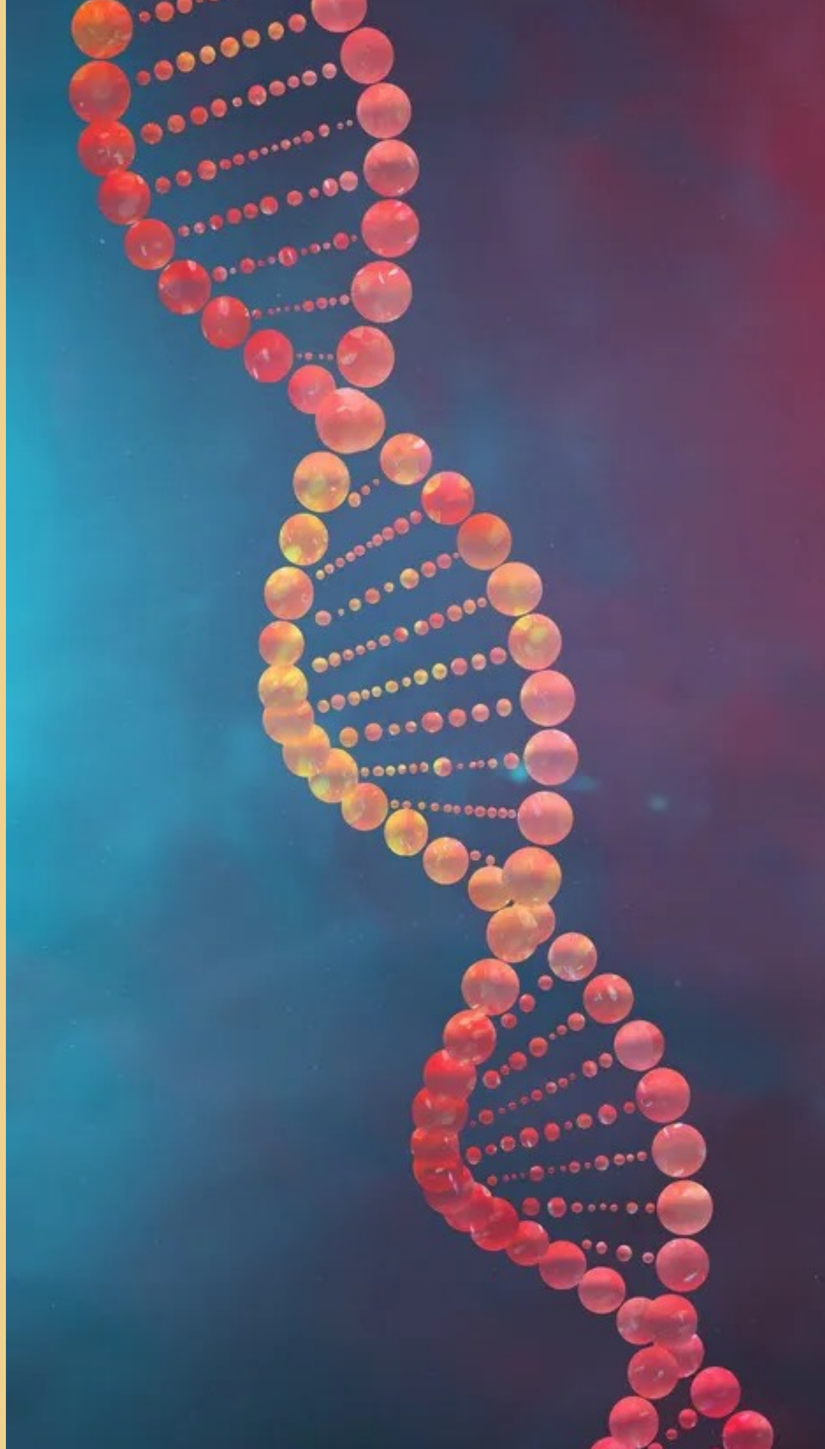
DAWG Workshop 3

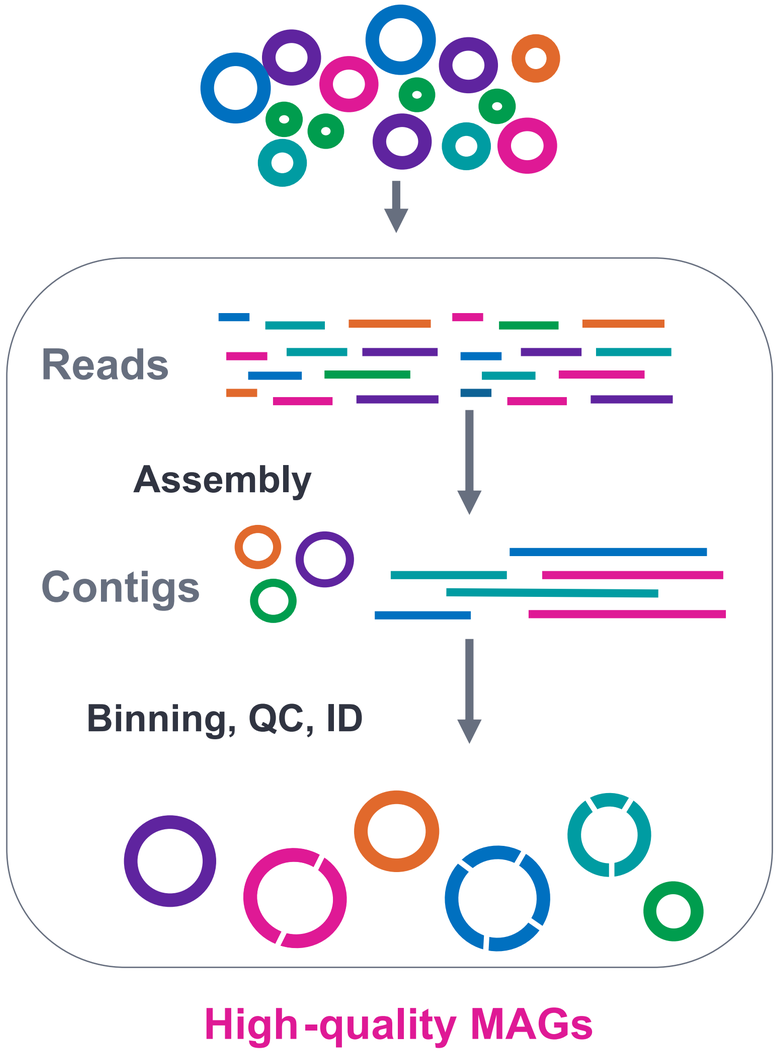
Metagenomic Sequencing
Tree of Life Workshop Series

Dec 5th, 2025



Workshop Pipeline

- QUALITY CONTROL (QC)
- ASSEMBLY
- MAPPING & COVERAGE
