# Phenotype and genotype input file preparation:

**For Phenotype analysis tab**

- The phenotype file should contain at least five columns such as Location, Season, Block, Accession, Trait1, Trait2 etc.
  Where,
  Trait1, Trait2 represents phenotype value of various traits and any number of traits can be added in the phenotype file for phenotype analysis.

**For 3k Rice (Default)**

| Location | Season | BLOCK | Accession | NPT | PL | GYKGPHA |
|---|---|---|---|---|---|---|
| ISARC | 20WS | 1 | POHHERLIMASION::IRGC 62025-1 | 5 | 28.9 | 2011.7 |
| ISARC | 21WS | 1 | POHHERLIMASION::IRGC 62025-1 | 8 | 35.33 | 3929.825 |
| ISARC | 21WS | 1 | NAN TEO 14::IRGC 7304-1 | 10 | 25.4 | 3512.069 |
| ISARC | 20WS | 1 | NAN TEO 14::IRGC 7304-1 | 6 | 19.6 | |
| ISARC | 21WS | 1 | URAIBOOL::IRGC 52785-1 | 7.67 | 28.9 | 2601.887 |
| ISARC | 20WS | 1 | URAIBOOL::IRGC 52785-1 | 5 | 25.8 | 3729.09 |
| ISARC | 20WS | 1 | IARI 11387::IRGC 19573-1 | 6 | 23.8 | 2347.06 |
| ISARC | 21WS | 1 | IARI 11387::IRGC 19573-1 | 7 | 20.53 | 3091.667 |
| ISARC | 21WS | 1 | HARRA::GERVEX 501-C1 | 6.67 | 18.4 | 749.1228 |
| ISARC | 20WS | 1 | HARRA::GERVEX 501-C1 | 7 | 12.5 | 478.33 |
| ISARC | 20WS | 1 | PL 3165::IRGC 62827-1 | 6 | 13.8 | 630 |
| ISARC | 21WS | 1 | PL 3165::IRGC 62827-1 | 8.67 | 18.13 | 1878.333 |
| ISARC | 20WS | 1 | ARC 18371::IRGC 42423-2 | 6 | 26.1 | 2444.26 |
| ISARC | 21WS | 1 | ARC 18371::IRGC 42423-2 | 5.67 | 21.3 | 3120.339 |

**For other diploid organism (external)**

| Trial | Genotype | Replication | TN | RN | ARP | RL | subpopulation |
|---|---|---|---|---|---|---|---|
| 2nd Expt | MG1 | 3 | 9 | 212 | 33.99 | 711.243 | xx |
| 2nd Expt | MG1 | 2 | 9 | 235 | 40.06 | 701.678 | xx |
| 2nd Expt | MG1 | 1 | 10 | 283 | 37.82 | 524.315 | xx |
| 2nd Expt | MG5 | 1 | 8 | 163 | 34.65 | 732.077 | xx |
| 2nd Expt | MG5 | 2 | 9 | 177 | 37.64 | 617.693 | xx |
| 2nd Expt | MG5 | 3 | 9 | 196 | 31.9 | 892.359 | xx |
| 2nd Expt | MG6 | 1 | 9 | 218 | 24.79 | 828.369 | xx |
| 2nd Expt | MG6 | 3 | 9 | 226 | 24.7 | 589.501 | xx |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 2nd Expt | MG6 | 2 | 9 | 239 | 25.79 | 507.984 | yy |

**For other analysis such as Genofile Extract, GWAS, Et-GWAS, Haplopheno tabs**

**External:** Accession/ IDs**,** should be same

**Default**

- First column name should be X.Phenotype. of phenotype file(.csv format).
- ID's in X.Phenotype. column should be in the following format e.g; IRIS_313-11586.
- No need to provide a subpopulation column.

**External**

- The first column name should be X.Phenotype of phenotype file.
- Accession/ IDs in X.Phenotype. column should be the same as the vcf file (genotype file) provided by the user.
- Subpopulation or group information should be provided in the third column of the phenotype file and name of the **column should be "subpopulation"**.

**For 3k Rice (Default)**

| X.Phenotype. | PL |
|---|---|
| IRIS_313-11586 | 27.23 |
| IRIS_313-11037 | 27.83 |
| IRIS_313-8061 | 19.63 |
| IRIS_313-8060 | 26.07 |
| IRIS_313-8057 | 26.83 |
| IRIS_313-9140 | 24.33 |
| IRIS_313-12185 | 27.5 |
| IRIS_313-11744 | 24.73 |
| IRIS_313-12057 | 29.07 |

**For other diploid organism (external)**

| X.Phenotype. | ARP(adjusted) | subpopulation |
|---|---|---|
| MG1 | 37.00889 | yq |
| MG10 | 29.31557 | yy |
| MG100 | 30.35416 | yy |
| MG101 | 30.87122 | xx |
| MG106 | 23.78305 | zz |
| MG109 | 21.66455 | yy |
| MG11 | 34.11961 | zz |
| MG110 | 35.22003 | xx |
| MG114 | 31.79448 | zz |

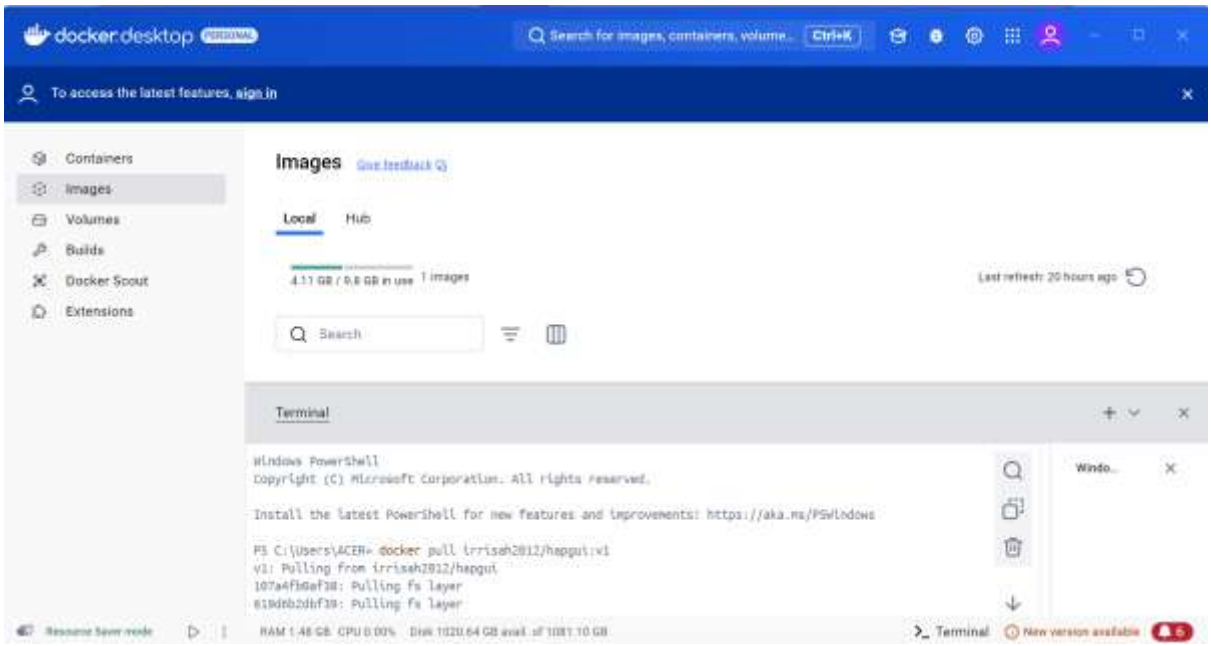
## Run HapGUI from docker hub: (follow video manual)

1. Download docker desktop based on operating system from https://www.docker.com/products/docker-desktop/
2. Follow the instructions for installing docker by following the provided video manual.

