

Joanna FOURQUET<sup>1</sup>, Céline NOIROT<sup>2</sup>, Christophe KLOPP<sup>2</sup>,  
Philippe PINTON<sup>3</sup>, Sylvie COMBES<sup>1</sup>, Claire HOEDE<sup>2</sup>, Géraldine PASCAL<sup>1</sup>

<sup>1</sup> GenPhySE, Université de Toulouse, INRAE, ENVT, F-31326, Castanet Tolosan, France

<sup>2</sup> INRAE, UR875 MIAT, PF Bioinfo GenoToul, F-31326, Castanet-Tolosan, France

<sup>3</sup> Toxalim, Université de Toulouse, INRAE, ENVT, INP-Purpan, UPS, F-31027 Toulouse, France

Corresponding Author: [geraldine.pascal@inrae.fr](mailto:geraldine.pascal@inrae.fr)




## Acknowledgements

ExpoMycoPig project and JF are funded by France Futur Elevage

We are grateful to the Genotoul bioinformatics platform Toulouse Occitanie (Bioinfo Genotoul, doi: 10.15454/1.5572369328961167E12) for providing computing and storage resources

## What bioinformatics solution exists to process whole metagenome shotgun data?


### nf-core/mag

- Incomplete
  - ✓ Metagenome assembly
  - ✓ Taxonomic affiliation of reads and bins
  - × Taxonomic affiliation of contigs
  - × Functional annotation
- Easy of use
  - ✓ Automated with **nextflow** [1]
- Highly reproducible  [2]
  - ✓ Comes with a Singularity container
- Modular
  - ✓ Different parameters
  - ✓ Possibility to skip certain steps
- Available
  - ✓ <https://github.com/nf-core/mag>

### ATLAS<sup>[3]</sup>

- Incomplete
  - ✓ Metagenome assembly
  - × Taxonomic affiliation of reads and contigs
  - ✓ Taxonomic affiliation of bins
  - ✓ Functional annotation
- Issues during use
  - ✓ Automated with **Snakemake** [4]
- Reproducible
  - ✓ Installation and run with **CONDA**
- Modular
  - ✓ Different parameters
  - ✓ Possibility to skip certain steps
- Available
  - ✓ <https://github.com/metagenome-atlas/atlas>

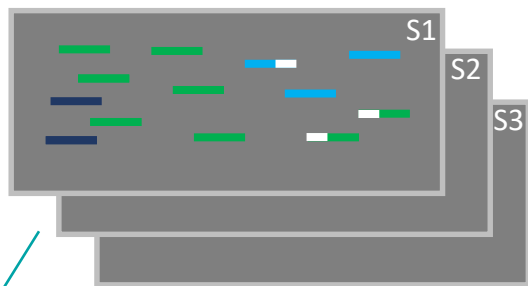
### Our solution: metagWGS

- Complete
  - ✓ Metagenome assembly
  - ✓ Taxonomic affiliation of reads, contigs and bins
  - ✓ Functional annotation
- Easy of use
  - ✓ Automated with **nextflow**
- Highly reproducible 
  - ✓ Comes with a Singularity container
- Modular
  - ✓ Different parameters
  - ✓ Possibility to skip certain steps
- Available
  - ✓ <https://forgemia.inra.fr/genotoul-bioinfo/metagwgs>

# metagWGS: preprocessing, assembly and annotation

## Raw data

- Adapter sequence (entire or truncated)
- High quality read
- Low quality read
- Host read



## Quality control

FastQC

[10]

.html ...

FastQC

[10]

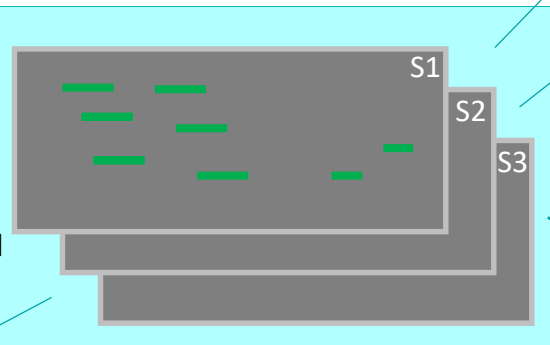
.html ...