- BINNING
- QUALITY ASSESSMENT
- ANNOTATION & VISUALIZATION



QC

FastP

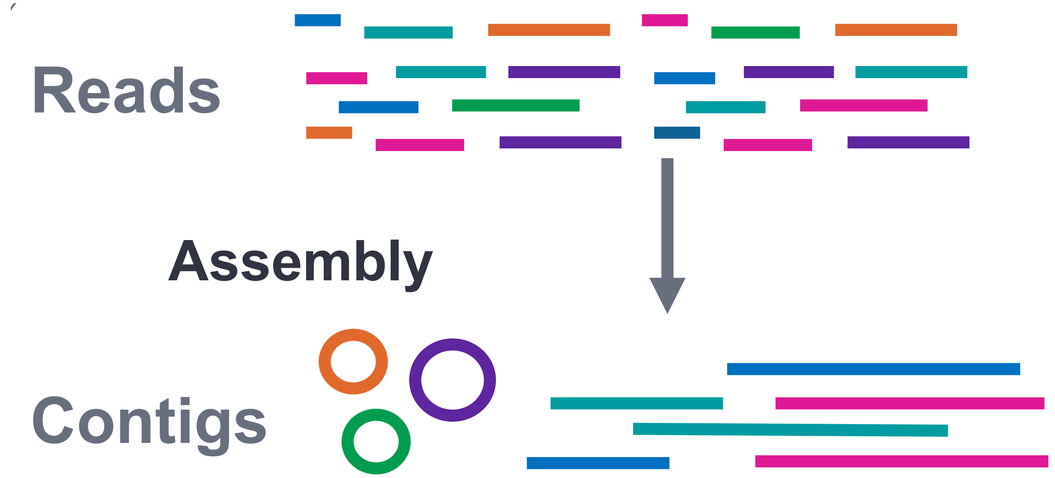
- Used for read trimming and quality control
- Removes adapter sequences
- Trims low-quality bases
- Removes reads that are too short

