SICSEQ – SINGLE CELL METAGENOMICS

Baptiste Avot – Meilee Ling – Judit Szarvas

Main goals

- Benchmark SC metagenomic sequencing
- Improve AMR traceability
- Taxonomic assignment of encapsulated microbes
- Improve genome quality/completeness by refining coassembly
- Identify novel microbes (dark matter) and associated genomes
- Compare with classic shotgun sequencing
- And many others (plasmid dissemination – clonal phylogeny)…

Schedule

- Presentation of the 2 datasets
- The tale of taxonomic assignment and gene traceability
- How deep/shallow SC compare with shotgun sequencing
- Results of spike sample analyses
- Quality assessment of the genomes we get

2 datasets

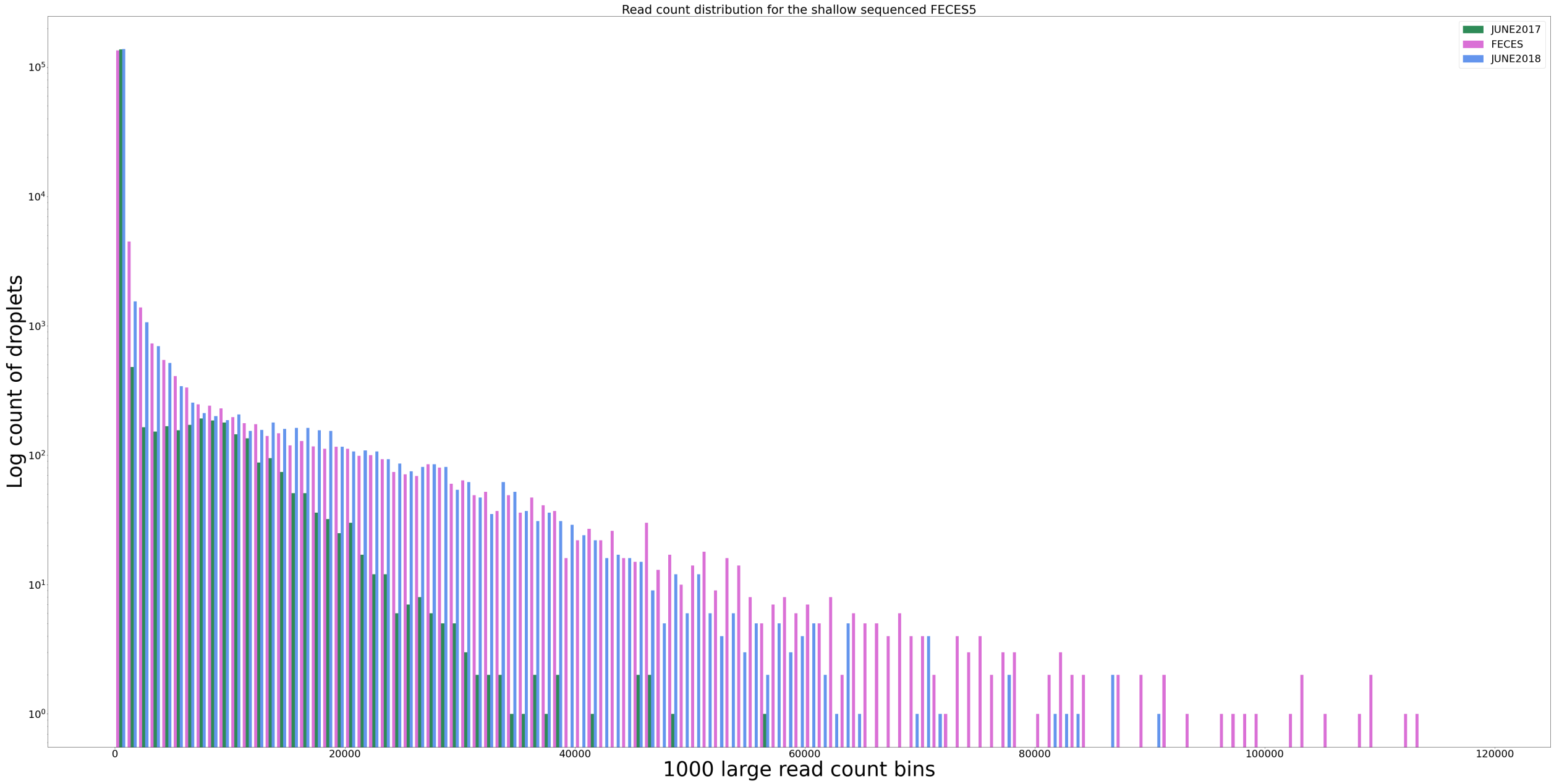
1. Pilot:
   - 3 samples (2 sewage and 1 feces)
     - June 2018 Bangladesh
     - June 2017 Chad
     - Feces: danish pig farm
2. Mock community:
   - 2 samples spiked (1 feces – 1 sewage)
     - Feces (same as in pilot)
     - Sewage (Bangladesh 2018)
   - Mock community: ZymoBIOMICS Gut Microbiome Standard