## Cleaning

cutadapt [5]

sickle [6]

bwa mem [7]



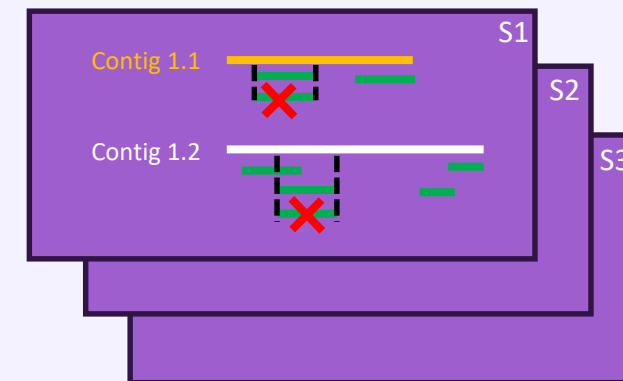
## Assembly

metaspades [11] or megahit [12]



## Reads deduplication

bwa mem, samtools markup [13]



## Assembly filter

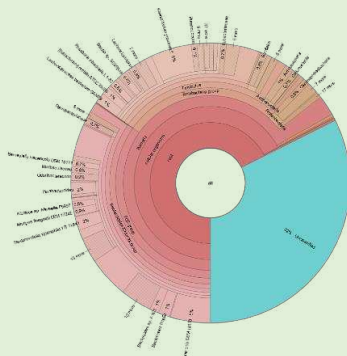
Filter\_contig\_per\_cpm.py



## Taxonomic affiliation of reads

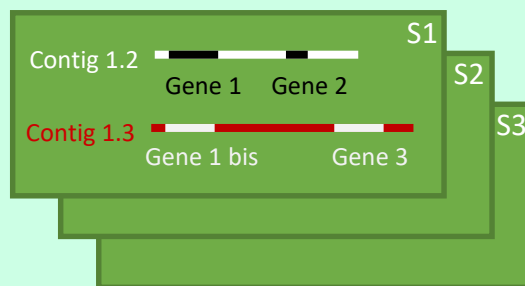
kaiju [8]

kronaTools [9]



## Annotation of genes

prokka [14]

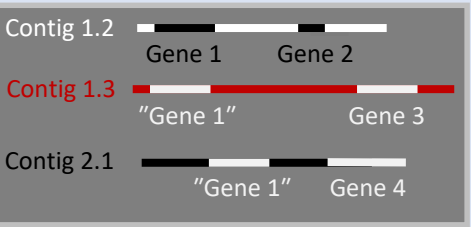


----- Possible skipping

# metagWGS: quantification and affiliation

## Clustering of genes

cd-hit-est [15]



