Metagenome assembly
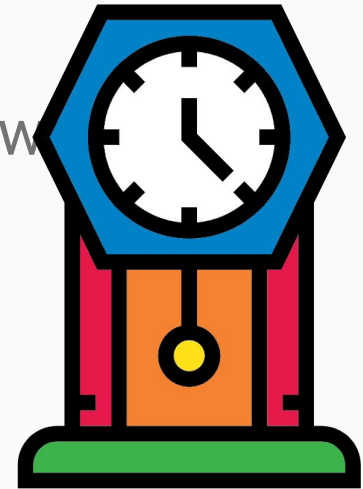
Format & Schedule

This intro

Bookdown

Theory

Practice

Exercises

Work at your own pace

We are here to help

Time with breaks in betw

- 10:00-11:15
- 11:30-12:30
- 13:30-14:45
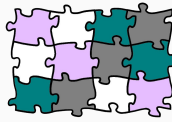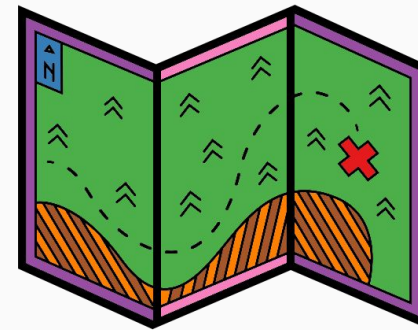- 15:00-16:00

Plan

Assembly approach

   Stitch reads, assembly, binning + annotation

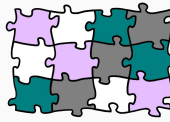Read vs Assembly approach
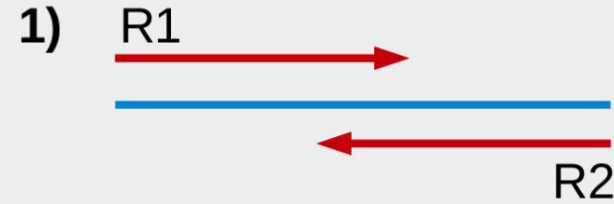
Visualisation?

Eukaryotic analysis
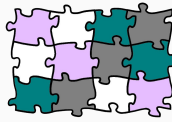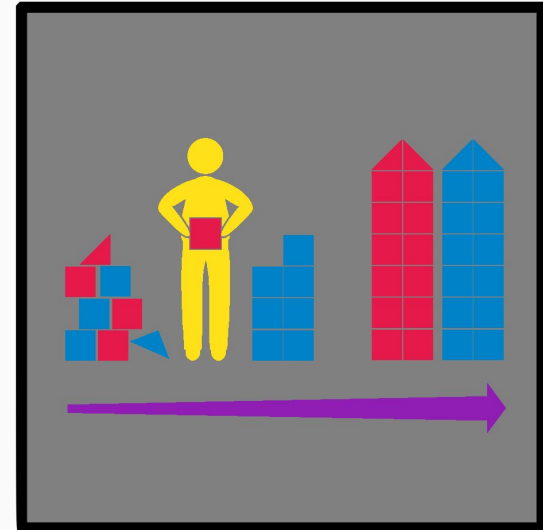
Short read vs long read

Stitch reads

- FLASH
- Get longer reads
- Only if reads have overlap
- Useful for assembly

Metagenome assembly

- MEGAHIT
- Iterations of assemblies
- Creates k-mer sets from reads
  - 29-mer
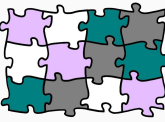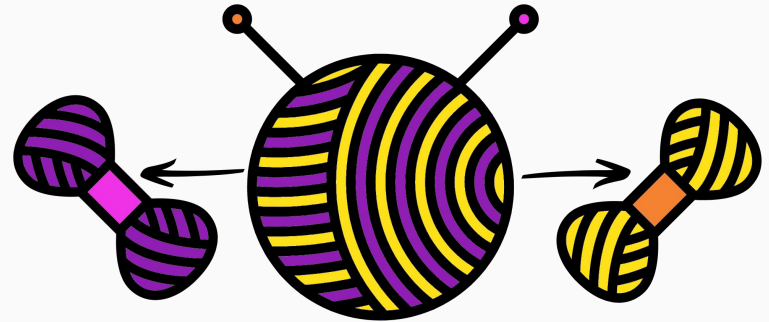 - 49-mer
  - 69-mer
  - 89-mer
  - Etc.

