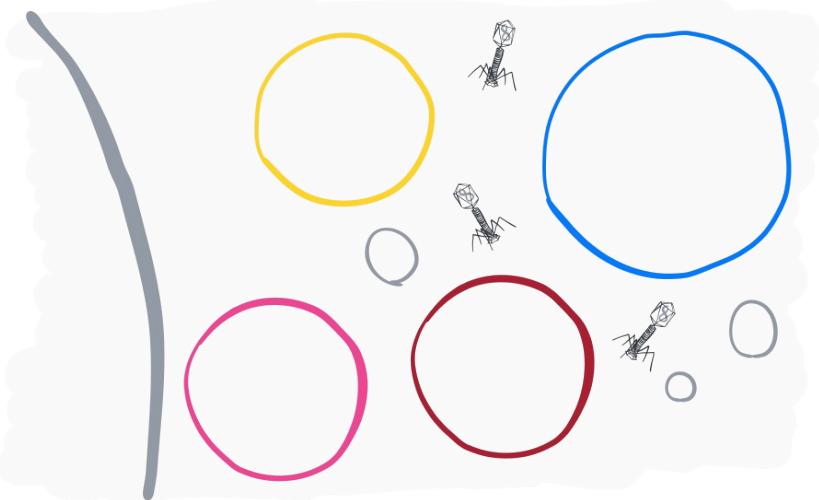
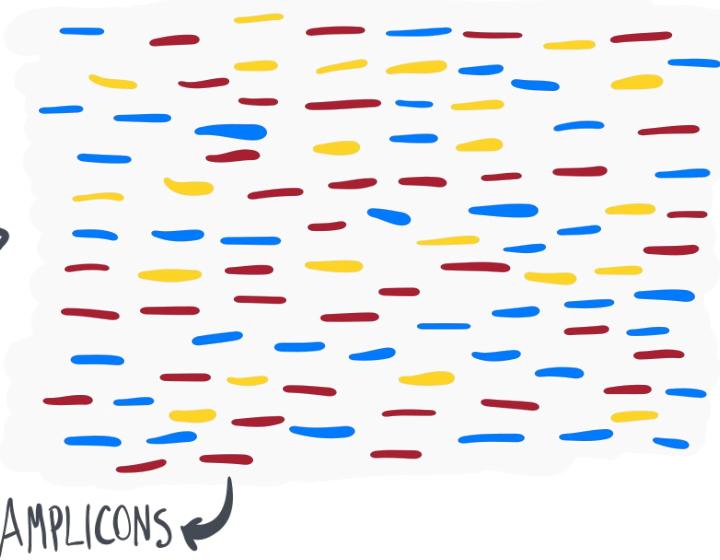


SHOTGUN
SEQUENCING

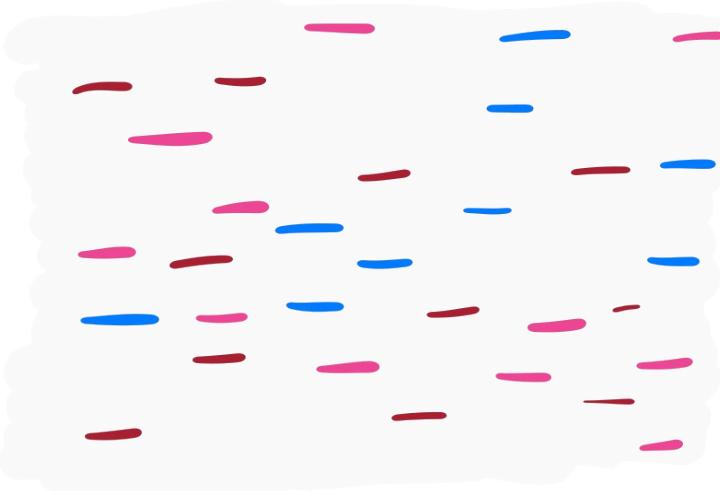


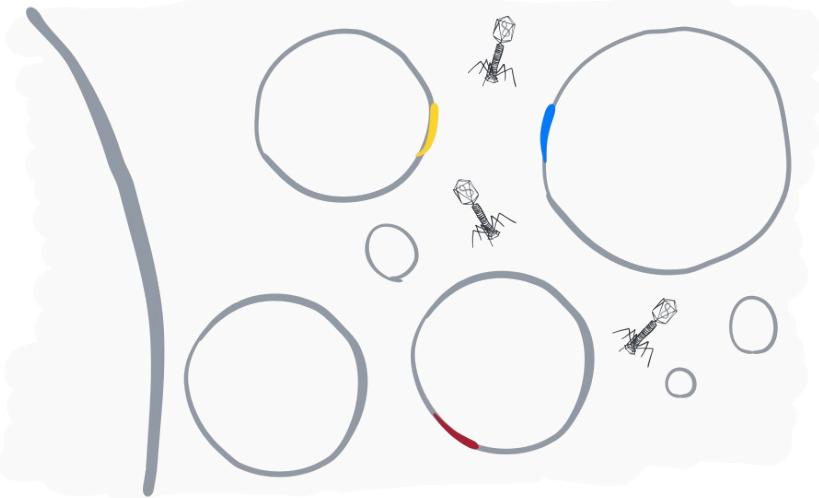


AMPLICON
SEQUENCING

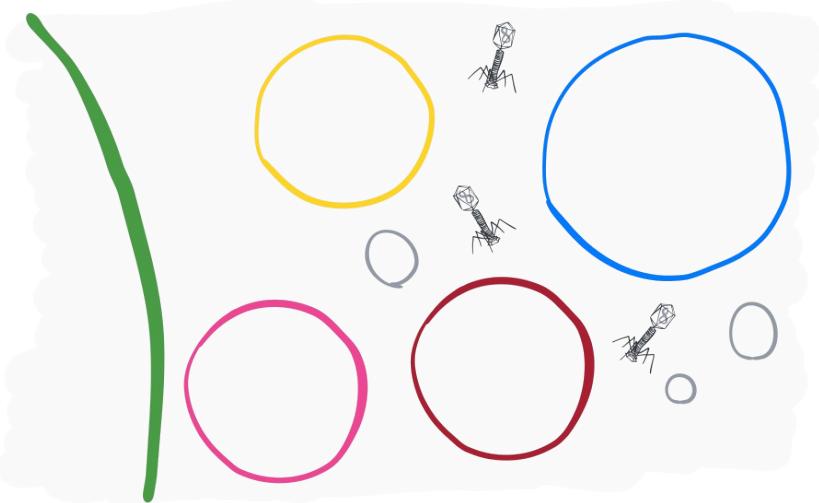
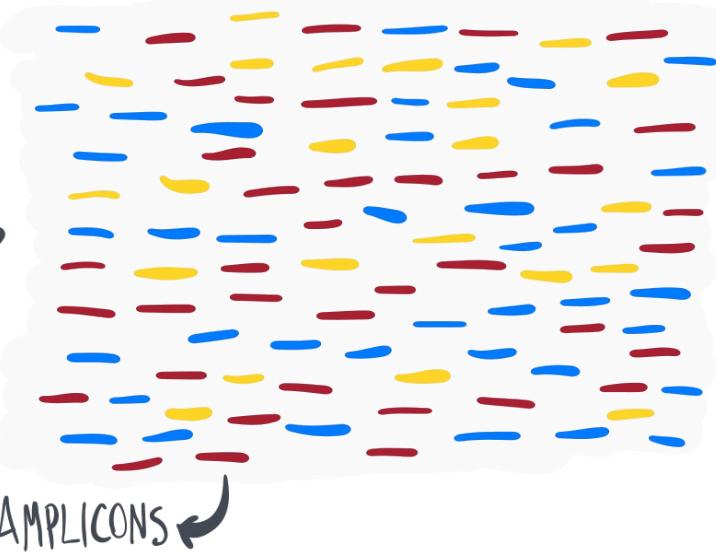


