

Deep dive into metagenomic data using

Welcome to

Open the webpage eccb.mmseqs.com

Say hallo via the chat!

Please answer the polls.

Do you have questions you want to have answered?

Deep dive into metagenomic data using

metagenome-atlas

Silas Kieser | University of Geneva,
Switzerland

Sofie Thijs | Hasselt University,
Belgium

MMseqs2

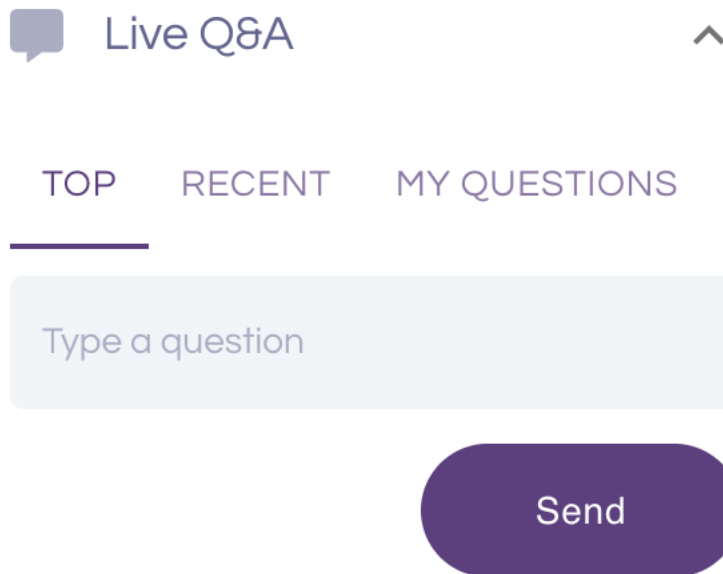
Martin Steinegger | Seoul National
University, Korea

Milot Mirdita | MPI for Biophysical
Chemistry, Germany



Ask questions!

Questions



The screenshot shows a user interface for asking questions. At the top, there is a purple speech bubble icon followed by the text 'Live Q&A' and a small upward-pointing arrow. Below this, there are three tabs: 'TOP', 'RECENT', and 'MY QUESTIONS'. The 'TOP' tab is selected, indicated by a purple underline. Below the tabs is a light blue input box with the placeholder text 'Type a question'. At the bottom right of the input box is a purple rounded button with the text 'Send'.

Chat Room

on eccb.mmseqs.com

- For bugs and larger messages.
- Available after the tutorial.

Content

Part 1: Metagenome Atlas

- Introduction to metagenomics
- Metagenome Atlas presentation
- break
- Metagenome Atlas Tutorial

Part 2: MMseqs2

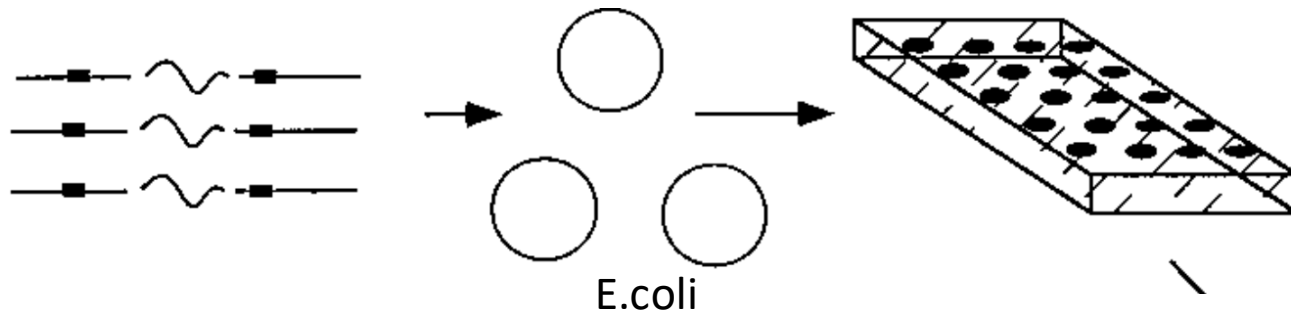
All links are on eccb.mmseqs.com

Metagenomics

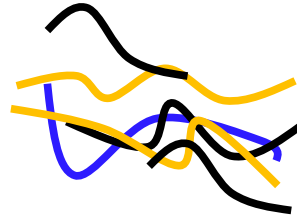
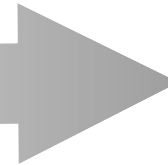
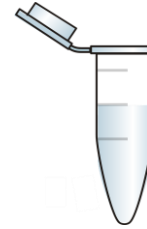
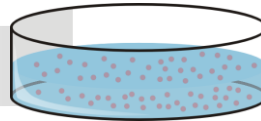
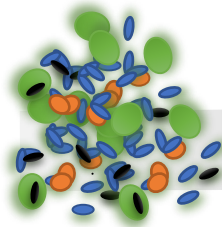
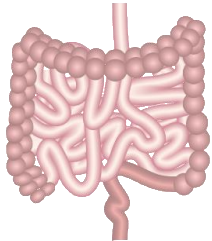


**Metagenomics is the
genomics of microbial
communities**

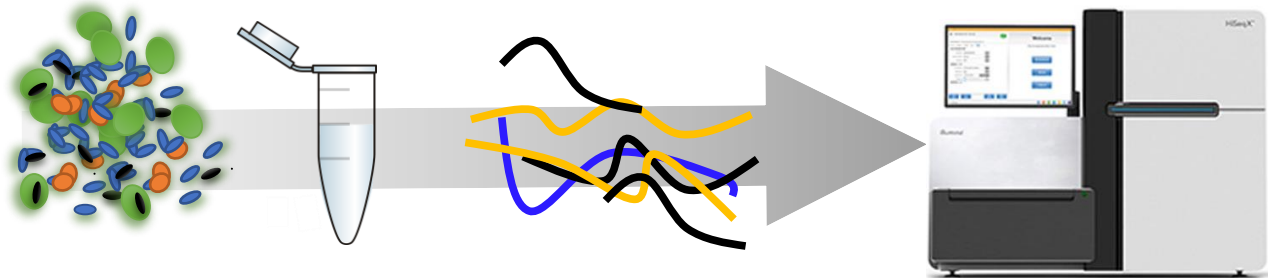
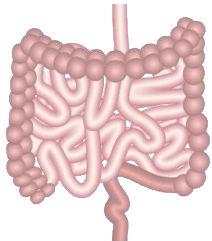
Metagenomics in 1996



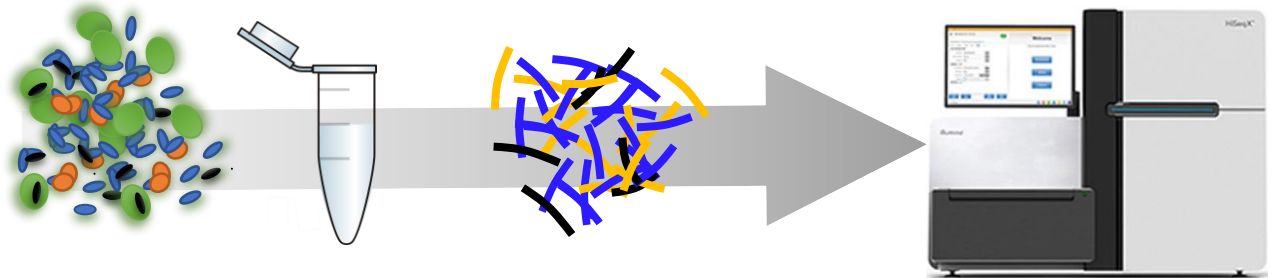
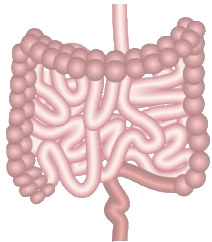
Study microbes