ASSEMBLY

MEGAHIT

INPUT: Raw reads (FastQ files)

OUTPUT: Assembled contigs (Fasta file)



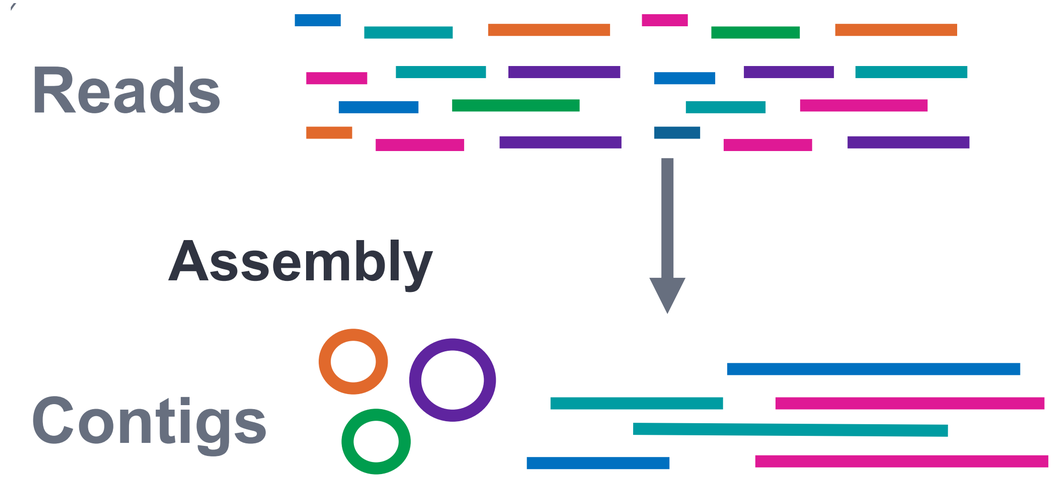
ASSEMBLY

MEGAHIT

INPUT: Raw reads (FastQ files)

OUTPUT: Assembled contigs (Fasta file)

Longer contiguous sequences



MAPPING

Bowtie2

INPUT: Raw reads (FastQ files) & Assembled contigs (Fasta files)

OUTPUT: Aligned reads (BAM file) and a subsequent depth/ coverage file

MAPPING

Bowtie2

INPUT: Raw reads (FastQ files) & Assembled contigs (Fasta files)

OUTPUT: Aligned reads (BAM file) and a subsequent depth/ coverage file

SAM file

convert to binary version using SAMTOOLS

BAM file

MAPPING

Bowtie2

INPUT: Raw reads (FastQ files) & Assembled contigs (Fasta files)