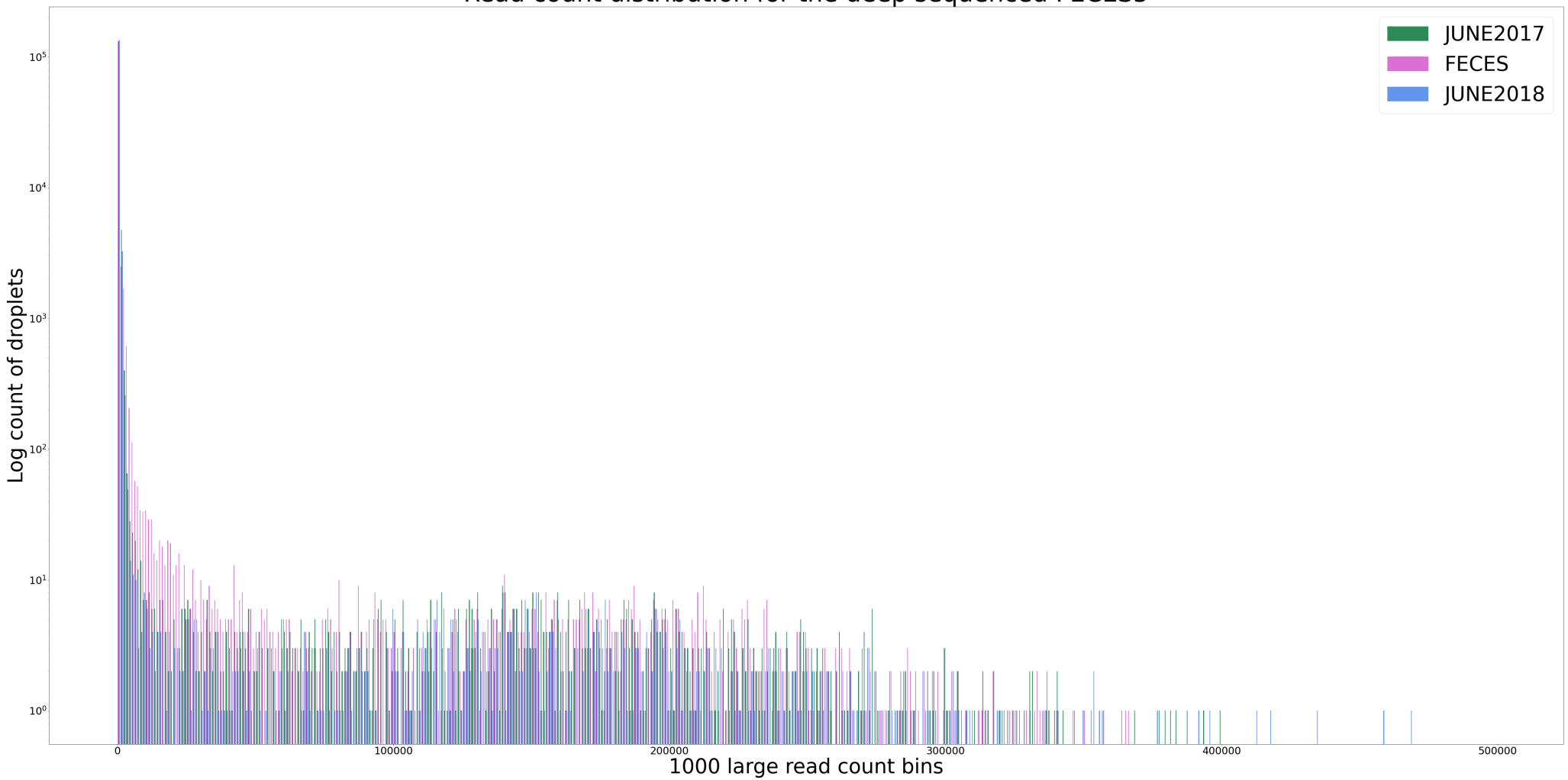
Pilot dataset

- Each sample:
  - Deep sequenced: > 100k reads
  - Shallow: > 10k reads
  - Single end – Illumina 135bps