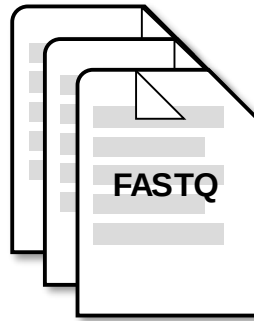


Metagenomics allows to study microbes where they are

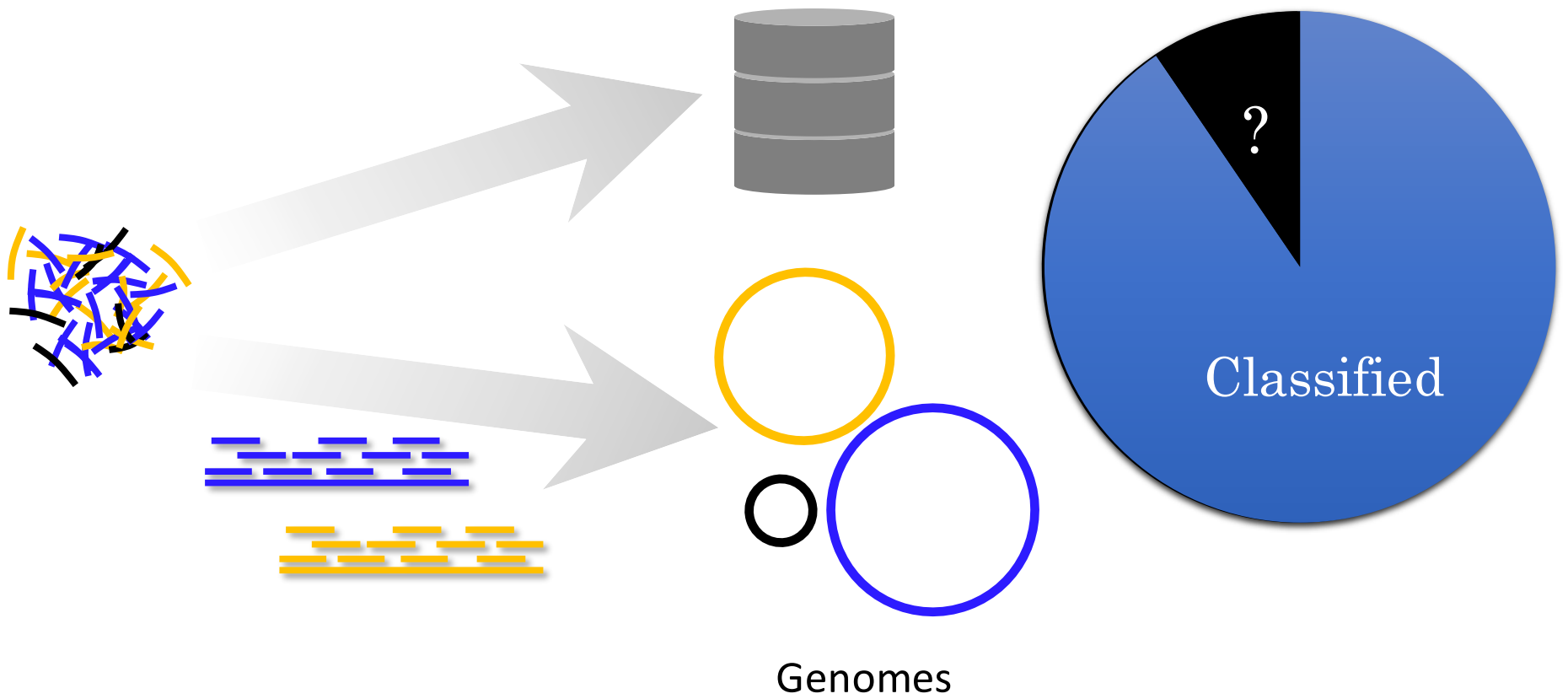


Metagenomics allows to study microbes where they are





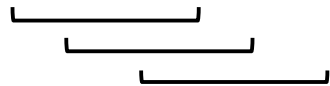
Analyze the data!



Metagenomic assembly

The Theory (De Bruijn graph)