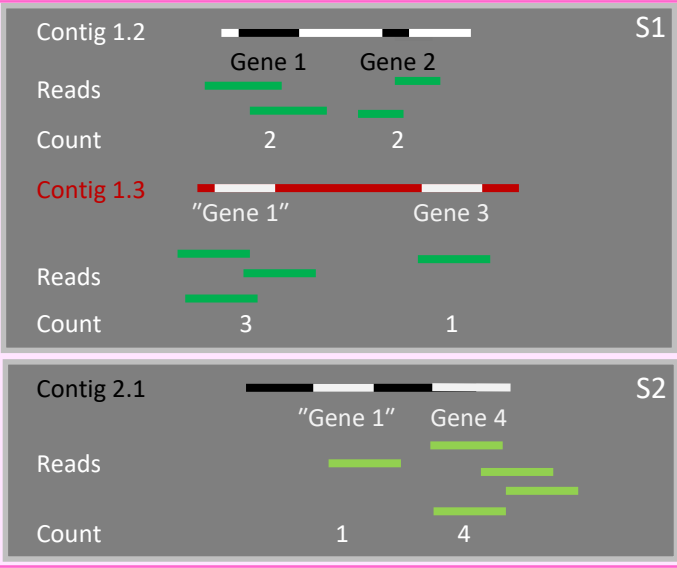
## Taxonomic affiliation of contigs

diamond [17], aln2taxaffi.py

| #contig | consensus taxid | consensus lineage  | S1 | S2 | S3 |
|---------|-----------------|--|----|----|----|
| S1_1    | 1263            | cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Clostridia; Clostridiales |    |    |    |
| S1_10   | 84108           | cellular organisms; Bacteria; Terrabacteria group; Actinobacteria; Coriobacteriia        |    |    |    |

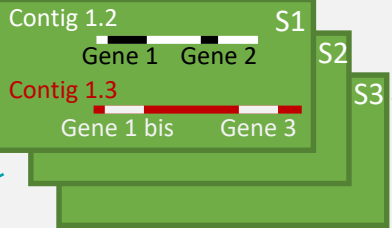
## Quantification of reads on genes

bwa mem, featureCounts [16]

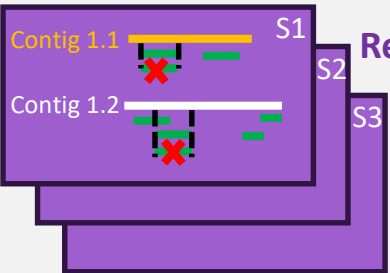


|        | S1 | S2 | ... | ... |
|--------|----|----|-----|-----|
| Gene 1 | 5+ | 1+ |     |     |
| Gene 2 | 2+ | 0+ |     |     |
| ...    |    |    |     |     |

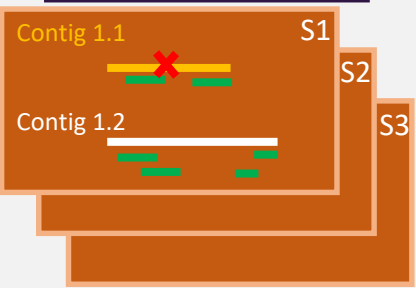
## Input from part 1



## Annotation of genes



## Reads deduplication

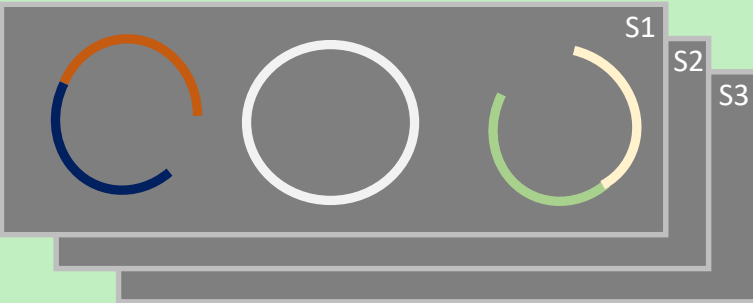


## Assembly filter

----- Possible skipping

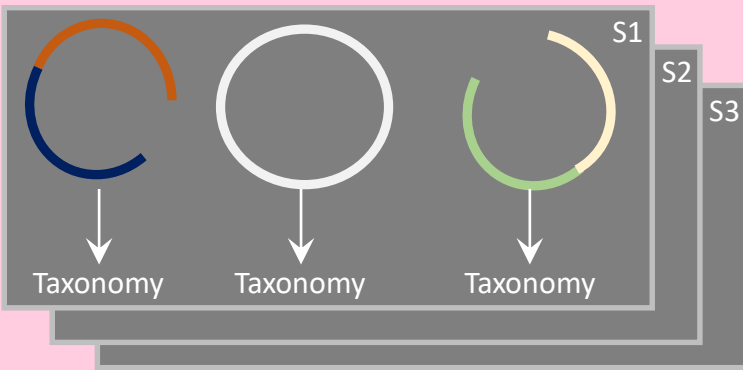
## Binning of contigs

bowtie2 [18], metabat2 [19]

