

# PRJNA297664

Maribel Alcoriza Balaguer and Guillermo Ayala

2024-04-02

## Table of contents

Download the sra files . . . . .	1
Generating the fastq files . . . . .	2
Download the index file for bowtie2 . . . . .	2
Generating the directories with the index files . . . . .	2
Short read alignment using bowtie2 . . . . .	2
From sam to bam using samtools and sorting the reads . . . . .	2
Creating the file bamfiles.txt with the names of the bam files . . . . .	3
Using Rsamtools for counting the reads aligned . . . . .	3
Adding identifiers . . . . .	4
Saving data set . . . . .	4

## Download the sra files

```
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP%2FSRP064%2FSRP
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP%2FSRP064%2FSRP
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP%2FSRP064%2FSRP
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP%2FSRP064%2FSRP
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP%2FSRP064%2FSRP
wget ftp://ftp-trace.ncbi.nlm.nih.gov/sra/sra-instant/reads/ByStudy/sra/SRP%2FSRP064%2FSRP
```

## Generating the fastq files

```
fastq-dump -I --split-files SRR2549634.sra
fastq-dump -I --split-files SRR2549636.sra
fastq-dump -I --split-files SRR2549638.sra
fastq-dump -I --split-files SRR2549635.sra
fastq-dump -I --split-files SRR2549637.sra
fastq-dump -I --split-files SRR2549639.sra
```

## Download the index file for bowtie2

```
wget ftp://igenome:G3nom3s4u@uscd-ftp.illumina.com/Saccharomyces_cerevisiae/Ensembl/R64-1-1-
```

## Generating the directories with the index files

```
gzip -d Saccharomyces_cerevisiae_Ensembl_R64-1-1.tar.gz
tar xvf Saccharomyces_cerevisiae_Ensembl_R64-1-1.tar
```

## Short read alignment using bowtie2

```
bowtie2 -x Saccharomyces_cerevisiae/Ensembl/R64-1-1/Sequence/Bowtie2Index/genome -U SRR254
bowtie2 -x Saccharomyces_cerevisiae/Ensembl/R64-1-1/Sequence/Bowtie2Index/genome -U SRR254
bowtie2 -x Saccharomyces_cerevisiae/Ensembl/R64-1-1/Sequence/Bowtie2Index/genome -U SRR254
bowtie2 -x Saccharomyces_cerevisiae/Ensembl/R64-1-1/Sequence/Bowtie2Index/genome -U SRR254
bowtie2 -x Saccharomyces_cerevisiae/Ensembl/R64-1-1/Sequence/Bowtie2Index/genome -U SRR254
bowtie2 -x Saccharomyces_cerevisiae/Ensembl/R64-1-1/Sequence/Bowtie2Index/genome -U SRR254
```

## From sam to bam using samtools and sorting the reads

```
samtools view -bS SRR2549634_1.sam | samtools sort - SRR2549634_1
samtools view -bS SRR2549636_1.sam | samtools sort - SRR2549636_1
samtools view -bS SRR2549638_1.sam | samtools sort - SRR2549638_1
samtools view -bS SRR2549635_1.sam | samtools sort - SRR2549635_1
samtools view -bS SRR2549637_1.sam | samtools sort - SRR2549637_1
samtools view -bS SRR2549639_1.sam | samtools sort - SRR2549639_1
```

## Creating the file bamfiles.txt with the names of the bam files

```
SRR2549634_1.bam
SRR2549636_1.bam
SRR2549638_1.bam
SRR2549635_1.bam
SRR2549637_1.bam
SRR2549639_1.bam
```

It can be done with

```
ls *.bam > bamfiles.txt
```

## Using Rsamtools for counting the reads aligned

```
library(Rsamtools)
library(GenomicFeatures)
sampleTable = read.table("bamfiles.txt")
dirActualData = paste(getwd(), "/", sep="")
fls = paste(dirActualData, sampleTable[,1], sep="")
bamLst = BamFileList(fls, index=character(), yieldSize=100000, obeyQname=TRUE)
gtfFile = "../Saccharomyces_cerevisiae/Ensembl/R64-1-1/Annotation/Genes/genes.gtf"
txdb = makeTxDbFromGFF(gtfFile, format="gtf")
genes = exonsBy(txdb, by="gene")
library(GenomicAlignments)
PRJNA297664 = summarizeOverlaps(features = genes, read=bamLst,
                                mode="Union",
                                singleEnd=TRUE, ## No son lecturas apareadas
                                ignore.strand=TRUE,
                                fragments=FALSE)
SampleName = c("GSM1900735", "GSM1900737", "GSM1900739", "GSM1900736",
               "GSM1900738", "GSM1900740")
Run = c("SRR2549634", "SRR2549636", "SRR2549638", "SRR2549635",
        "SRR2549637", "SRR2549639")
treatment = c(0, 0, 1, 0, 1, 1)
treatment = factor(treatment, levels=0:1, labels=c("Wild", "SEC66 deletion"))
replication = c(1, 3, 2, 2, 1, 3)
colData(PRJNA297664) = DataFrame(SampleName, Run, treatment, replication)
```

## Adding identifiers

```
a = AnnotationDbi::select(org.Sc.sgd.db,keys=rownames(PRJNA297664),
                           columns=c("ORF","ENTREZID","ENSEMBL"),keytype="ORF")
b = match(rownames(PRJNA297664),a[, "ORF"])
rowData(PRJNA297664) = a[b,]
PRJNA297664 = PRJNA297664[which(!is.na(rowData(PRJNA297664)[, "ORF"])),]
sel = match(unique(rowData(PRJNA297664)[, "ORF"]),rowData(PRJNA297664)[, "ORF"])
PRJNA297664 = PRJNA297664[sel,]
```

## Saving data set

```
save(PRJNA297664,file="PRJNA297664.rda")
```