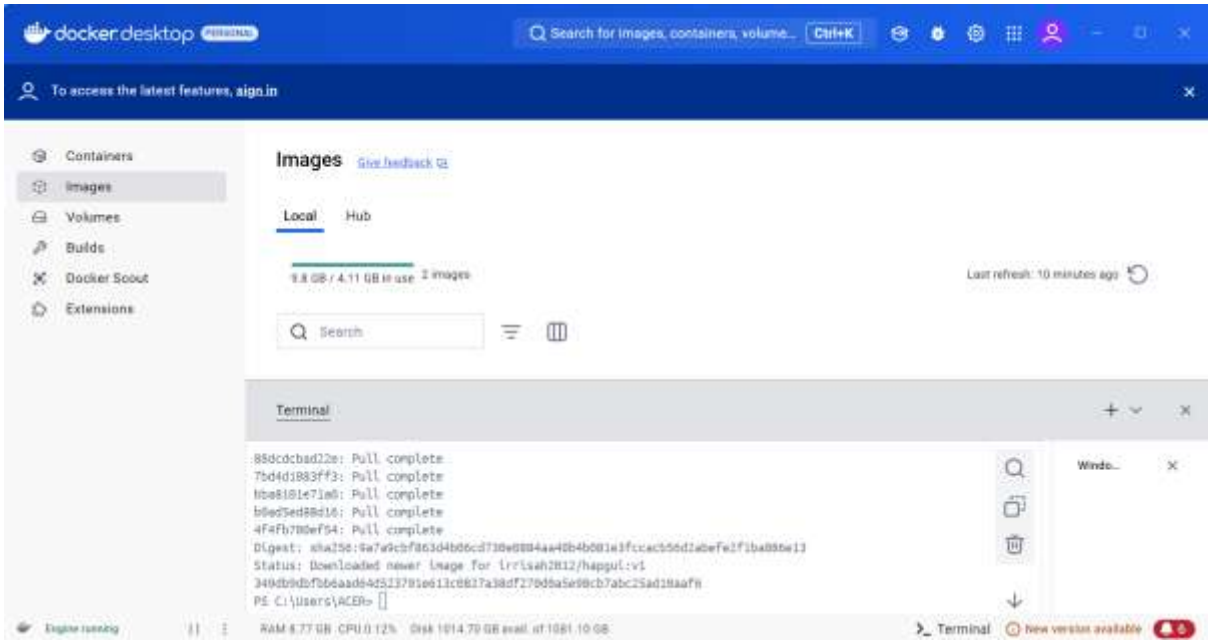
    - While installation of docker, if following window is displayed uncheck above two option and only check third option (Add shortcut to desktop). Everything else will be chosen default as shown in the video manual.


3. To run HapGUI, users need to follow the following instructions just once.
    - Then open the terminal of docker, type the following commands and run one by one.

docker pull irrisah2012/hapgui:v1

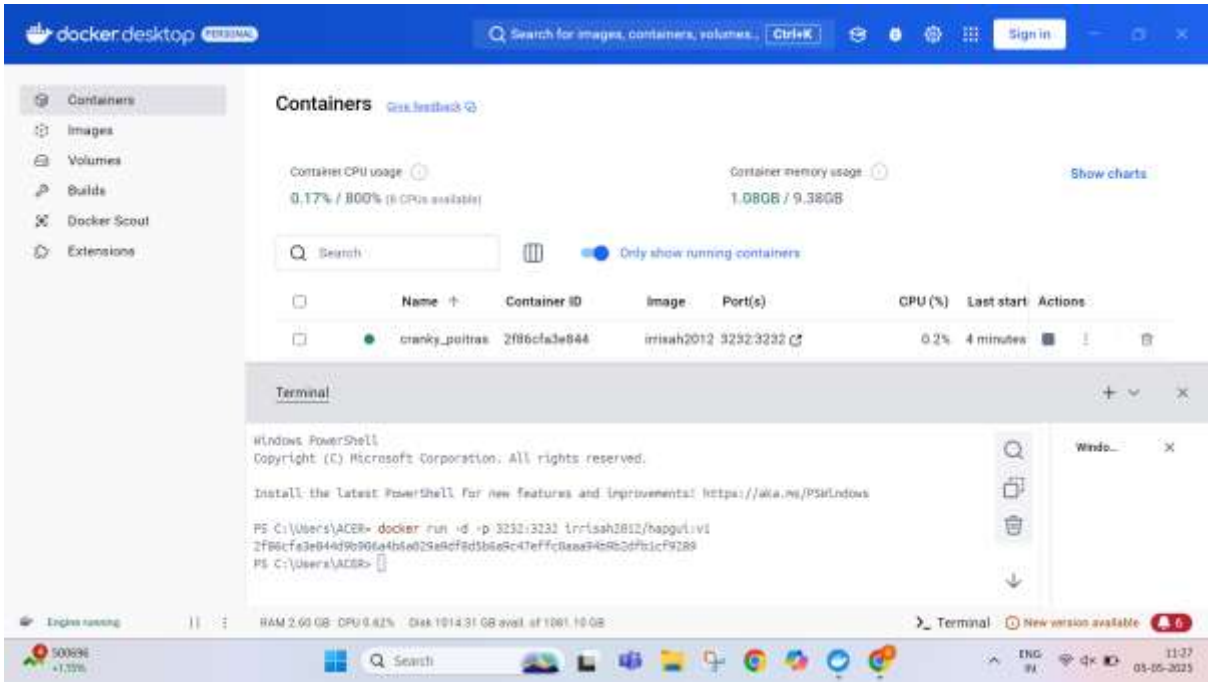


- After the pull completes the docker terminal will look like the following.



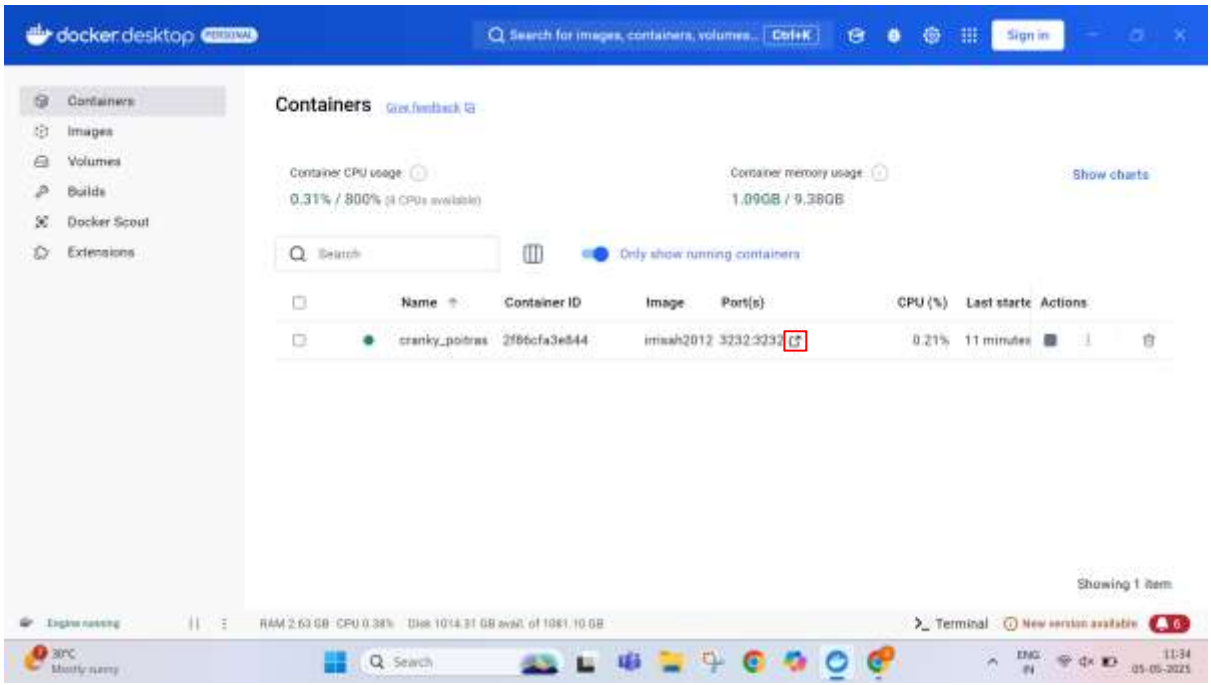

- Then type following command and run.

