

DIFFERENTIAL GENE EXPRESSION ANALYSIS MODULE -GENOME FILE (FASTA) AND ANNOTATION FILE (GFF) DOWNLOAD

DOWNLOAD GENOME AND ANNOTATION FILES FROM PUBLIC DATABASES

DOWNLOAD LATEST HUMAN GENOME ASSEMBLY/REFERENCE FASTA FILE AND GENOME ANNOTATION GFF3 FILE

```
#Human Genome can be downloaded from several public web portals.
```

```
# Gencode is an option - Updated with the most recent release.
```

```
# Also very clean, clear and easy to download.
```

```
# Gencode website link
```

```
# https://www.gencodegenes.org/
```

```
# https://www.gencodegenes.org/human/
```

```
# To download the genome sequence in FASTA file format
```

```
cd /home/username/DGE_Virtual_Jan2021/
```

```
mkdir human_reference/
```

```
cd human_reference/
```

To get the download link to use in the “wget” command, do the following steps,

STEP 1: Go the relevant web page from where the genome fasta should be downloaded

The screenshot shows the NCBI Genome database search results for "yeast". The search results section is highlighted with a blue oval. The search results list is highlighted with a blue oval. A blue arrow points from the text "Hover your mouse over the link" to the search results list. Another blue arrow points from the text "You should see the complete link" to the search results list. A third blue oval highlights the search results list.

NCBI Resources How To Sign in to NCBI

Genome yeast Search

Create alert Limits Advanced Help

COVID-19 is an emerging, rapidly evolving situation.
Get the latest public health information from CDC: <https://www.coronavirus.gov>.
Get the latest research from NIH: <https://www.nih.gov/coronavirus>.

Saccharomyces cerevisiae (baker's yeast)
Reference genome: **Saccharomyces cerevisiae S288C (assembly R64)**
Download sequences in FASTA format for **genome, transcript, protein**.
Download genome annotation in **GFF, GenBank or tabular** format
BLAST against Saccharomyces cerevisiae **genome, transcript, protein**
All 815 genomes for species:
Browse the **list**
Download sequence and annotation from **RefSeq or GenBank**
Try **NCBI Datasets** - a new way to download genome sequence and annotation we're testing in NCBI Labs

Filters: [Manage Filters](#)

Find related data
Database: Select
Find items

Search details
yeast[All Fields]
Search See more...

See **UBE2E3 (YEAST) ubiquitin conjugating enzyme E2 E3** in the Gene database

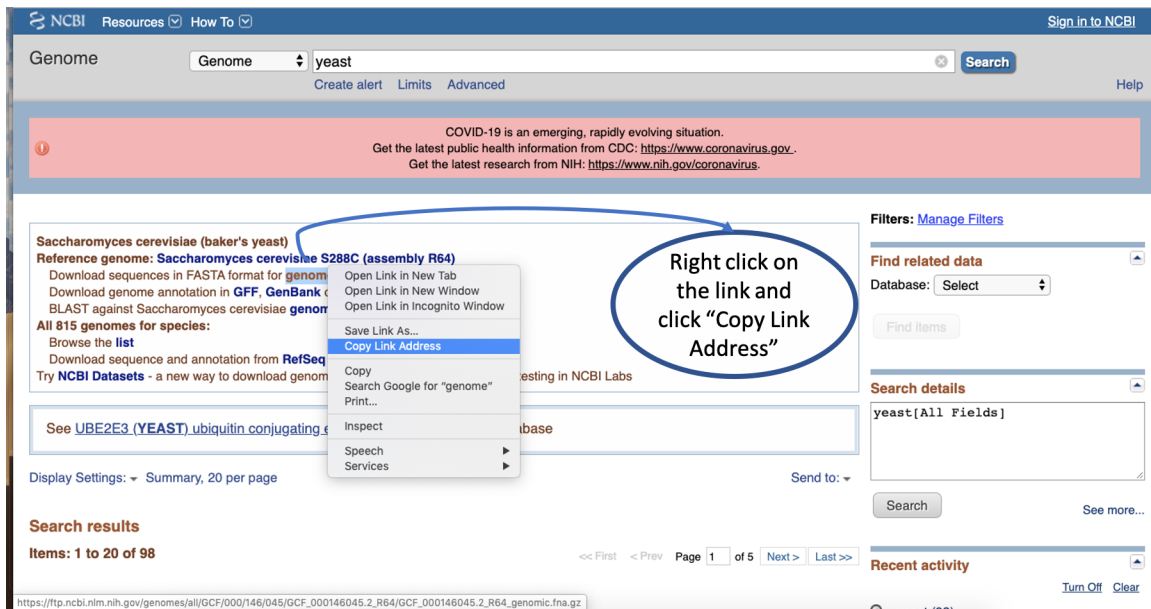
Display Settings: Summary, 20 per page Send to:

Search results
Items: 1 to 20 of 98

Page 1 of 5

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

Recent activity
Turn Off Clear



STEP 2: Once the link has been copied go to the server and execute the wget command

```
asundara@logrus:~/DGE_Virtual_May2020/human_reference
[asundara@logrus human_reference]$ wget "https://ftp.ncbi.nlm.nih.gov/genomes/all/GCF/000/146/045/GCF_000146045.2_R64/GCF_000146045.2_R64_genomic.fna.gz"
```

#RUN DOWNLOADS ON A SCREEN

```
wget "ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_29/GRCh38.p12.genome.fa.gz"
```

#wget command enables you to download files from the web directly to your server.
#Useful command to download genome fasta and annotation files from publicly available databases.

```
gunzip GRCh38.p12.genome.fa.gz
```

#gunzip is used to extract the contents within a zipped file (uncompress).

The above file will contain Nucleotide sequence of the GRCh38.p12 genome
assembly version on all regions, including reference chromosomes, scaffolds,
assembly patches and haplotypes

#To download the genome annotation in a GFF3 file format

```
wget "ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_29/\ngencode.v29.chr_patch_hapl_scaff.annotation.gff3.gz"
```

#wget command enables you to download files from the web directly to your server.
#Useful command to download genome fasta and annotation files from publicly available databases.

```
gunzip gencode.v29.chr_patch_hapl_scaff.annotation.gff3.gz
```

#gunzip is used to extract the contents within a zipped file (uncompress).

The above file will contain comprehensive gene annotation on the reference chromosomes,
scaffolds, assembly patches and alternate loci (haplotypes)

IN-CLASS EXERCISE TO NAVIGATE THROUGH SOME DATABASES TO DOWNLOAD GENOME AND ANNOTATION FILES

There are several publicly available databases that store genome and annotation files for all organisms sequenced so far. While some genomes are completely resolved at chromosome level, many aren't. Model organisms have complete genomes down to chromosome level resolution whereas several others just have a draft assembly. Examples of some databases include NCBI, ENSEMBL, TAIR, Flybase, Gencode etc.

ACTIVITY 1 (download assigned genome and annotation files from NCBI)

In the following activity, you will download genomes and the associated annotation file for the assigned organism from NCBI using the command `wget`.

```
#Create a folder in your workspace called activity_1

cd /home/username/

mkdir activity_1

cd activity_1

#make a directory called reference and download your genomes into that folder

mkdir reference

cd reference

wget "paste the link here"
```

Student	Genome	Database/Link
Amanda	Zebra fish	https://www.ncbi.nlm.nih.gov/genome/?term=zebra+fish
Gail	Arabidopsis thaliana	https://www.ncbi.nlm.nih.gov/genome/4
Stephanie	Yersinia pestis	https://www.ncbi.nlm.nih.gov/genome/?term=yersinia+pestis
Irene	Entamoeba histolytica	https://www.ncbi.nlm.nih.gov/genome/?term=Entamoeba+histolytica+HM-1%3AIM
Randall	Aedes aegypti	https://www.ncbi.nlm.nih.gov/genome/44
Joey	Oryza sativa	https://www.ncbi.nlm.nih.gov/genome/10
Kevin	Apis mellifera	https://www.ncbi.nlm.nih.gov/genome/?term=honey+bee
Roger	Staphylococcus aureus	https://www.ncbi.nlm.nih.gov/genome/154
Mahgol	Drosophila melanogaster	https://www.ncbi.nlm.nih.gov/genome/?term=Drosophila+melanogaster
Caleb	Xenopus tropicalis	https://www.ncbi.nlm.nih.gov/genome/?term=xenopus+tropicalis
Francisco	Caenorhabditis elegans	https://www.ncbi.nlm.nih.gov/genome/?term=Caenorhabditis+elegans
Diana	Chlamydomonas reinhardtii	https://www.ncbi.nlm.nih.gov/genome/?term=chlamydomonas+reinhardtii
John	Tetrahymena thermophila	https://www.ncbi.nlm.nih.gov/genome/222