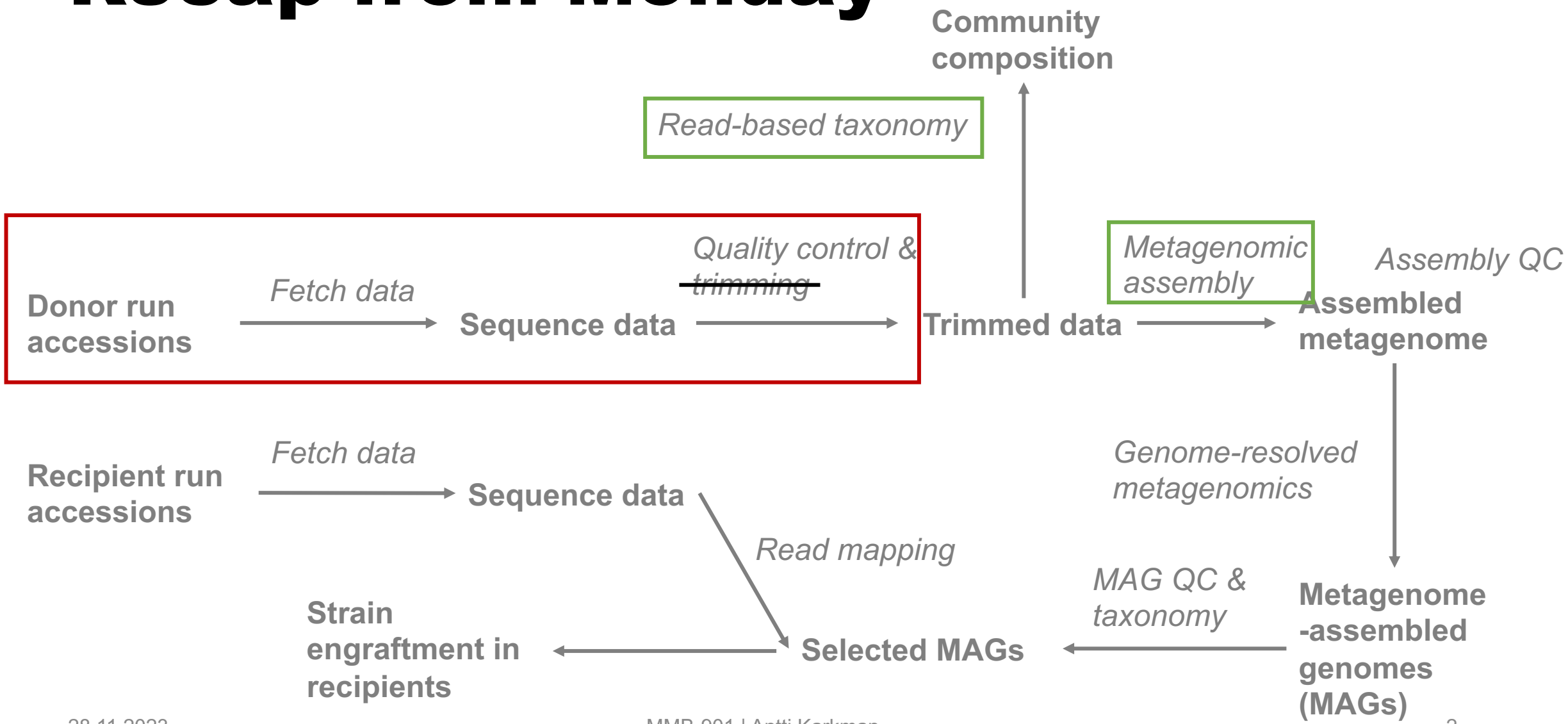
Microbial metagenomics

MMB-901

Recap from Monday

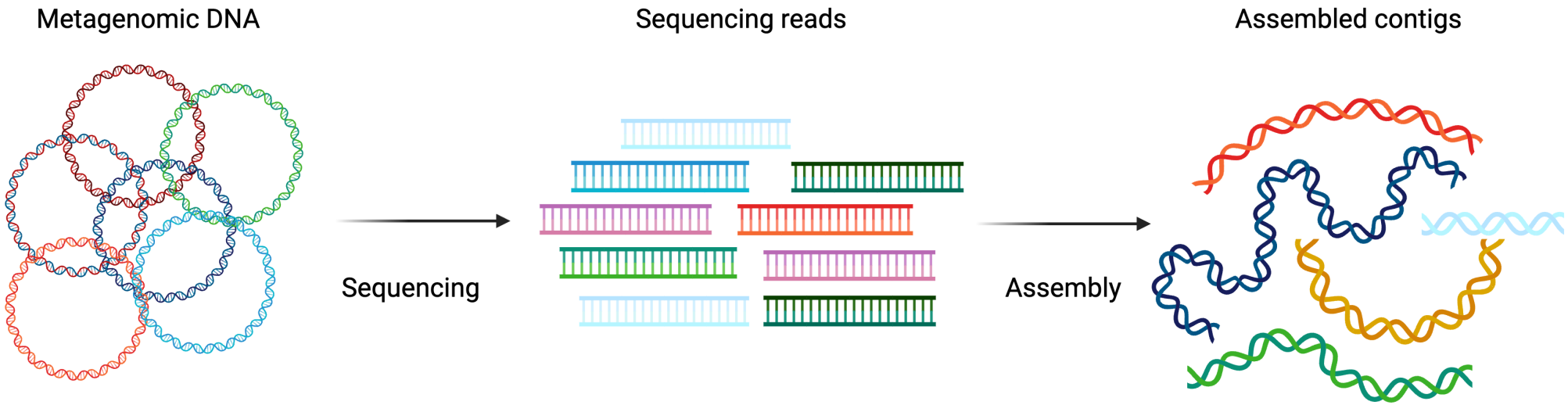


Metagenomic assembly

Approaches and challenges

Metagenomic assembly

- Combine sequencing reads into contiguous sequences (contigs)
- kmers and deBruijn graphs