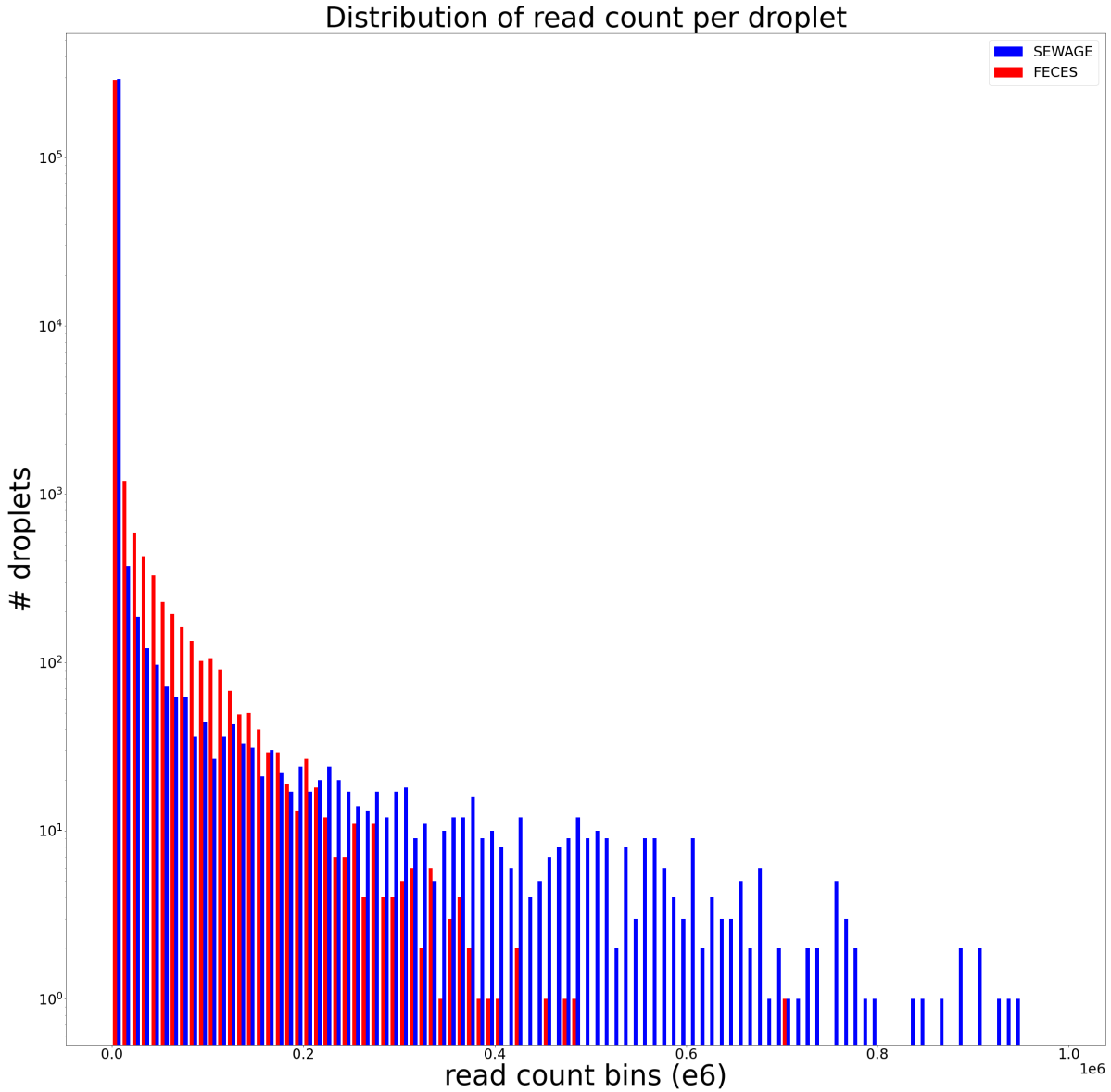
Demultiplexing

- Each read has 1 barcode
- 1 barcode = 1 droplet (1 SNP tolerance)
- 16 bps barcodes


Read count distribution for the shallow sequenced FECES5
JUNE2017
FECES
JUNE2018
Log count of droplets
10^5
10^4
10^3
10^2
10^1
10^0
0
20000
40000
60000
80000
100000
120000
1000 large read count bins


Read count distribution for the deep sequenced FECES5
JUNE2017
FECES
JUNE2018
Log count of droplets
10^5
10^4
10^3
10^2
10^1
10^0
0
100000
200000
300000
400000
500000
1000 large read count bins

