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/usrname/DGE_Virtual_Jan2021/

mkdir human_reference/

cd human_reference/
```
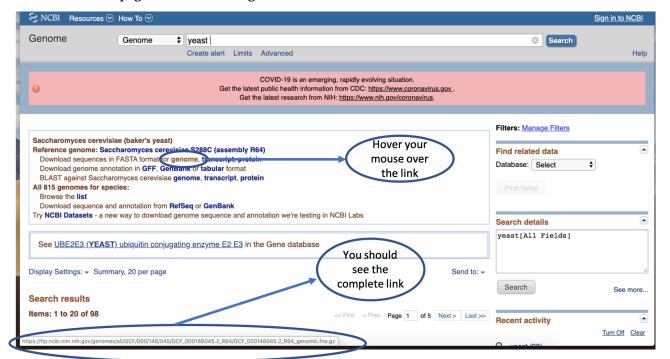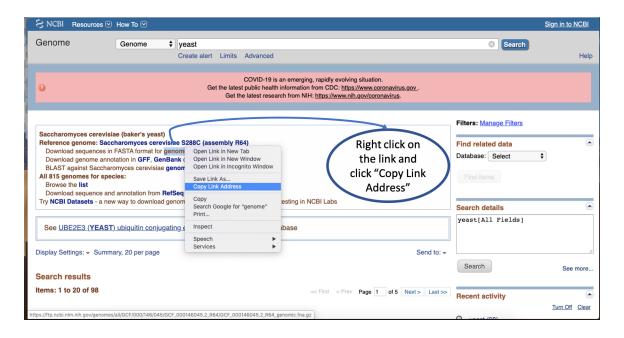
To get the download link to use in the "wget" command, do the following steps,

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

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


```
#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/\
gencode.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/usrname/

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
Franscisco	Caenorhabditis elegans	https://www.ncbi.nlm.nih.gov/genome/?term=Caenorhabditis+elegans
Diana	Chalmydomonas reinhardtii	https://www.ncbi.nlm.nih.gov/genome/?term=chlamydomonas+reinhardtii
John	Tetrahymena thermophila	https://www.ncbi.nlm.nih.gov/genome/222