  docker run -d -p 3232:3232 irrisah2012/hapgui:v1

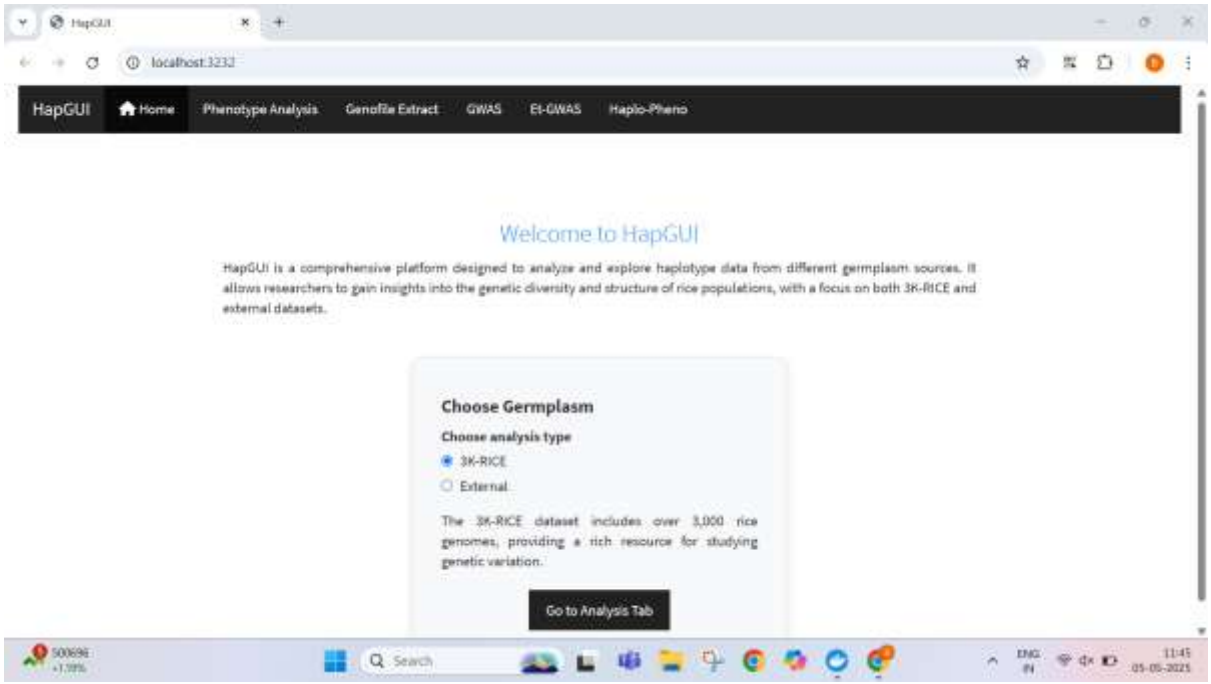



4. Then click on the highlighted symbol, after that on the browser (chrome, firefox etc.) of the user's PC HapGUI will open as provided in the screenshort below.

   Else, copy the link (http://localhost:3232/) and paste on any browser and press enter.

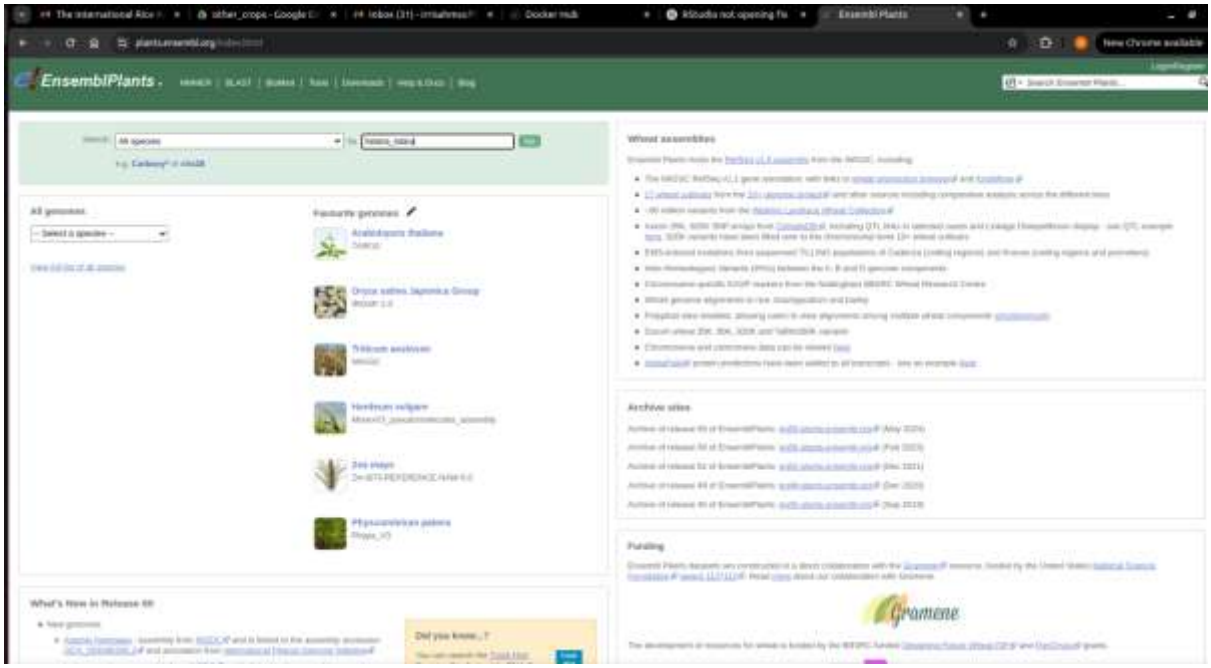


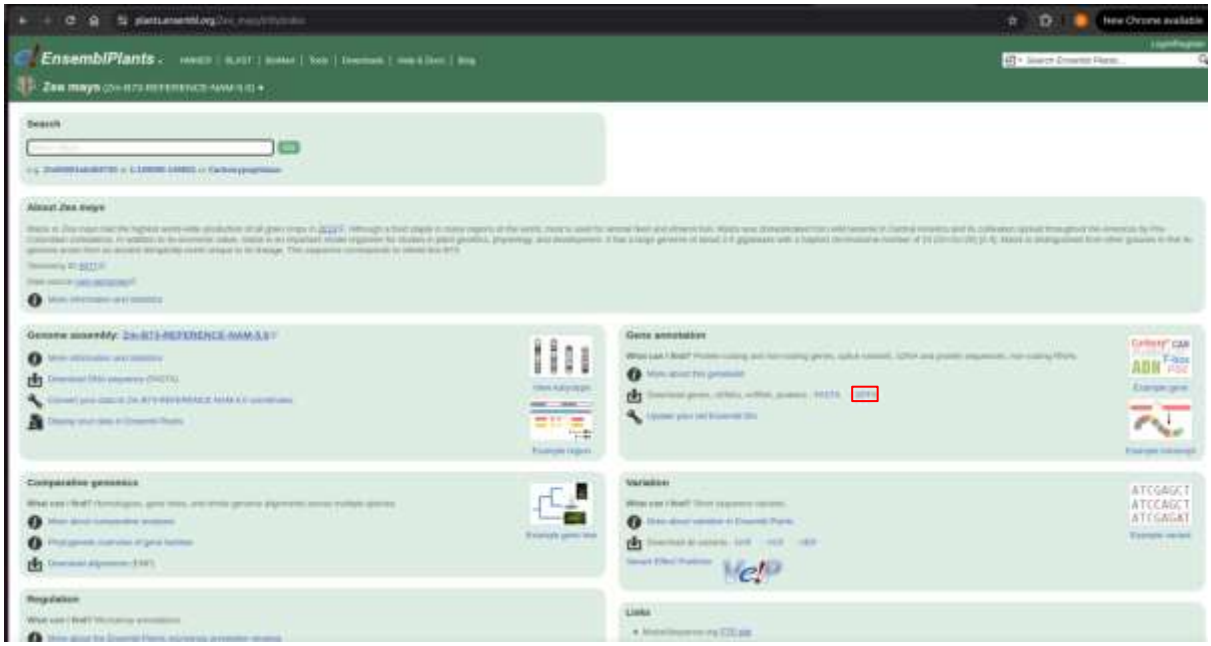

Then HapGUI interface will open as shown below.