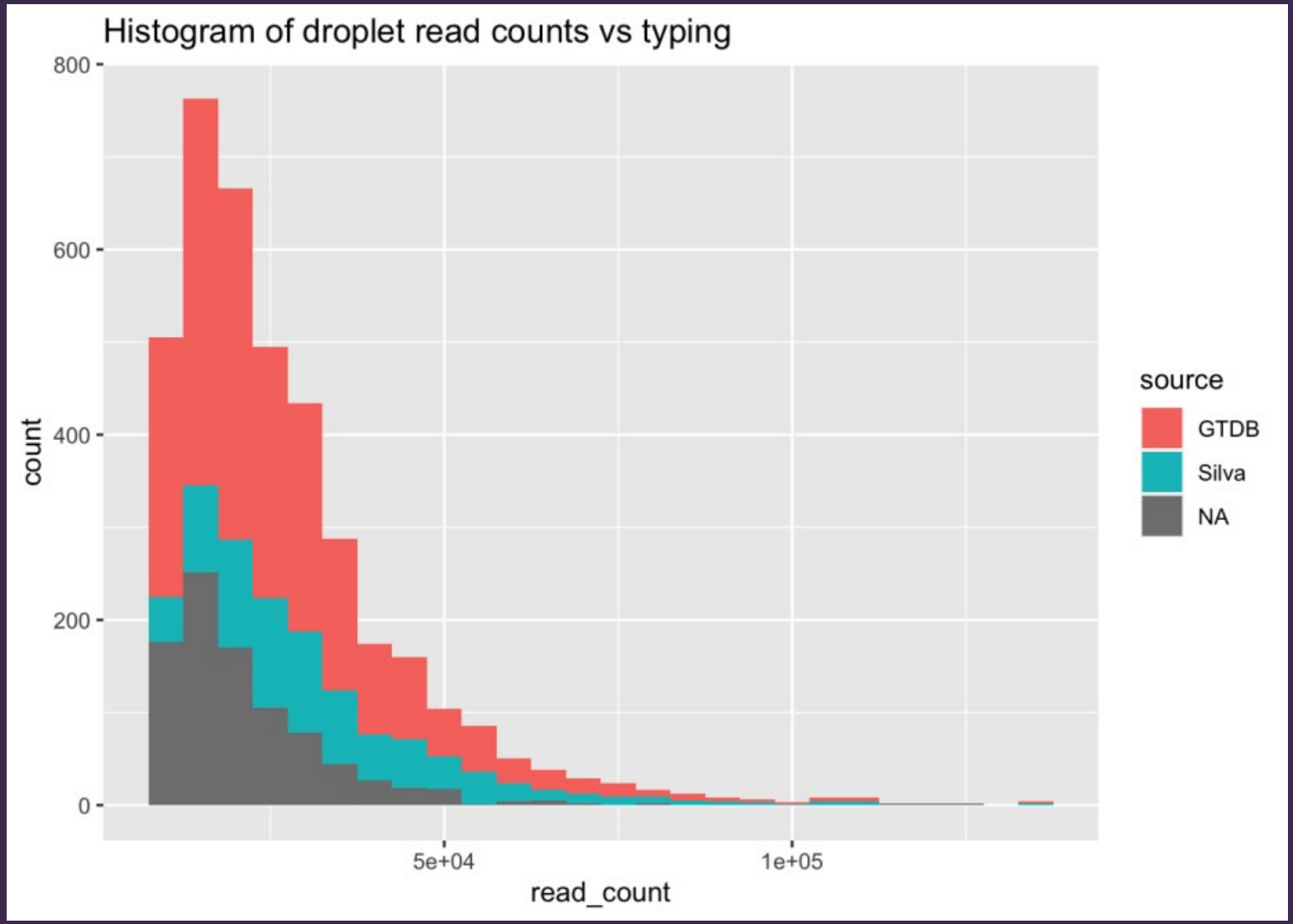
Spiked samples



The taxonomic assignment story

- Marker gene based:
  - 16S rRNA: SILVA
  - 120 marker genes: GTDB
- Genomic database (too computationally expensive)