Genome binning

Clusters contigs
- Coverage
- Composition

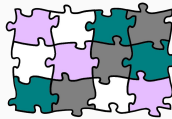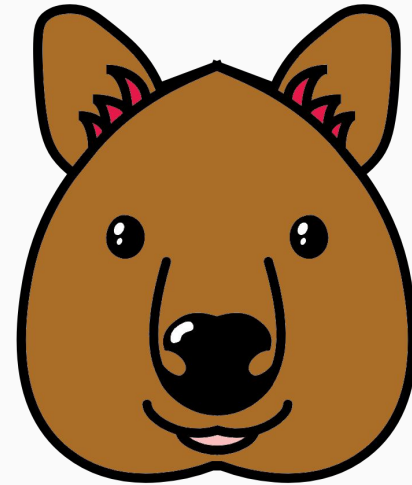Metagenome-Assembled Genomes (MAGs)

Dark matter

Functional annotation

Prokka
- Bacteria, Archaea, & Virus
- GFF3

MinPath
- EC (Enzyme Commission) numbers
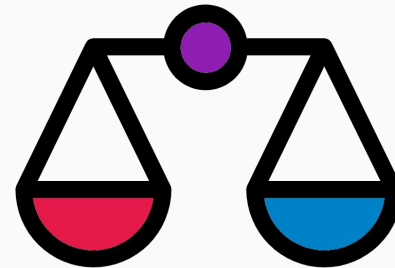- MetaCyc

Read vs Assembly

Read

- Less resources
- Relies on information in databases
- Similar taxonomic annotation as assembly
- More functions retrieved, more over predictions.

Assembly

- Requires large amount of read for good results
- Assembles genes
- Discover microbial dark matter
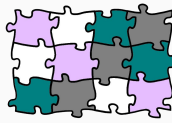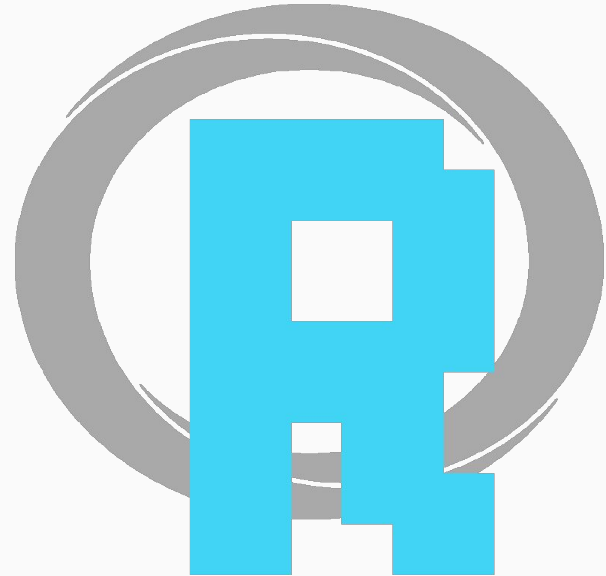
Visualisation?

No QIIME2

MetaWrap is closest
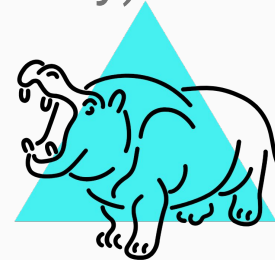
Dependant on dataset

R

ggplot2, Phyloseq

Eukaryotic analysis

- Taxonomic annotation
  - Kraken2 (reads or metagenome assembly)
    - Custom databases
  - Other methods
- Functional annotation
  - Highly complex
  - Resource intensive
  - DB completeness
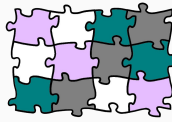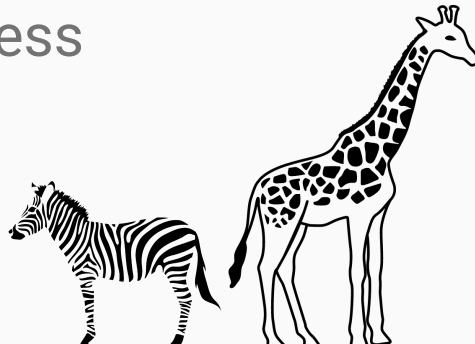 - MetaEuk

Short reads and long reads

Short reads

- Illumina
- Good quality and depth
- One read will only partially cover a gene
- Difficulty in assemblies
- Many tools
- Easier to process

Long reads

- PacBio and ONT
- Lower quality and/or depth
- One read may cover multiple genes or even the majority of a genome
- Massively helps assemblies
- Very few tools