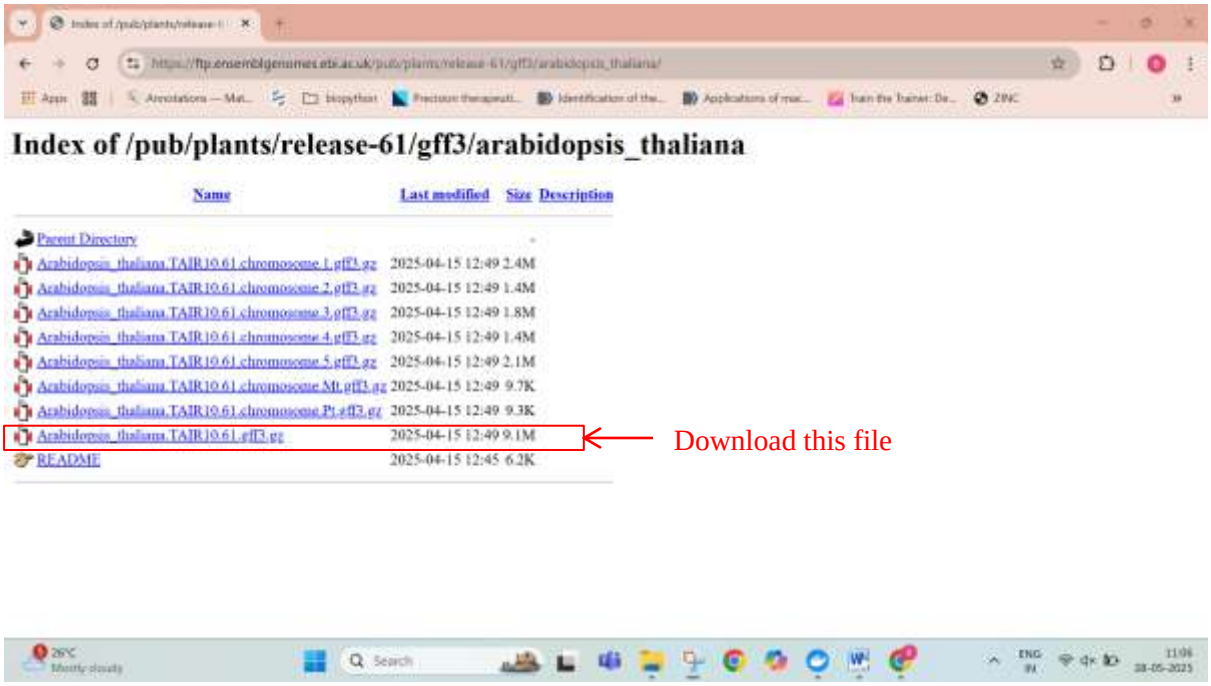

Before running other analysis for external organism Please download the gff3 file as demonstrated below.

Before performing Genofile Extract (Genotype file extraction) for GWAS, Et-GWAS and Haplopheno analysis for other panels of Rice as well as other diploid organism gff3 file (General Feature Format) need to be downloaded from ensemble database (https://ensemblgenomes.org/) as explained in the video manual.




- Be careful while downloading the gff3 file. Because the gff3 file will be available separately for each chromosome and one gff3 file will be with all chromosomes. So, we need the gff3 file with all chromosomes as mentioned below (e.g., for Arabidopsis_thaliana).
- After downloading extract the gff3 file and use in the analysis.

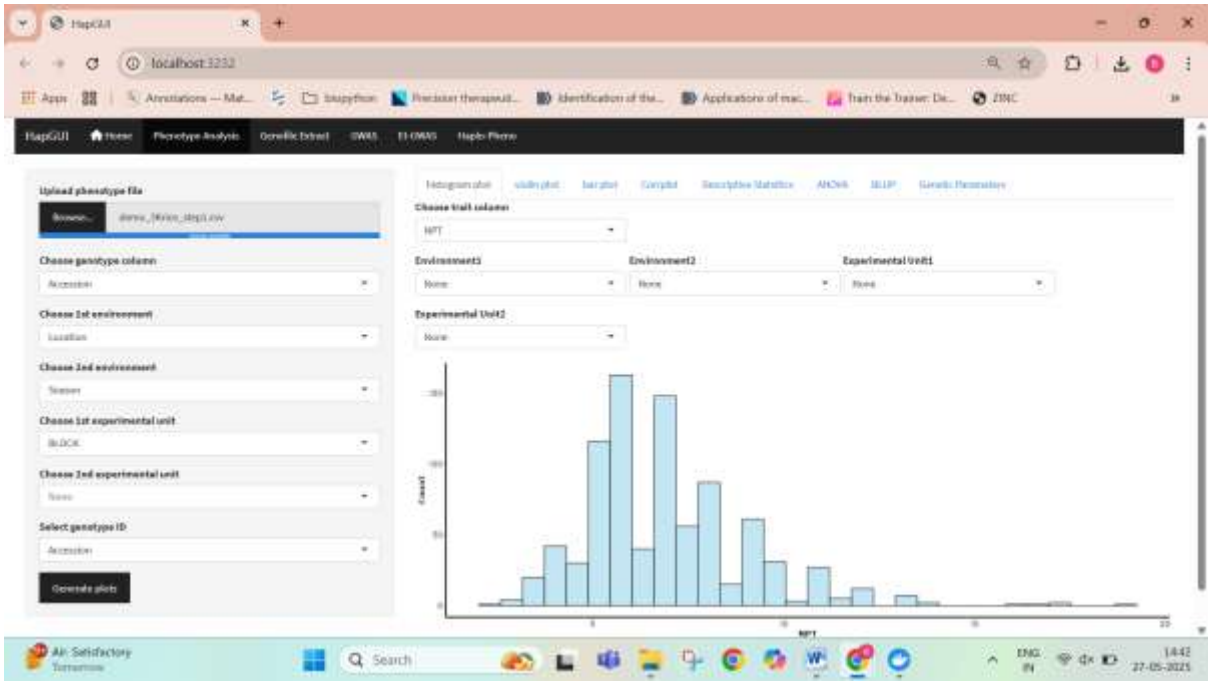



## Phenotype analysis:

i. Choose phenotype file.
ii. Choose genotype column.
iii. Choose 1st environment. If user has 2nd environment, then choose (if available in your phenotype data).
iv. Next choose 1st and 2nd experimental unit (if available in your phenotype data).
v. Select genotype ID.
vi. Click on generate plots.

**<u>Histogram plot</u>**

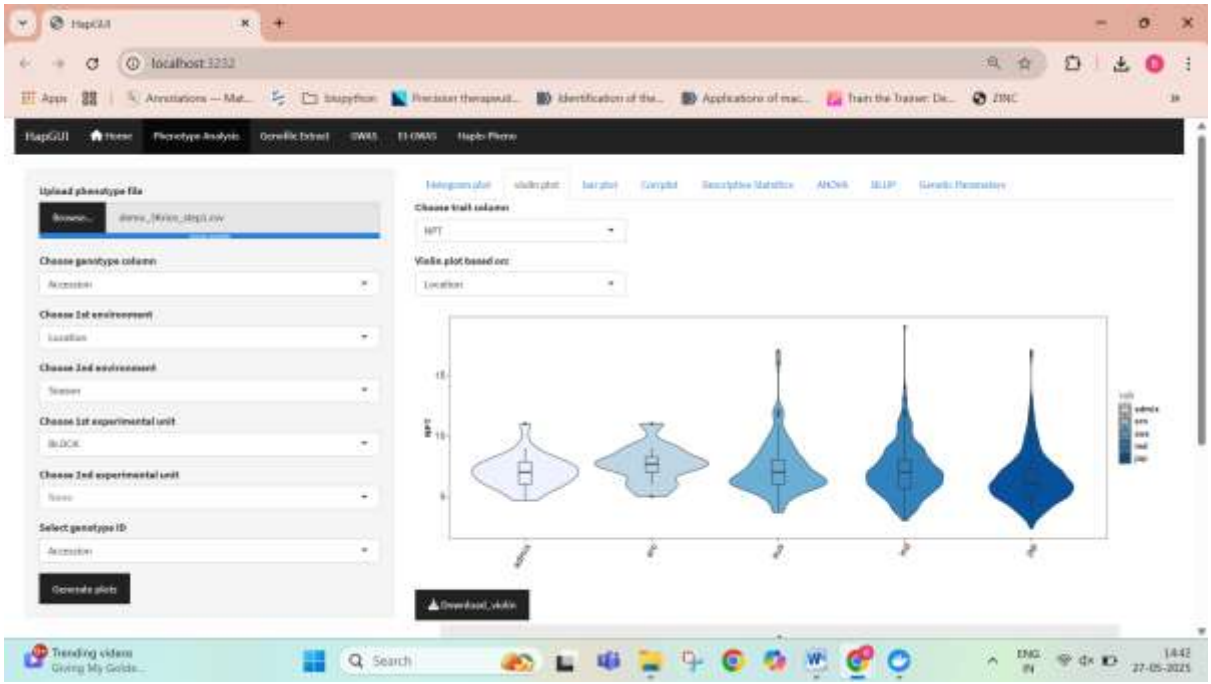
i. Choose trait column that needs to be visualized.
ii. The histogram plot can be downloaded by clicking on “Download_histogram”.