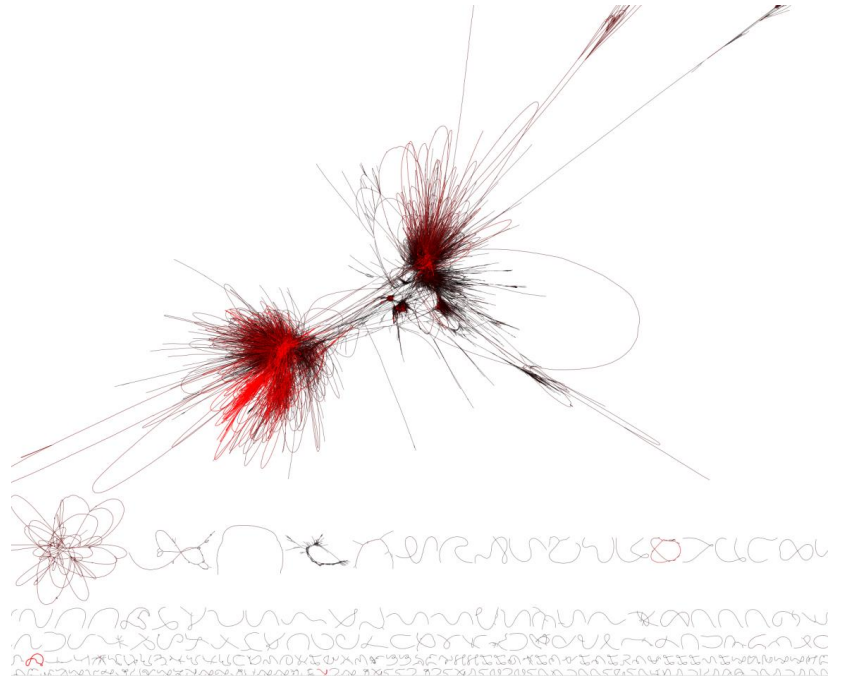
ATCGTA



$K = 4$

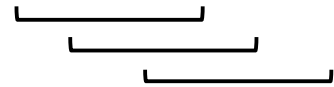


The practice



Assembly: De-Bruijn graph

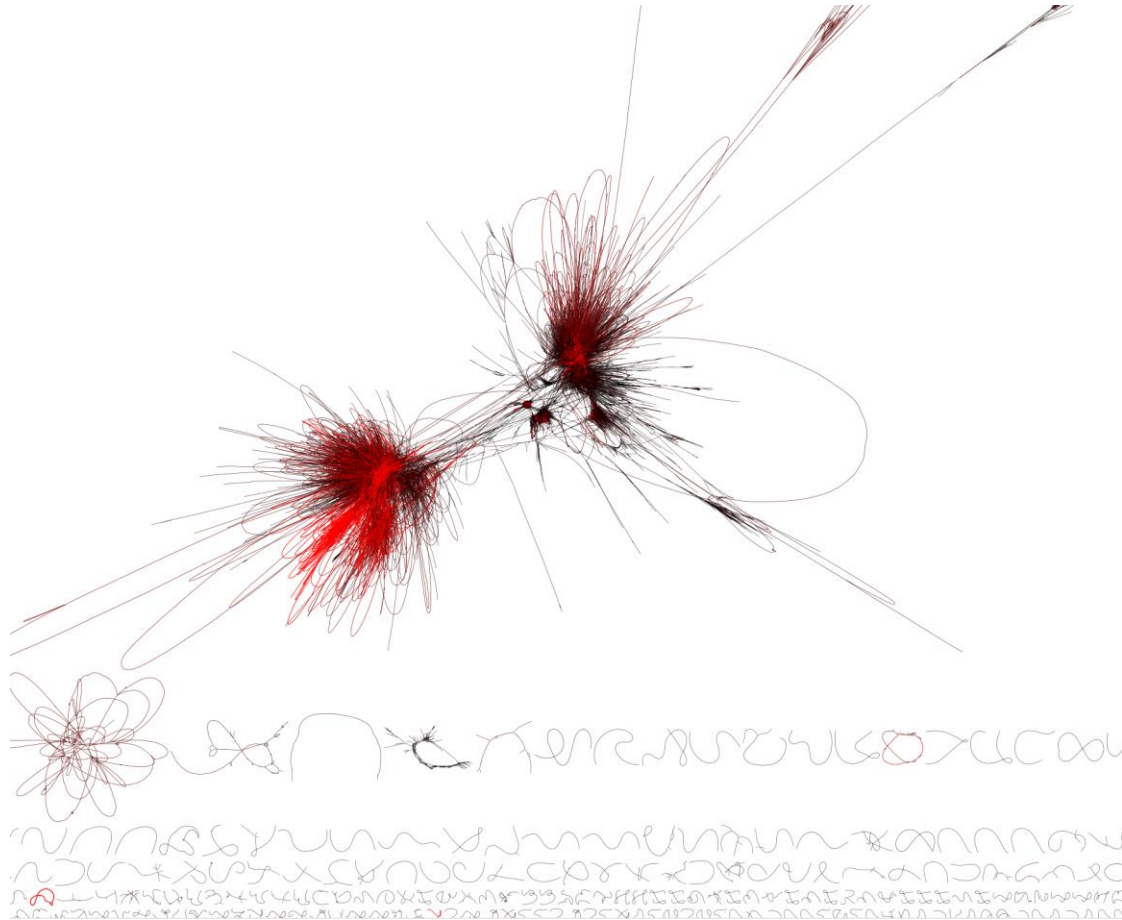
ATCGTA



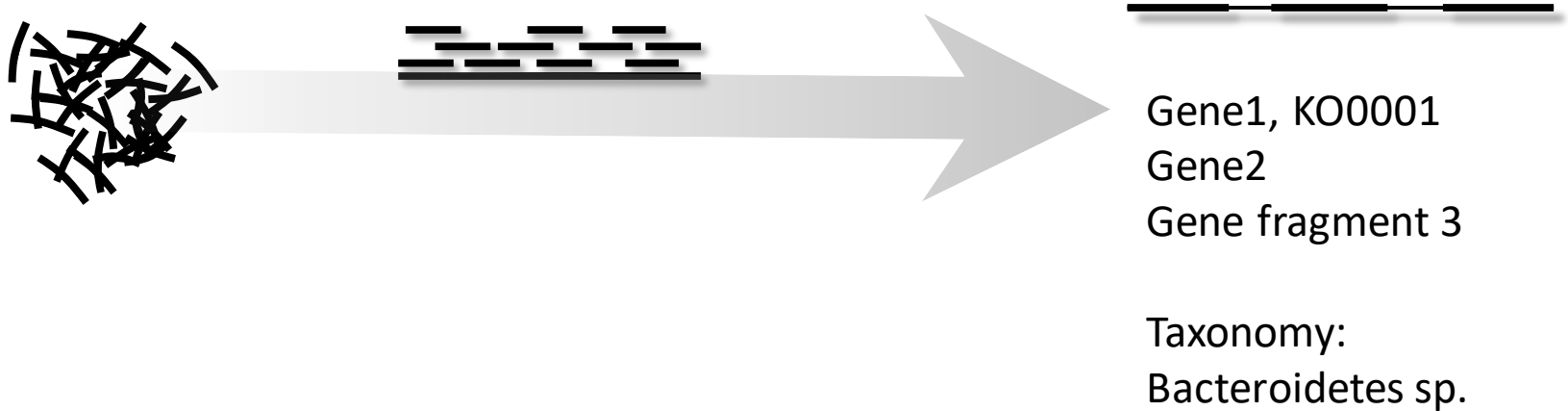
$K = 4$



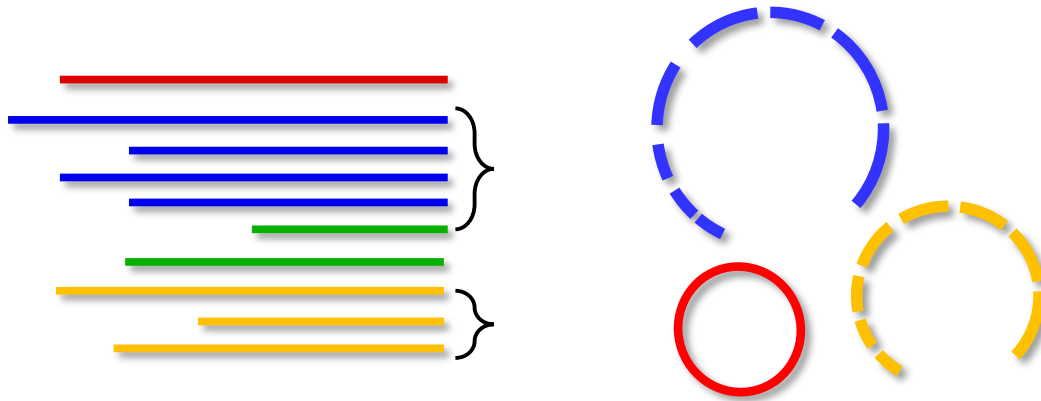
Assembly is challenging



Annotate genes on contigs



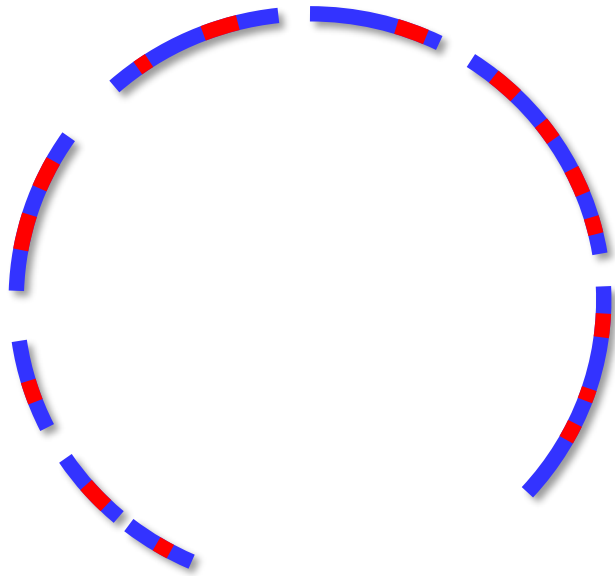
Binning contigs to metagenome assembled genomes (MAGs)



Based on:

- Sequence composition
- Abundance

Quality assessment of MAGs using sets of marker genes

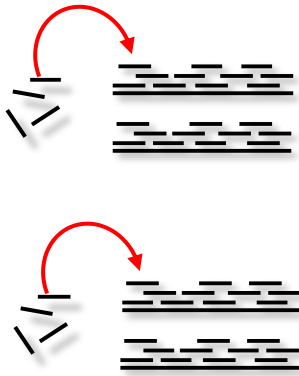


Completeness= present / expected

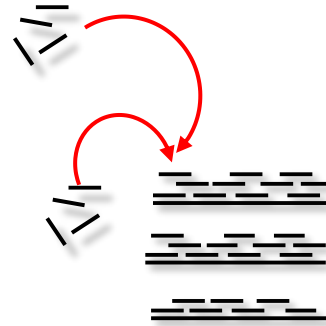
Contamination = duplicated / expected

Multiple samples

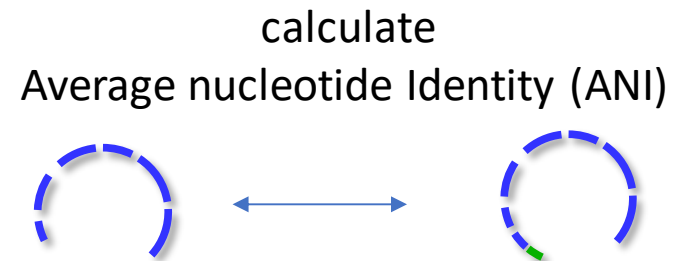
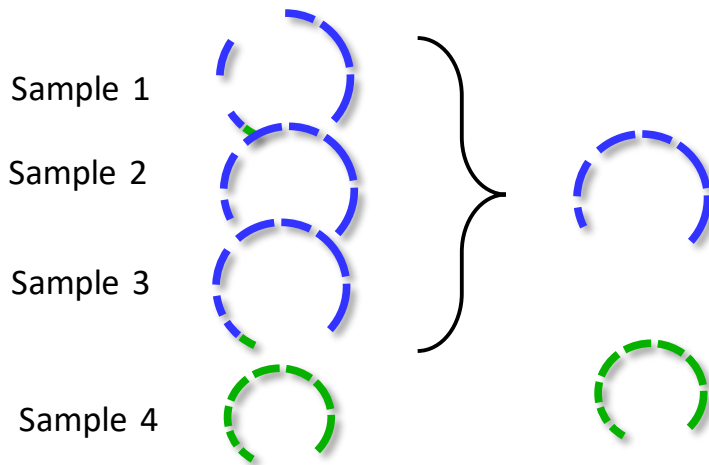
Single-sample
assembly

