

Downloading sequence datasets for Genome Assembly

```
#!/bin/bash
#SBATCH --job-name=yeast_download
#SBATCH --output=yeast_download.log
#SBATCH --ntasks=1
#SBATCH --cpus-per-task=4
#SBATCH --time=6:00:00
#SBATCH --mem=8G
#SBATCH --partition=msc_appbio

module load sra-tools/3.0.3-gcc-13.2.0

# This script downloads all sequencing datasets used in the study
# for both yeast strains (SY14 and BY4742).
# Data are organised by strain and sequencing technology
# (PacBio long reads and Illumina short reads).

echo "Starting downloads at $(date)"

# -----
# SY14 long-read data
# -----
# Download PacBio sequencing data for the SY14 strain,
# which will be used for de novo genome assembly.
cd SY14/pacbio
prefetch SRR6823435
fastq-dump --split-files SRR6823435
gzip *.fastq
cd ../..

# -----
# SY14 short-read data
# -----
# Download Illumina sequencing data for SY14,
# used for polishing and validation of the genome assembly.
cd SY14/illumina
prefetch SRR6825081
fasterq-dump --split-files SRR6825081
gzip *.fastq
cd ../..

# -----
# BY4742 long-read data
# -----
# Download PacBio sequencing data for the wild-type BY4742 strain.

cd BY4742/pacbio
prefetch SRR6823437
fastq-dump --split-files SRR6823437
gzip *.fastq

prefetch SRR6823436
fastq-dump --split-files SRR6823436
gzip *.fastq
cd ../..

# -----
# BY4742 short-read data
# -----
# Download Illumina sequencing data for BY4742,
# used for assembly polishing and variant analysis.
cd BY4742/illumina
prefetch SRR6825082
fasterq-dump --split-files SRR6825082
gzip *.fastq
cd ../..

echo "All downloads completed at $(date)"
```