## Violin plot

Select appropriate option of interest from the steps mentioned below:

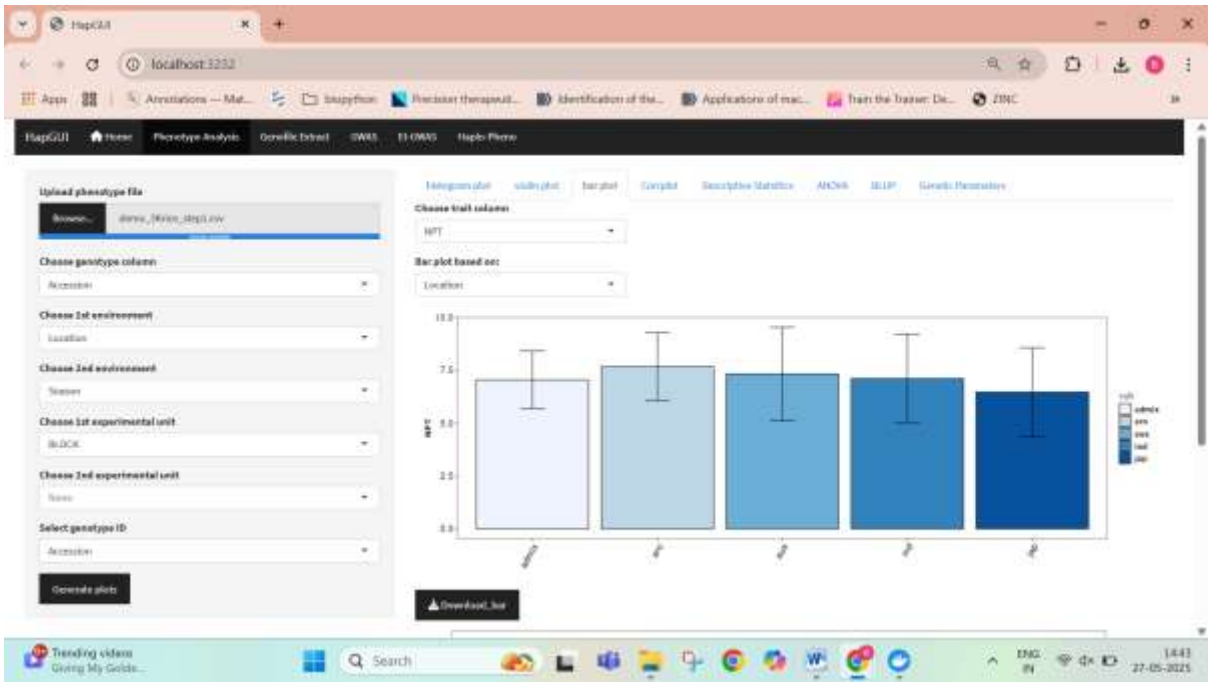
i. Choose trait column that needs to be visualised.
ii. Select “Violin plot based on” option that needs to be visualized.
iii. The histogram plot can be downloaded by clicking on “Download_violin”.



## Box plot

Select appropriate option of interest from the steps mentioned below:

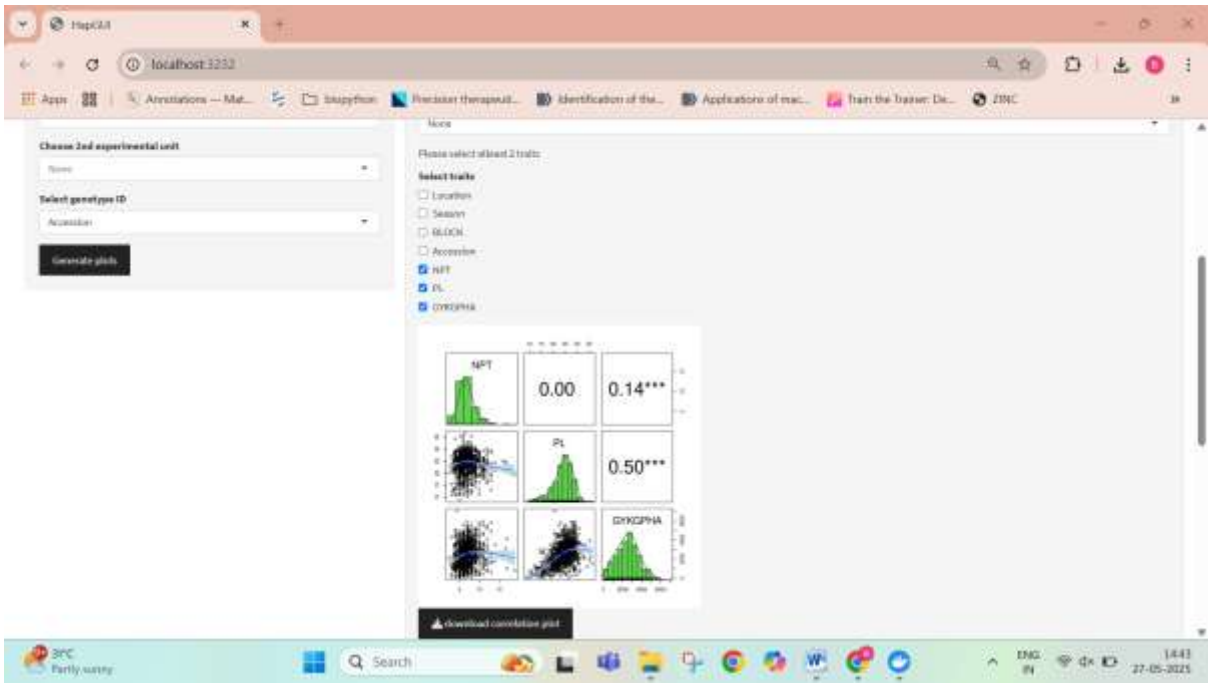
i. “Choose trait column”.
ii. “Bar plot based on”.
iii. The bar plot can be downloaded by clicking on “Download_bar”.



## Correlation plot

Select appropriate option of interest from the steps mentioned below:

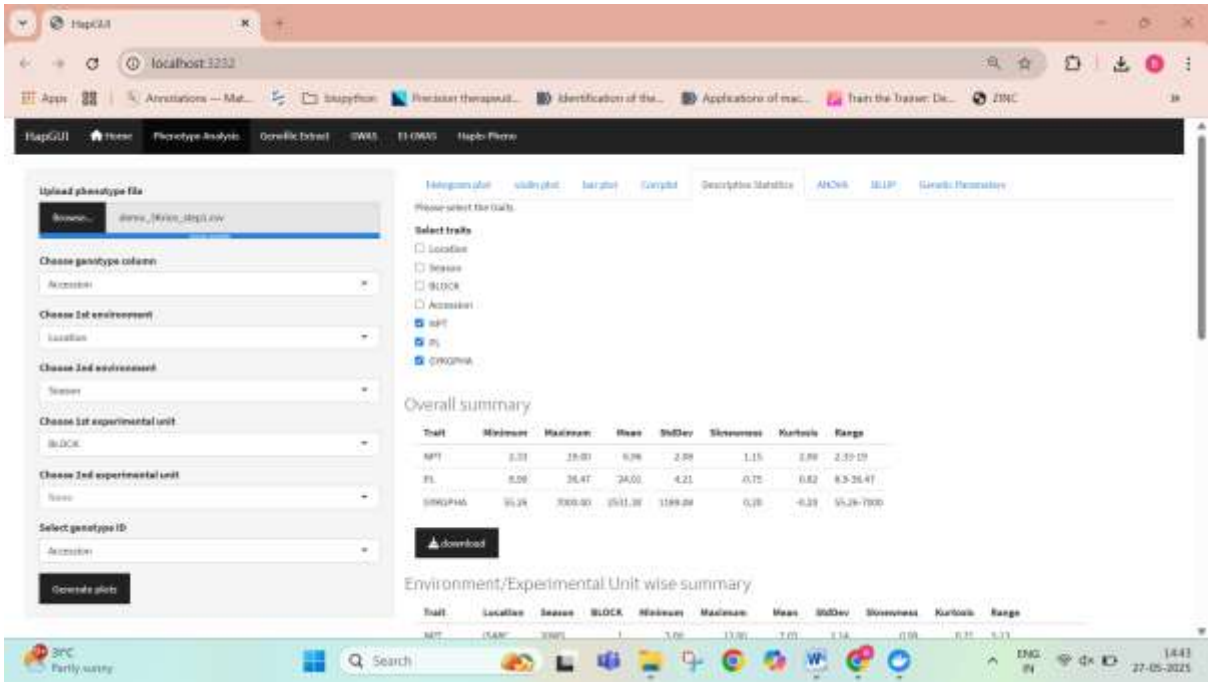
i. Please select more than 2 traits for getting correlation plot.
ii. The correlation plot can be downloaded by clicking on “Download_correlation plot”.