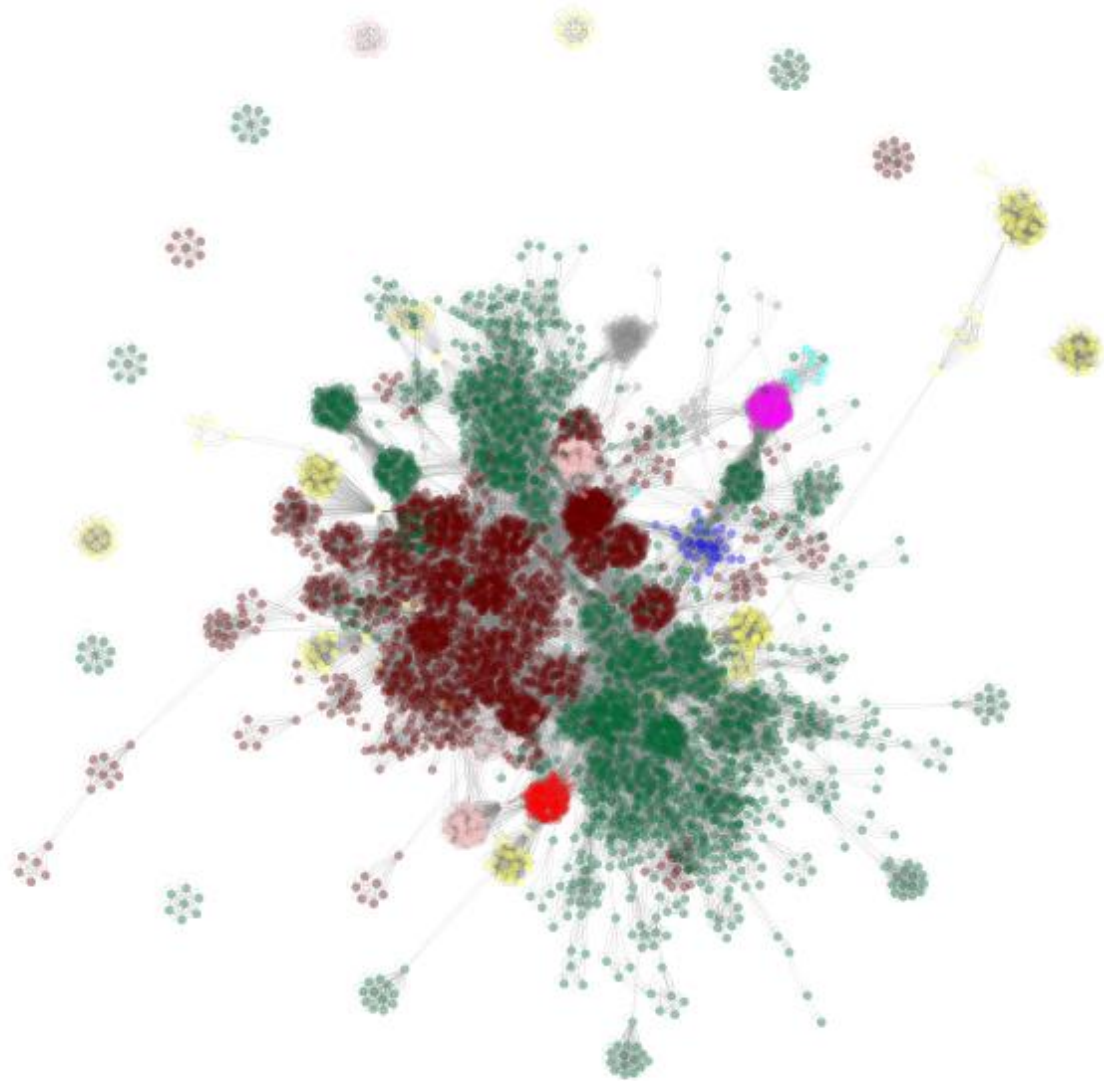


Co-assembly

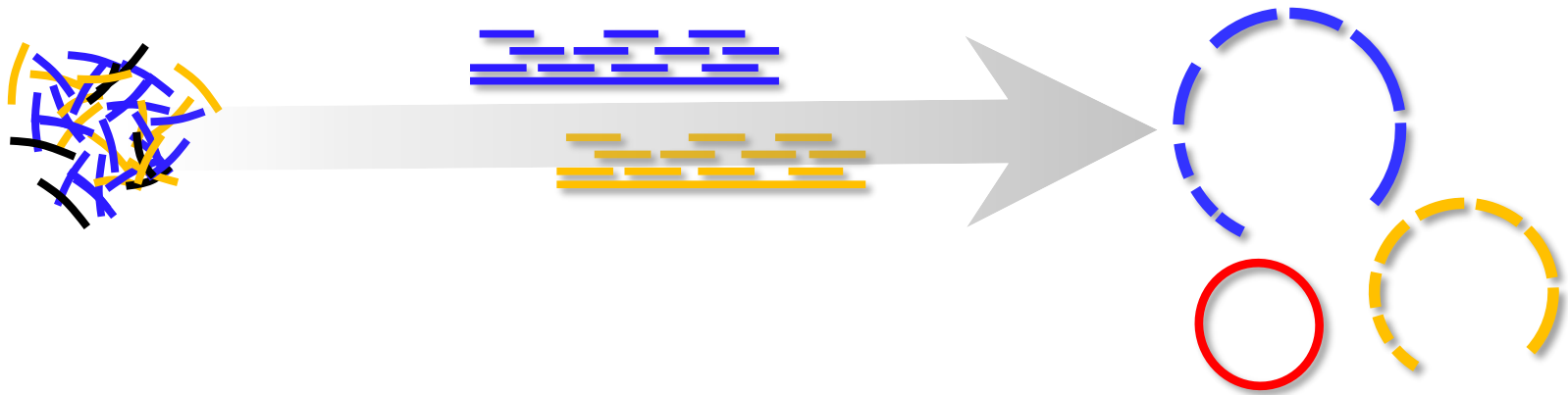


Remove redundant genomes





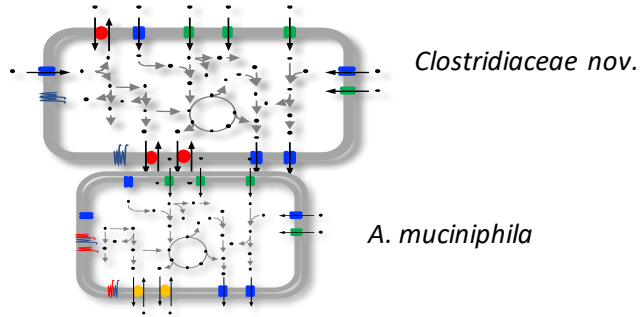
Great, we have genomes.



What is next?