- Contigs should represent pieces of original DNA in sample



Created with Biorender.com

Challenges in metagenomic assembly

No reference

Repetitive regions

Missing pieces

Multiple copies

Worn-out or dirty pieces

Lots of pieces

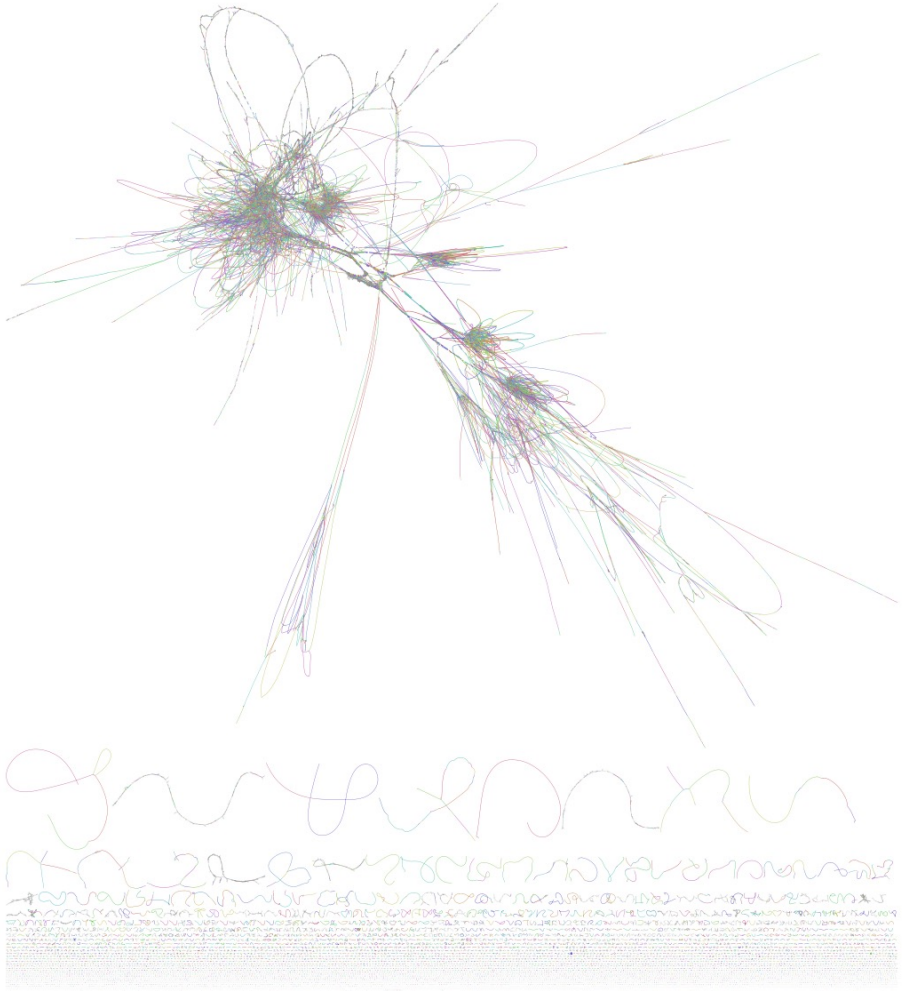
https://en.clementoni.com/products/disney-the-lion-king-60-pieces