SHOTGUN
SEQUENCING

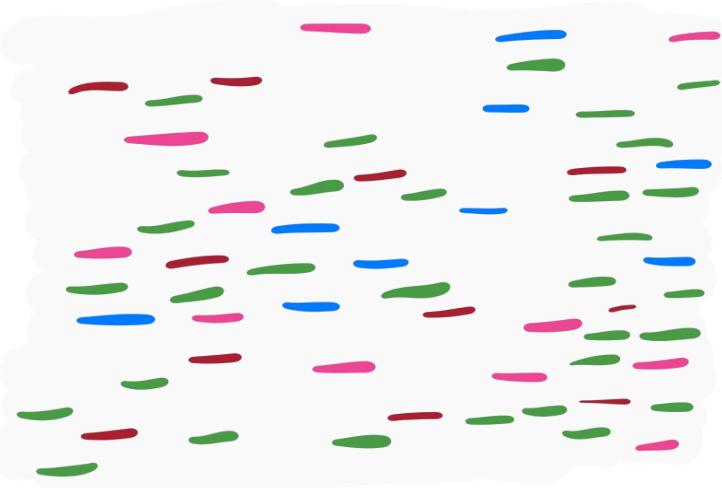


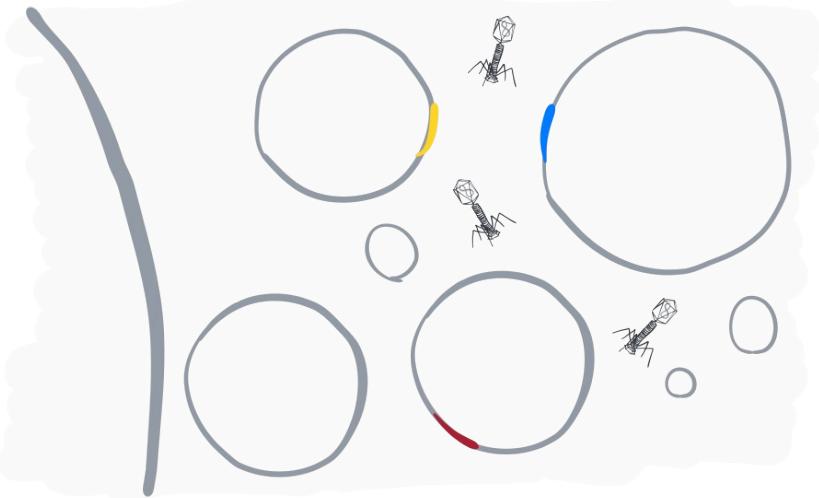


AMPLICON
SEQUENCING

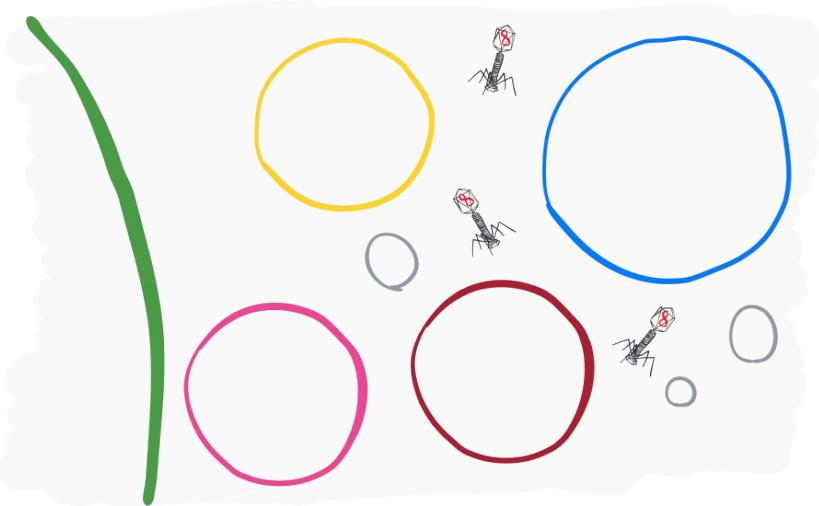
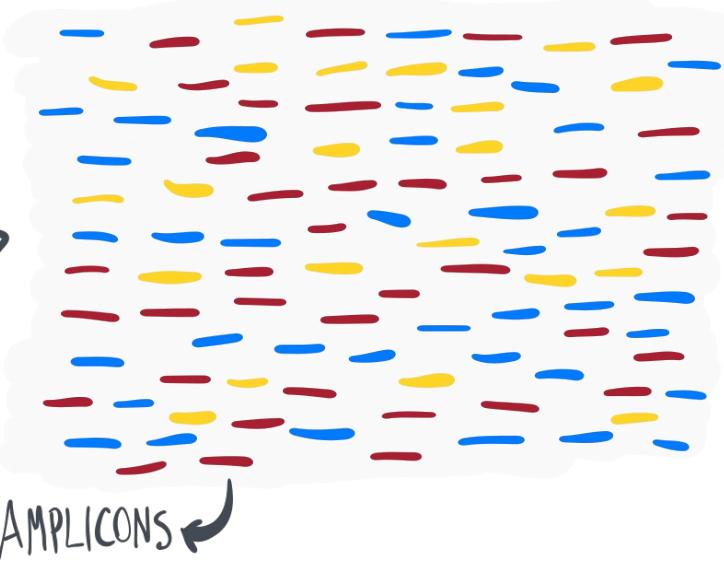


