

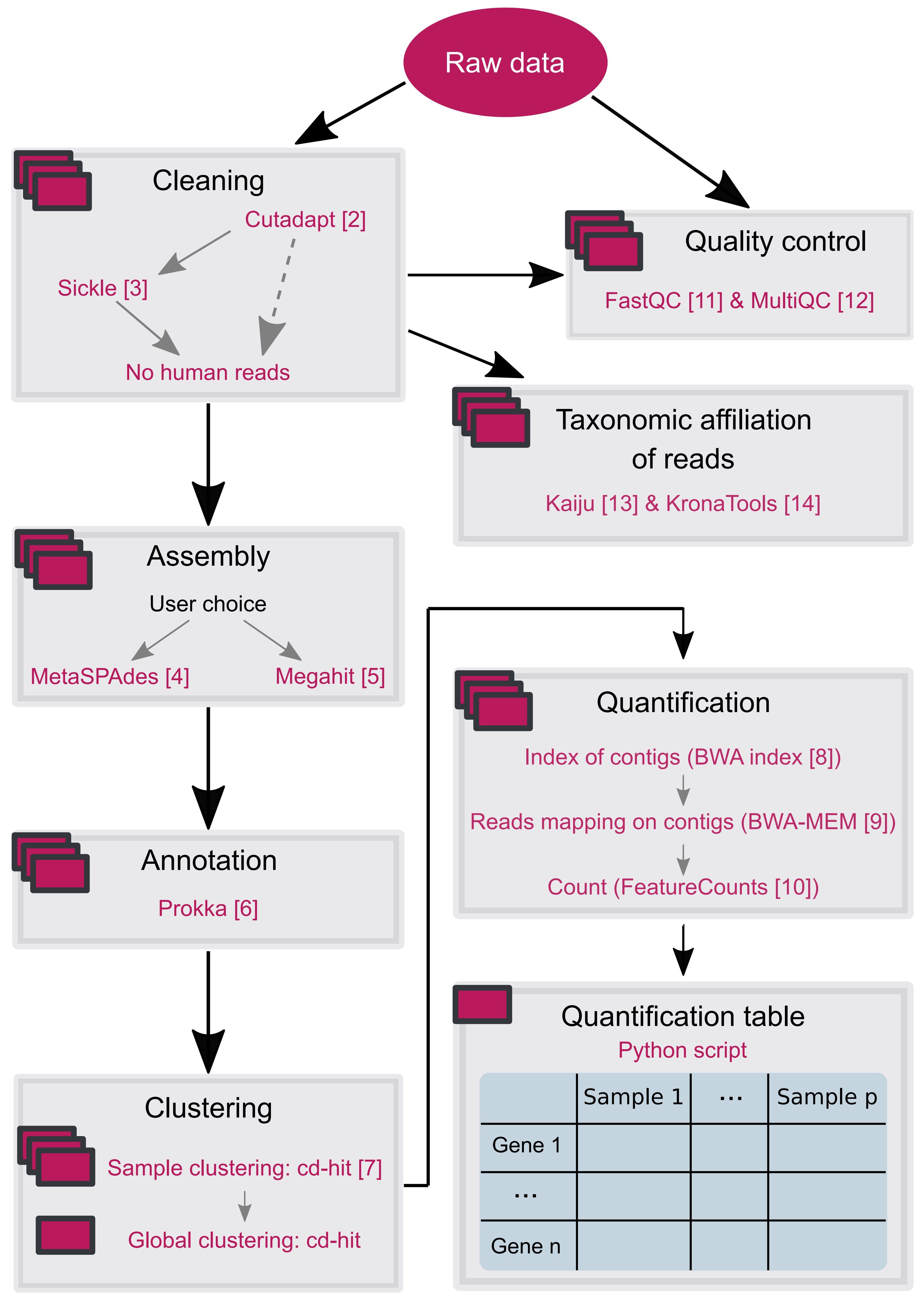
Joanna FOURQUET¹, Adeline CHAUBET², Hélène CHIAPPELLO³, Christine GASPIN¹, Marisa HAENNI⁴, Christophe KLOPP¹, Agnese LUPO⁴, Jean MAINGUY¹, Céline NOIROT¹, Tony ROCHEGUE⁴, Matthias ZYTNICKI⁵, Tristan FERRY⁶, Claire HOEDE¹

- ¹ MIAT, PF Bioinfo GenoToul, Université de Toulouse, INRA, 24 Chemin de Borde Rouge, 31320 Castanet-Tolosan, France
- ² INRA, US 1426, GeT-PlaGe, Genotoul, 31320 Castanet-Tolosan, France
- ³ MalAGE, INRA, Université Paris-Saclay, 78350 Jouy-en-Josas, France
- ⁴ Université de Lyon, ANSES-Laboratoire de Lyon, Unité Antibiorésistance et Virulence Bactériennes, 69007 Lyon, France
- ⁵ MIAT, Université de Toulouse, INRA, 31320 Castanet-Tolosan, France
- ⁶ Hospices Civils de Lyon, Université Claude Bernard Lyon 1, International Center for Research in Infectiology, CIRI, INSERM U1111, CNRS UMR5308, ENS de Lyon, 69364 Lyon, France