- Annotation

- Taxonomy
- Functions



- Quantification

- Compare

Content

Part 1: Metagenome Atlas

- Introduction to metagenomics
- Metagenome Atlas presentation
- break
- Metagenome Atlas Tutorial

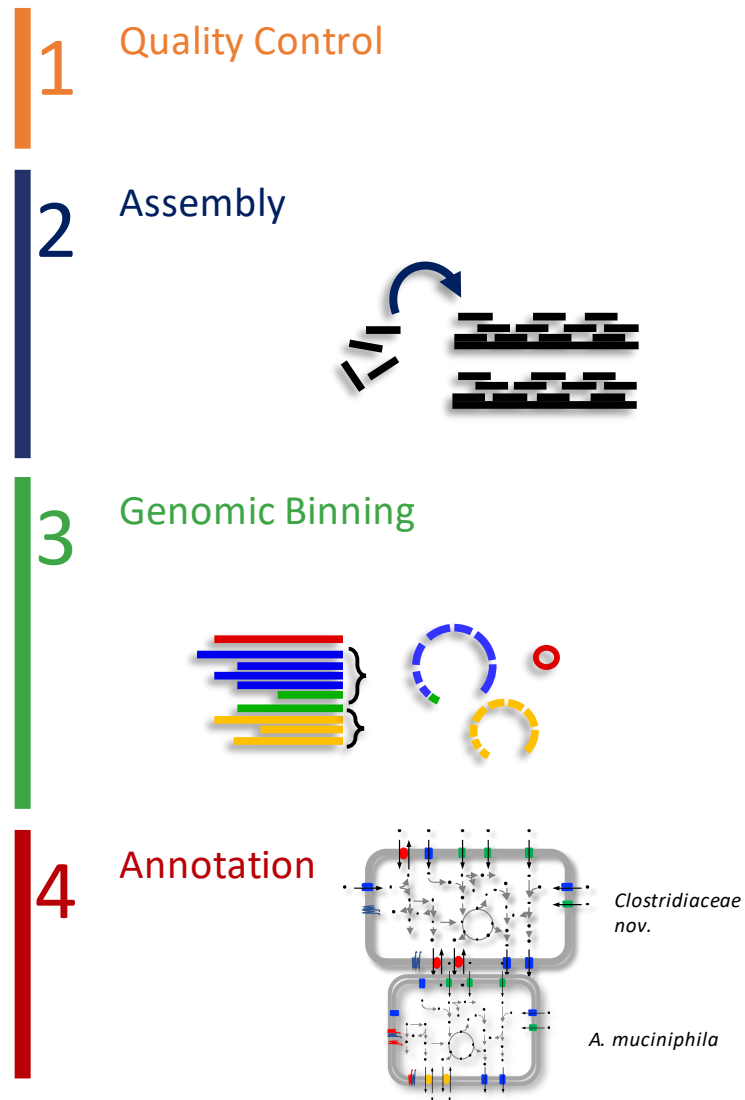
Part 2: MMseqs2

Metagenome-Atlas

The All-in-one metagenome pipeline.

Metagenome-Atlas

The All-in-one metagenome pipeline.



Start in three commands!

```
conda install metagenome-atlas  
atlas init path/to/fastq  
atlas run all
```



Automatic submit to clusters

```
[Sat Jun 6 12:51:00 2020]
rule get_read_stats:
  input: sample1/sequence_quality_control/sample1_deduplicated_R1.fastq.gz, sample1/sequence_quality_control/sample1_deduplicated_R2.fastq.gz
  output: sample1/sequence_quality_control/read_stats/deduplicated.zip, sample1/sequence_quality_control/read_stats/deduplicated_read_count.txt
  log: sample1/logs/QC/read_stats/deduplicated.log
  jobid: 42
  wildcards: sample=sample1, step=deduplicated
  priority: 30
  threads: 4
  resources: mem=10, java_mem=8, time=0.5

submit command: sbatch --parsable --job-name=get_read_stats -n 4 --mem=10g --partition=mono-shared-EL7 --time=30 -N 1 /srv/beegfs/scratch/user/h/snakejob.get_read_stats.42.sh
Submitted job 42 with external jobid '34361241'.
[Sat Jun 6 12:51:09 2020]
Finished job 36.
10 of 76 steps (13%) done
```

Wrapper for LSF, Pbs-torque and Slurm, ...