SHOTGUN
SEQUENCING

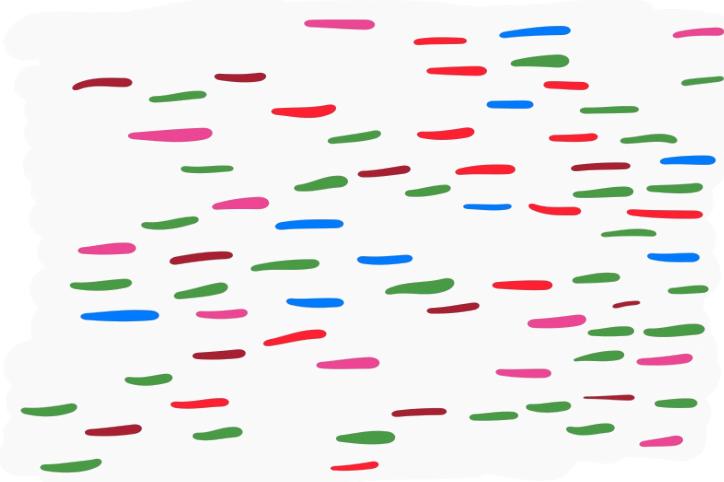


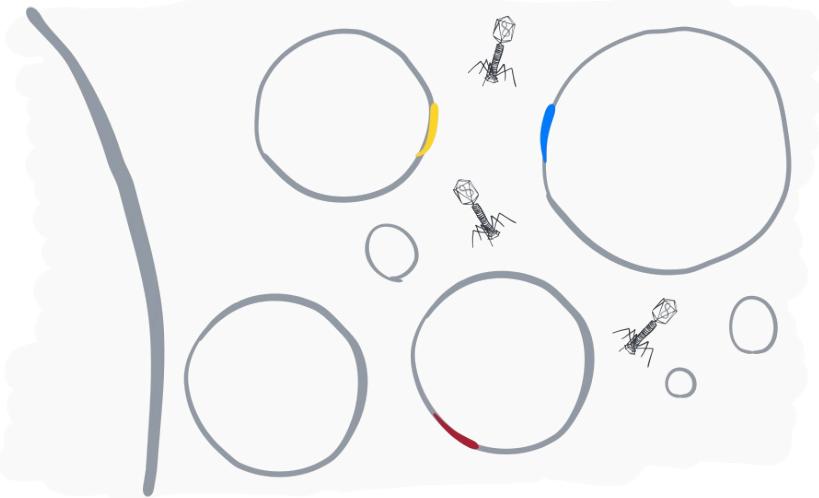


AMPLICON
SEQUENCING

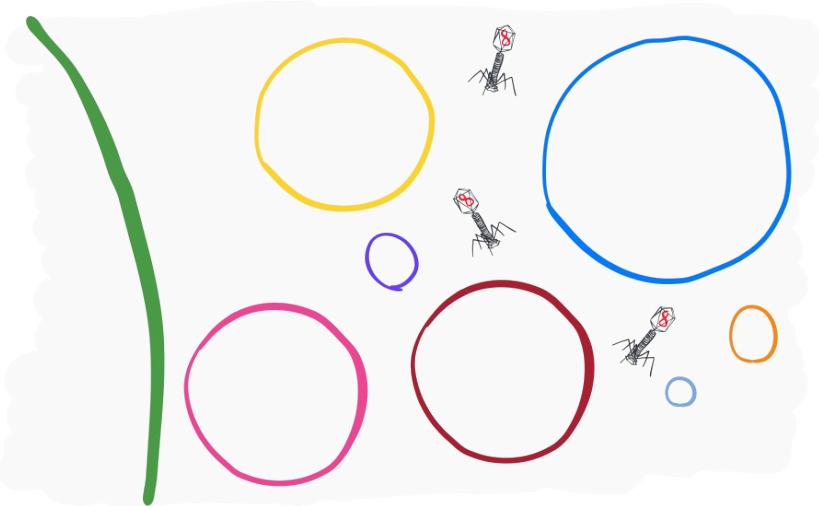
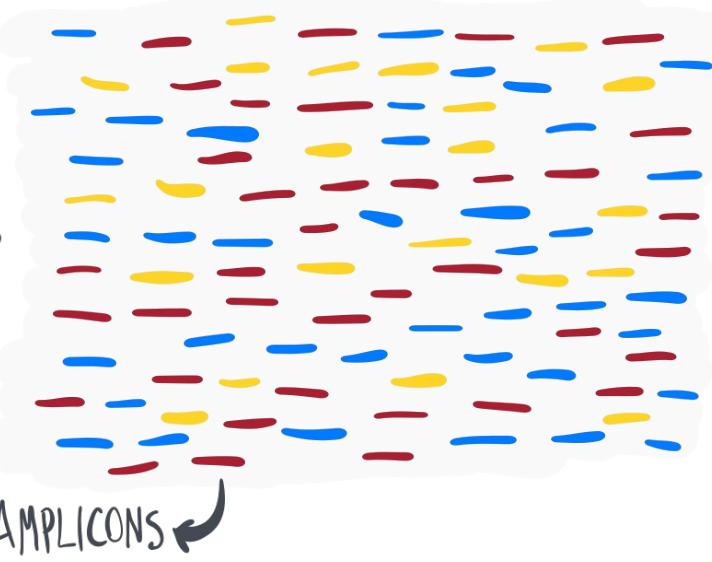


SHOTGUN
SEQUENCING

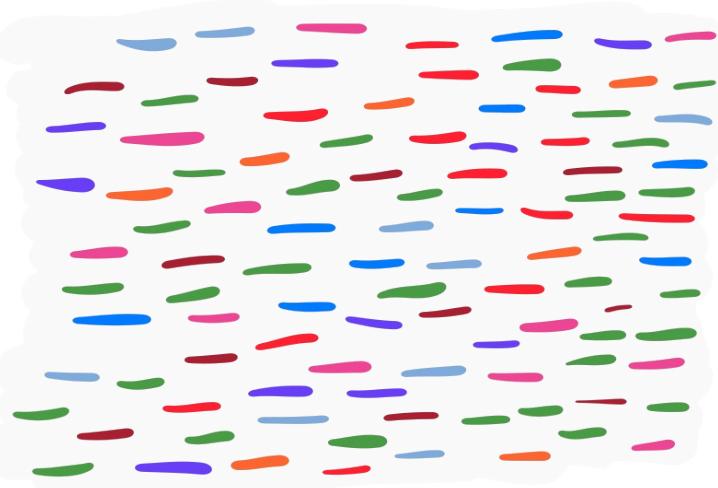


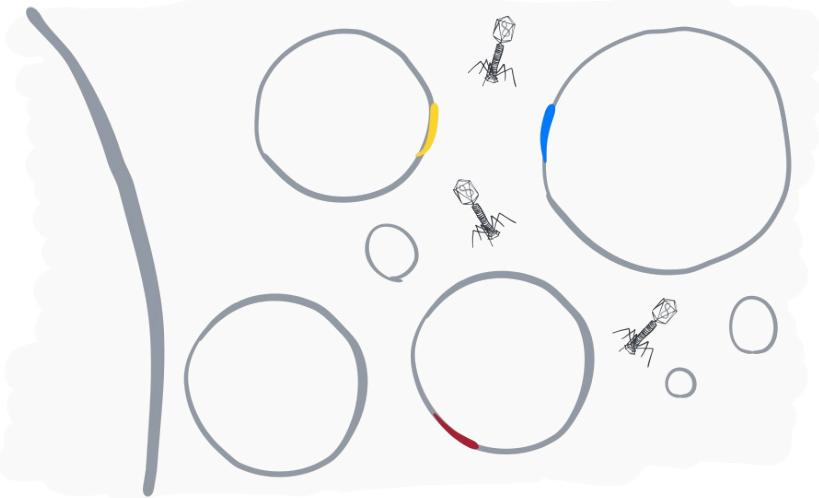


AMPLICON
SEQUENCING

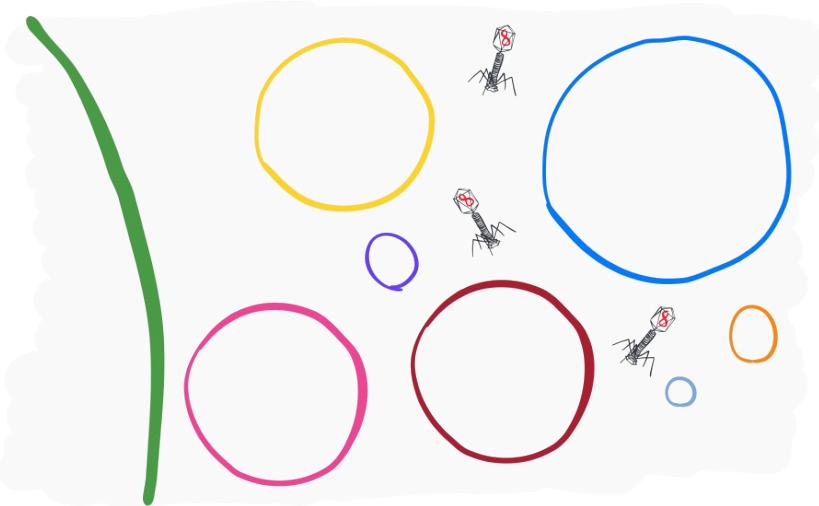
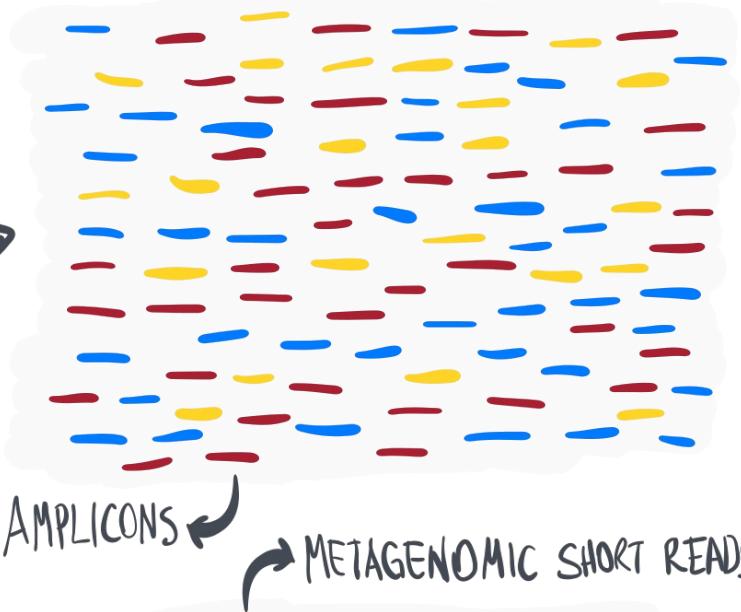


