

7.3 SY14 Assembly - Initial Attempt (Script 05)

Script: 05_sy14_canu.sh

This was the initial assembly attempt for SY14 strain.

```
#!/bin/bash
#SBATCH --job-name=sy14_final
#SBATCH --output=logs/sy14_final_%j.out
#SBATCH --error=logs/sy14_final_%j.err
#SBATCH --cpus-per-task=4
#SBATCH --mem=8G
#SBATCH --time=1:00:00
#SBATCH --partition=msc_appbio

module load anaconda3
conda activate canu_env

cd /scratch/grp/msc_appbio/Group2_ABCC/Gene_Assembling/assembly

canu \
-p SY14 \
-d SY14 \
genomeSize=12m \
-pacbio /cephfs/volumes/hpc_data_grp/msc_appbio/70e465a3-ea70-4bd6-b919-9da7af124271/Group2_ABCC/
    Gene_Assembling/data/SY14/pacbio/SRR6823435_1.fastq.gz \
maxThreads=4 \
maxMemory=8
```

Status: Partially successful – assembly pipeline completed but final output not generated.

What completed successfully:

- Read correction (2.4 Gb input, 278,658 reads)
- Read trimming
- Unitigging (contig construction)
- Consensus sequence generation

Error/Issue: Final FASTA file generation failed due to SLURM job dependency error.

Intermediate consensus files existed but were not combined into final FASTA format.

Evidence of partial success:

- All intermediate files present in unitigging/ directories
- SY14.report contained complete assembly statistics
- Expected output files (SY14.contigs.fasta, SY14.unassembled.fasta) missing

Why this occurred: SLURM's job dependency mechanism failed to properly trigger the subsequent job. This is an infrastructure issue, not a scientific or algorithmic failure.