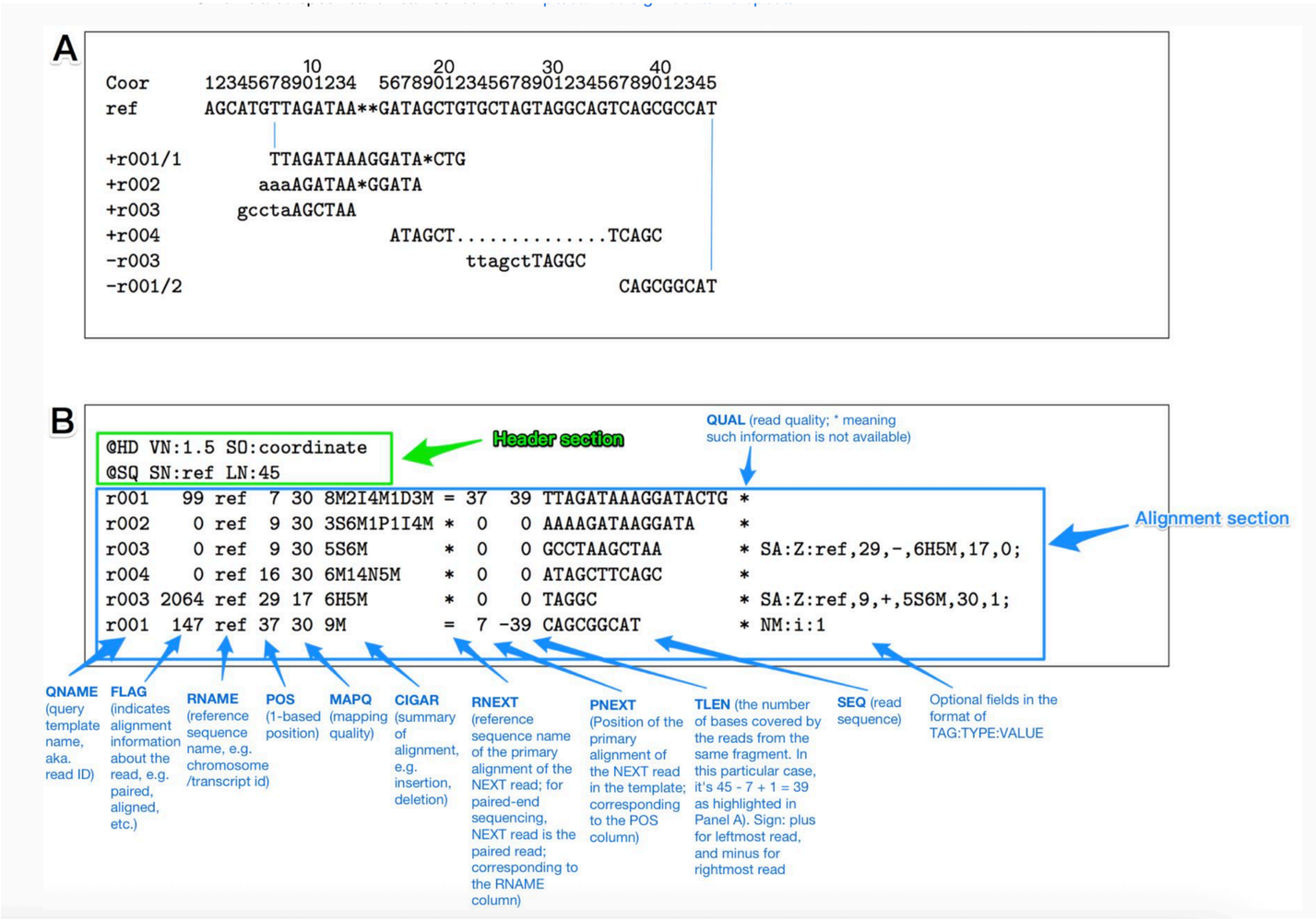
OUTPUT: Aligned reads (BAM file) and a subsequent depth/ coverage file

convert to binary version using SAMTOOLS

SAM file

BAM file

SAM file



COVERAGE

jgi_summarize_bam_contig_depths