## Descriptive statistics

Select appropriate option of interest from the steps mentioned below:

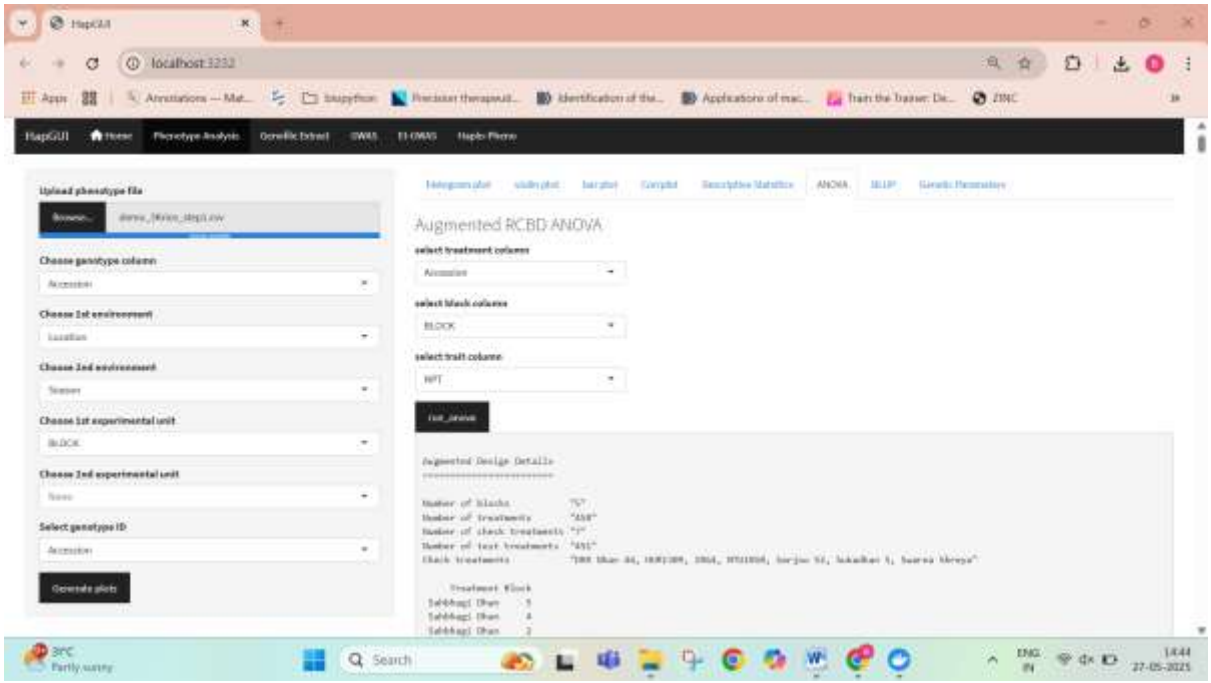
i. Select traits of interest
ii. The descriptive statistics can be downloaded by clicking on “download” button.




## ANOVA

Select appropriate option of interest from the steps mentioned below:

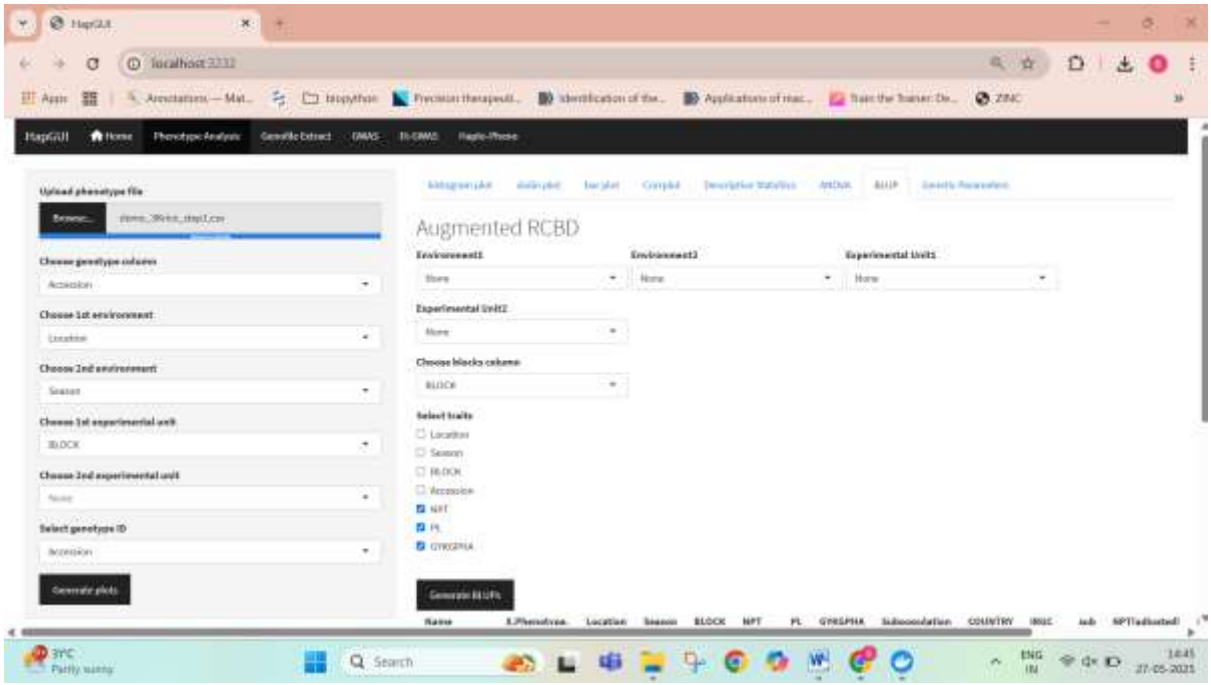
i. Select treatment column (e.g, Accession).
ii. Select block column.
iii. Select trait column.
iv. Click “Run_anova”.
v. The results will be displayed as given below.



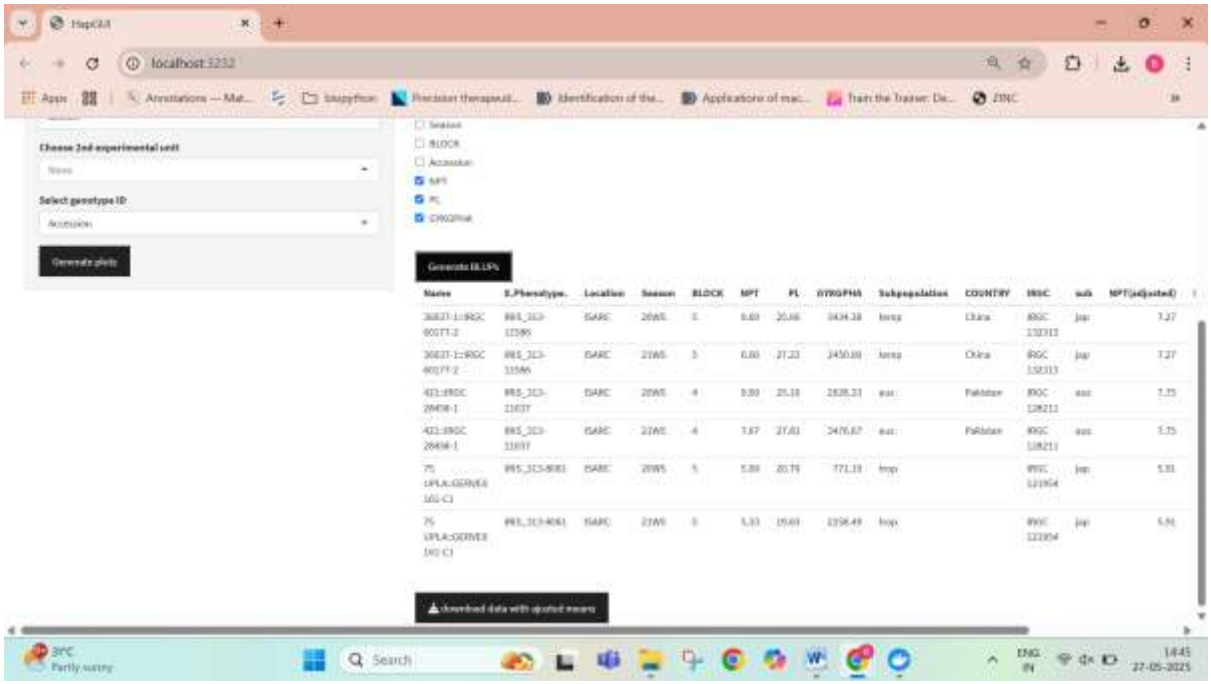

**BLUP**

Select appropriate option of interest from the steps mentioned below:

i. "Select traits" (check on traits of interest).
ii. Click "Generate BLUPs".
iii. Blup results



