

```

# Reference genome and annotation download (HPC workflow)
# This script downloads the reference genome and gene annotation files for Saccharomyces
cerevisiae and prepares an RSEM
reference for RNA-seq quantification.
# Reference genome: Organism: Saccharomyces cerevisiae (yeast), Strain: S288C, Assembly: R64,
GenBank assembly accession: GCF_000146045.2
# Download the reference genome and gene annotation files from NCBI using the provided GCF
(GenBank assembly) accession number.

# -----
# 1) create directory for reference genome files.
# -----
mkdir ref_genome
cd ref_genome

# -----
# 2) Download the genomic DNA sequence in FASTA format.
# -----
# File extension: .fna.gz = compressed FASTA file containing the reference genome sequence

wget
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.fna.gz

# -----
# 3) Download the gene annotation in GTF format.
# -----
# File extension: .gtf.gz → compressed GTF file containing gene, transcript, exon coordinates

wget
https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.gtf.gz

# -----
# 4) Decompress downloaded files
# -----
gunzip *.gz

# Uncompress all .gz files in the current directory.
# Resulting files:
# GCF_000146045.2_R64_genomic.fna
# GCF_000146045.2_R64_genomic.gtf

# -----
# Building an RSEM reference for yeast
# -----
# Building an RSEM reference for yeast using:
#   genome FASTA: GCF_000146045.2_R64_genomic.fna
#   annotation GTF: GCF_000146045.2_R64_genomic.gtf
#   and generate Bowtie2 indices for read alignment and quantification within RSEM.
#   Run this in the directory that contains the .fna and .gtf files.

# -----
# 5) Set up software for RSEM reference preparation
# -----


# Load Conda

module avail conda      # shows available conda/anaconda modules
module load anaconda3

# Configure conda channels

conda config --add channels defaults

```

```

conda config --add channels bioconda
conda config --add channels conda-forge

# -----
# 6) Create and activate an environment containing RSEM
# -----

# Create and activate new conda environment named rsem_env and install the package "rsem"
conda create -n rsem_env rsem
conda activate rsem_env

# Verify RSEM installation and version
rsem-calculate-expression --version

# -----
# 7) Load Bowtie2 aligner
# -----

# used by RSEM for indexing/alignment

module avail bowtie2
module load bowtie2

# Check Bowtie2 version

bowtie2 --version

# -----
# 8) Prepare the RSEM reference (build transcript reference + Bowtie2 indices)
# -----


#This command:
# Builds transcript reference files from the genome and GTF
# Generates Bowtie2 index files
# Produces files required for RNA-seq quantification with RSEM

rsem-prepare-reference \
    --bowtie2 \                                # tells RSEM to specifically build Bowtie2 indices
    --gtf GCF_000146045.2_R64_genomic.gtf \    # provides gene annotation in GTF format
    GCF_000146045.2_R64_genomic.fna \          # reference genome FASTA
    yeast_rsem                                # output reference prefix (creates multiple files)

# Output:
# Multiple reference and index files with prefix "yeast_rsem"
# transcript reference files derived from the genome + GTF
# Bowtie2 index files needed for alignment-based quantification
# These files will be used in downstream RNA-seq expression analysis

```