jgi_summarize_bam_contig_depths

Script used to calculate the depth of each contig across multiple BAM files.

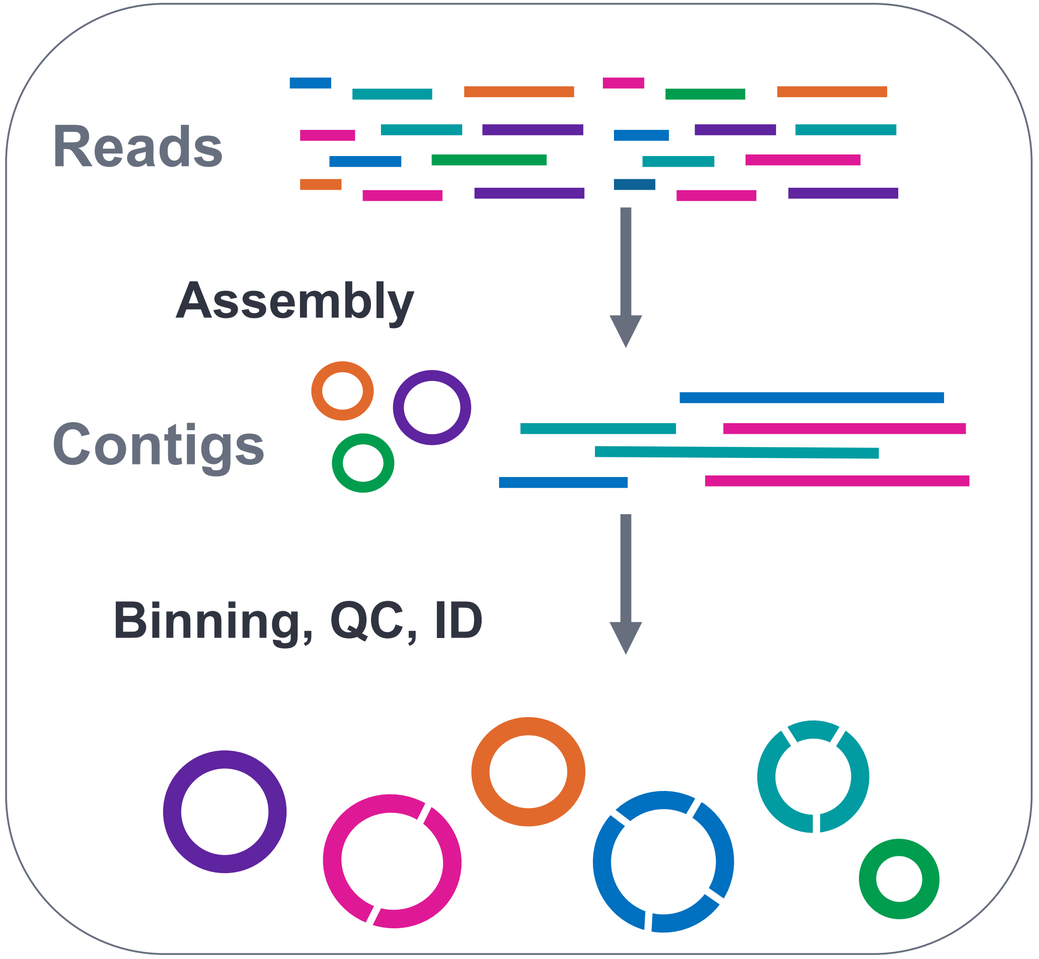
It determines the average number of reads that align to each contig.