SHOTGUN
SEQUENCING

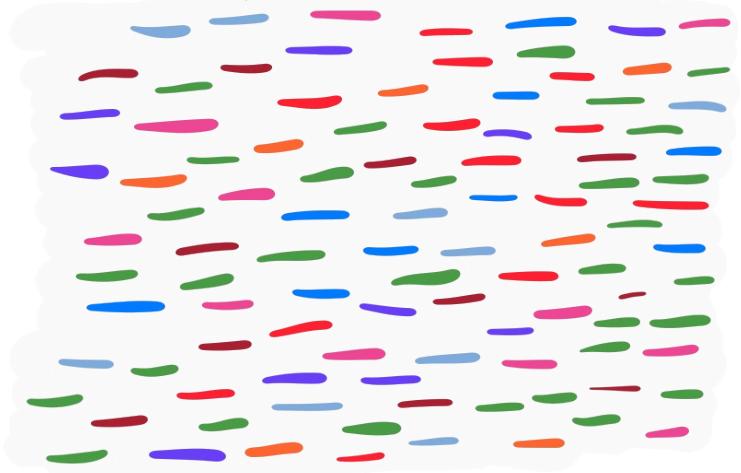




AMPLICON
SEQUENCING

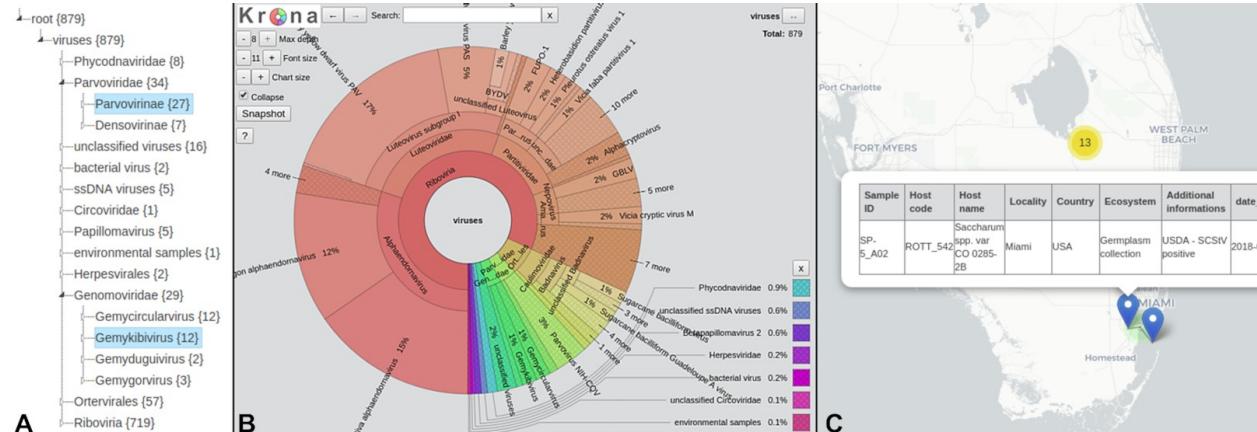
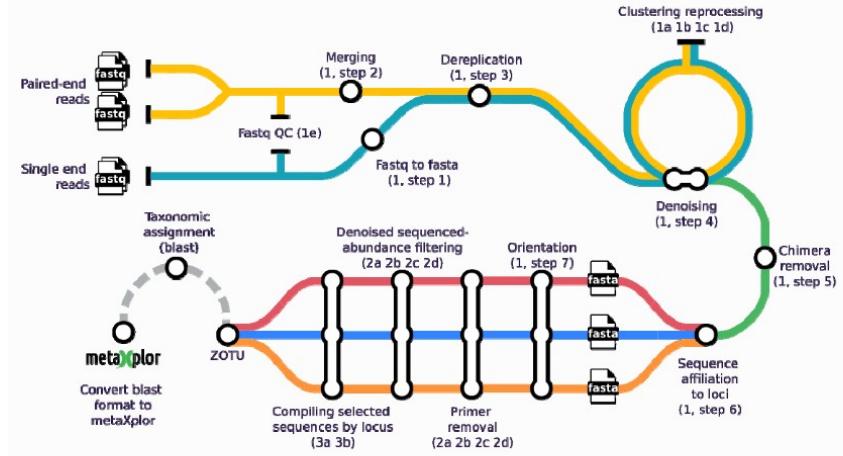


SHOTGUN
SEQUENCING



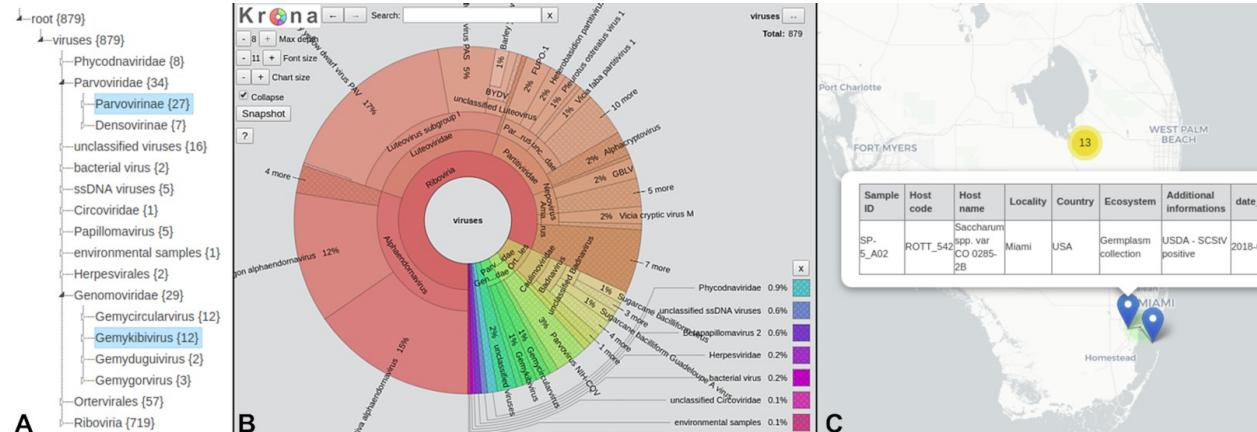
Amplicon sequencing

- Cheap
- Quick
- Resolution to the genus level
- Very good pipelines, Dada2, Qiime, Metaphlan, Vsearch (mbctools package)
- Do not requires lots of CPU and RAM