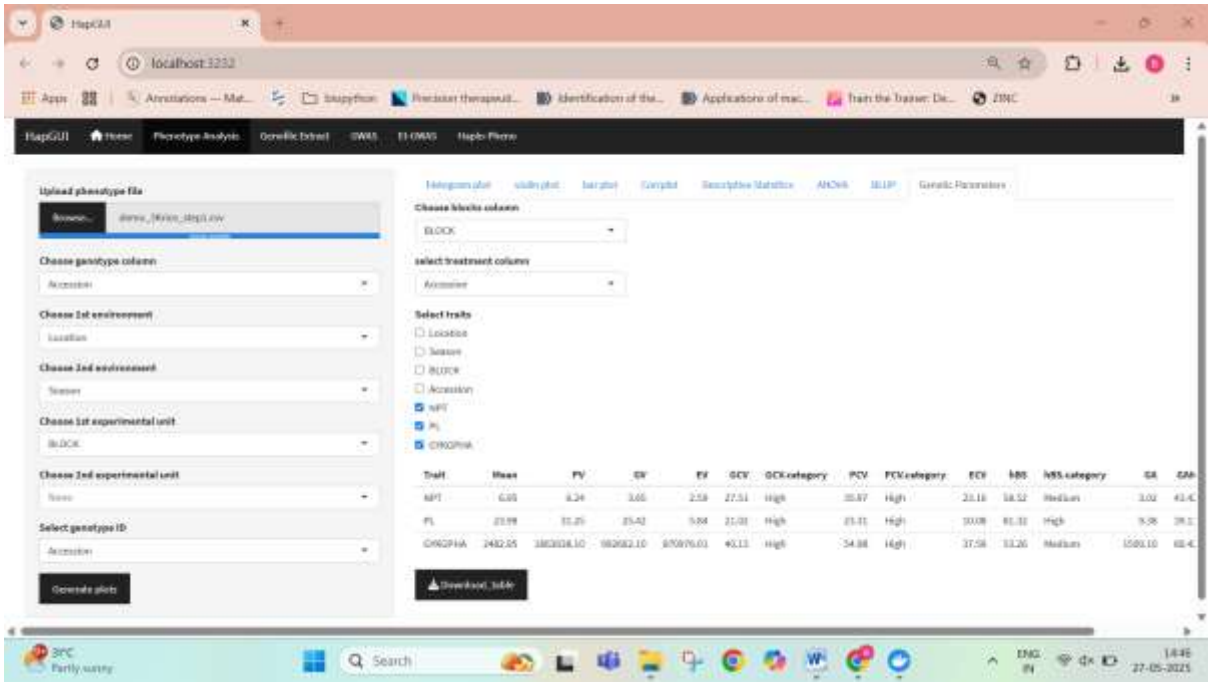

- Displaying the result of BLUP.



## Genetic parameters

Select appropriate option of interest from the steps mentioned below:

i. Choose blocks column
ii. Select treatment column
iii. “Select traits” Check on traits for those BLUP values needs to be generated.



## Genofile Extract

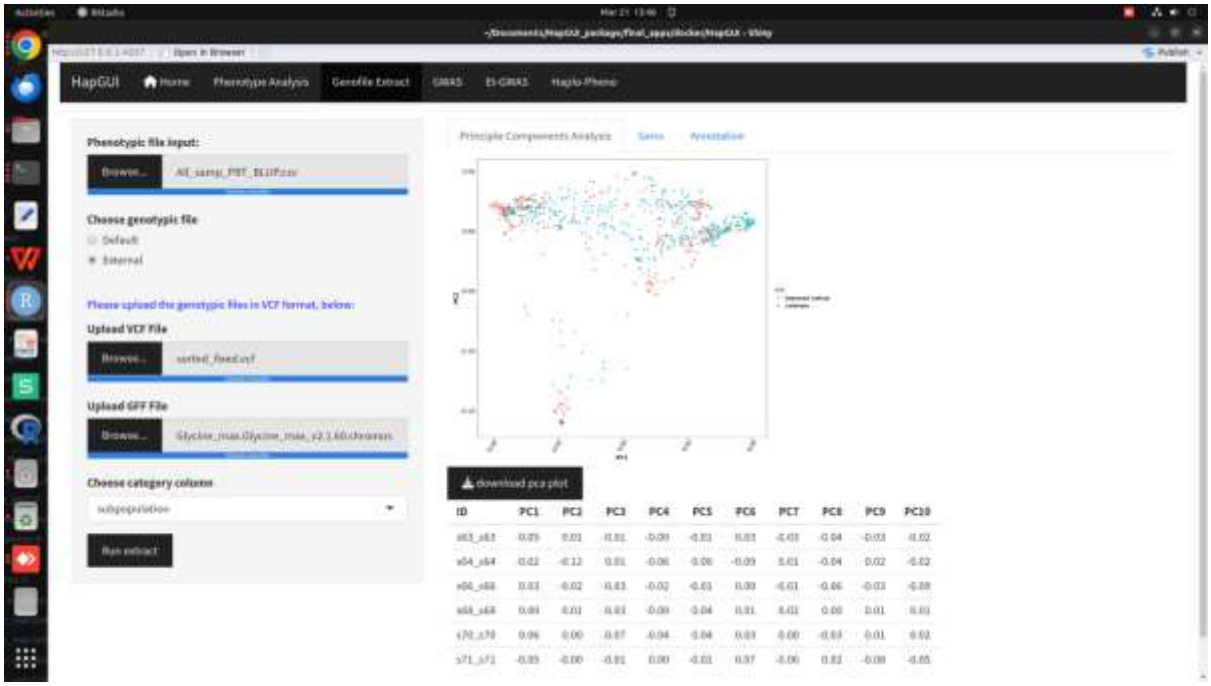
### For 3k Rice panel

i. Upload your phenotype file in appropriate format as mentioned in the page no. 2

ii. Select “Default” in “Choose genotypic file” section.