This depth information is crucial for binning -> grouping contigs into potential genome bins based on coverage similarity. Input for **METABAT**.

BINNING

METABAT2

INPUT: Contigs (FASTA files) and coverage/depth information

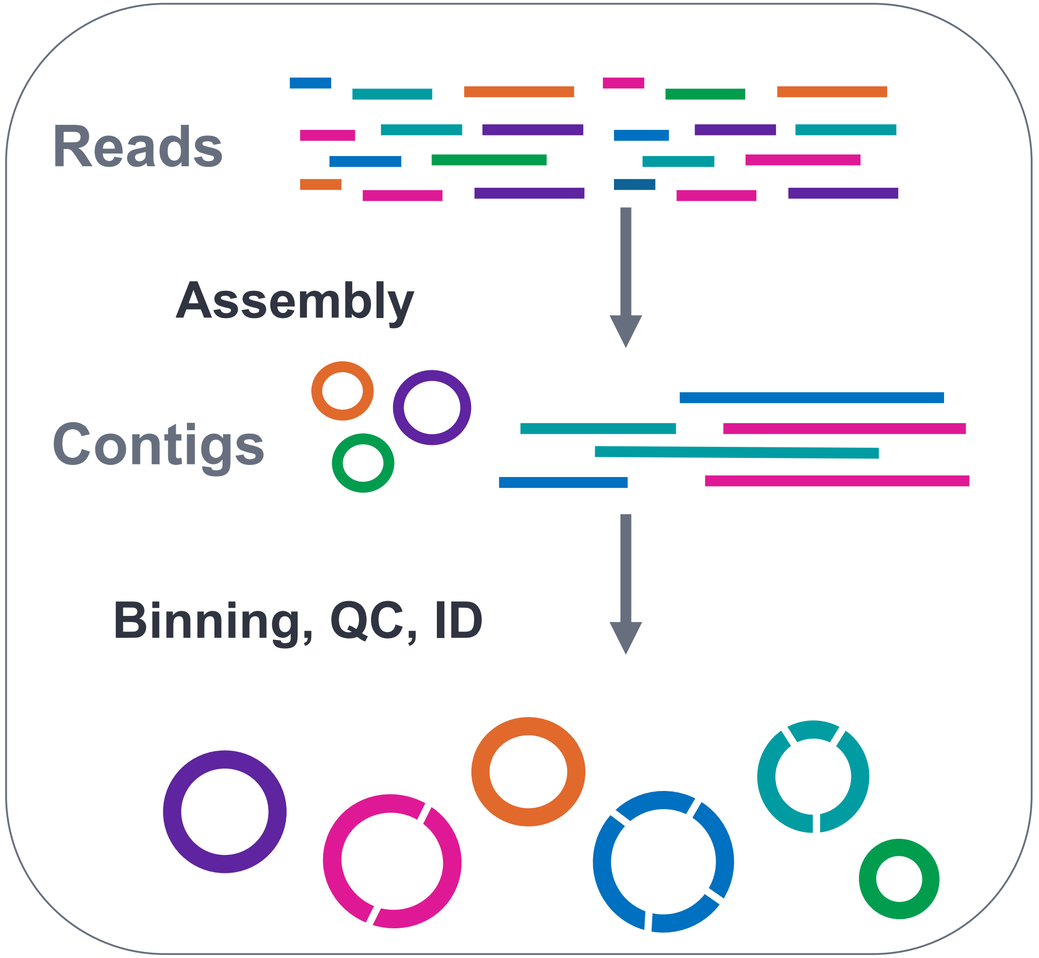


BINNING

METABAT2

INPUT: Contigs (FASTA files) and coverage/depth information

OUTPUT: Draft MAGs (multiple FASTA files, one for each "bin")



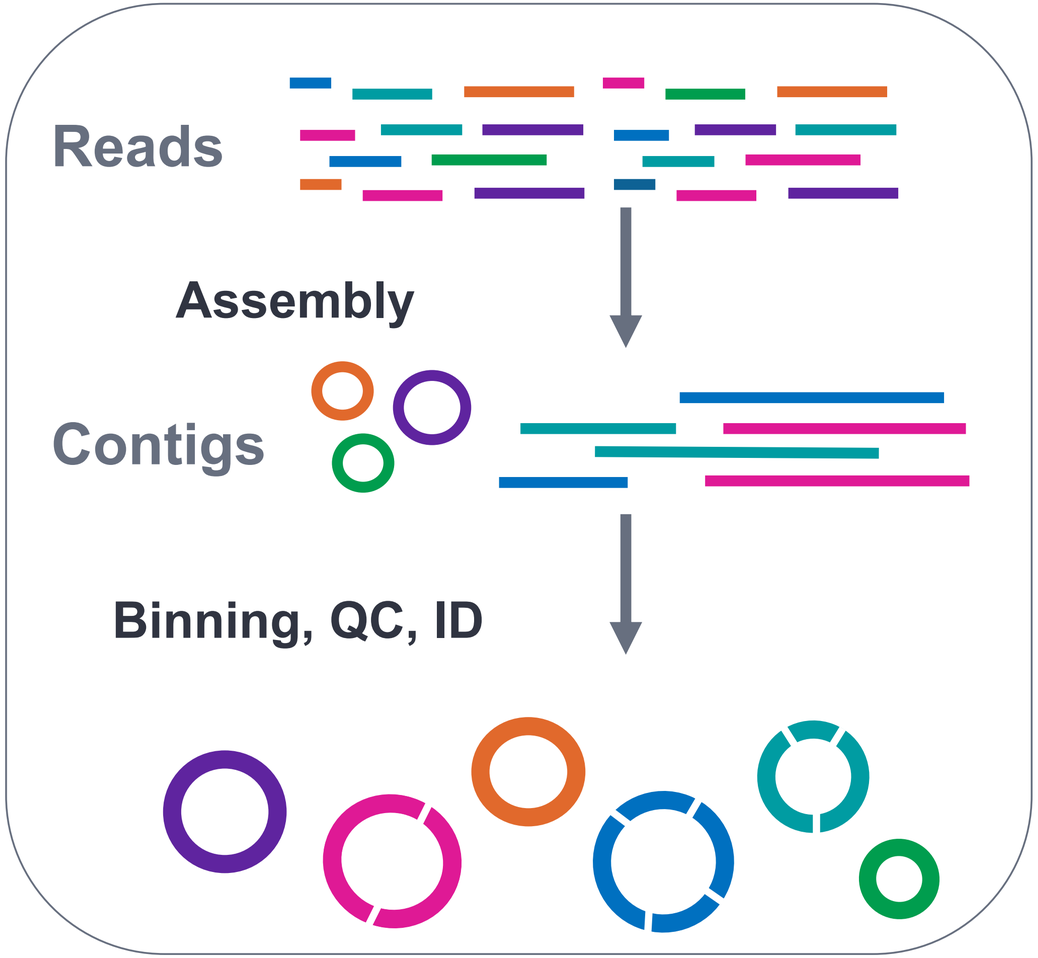
BINNING

METABAT2

INPUT: Contigs (FASTA files) and coverage/depth information

OUTPUT: Draft MAGs (multiple FASTA files, one for each "bin")