GTDB 120 marker genes

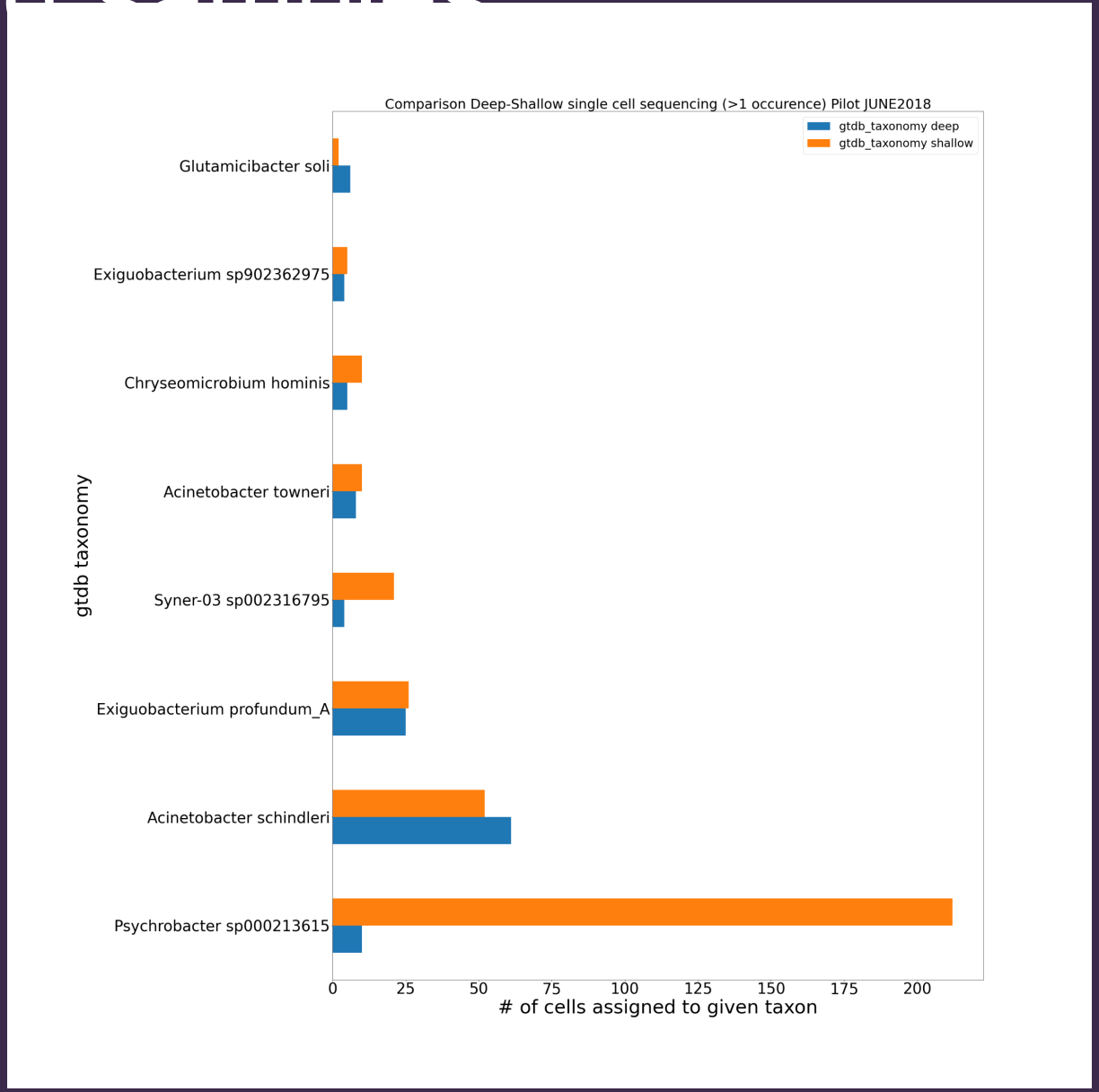


Credits to Judit

- Shotgun sequencing picked up: one