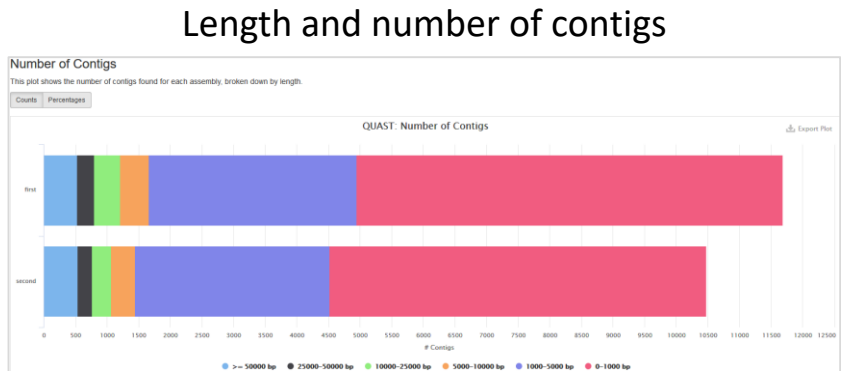
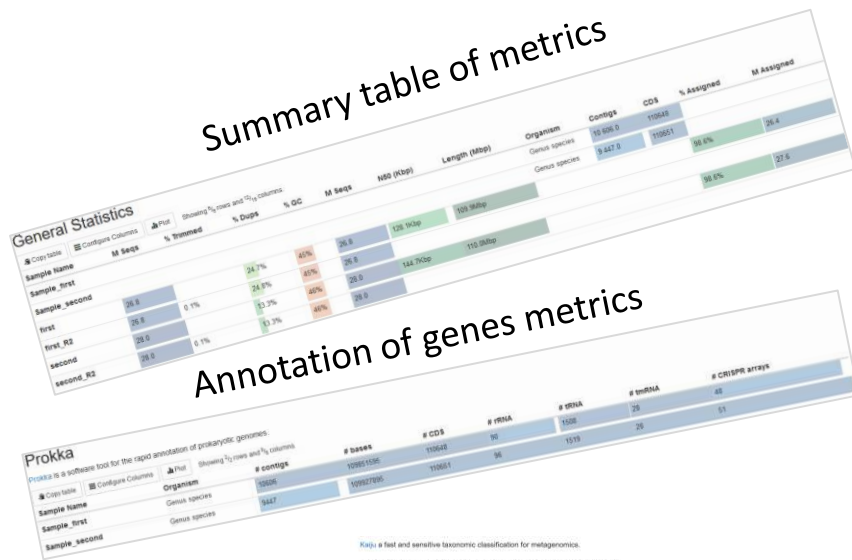


## Taxonomic affiliation of bins

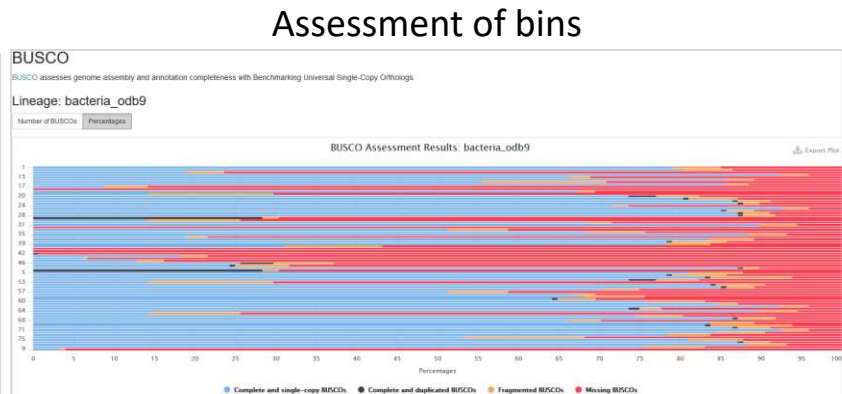
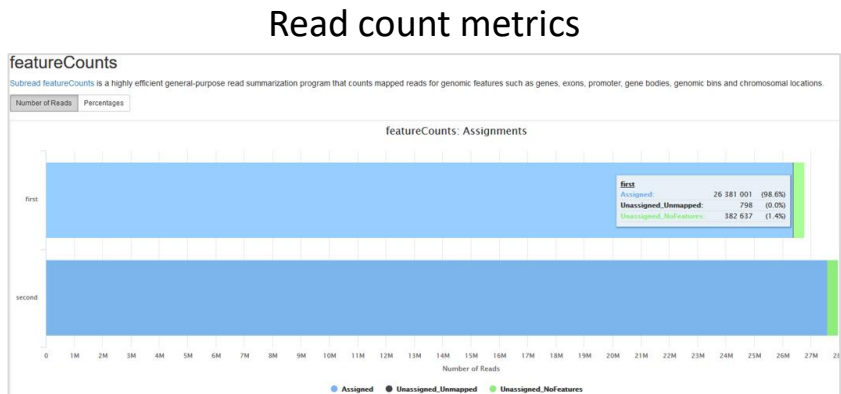
BAT [20]