Hybrid approach

- Read approach
  - Just Illumina
- Assembly approach
  - Hybrid using short + long reads
  - Long reads only used for assembly step
- Long read usage still experimental
  - Will improve
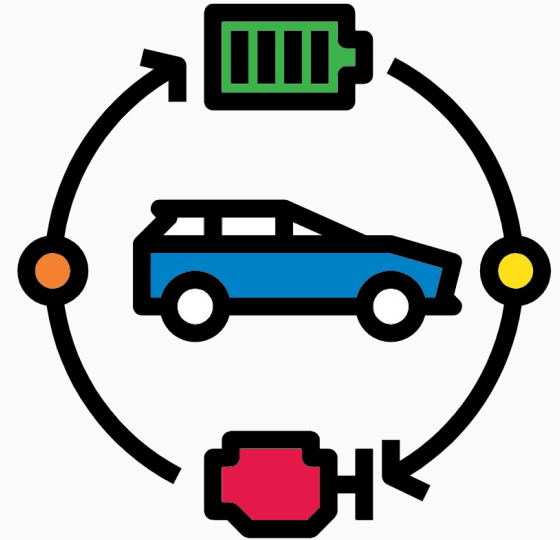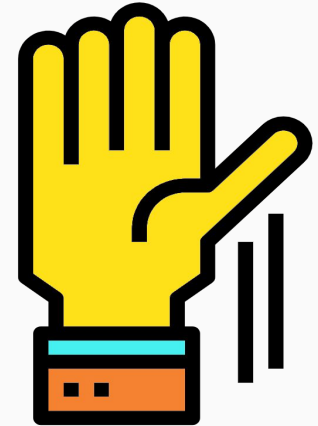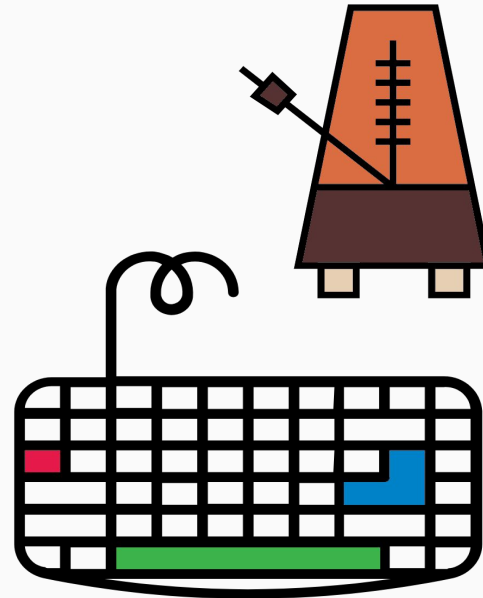
Reminders and Tips

Work at your own pace

Typos

Ask questions

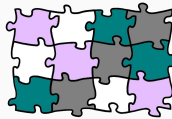Breaks are important

Tab, space, and enter

During sessions

Zoom - Ask via microphone if no question currently being asked/answered

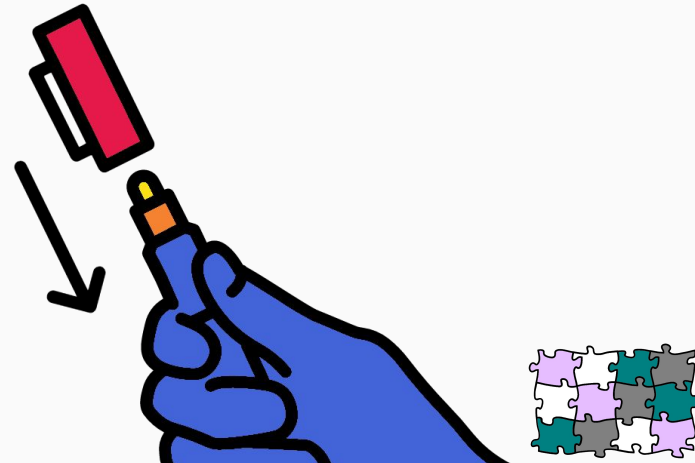Slack - Ask questions via the channel or ask to go into a zoom breakout room with one of us

WebVNC - We can connect to your webVNC to see and help with issues.

Breakout rooms upon request

Recap

- Assembly approach
  - Stitch reads, assembly, binning + annotation
- Read vs Assembly approach
- Visualisation?
- Eukaryotic analysis
- Short read vs long read

Thank you!

Questions?