AMR gene in one Acinetobacter

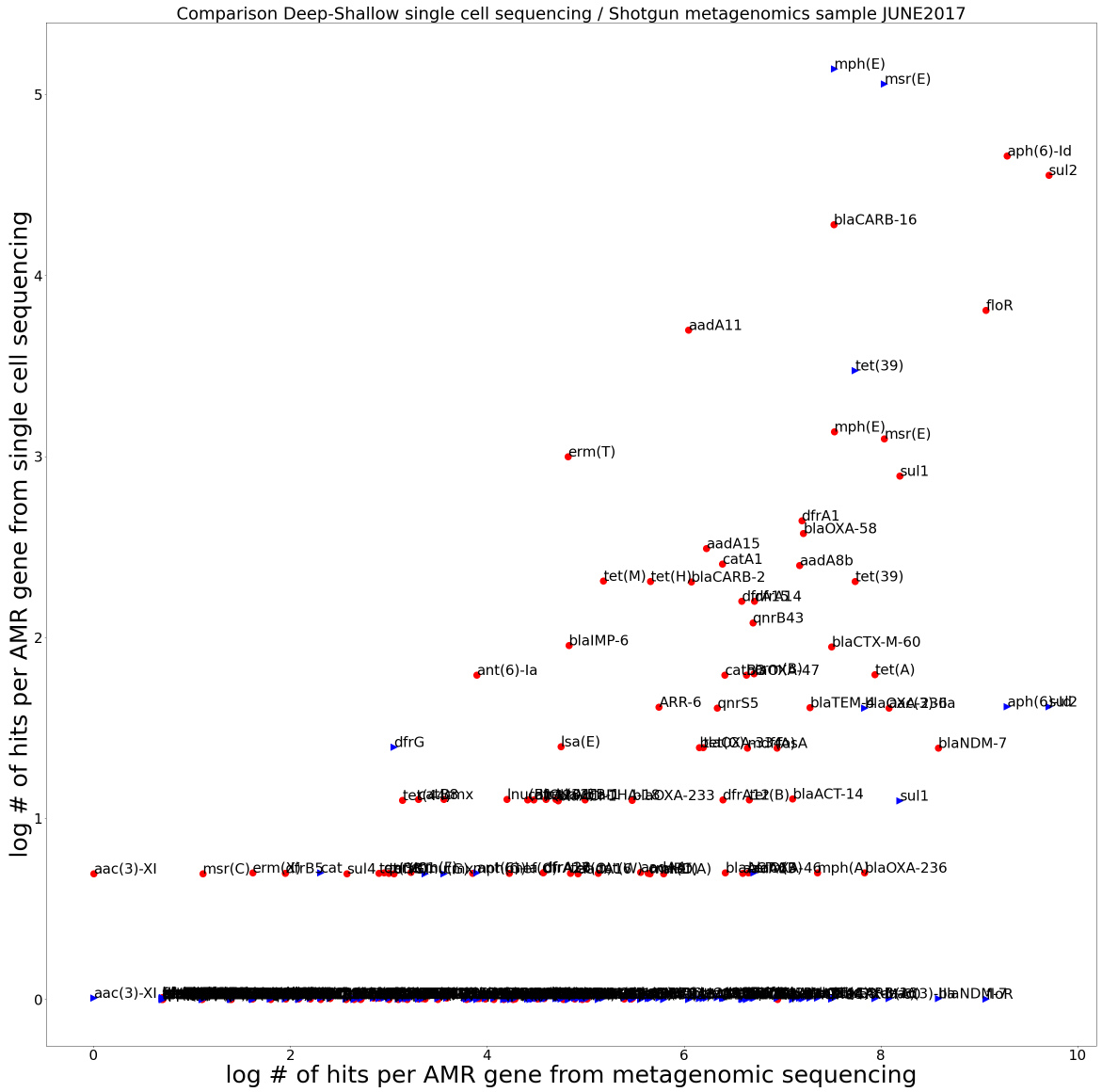
How to deep and shallow sequencing compare?

