Project Plan: Genome Analysis of Oxytetracycline Production in *Streptomyces rimosus*

Jakob Siljebo

March 28, 2024

1 Project Overview

The project aims to analyze the genome of *Streptomyces rimosus* to understand the mechanisms behind oxytetracycline production. The analysis will involve genome assembly, transcriptomics, differential expression analysis, and annotation.

2 Objectives

1. Conduct de novo assembly of long reads acquired through Nanopore sequencing, followed by refinement using short Illumina reads.
2. Perform transcriptomics and comparative differential expression analyses, comparing the hyperproducer S. rimosus strain with the wild-type strain.
3. Evaluate the quality of genome assembly, annotate the genome, and assess synteny with the wild-type genome.
4. Identify candidate genes and pathways associated with antibiotic hyperproduction.
5. Document findings and methodologies for future reference.

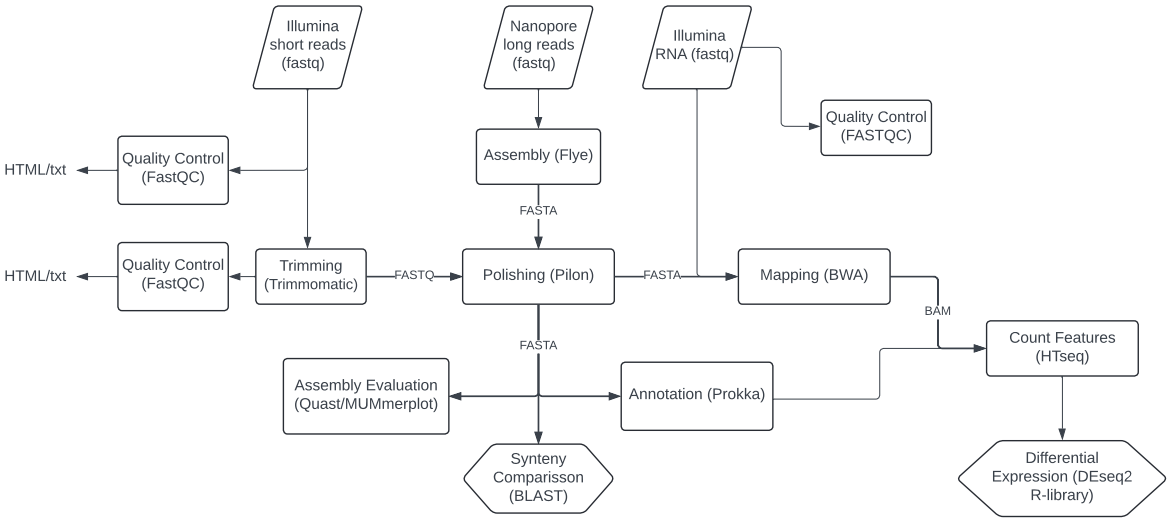


Figure 1: Overview of the project plan.

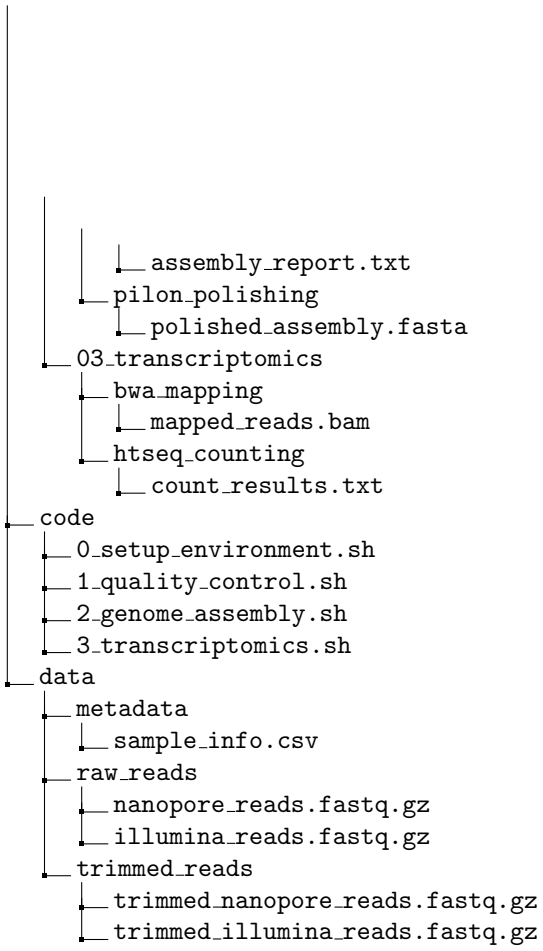
3 Timeline

Checkpoints	Tasks
28/3	Project Setup - Create an UPPMAX account - Create a GitHub repository
8/4	Quality Control and Genome Assembly - Obtain raw Nanopore and Illumina sequencing data for hyperproducer *S. rimosus* (HP126) - Perform quality control using FastQC - Assemble Nanopore reads using Flye to generate initial contigs - Trim Illumina reads using Trimmomatic to remove adapters and low-quality bases
16/4	Polishing - Polish the assembly with Illumina reads using Pilon for error correction and accuracy improvement
18/4	Genome Annotation - Evaluate assembly quality using Quast and/or MUMmerplot - Annotate the assembled genome using Prokka to identify genes, regulatory elements, and other features
23/4	Synteny Comparison - Conduct synteny comparison between the hyperproducer strain and the wild-type strain using BLAST to identify structural variations and rearrangements
25/4	RNA Quality Control - Control Illumina RNA reads with FastQC
2/5	RNA mapping - Map RNA-Seq reads to the assembled genome using BWA
6/5	Transcriptomics and Differential Expression Analysis - Quantify gene expression levels and perform differential expression analysis using HTseq and DESeq2 R-library
15/5	Data Integration and Final Analysis - Integrate results from genome annotation, synteny comparison, and differential expression analysis to identify candidate genes and pathways associated with antibiotic hyperproduction - Prepare project presentation, summarizing findings, methodologies, and conclusions

4 Data Management

In order to maintain a clear structure of all data throughout the project, the data will be categorized into three main sections: Data, Code, and Analysis. Below is an example outlining the structure:

```
genome_analysis_project/
└── analyses
    ├── 01_quality_control
    │   ├── nanopore_fastqc
    │   │   └── fastqc_report.txt
    │   └── illumina_fastqc
    │       └── fastqc_report.txt
    └── 02_genome_assembly
        └── flye_assembly
```


assembly_report.txt
pilon_polishing
polished_assembly.fasta
03_transcriptomics
bwa_mapping
mapped_reads.bam
htseq_counting
count_results.txt
code
0_setup_environment.sh
1_quality_control.sh
2_genome_assembly.sh
3_transcriptomics.sh
data
metadata
sample_info.csv
raw_reads
nanopore_reads.fastq.gz
illumina_reads.fastq.gz
trimmed_reads
trimmed_nanopore_reads.fastq.gz
trimmed_illumina_reads.fastq.gz