Corresponding author: claire.hoede@inra.fr

A Nextflow pipeline for whole metagenome assembly and gene quantification

- Type of NGS data:** whole genome shotgun sequencing (Illumina HiSeq3000, paired, 2*150bp)
- Workflow:** a scalable and reproducible metagenomic analysis with a **nextflow** [1] pipeline
- Availability:** <https://forgemia.inra.fr/genotoul-bioinfo/metagwgs>



Biological data

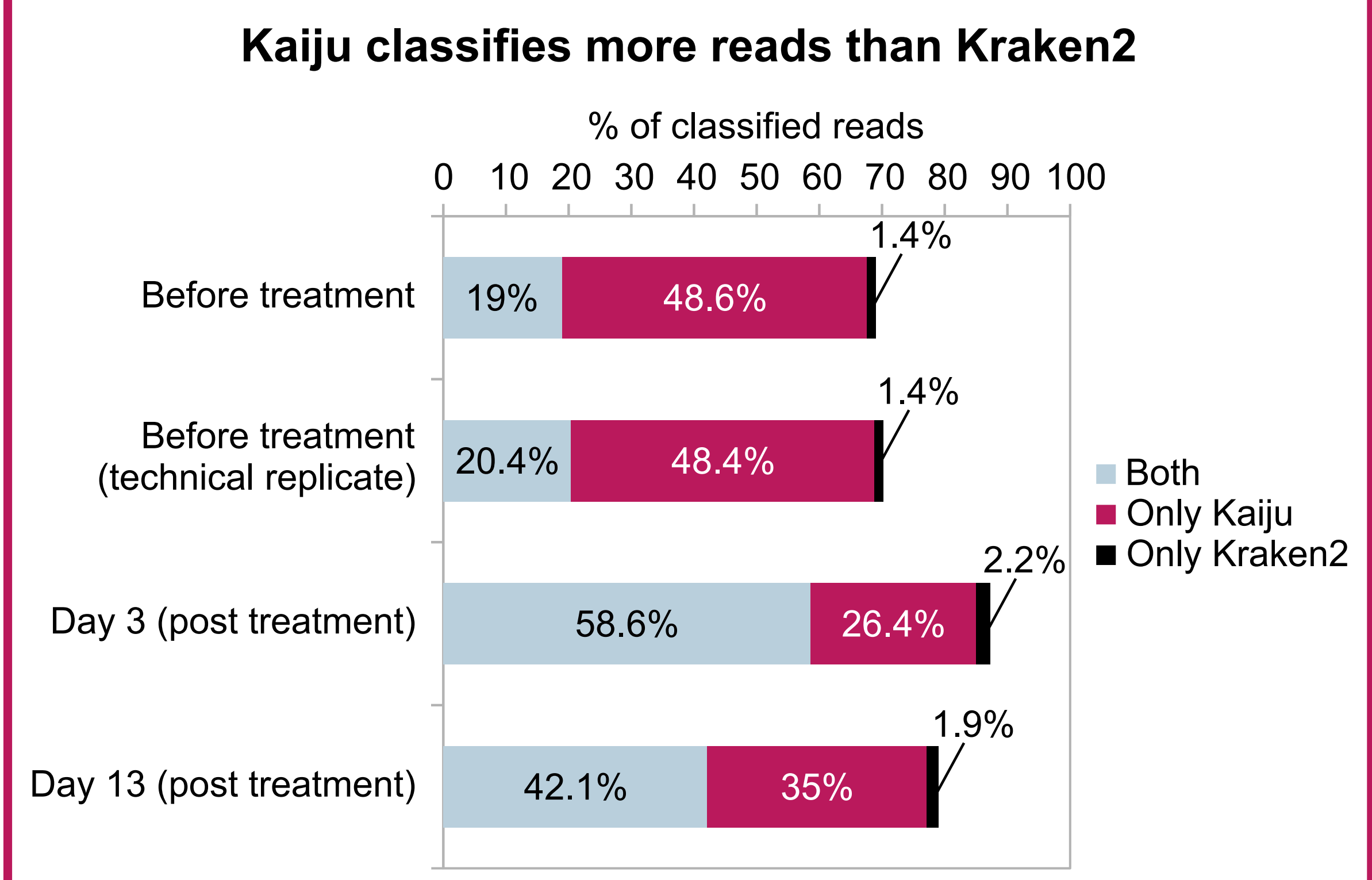
Objective: analyze eventual antibiotic resistance development during antibiotic therapy