Metagenome-Atlas in detail



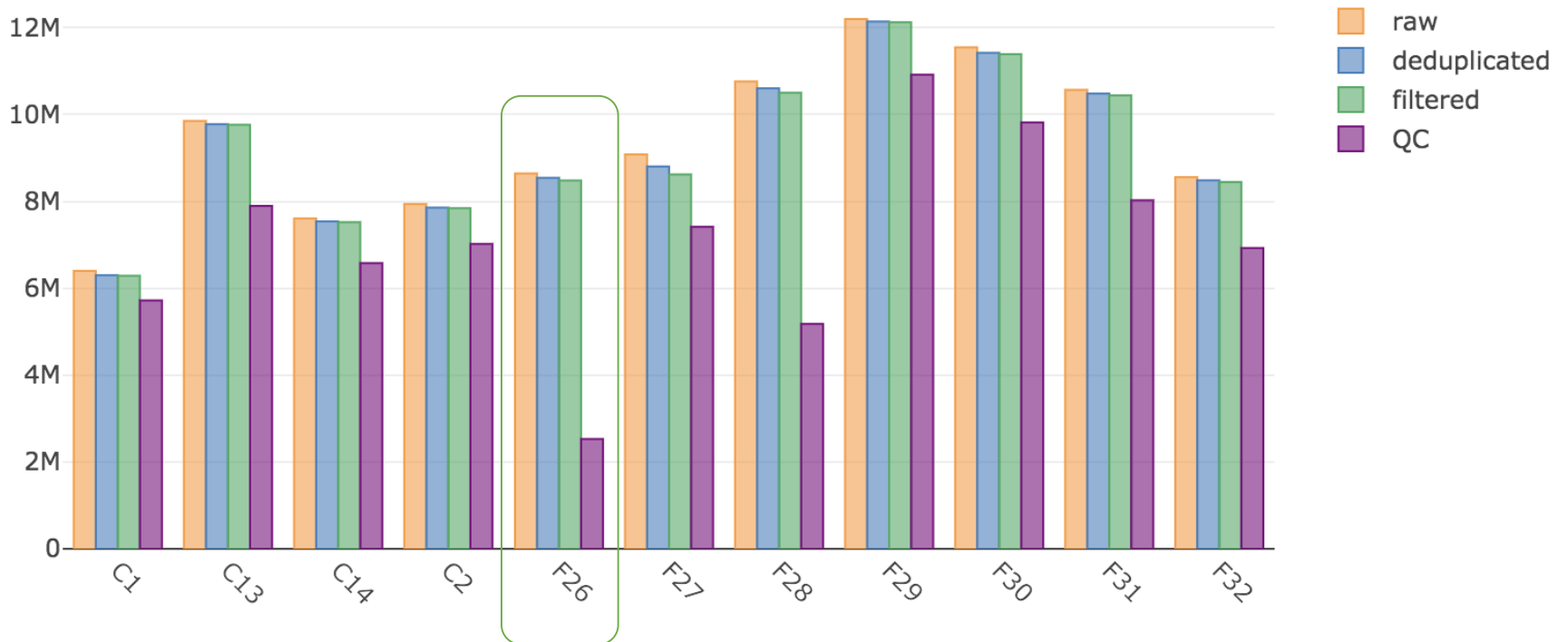
1. Quality control

- Remove low quality bases
- Common contaminant removal
- Host removal

➤ **Good-quality reads**

Quality report

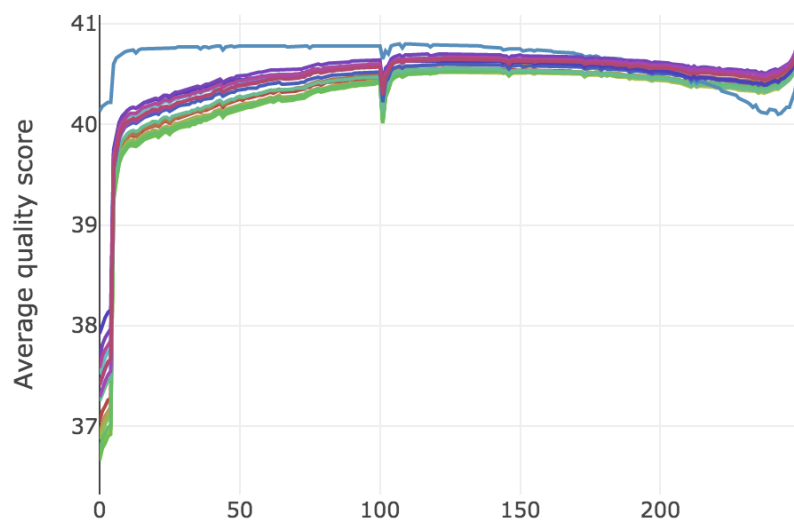
Total reads per sample



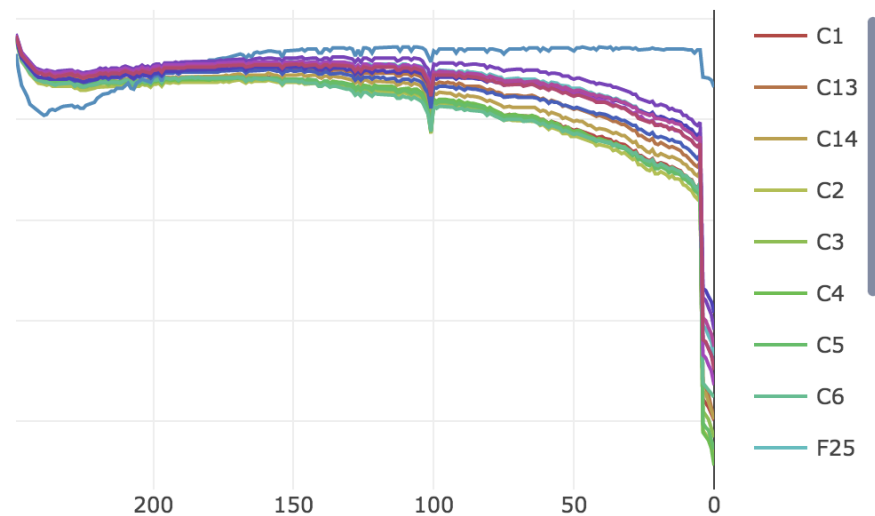
Quality report

Reads quality after QC

Paired end



Position forward read



Position reverse read

C1
C13
C14
C2
C3
C4
C5
C6
F25

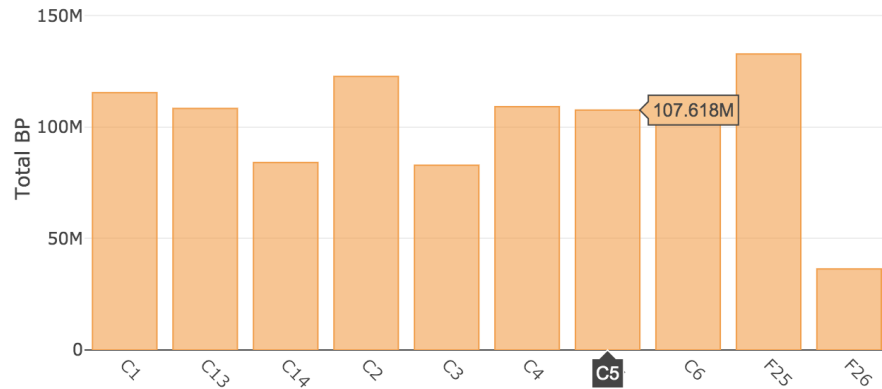
2. Assembly

- Uses metaSpades or megahit
- Error correction
- Paired-end merging (pre-assembly)
- Hybrid assembly supported.

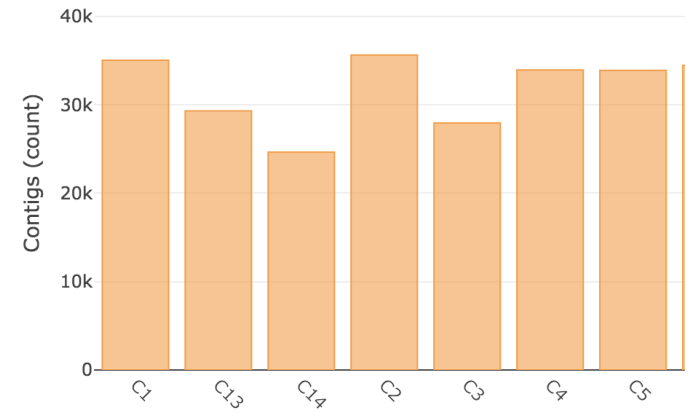
➤ **Scaffolds**

Assembly report

Assembly Length



Number of Contigs

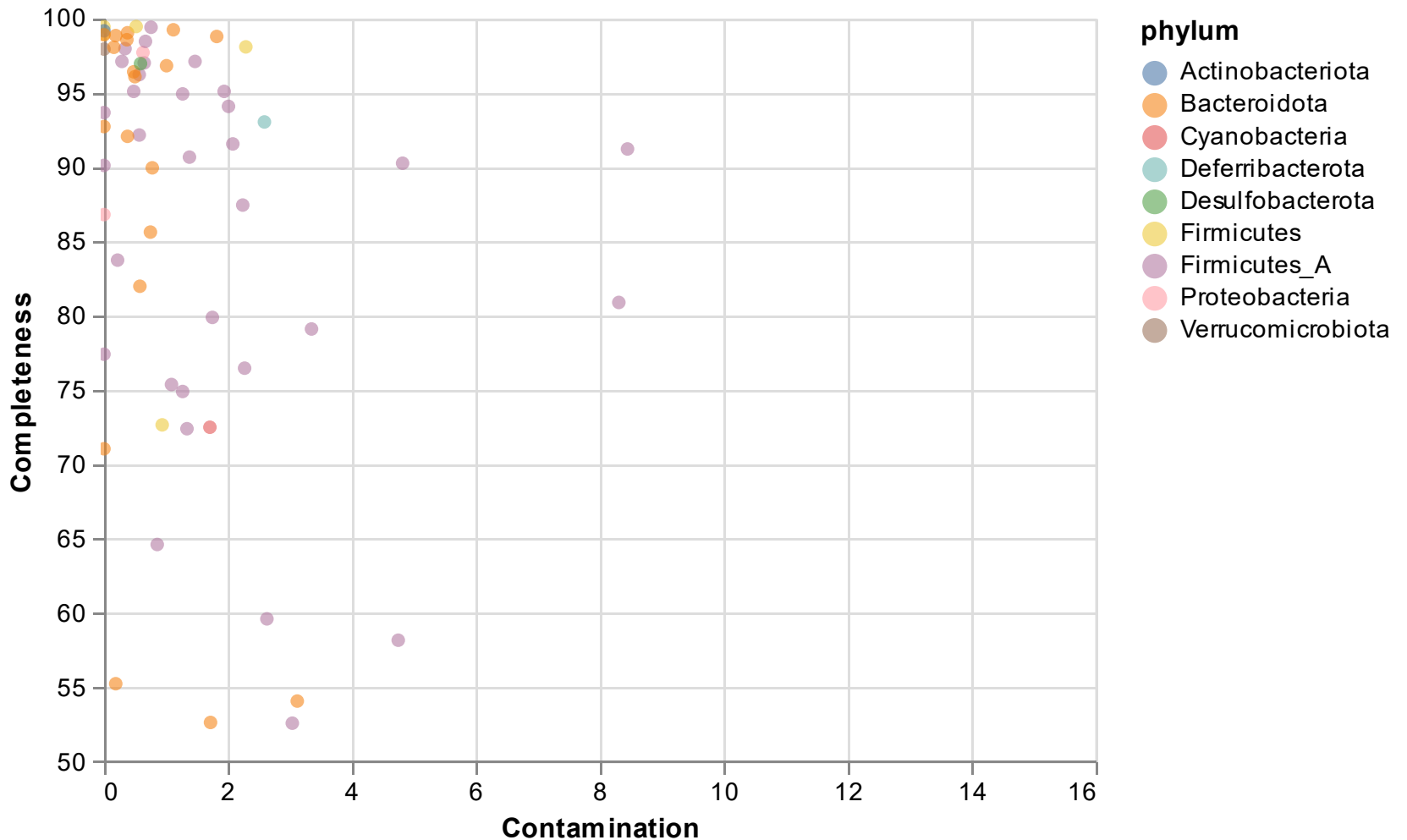


3.a Binning

- Uses DAS Tool
 - Metabat2
 - maxbin2
- Quality Assessment (checkM)
- Co-binning supported
- Dereplication