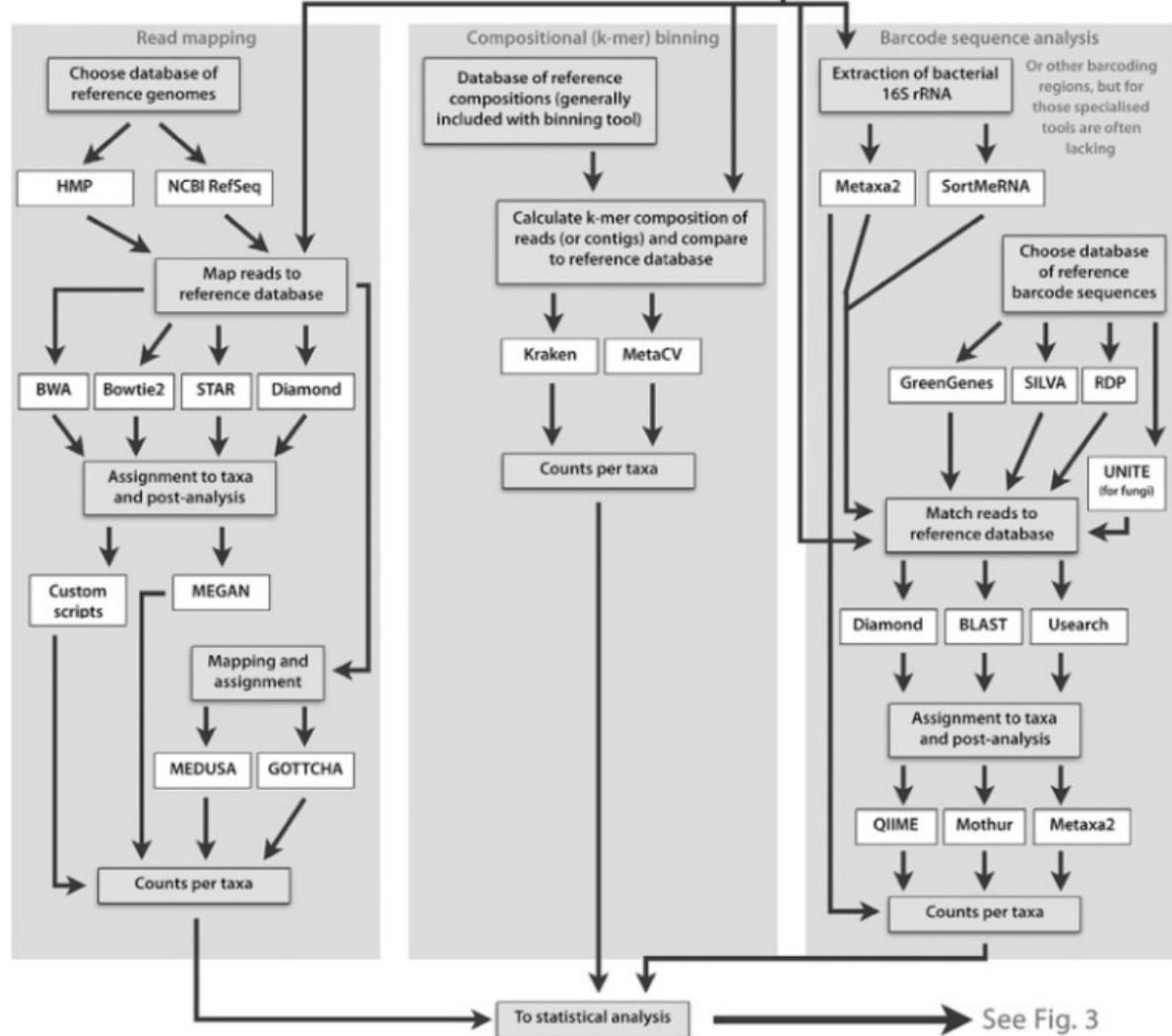
mbctools: A User-Friendly Metabarcoding and Cross-Platform Pipeline for Analyzing Multiple Amplicon Sequencing Data across a Large Diversity of Organisms

<https://doi.org/10.1101/2024.02.08.579441>



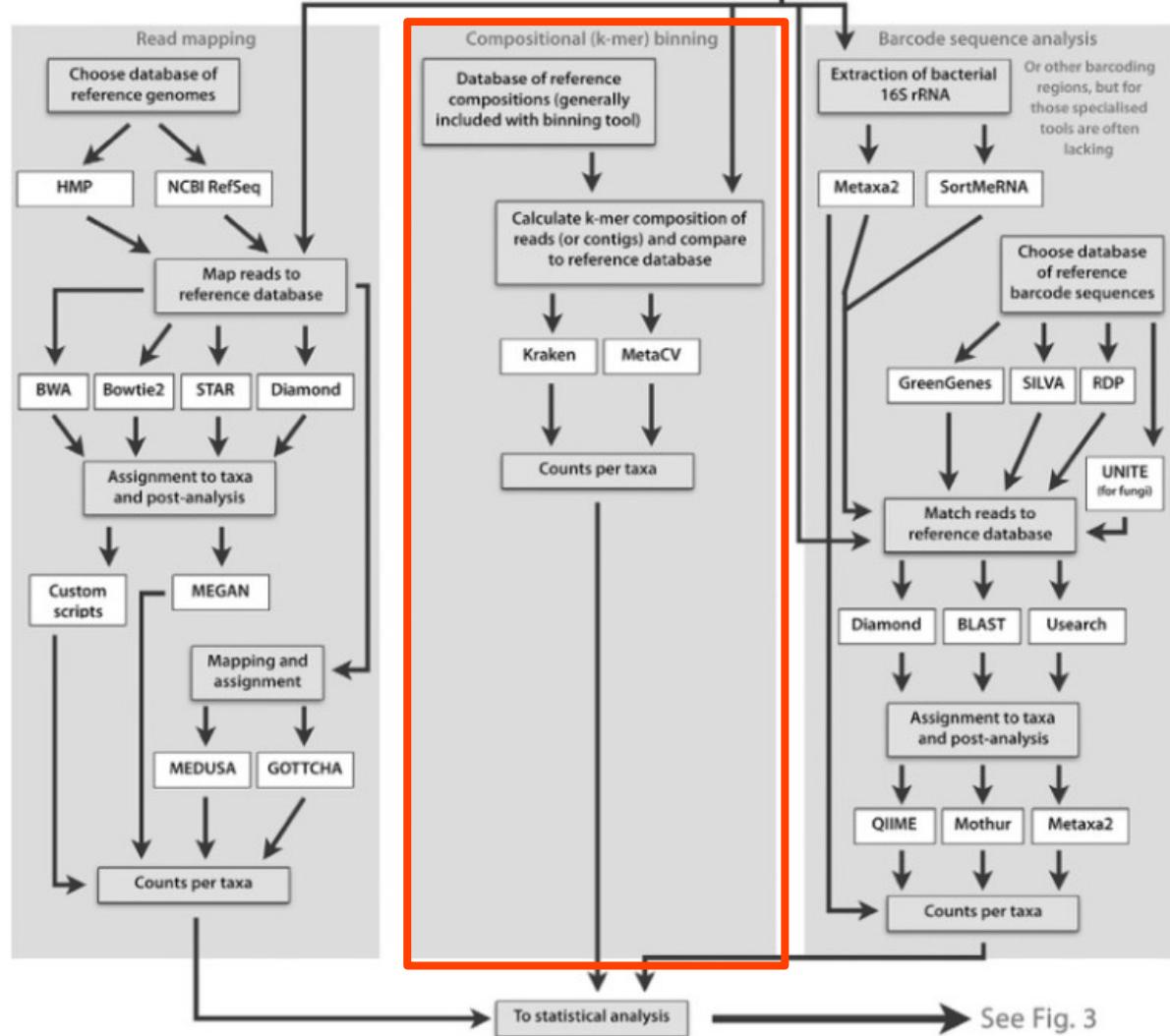
Shotgun sequencing – Read based

- Read mapping and compositional binning
- Analysis of all reads
- Reference database of sequenced genomes
- Mapping: slow, requires lots of CPU and RAM
- Compositional binning: faster but less accurate