What makes assembly tricky

- Many pieces (computational)
- Errors in sequences (which one is correct?)
- Missing fragments (sequencing depth)
- Repetitive fragments (tandem repeats)
- Multiple copies of the same gene (e.g. 16S rRNA gene)
- Conserved genes
- Within species diversity

What to expect from an assembly

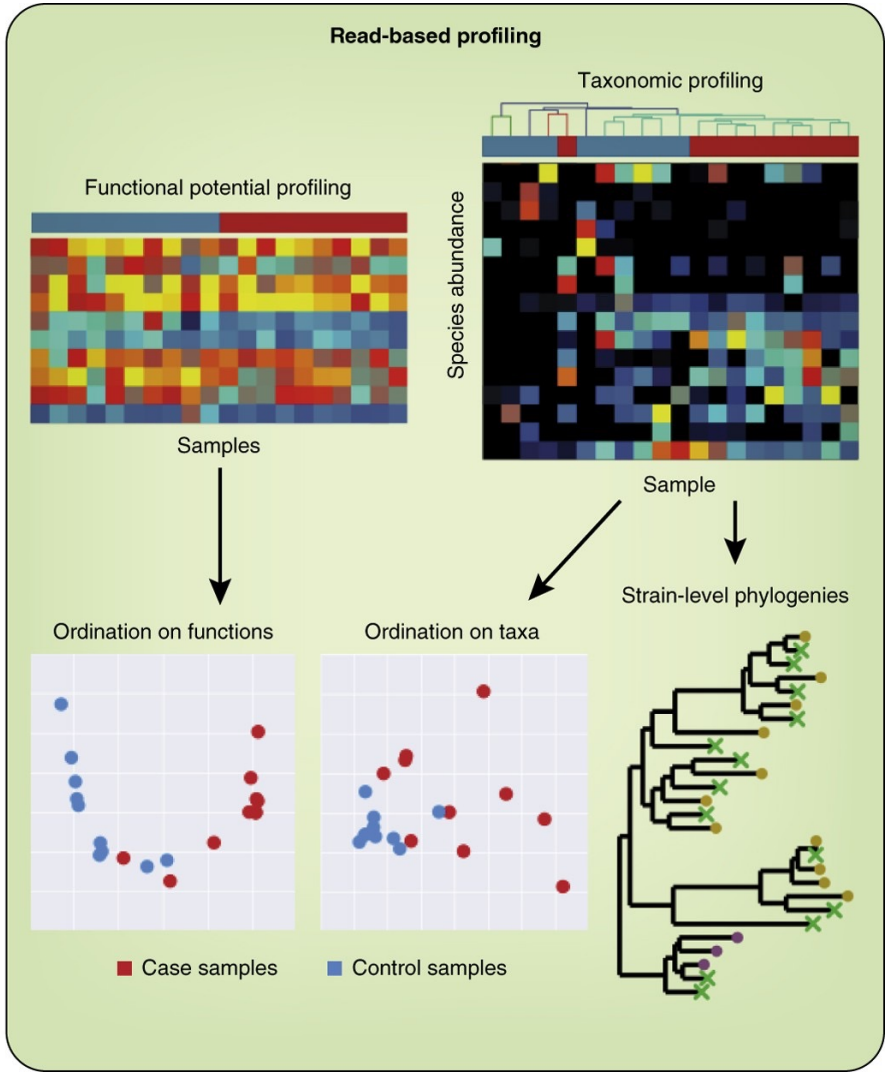
- Only fraction of reads assemble
- Can (and will) contain errors
- Different assemblers will give (slightly) different results



Read-based approaches

Read-based profiling

- Map (compare) reads against a database
  - 16S rRNA or marker-gene-based taxonomic profiling
- Metaphlan4
  - 5.1 M clade-specific markers from ~1 M microbial genomes (isolates & MAGs)
  - species-level annotation
- Functional profiling
  - Broad scope databases (e.g. KEGG)
  - Speficic databases (e.g. AMR)



Let’s get to work

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