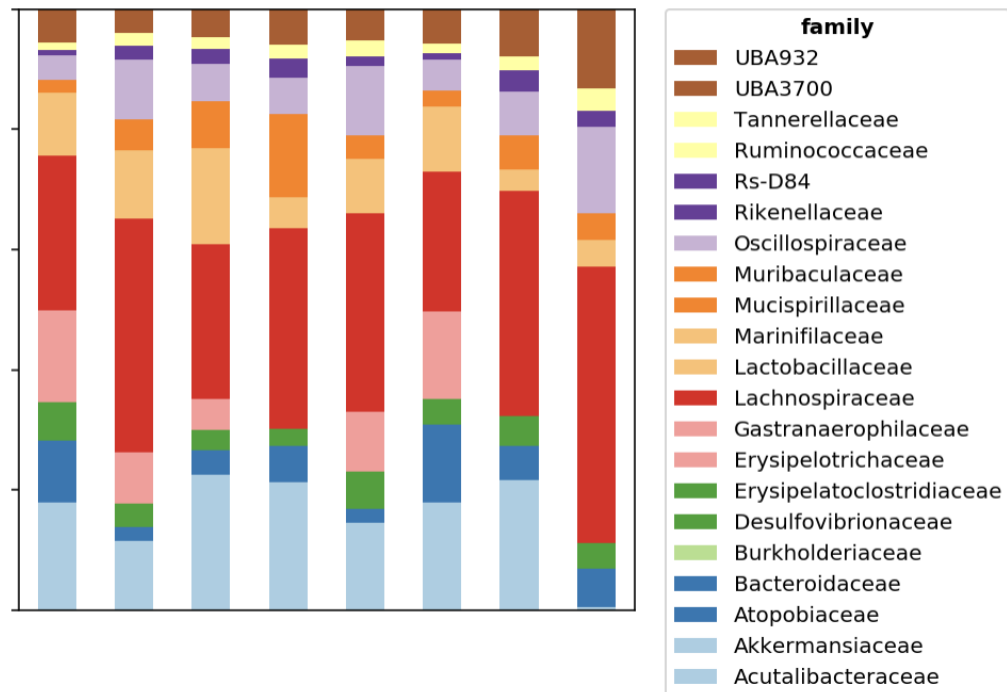
➤ Genomes

Quality of MAGs



3.b Quantification

- Bbmap
- Median coverage over the genomes
- Raw counts

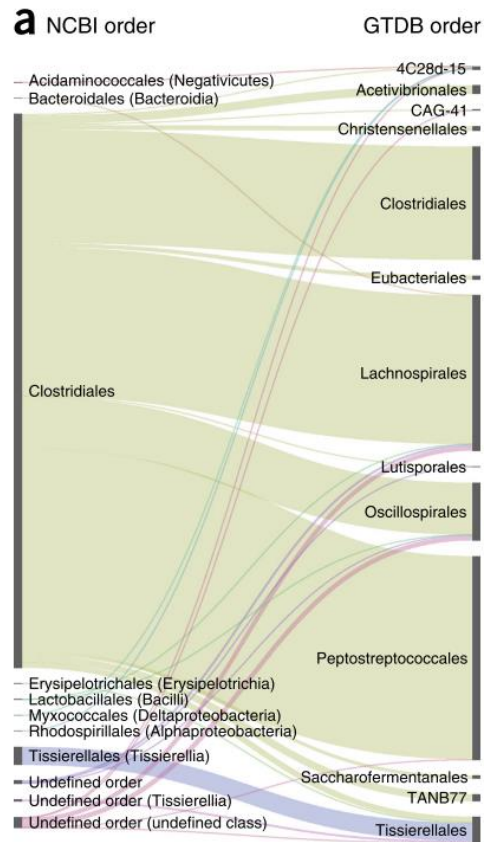


4.a Taxonomic annotation

- Genome Taxonomy database (GTDB)
- Taxonomy based on genomes and MAGs
- Split taxon that contain multiple lineages

