

Assignment 1 (AGRO 932-Selection Scan Project)

For this study the article titled “Genome sequencing reveals evidence of adaptive variation in the genus *Zea*.” by Chen et.al. (2022) was selected. In this study, 237 teosinte accessions were collected from CIMMYT, USDA and other collaborators. Young leaves from each accessions were used to extract DNA and sequencing was done using Illumina HiSeq platform and NocaSeq 6000 platform. The data is available in the NCBI Sequence Read Archive with the accession numbers PRJNA641489, PRJNA816255, PRJNA816273 and PRJNA645739. Similarly, this study used already available DNA sequencing data of maize by downloading from NCBI Sequence Read Archive database on accession number “PRJNA531553”. The genome size of maize is about 2.3 Gb and of teosinte is about 2.1 Gb.

For our assignment we only download 10 accessions of each teosinte and maize. The accession number “PRJNA641489” and “PRJNA531553” was used to search for the teosinte and maize data respectively. Using the code “wget” for each individual data link, the data for teosinte for 10 genotypes was downloaded in HCC. The example codes for downloading 3 genotypes data is shown in code chunk below.

The data of teosinte is sotred in HCC in path: “/work/agro932/niranjan27/teosinte_data”

The data of maize is sotred in HCC in path: “/work/agro932/niranjan27/maize_data”

```
wget https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos2/sra-pub-zq-18/SRR012/12460/SRR12460347/SRR12460348
wget https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos2/sra-pub-zq-18/SRR012/12460/SRR12460349/SRR12460350
wget https://sra-downloadb.be-md.ncbi.nlm.nih.gov/sos2/sra-pub-zq-18/SRR012/12460/SRR12460350/SRR12460351
```

Similarly, the reference genome and gene annotation files of maize is stored in HCC with title “Reference genome and annotation files” and path is :/work/agro932/niranjan27

```
module load SRAToolkit/2.11 #to load SRA toolkit module
fastq-dump --split-3 SRR8906778.sralite.1 # to unzip the SRR file
```

The DNA sequencing data can be unzipped using “SRAToolkit/2.11” and using codes shown in chunks below.