Samples: fecal samples from gut microbiota of a healthy volunteer

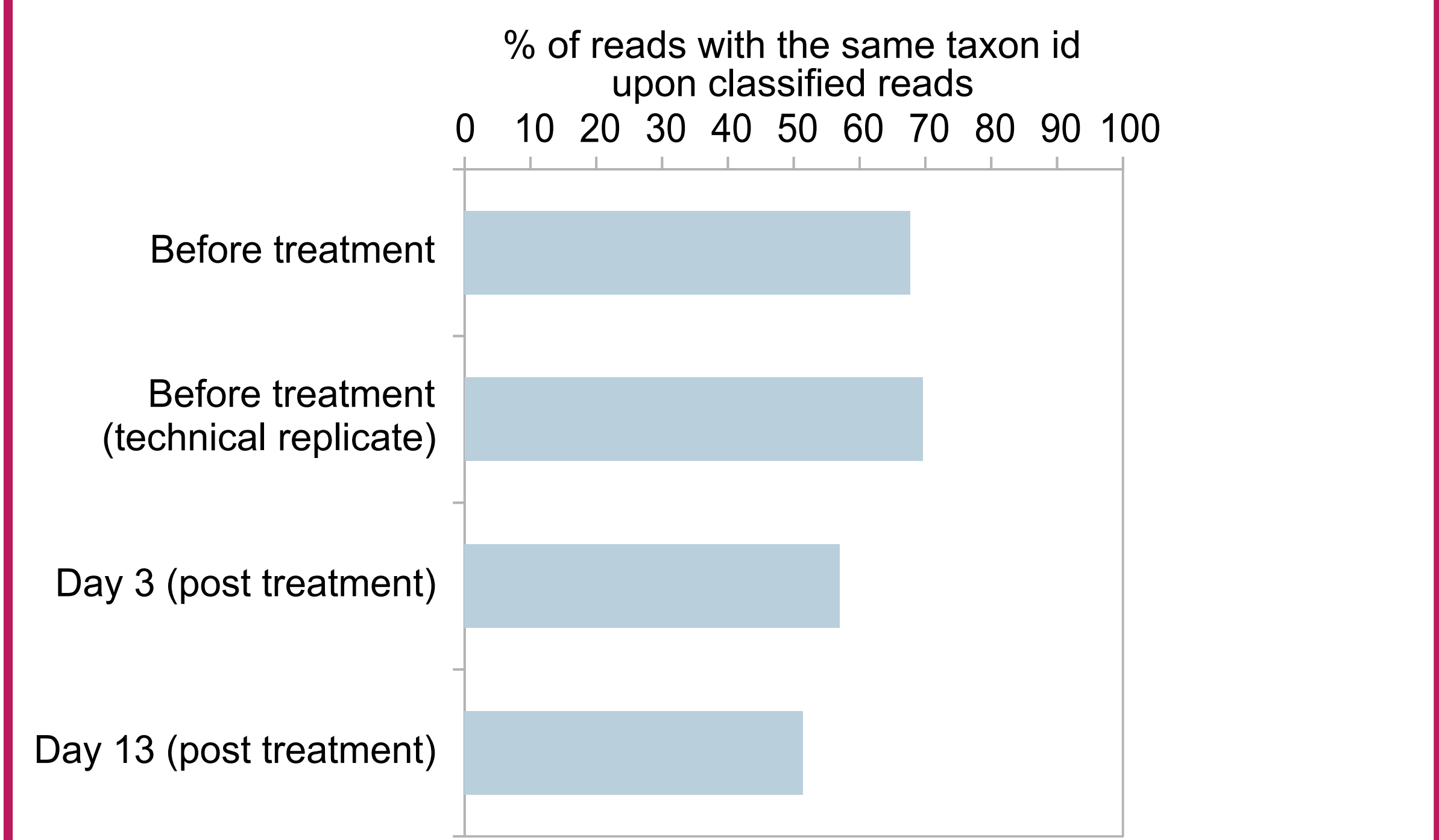
Patient treatment: combination of 2 antibiotics for 3 days