AIM: group contigs into species



QUALITY ASSESSMENT

CheckM

INPUT: Draft MAGs/bins (FASTA files)

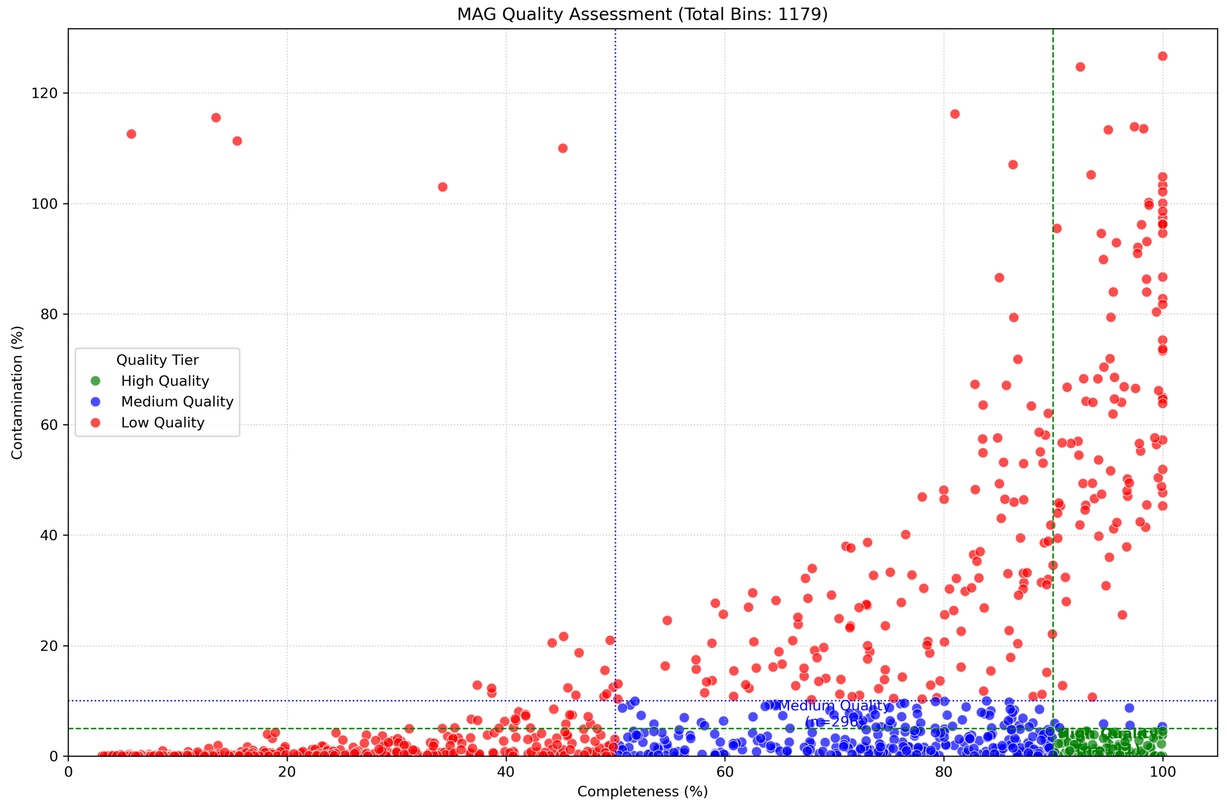
OUTPUT: Summary table with quality scores.

QUALITY ASSESSMENT

CheckM

INPUT: Draft MAGs/bins (FASTA files)

OUTPUT: Summary table with quality scores.



