7.4 SY14 Assembly - Successful Rerun (Script 06)

Script: `06_sy14_assemblyrun2_recovered.sh`

After identifying the SLURM dependency issue, the identical Canu command was resubmitted.

```
#!/bin/bash
#SBATCH --job-name=sy14_assemblyrun2_recovered
#SBATCH --output=/scratch/grp/msc_appbio/Group2_ABCC/Gene_Assembling/logs/canu_%j.out
#SBATCH --error=/scratch/grp/msc_appbio/Group2_ABCC/Gene_Assembling/logs/canu_%j.err
#SBATCH --cpus-per-task=4
#SBATCH --mem=120G
#SBATCH --time=8:00:00
#SBATCH --partition=msc_appbio

# Load conda environment with Canu installed
module load anaconda3/2022.10-gcc-13.2.0
source activate canu_env

# Run Canu
canu \
  -p SY14 \
  -d /scratch/grp/msc_appbio/Group2_ABCC/Gene_Assembling/assembly/SY14 \
  genomeSize=12m \
  -pacbio-raw /scratch/grp/msc_appbio/Group2_ABCC/Gene_Assembling/data/SY14/pacbio/*.fastq.gz \
  maxThreads=4 \
  maxMemory=120G
```

Status: Fully successful – assembly completed with all output files generated.

How Canu's checkpoint recovery worked:

- Canu detected existing intermediate files from Script 05
- Automatically skipped completed stages (correction, trimming, unitigging, consensus)
- Only executed the missing final output generation step

Final assembly statistics:

- **Total contigs:** 7
- **Total assembled length:** 11.78 Mb (98.5% of expected 12 Mb genome)
- **N50:** 11.8 Mb
- **Largest contig:** 11.8 Mb (essentially the entire fused chromosome)
- **Unassembled sequences:** 330 sequences, 1.9 Mb (1.5% of genome)

Assembly quality assessment:

- Excellent contiguity (N50 = 11.8 Mb indicates one major contig)
- High genome coverage (98.5% assembled)
- Consistent with expected SY14 phenotype (single-chromosome strain)
- Successfully used for downstream Hi-C read alignment

Output files generated:

- `SY14.contigs.fasta` – Final assembled contigs
- `SY14.unassembled.fasta` – Unassembled reads
- `SY14.report` – Detailed assembly statistics
- `SY14.contigs.layout.tigInfo` – Contig metadata
- `SY14.contigs.bam` – Read alignments to contigs