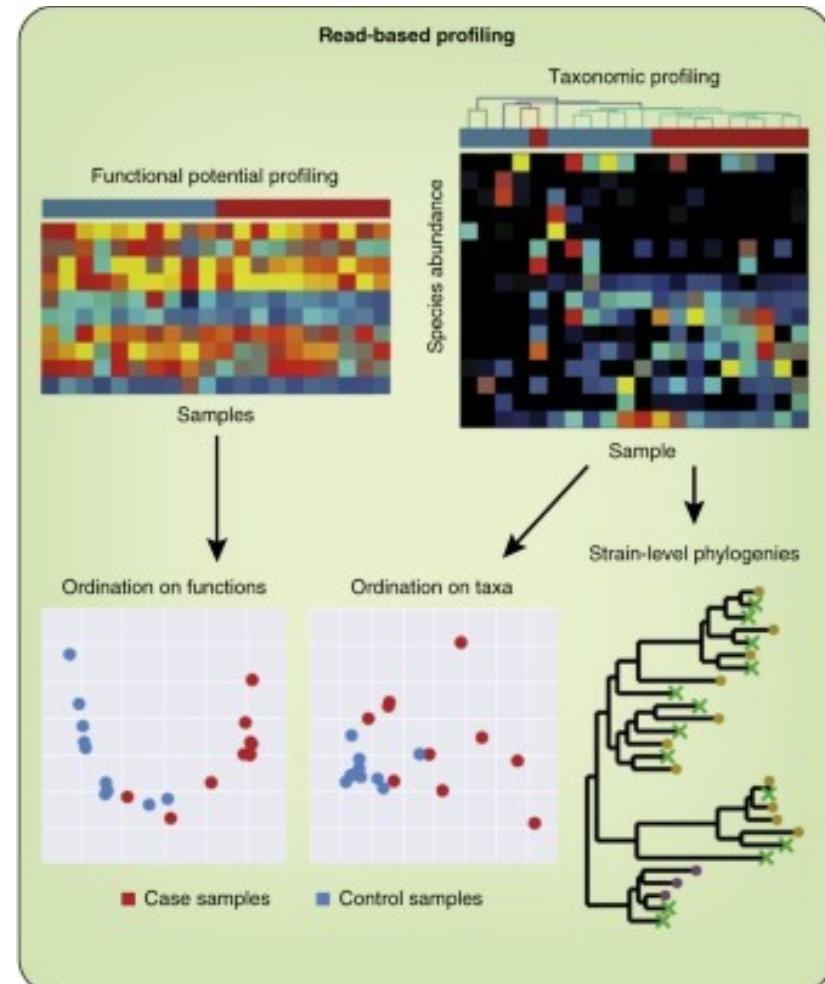
Shotgun sequencing – Read based

- Read mapping and compositional binning
- Analysis of all reads
- Reference database of sequenced genomes
- Mapping: slow, requires lots of CPU and RAM
- Compositional binning: faster but less accurate



Read-based profiling

- Fast
- Quantitative
- Can give interesting preliminary insights
- Usually done as a "quick-and-dirty" estimate prior to assembly



Approaches to taxonomic profiling: how to choose?

Analysis of **all reads** suffer from limited databases of reference genomes

- More suitable for environments that are better described (e.g. human gut)

Analysis of **barcode genes** suffer from lower resolution

- More suitable for environments with a high fraction of unknown microorganisms (e.g. soil)

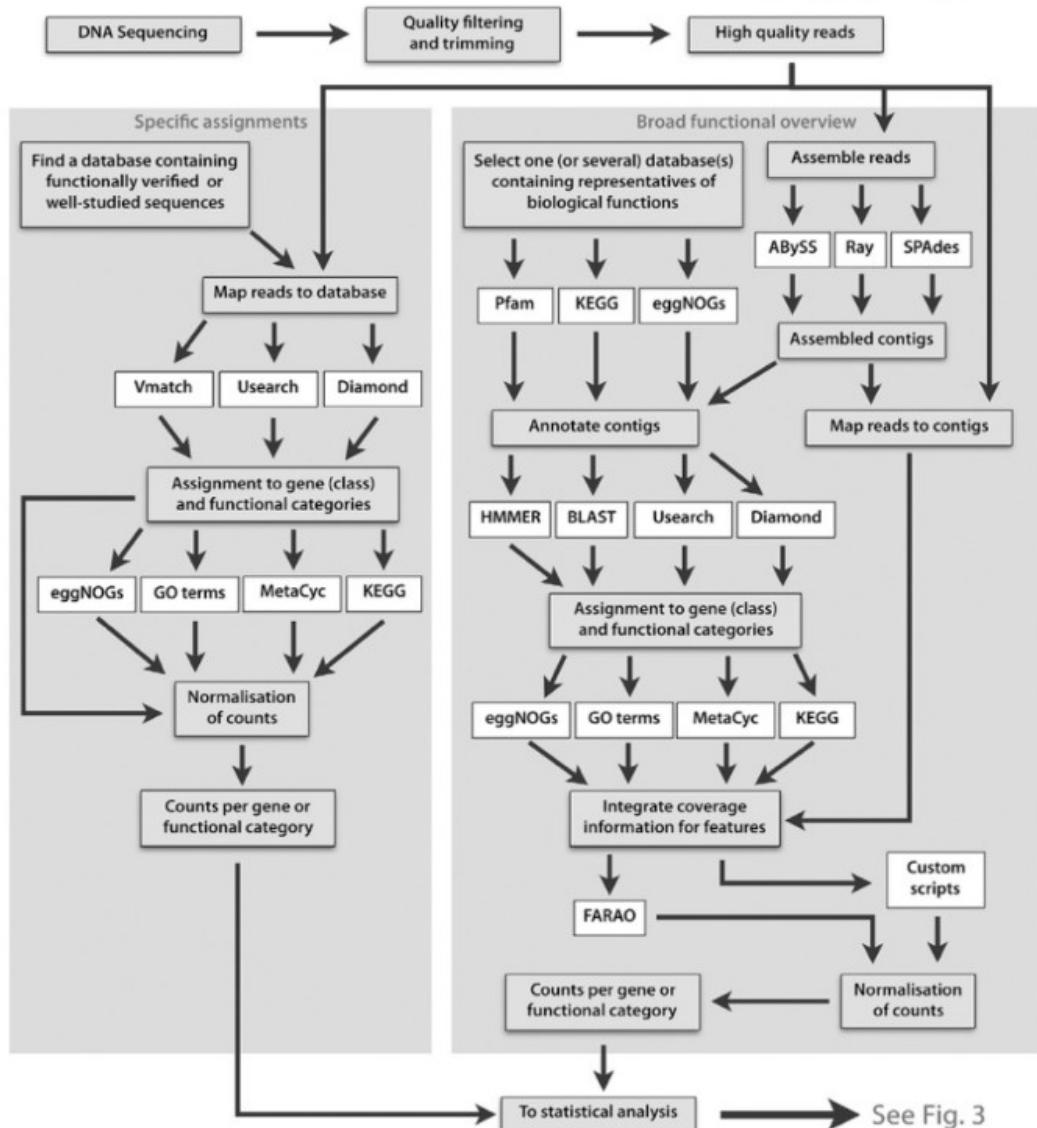
Approaches to functional profiling – genome resolved metagenomics

Broad versus specific profiling

- Broad DBs: entire functional universe (e.g. KEGG, PFAM)
- Specific DBs: focusing on one or few processes (e.g. CAZy, CARD)

CAZy is a database of Carbohydrate-Active enZYmes.

CARD = The Comprehensive Antibiotic Resistance Database



Approaches to functional profiling: how to choose?