- [x] ANNOTATION and VISUALIZATION

☑ ANNOTATION and VISUALIZATION

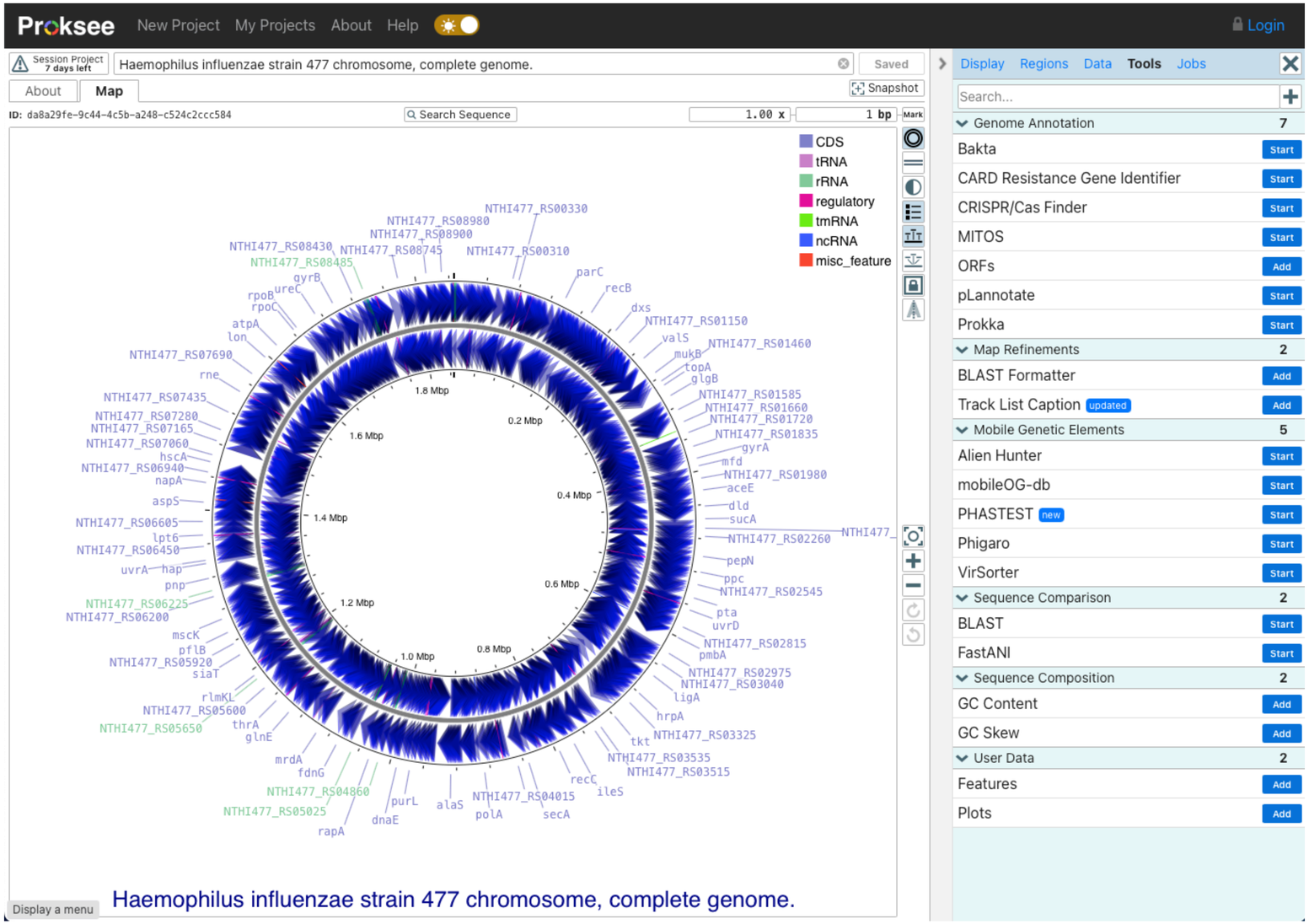
BAKTA & Proksee

☑ ANNOTATION and VISUALIZATION

BAKTA & Proksee

INPUT: High-quality MAGs (FASTA files)

☑ ANNOTATION and VISUALIZATION



BAKTA & Proksee

INPUT: High-quality MAGs (FASTA files)

☑ ANNOTATION & VISUALIZATION

ANNOTATION & VISUALIZATION

Anvi'o

☑ ANNOTATION & VISUALIZATION

Anvi'o

- Check bin quality
- Assembles reads into contigs
- Annotates MAGs
- Explore data
- Allows users to manually curate and refine genome bins and performs various other 'omics analyses, including comparative genomics and phylogenomics

ANNOTATION & VISUALIZATION

Anvi'o

- Check bin quality
- Assembles reads into contigs
- Annotates MAGs
- Explore data
- Allows users to manually curate and refine genome bins and performs various other 'omics analyses, including comparative genomics and phylogenomics