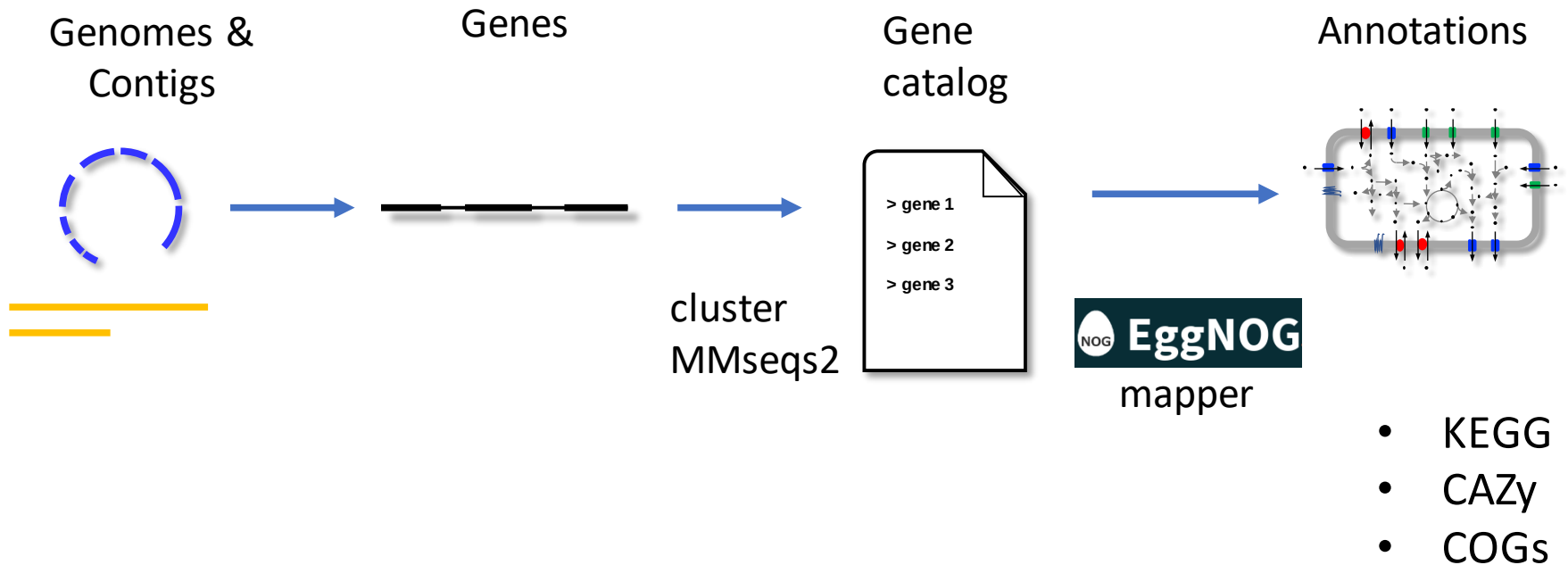
➤ **Robust taxonomy**

Proposed rearrangements

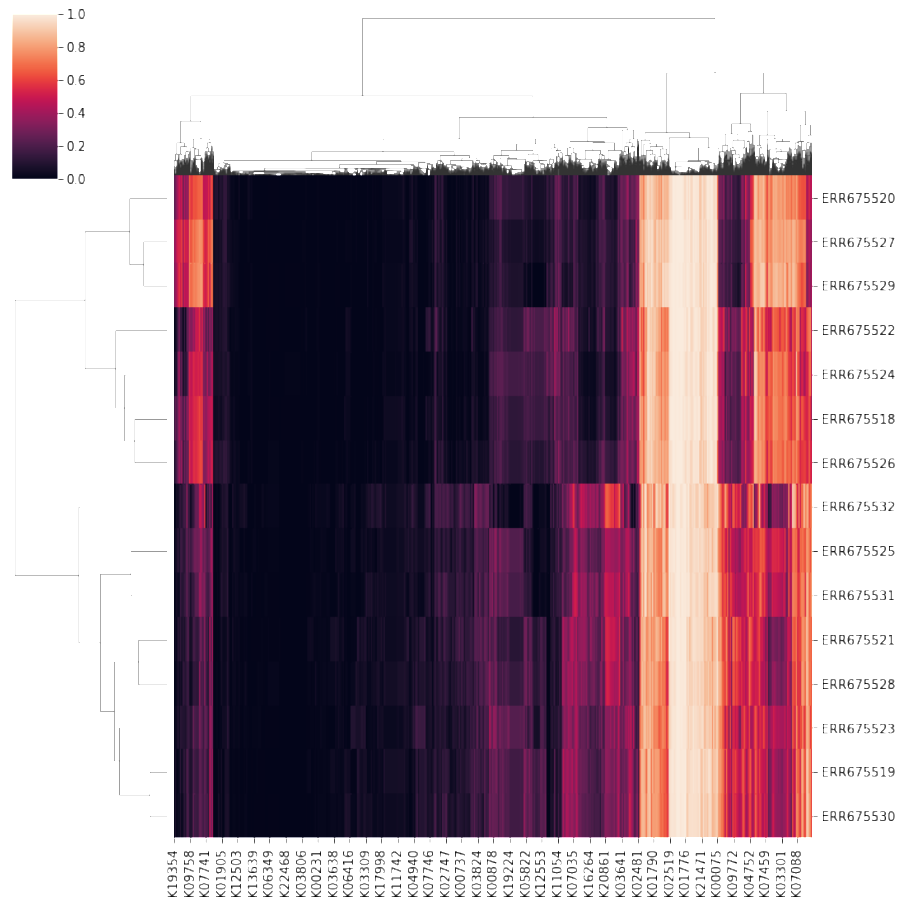


Parks et al. 10.1038/nbt.4229.

4.b Functional annotation



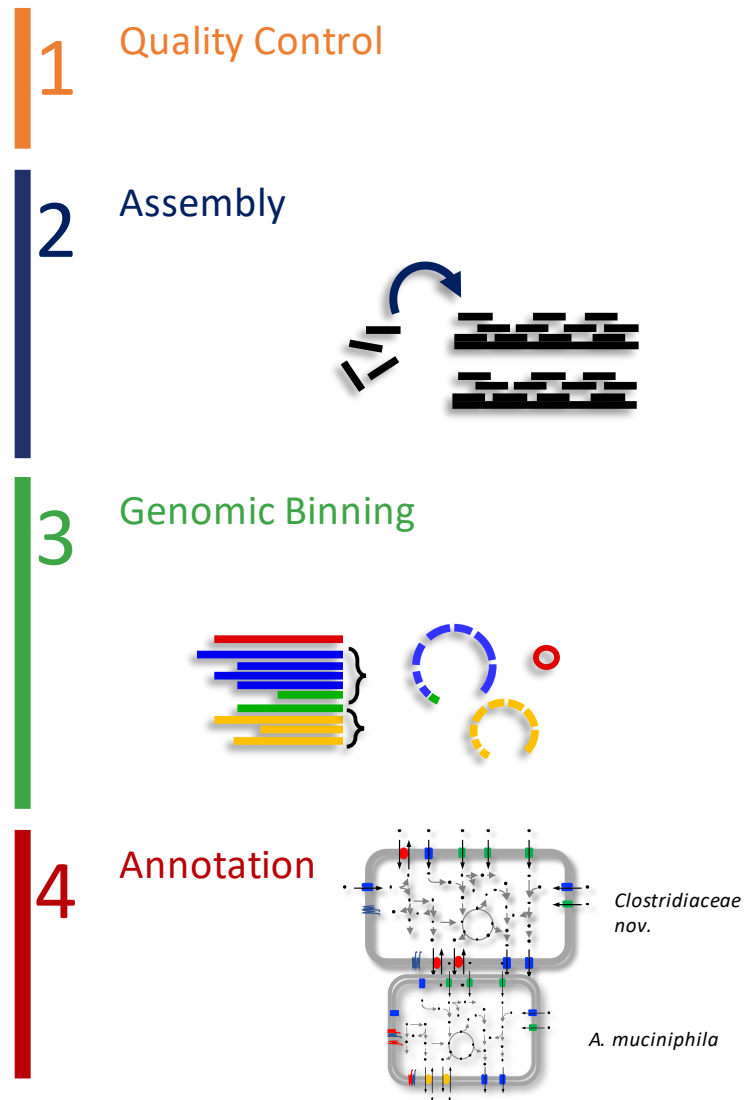
Output of Annotation



KEGG Orthologs

Metagenome-Atlas

The All-in-one metagenome pipeline.



Tutorial

1. Analyze the output of Metagenome-atlas
2. Run the three commands on a dummy dataset

github.com/metagenome-atlas/Tutorial

How does high fat diet impact the microbiome of mice?

How does high fat diet impact the microbiome of mice?



```
In [2]: #go to output file
        %cd ../Example/

        /Users/silas/Documents/GitHub/Tutorial/Example
```

```
In [3]: # load metadata
        metadata= pd.read_table('metadata.txt',index_col=0)
        metadata.head()
```

```
Out[3]:
```

	Age	Air_temperature	Diet	Food_name	Country	Provider	Sex	Strain	Body_weight
ERR675518	17.0	30.0	chow	D12450B	Norway	Taconic in Denmark	male	C57/BL6	25.62

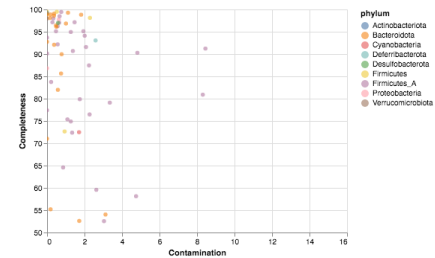
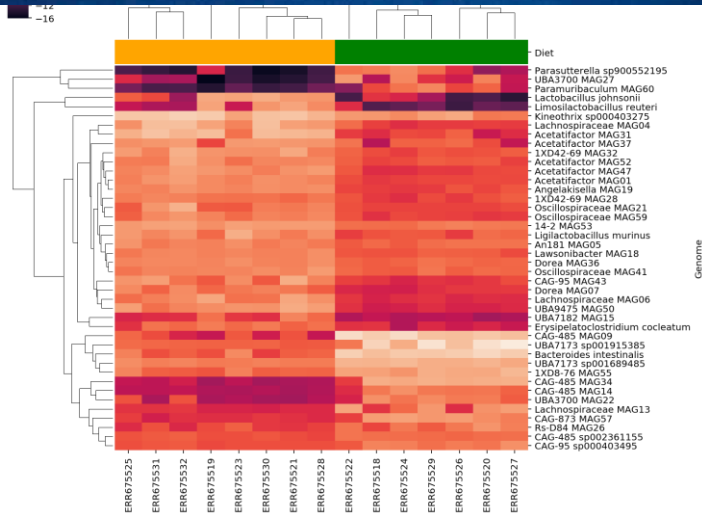
Tutorial

Have a look at the Tutorial

Ask questions if you have!

github.com/metagenome-atlas/Tutorial

1. Analyze the output of Atlas



Part 2: Set up Atlas on your server/computer

```
conda install metagenome-atlas  
atlas init path/to/fastq  
atlas run all
```