# MultiQC<sup>[21]</sup> graphic outputs of metagWGS



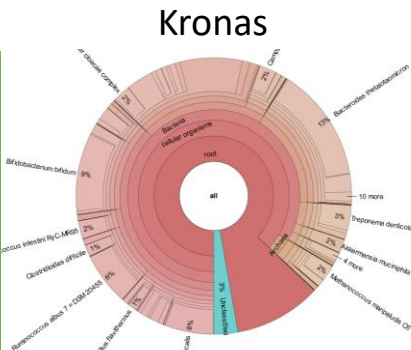
## Taxonomic affiliation of reads metrics



# Output tables and other graphs of metagWGS

## Taxonomic classification of contigs

| #contig    | consensus taxid | consensus lineage  |
|------------|-----------------|--|
| first1     | 210             | cellular organisms; Bacteria; Proteobacteria; delta/epsilon subdivisions; Epsilonproteobacteria; Campylobacterales; Helicobacteraceae; Helicobacter; Helicobacter pylori |
| first10    | 1681            | cellular organisms; Bacteria; Terrabacteria group; Actinobacteria; Actinobacteria; Bifidobacteriales; Bifidobacteriaceae; Bifidobacterium; Bifidobacterium bifidum       |
| first100   | 1358            | cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Bacilli; Lactobacillales; Streptococcaceae; Lactococcus; Lactococcus lactis                               |
| first1000  | 1322            | cellular organisms; Bacteria; Terrabacteria group; Firmicutes; Clostridia; Clostridiales; Lachnospiraceae; Blautia; Blautia hansenii                                     |
| first10004 | 644             | cellular organisms; Bacteria; Proteobacteria; Gammaproteobacteria; Aeromonadales; Aeromonadaceae; Aeromonas; Aeromonas hydrophila  |