Bacterial abundance in deep and shallow sequenced samples



Comparing SC results with traditional shotgun

AMR SC vs shotgu



Focus on spike samples

- Mock community contains 25 species in different proportions:
- Goal:
  - See how much of these we can capture
  - How complete are the genomes we get.

Bacteria species	Estimated abundance (Percentage)	Number of cells found in sewage
Faecalibacterium prausnitzii	14	1
Veillonella rogosae	14	1
Roseburia hominis	14	47
Bacteroides fragilis	14	0
Prevotella corporis	6	0
Bifidobacterium adolescentis	6	45
Fusobacterium nucleatum	6	0
Lactobacillus fermentum	6	19
Methanobrevibacter smithii	0.1	0
Salmonella enterica	0.01	0
Enterococcus faecalis	0.001	0
Clostridium perfringens	0.0001	0
Akkermansia muciniphila	1.5	0
Clostridioides difficile	1.5	0
Escherichia coli (B-766)	2.8	0
Escherichia coli (B-2207)	2.8	0
Escherichia coli (B-3008)	2.8	0
Escherichia coli (B-1109)	2.8	0
Escherichia coli (JM109)	2.8	0
Candida albicans	1.5	0
Saccharymeces cerevisiae	1.4	0

Coassembly

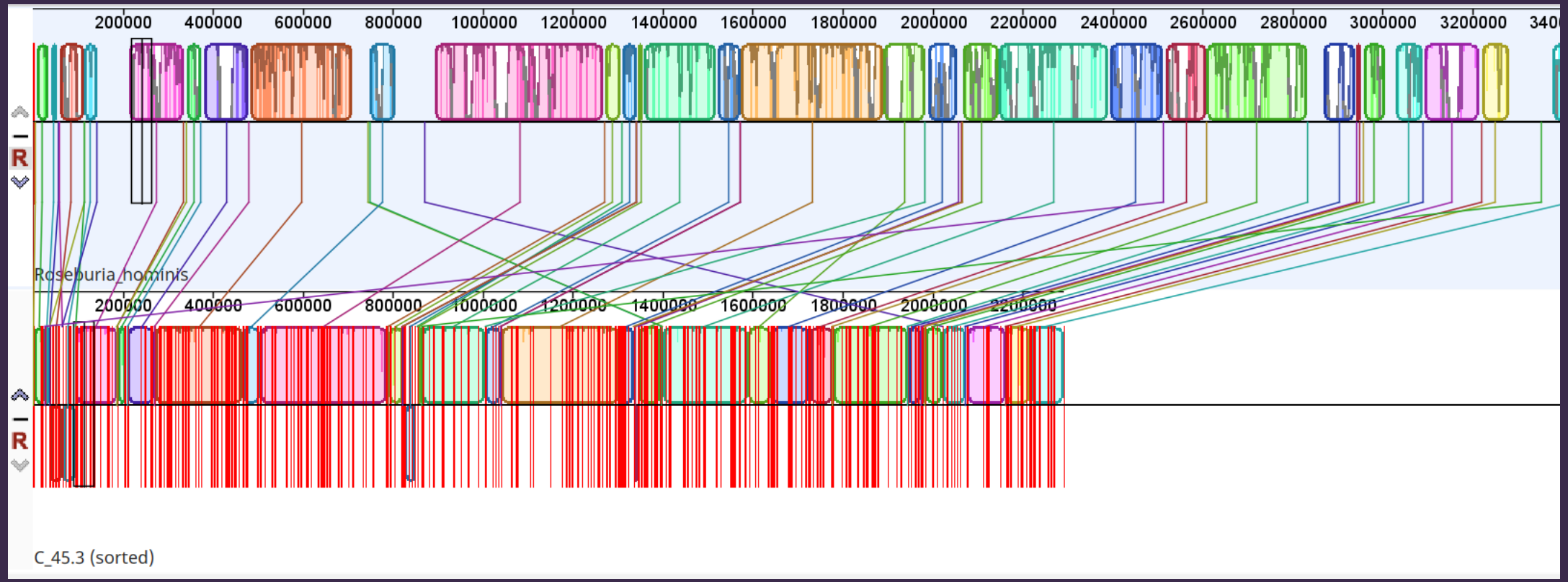
- Aggregating all the droplets that were assigned to the same taxon and assembling to get more complete genome
- Aim: get higher quality genomes and validate database free method to get novel genomes

Species	Completeness	Contamination	N50	Genome size	Genome size (NCBI)
Bifidobacterium adolescentis	91.66	7.54	13,512	2.174 MB	2.2 Mb
Roseburia hominis	92.93	38.37	9,897	4.269 MB	3.6 Mb

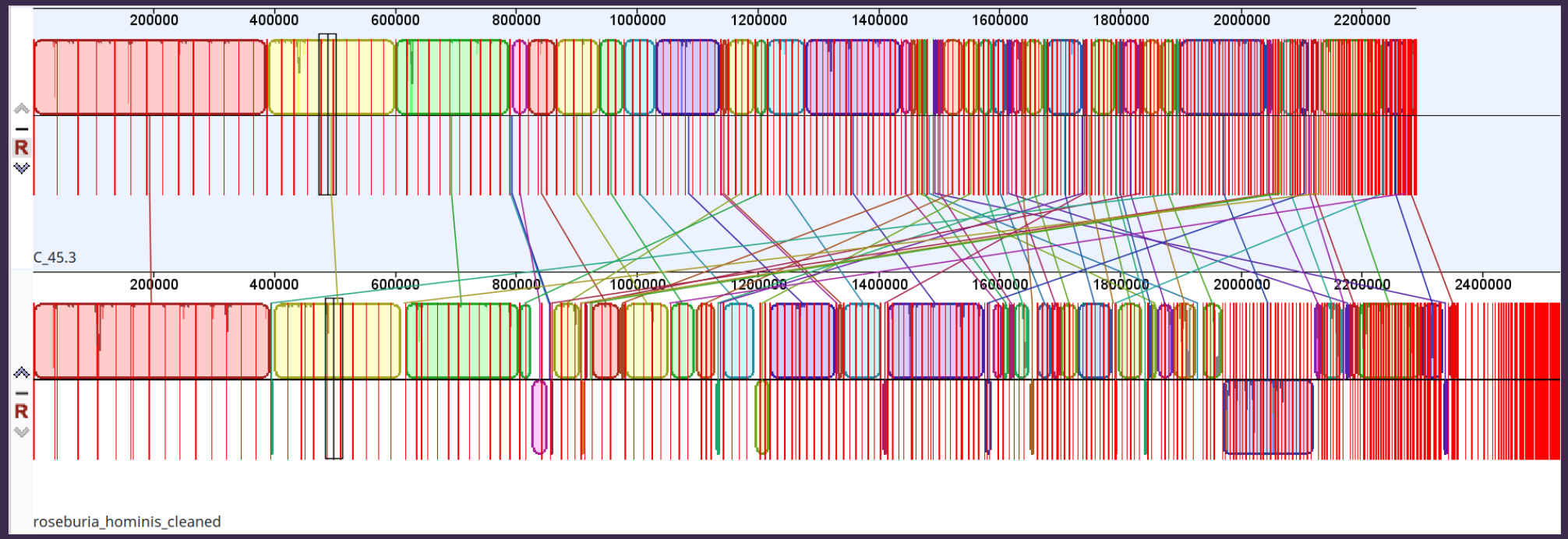
Need to get rid of contamination

- 2 methods:
  - Binning (database free)
  - Taxonomic assignment of contigs (Kraken)