Time points: 3 critical time points

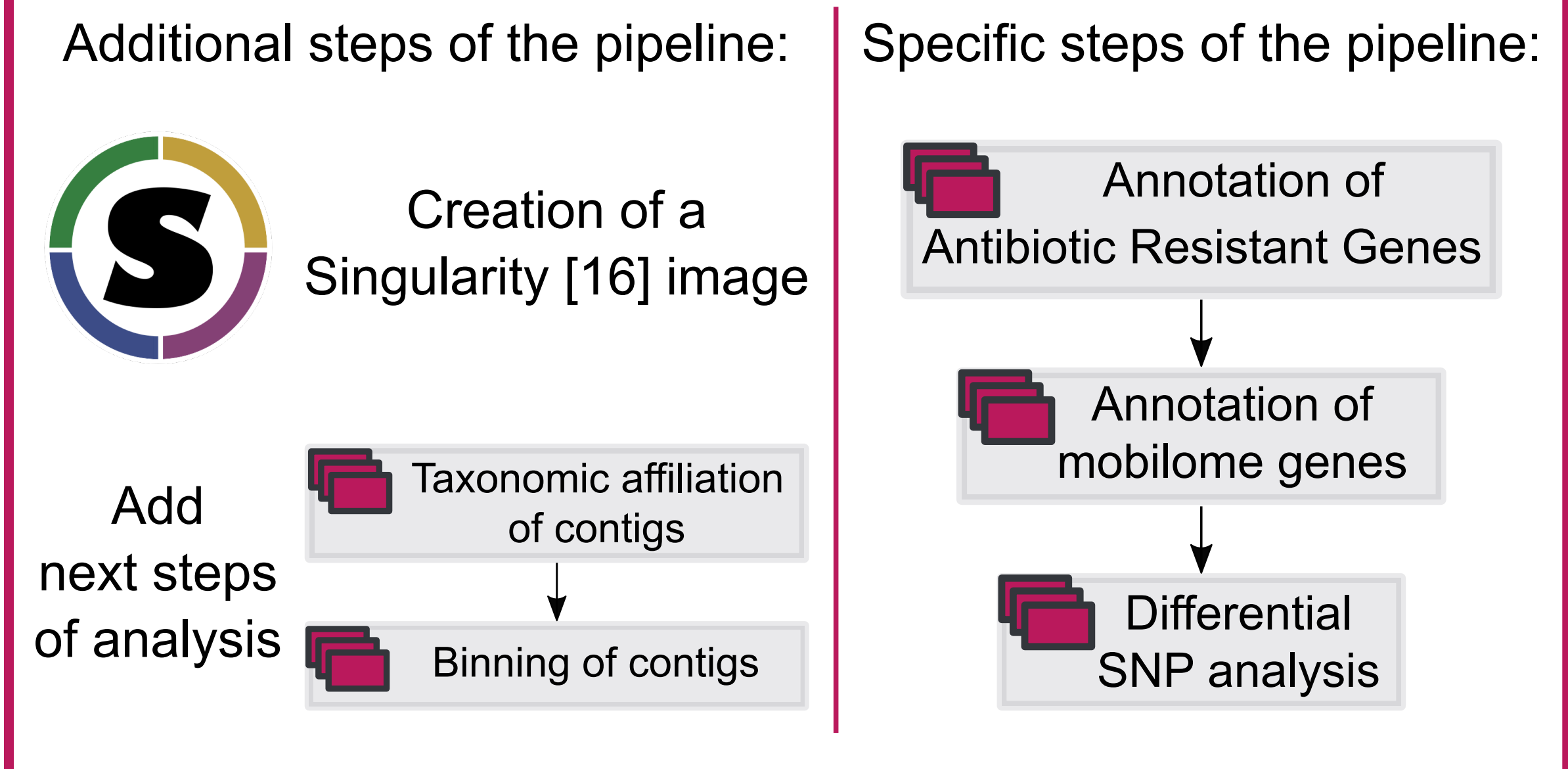
Benchmark: Kaiju MEM vs Kraken2 [15]



Kaiju and Kraken2 results are congruent



Perspectives



[1] P. Di Tommaso et al. Nextflow enables reproducible computational workflows. *Nat Biotechnol.*, 35(4):316-319, 2017.
 [2] M. Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet journal*, 17(1):10-12, 2011.
 [3] NA Joshi, JN Fass. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files [Software]. Available at <https://github.com/najoshi/sickle>, 2011.
 [4] S. Nurk et al. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.*, 27(5):824-834, 2017.
 [5] D. Li et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 31(10):1674-6, 2015.
 [6] T. Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14):2068-9, 2014.
 [7] L. Fu et al. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150-2, 2012.
 [8] H. Li et al. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14):1754-60, 2009.
 [9] H. Li. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv:1303.3997v2*, 2013.
 [10] Y. Liao et al. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 30(7):923-30, 2014.
 [11] S. Andrews. FastQC: a quality control tool for high throughput sequence data. Available at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, 2010.
 [12] P. Ewels et al. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 32(19):3047-8, 2016.
 [13] P. Menzel et al. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun.*, 7:11257, 2016.
 [14] BD Ondov et al. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 12(1):385, 2011.
 [15] D. E Wood and S. L Salzberg. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biology*, 15(3):R46, 2014.
 [16] GM Kurtzer et al. Singularity: Scientific containers for mobility of compute. *PLoS One*, 12(5), 2017.