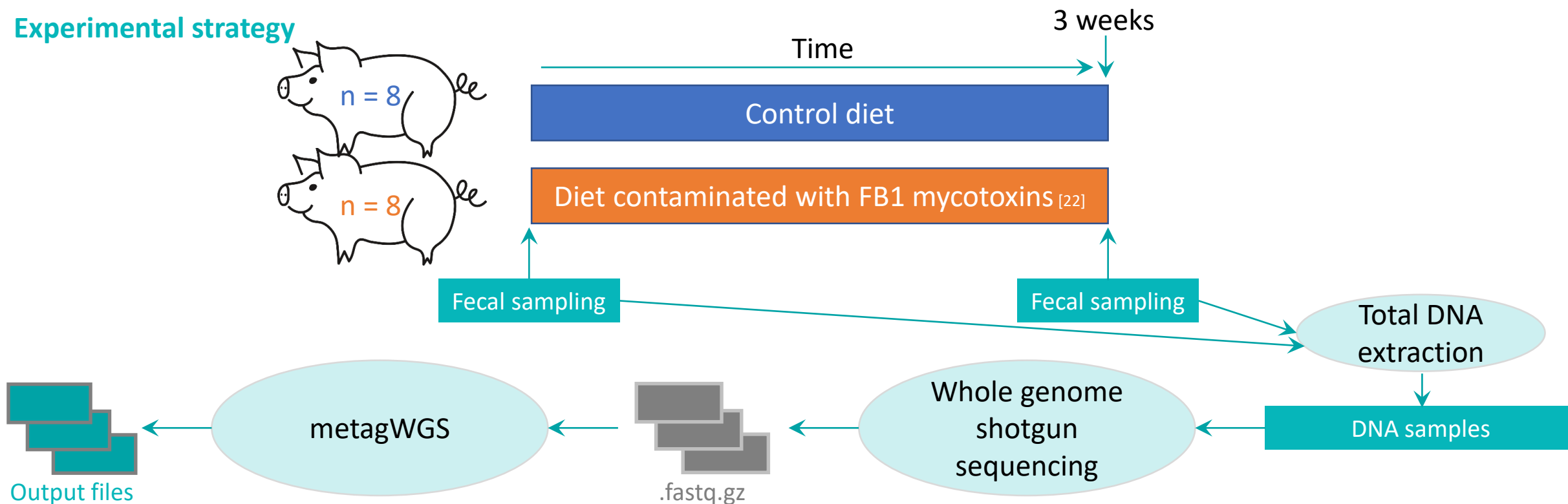
## Quantification table of reads on genes

| id_cluster          | first.featureCounts.tsv | second.featureCounts.tsv |
|---------------------|-------------------------|--------------------------|
| first59.Prot_28193  | 844                     | 887                      |
| first82.Prot_34066  | 847                     | 891                      |
| first992.Prot_96159 | 4092                    | 4389                     |
| first8.Prot_06503   | 5279                    | 3702                     |
| first21.Prot_13879  | 584                     | 611                      |

# metagWGS will be used to analyze ExpoMycoPig data

How can we characterize microbiome digestive ecosystems for pig exposed to mycotoxins?

## Experimental strategy



## References

- [1] Di Tommaso et al. Nextflow enables reproducible computational workflows. *Nat Biotechnol.*, 2017.
- [2] Kurtzer et al. Singularity: Scientific containers for mobility of compute. *PLoS One*, 2017.
- [3] Kieser et al., ATLAS: a Snakemake workflow for assembly, annotation, and genomic binning of metagenome sequence data, bioRxiv, 2019.
- [4] Köster et al., Snakemake - A scalable bioinformatics workflow engine. *Bioinformatics*, 2012.
- [5] Martin. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal*, 2011.
- [6] Joshi and Fass. Sickle: A sliding-window, adaptive, quality-based trimming tool for FastQ files [Software]. Available at <https://github.com/najoshi/sickle>, 2011.
- [7] Li and Durbin. Fast and accurate short read alignment with Burrows-Wheeler Transform. *Bioinformatics*, 2009.
- [8] Menzel et al. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun.*, 2016.
- [9] Ondov et al. Interactive metagenomic visualization in a Web browser. *BMC Bioinformatics*, 2011.
- [10] Andrews. FastQC. Available at <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>, 2010.
- [11] Nurk et al. MetaSPAdes: a new versatile metagenomic assembler. *Genome Res.*, 2017.

- [12] Li et al. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics*, 2015.
- [13] Li et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 2009.
- [14] Seemann. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 2014.
- [15] Fu et al. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 2012.
- [16] Liao et al. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, 2014.
- [17] Buchfink et al. Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, 2015.
- [18] Langmead and Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, 2012.
- [19] Kang et al. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ*, 2019.
- [20] von Meijenfeldt et al. Robust taxonomic classification of uncharted microbial sequences and bins with CAT and BAT. *Genome Biology*, 2019.
- [21] Ewels et al. MultiQC: Summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*, 2016.
- [22] Pinton and Oswald. Effect of deoxynivalenol and other Type B trichothecenes on the intestine: a review. *Toxins*, 2014.