Broad databases give an overview of the functional potential of microbial communities

- Suitable for investigating major differences across environments

Specific databases are often highly curated and can give substrate-level information

- Suitable for investigating e.g. gene variants across environments

Making sense of read-based analyses

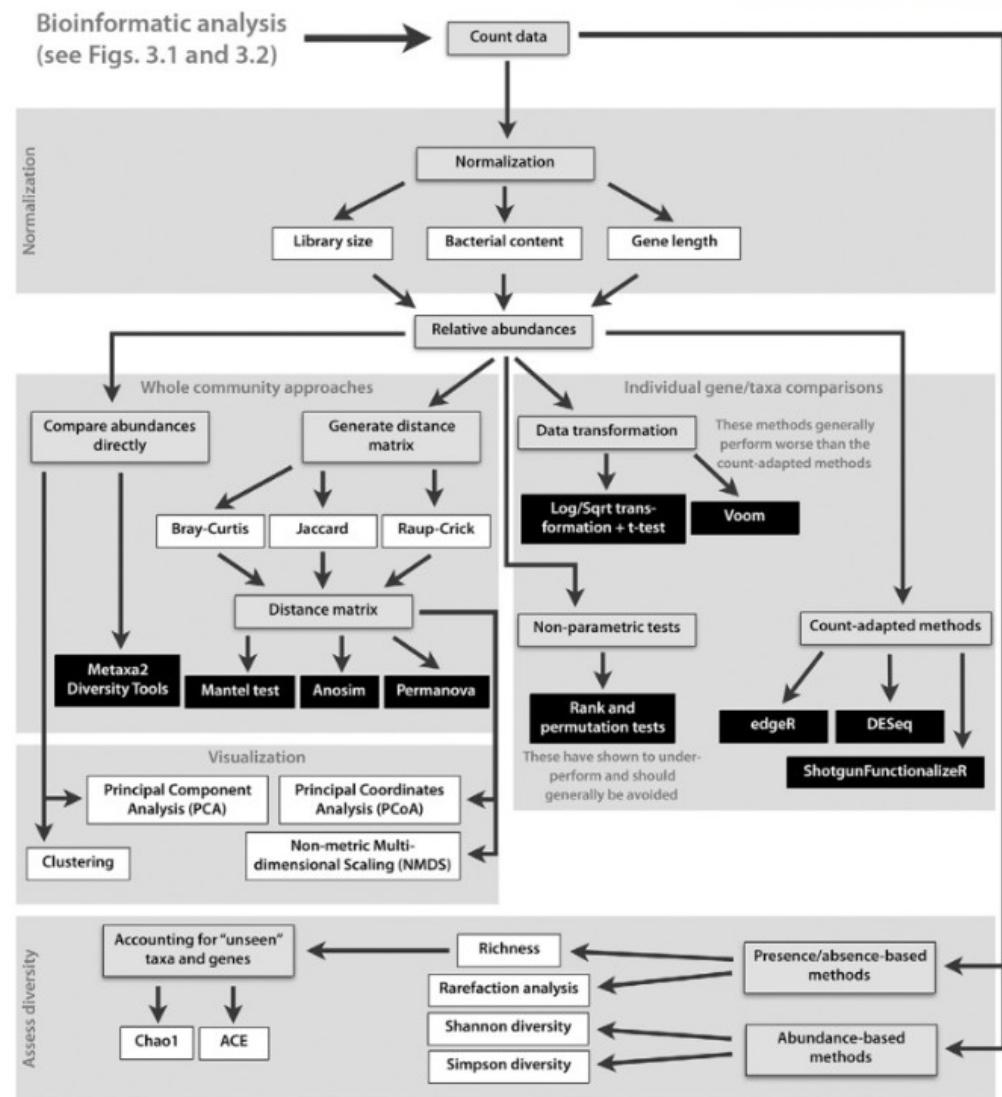
Comparative analyses

Statistics

- Univariate (e.g. ANOVA of specific genes and taxa)
- Multivariate (e.g. PERMANOVA, ordination/clustering, Mantel test)

Normalization!

- Library size
- Bacterial content



Making sense of read-based analyses

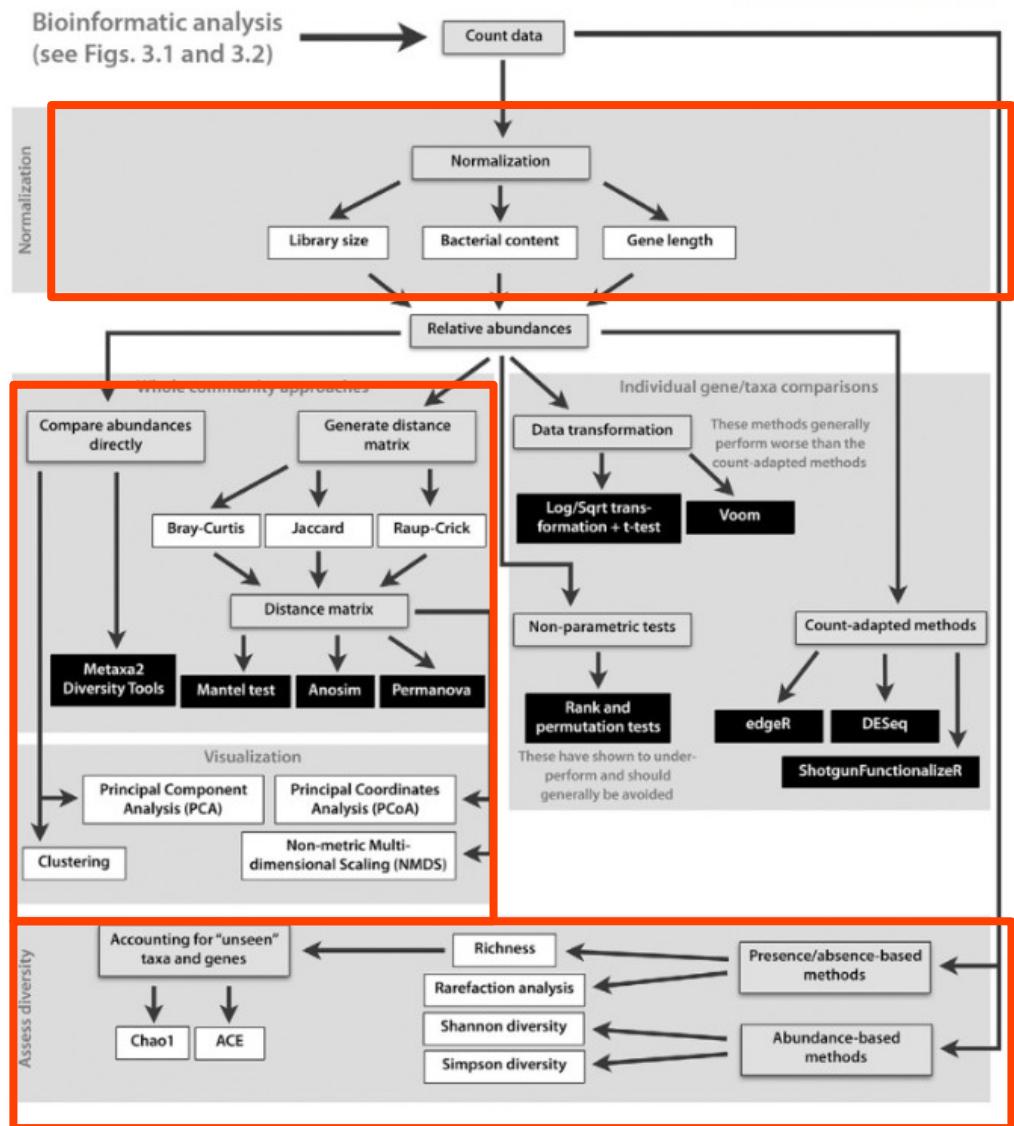
Comparative analyses

Statistics

- Univariate (e.g. ANOVA of specific genes and taxa)
- Multivariate (e.g. PERMANOVA, ordination/clustering, Mantel test)

Normalization!