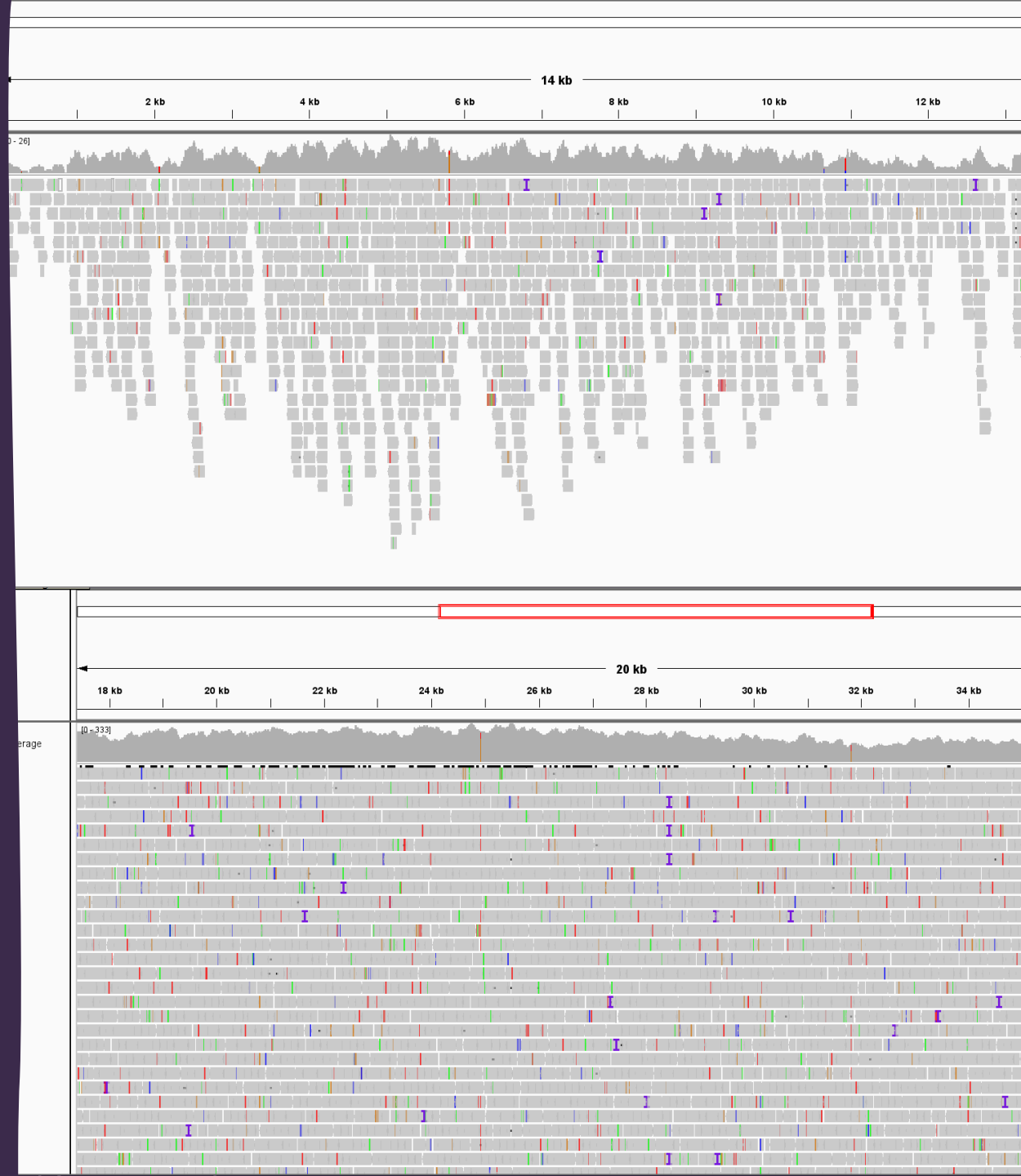
Zymobiomics Roseburia hominis (strain OB EAV1 11 DCM)



Kraken cleaned vs Binned genome



Main Reason: MDA



Improving bin quality by using several metagenomic binning tools

- Goal: get rid of contamination
- DAS Tool: Scores bins from different binners
- Binning tools used:
  - Metabat2
  - Maxbin2
  - VAMB

Conclusions

- We can trace AMR among different taxa within a microbial community
- We’re getting better at obtaining higher quality genomes
- Need more genome coverage
- Need more droplets:
  - Rarefaction curves
  - Species not picked up in mock community

Plans for the future

- Longitudinal Pig feces
- Clustering of unassigned droplets for genome discovery
- Screening for phages/viruses/archea/…
- Plasmid/AMR dissemination (improvement in ML tools)
- Automating workflow (when data comes in a fixed format)