- Library size
- Bacterial content



Pitfalls of read-based analyses

Curation level of the database

- Are sequences verified experimentally to perform the expected function?

Comprehensiveness of the database

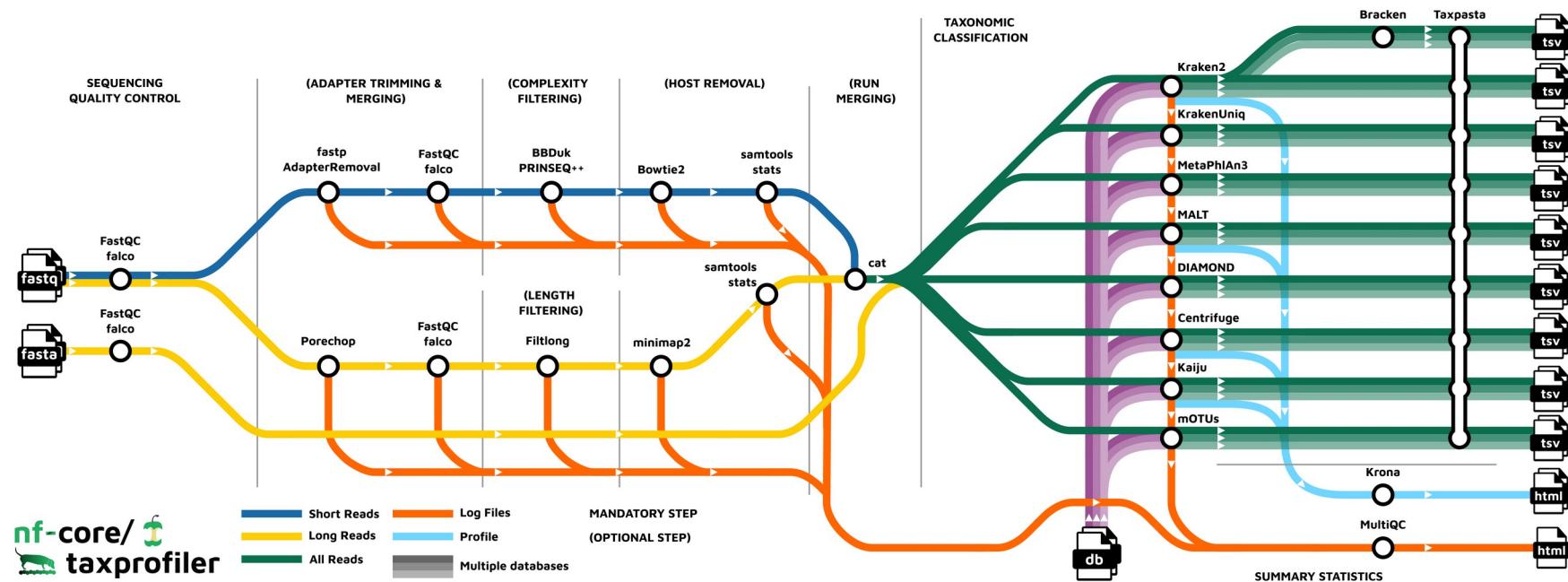
- Both taxonomic- and functionally

Remember: always sanity check!

Especially for unexpected findings:

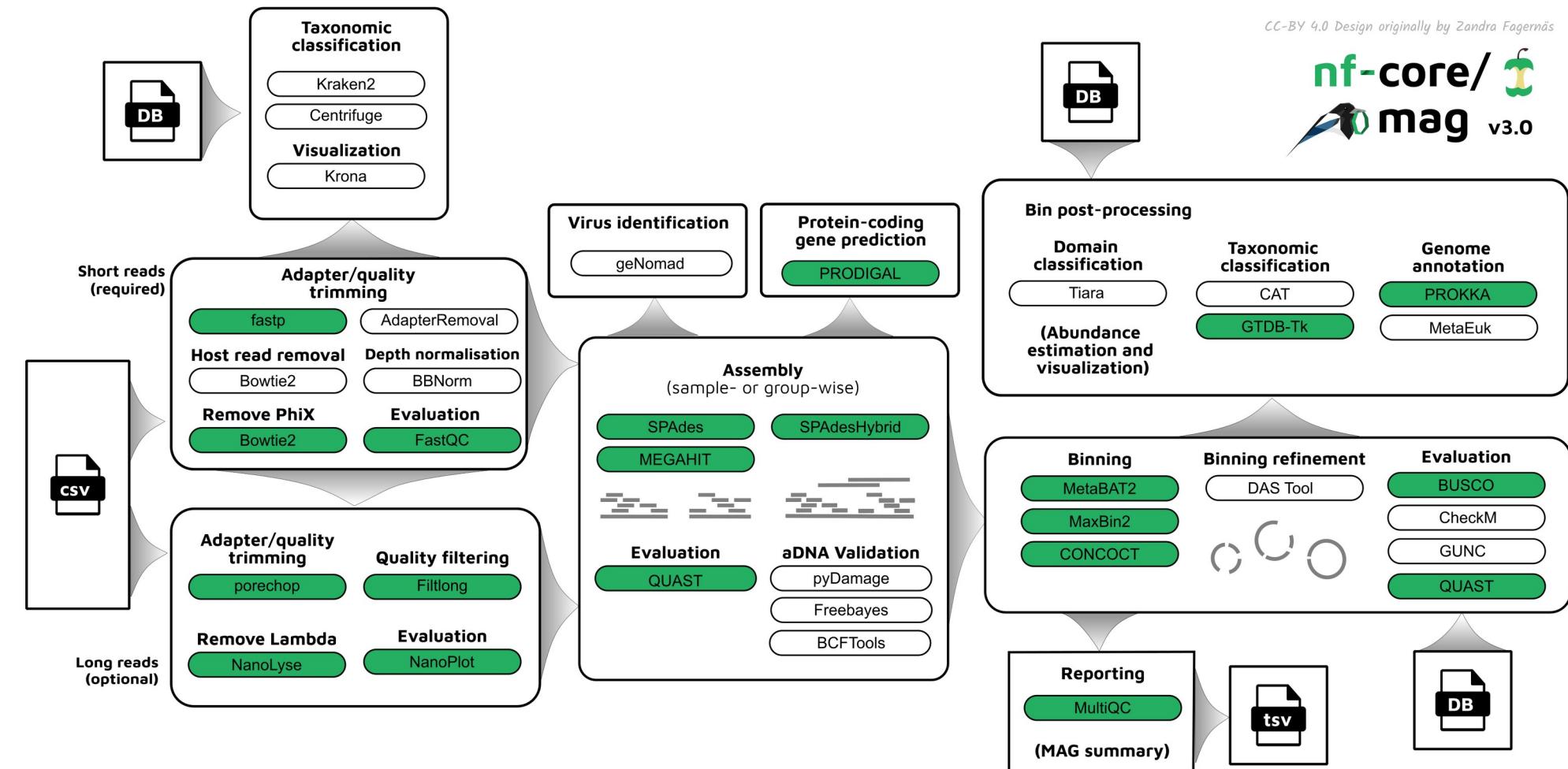
- Redo with more strict thresholds
- Redo with a different tool and database
- Investigate other genes belonging to the same pathway

Pipeline: read-based analyses



Pipeline : genome resolved metagenomics

CC-BY 4.0 Design originally by Zandra Fagernäs



Pipeline : genome resolved metagenomics