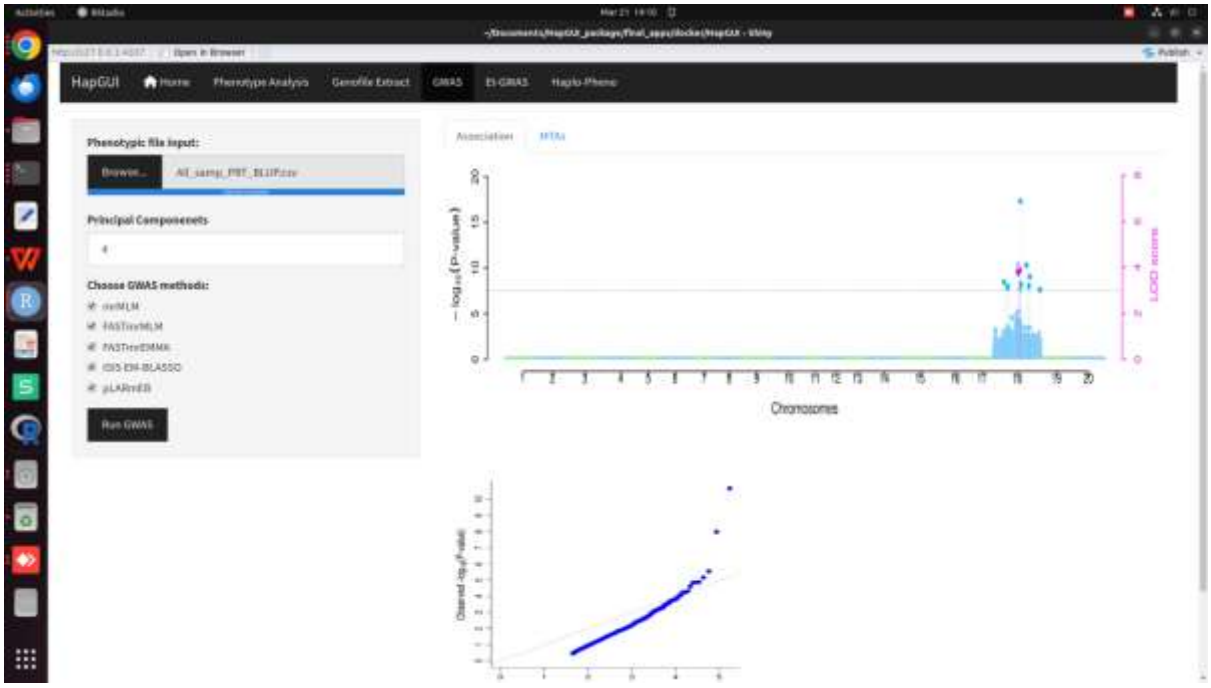

**For external diploid organism**

i. Upload your phenotype file in appropriate format as mentioned in the page no. 2.
ii. Select “External” in “Choose genotypic file” section.
iii. Upload genotype file in vcf format in “Upload VCF File” section.
iv. Upload General Feature Format file in gff3 format in “Upload GFF File” section.
v. In “Choose category column” choose “subpopulation”.



**GWAS**

i. Upload your phenotype file in appropriate format as mentioned in the page no. 2.
ii. Insert a number, in Principal Components. Insert more than three till 10.
iii. In “Choose GWAS methods”, choose your model of interest.
iv. Then Click on “Run GWAS”.

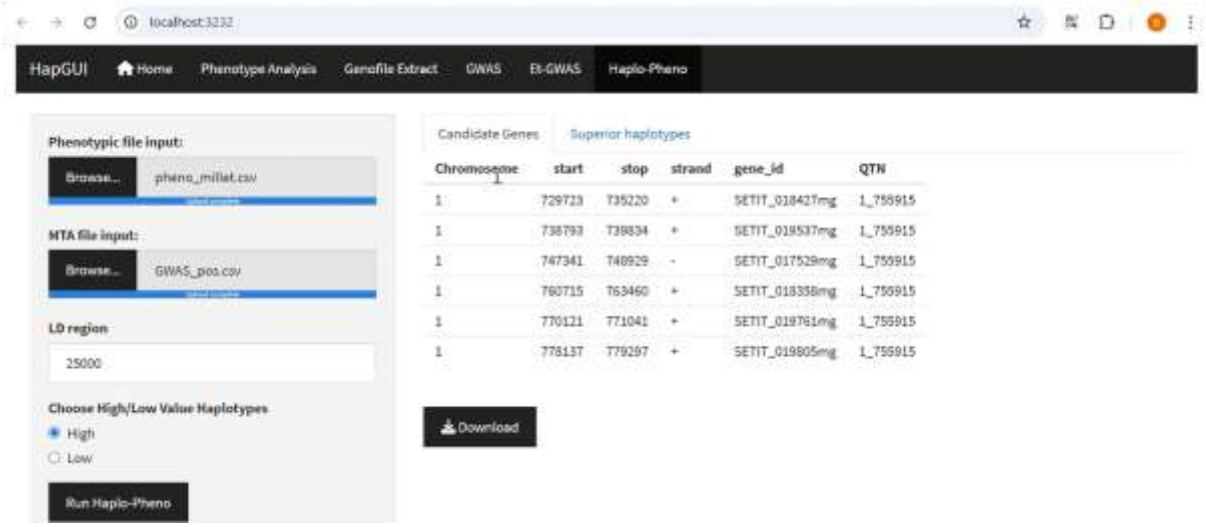


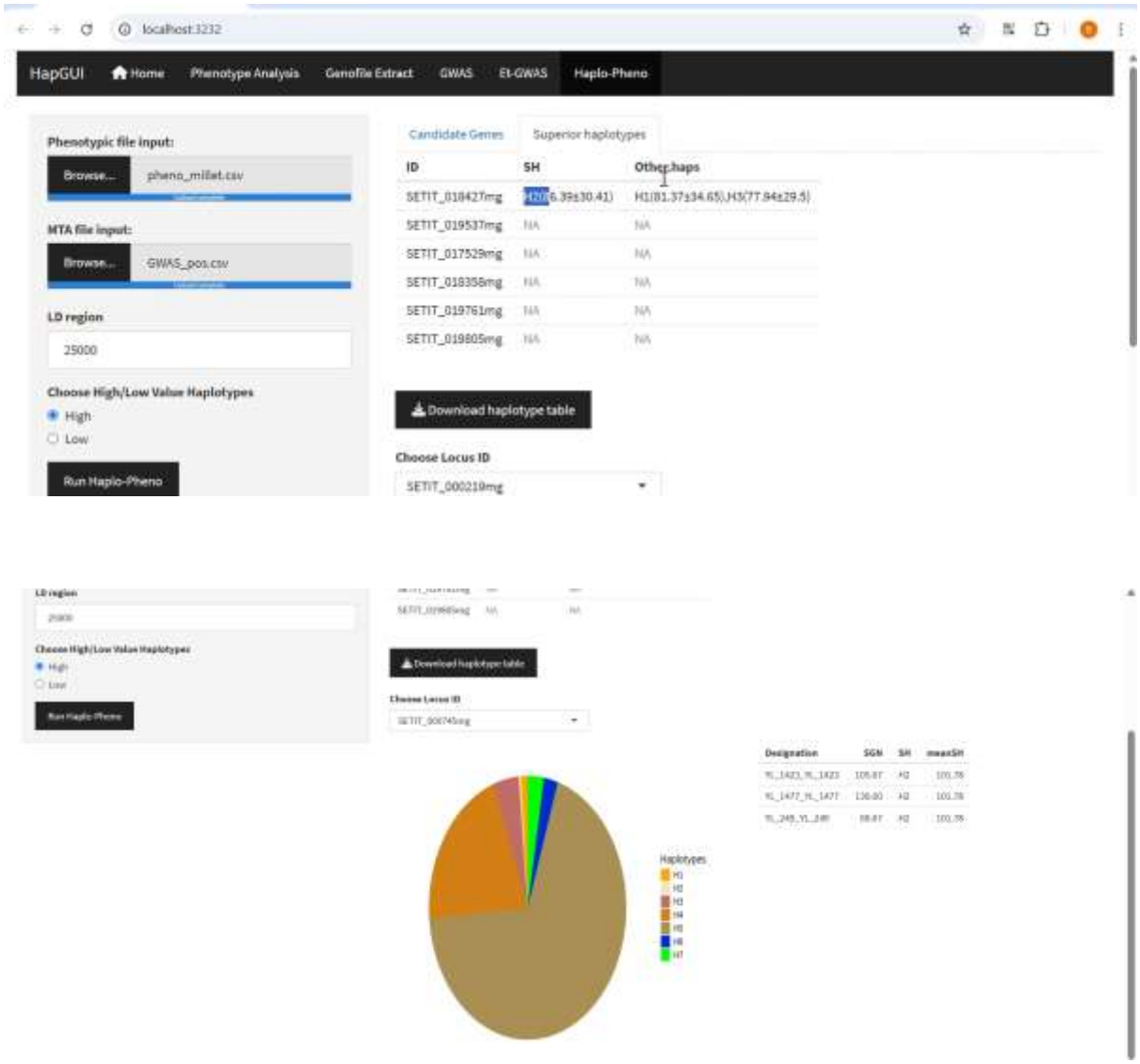

**Et-GWAS**

i. Select “Bulk size:”
ii. Select “Trait”
iii. Upload your phenotype file in appropriate format as mentioned in the page no. 2.
iv. Click “Run”

## Haplo-pheno

i. Upload your phenotype file in appropriate format as mentioned in the page no. 2.
ii. Upload the MTA file with extention “GWAS_pos” or “Et-GWAS_pos” from GWAS results, Et-GWAS_results folder.
iii. Write LD region (only 0 >= 250000) for identification of candidate genes.
iv. Choose “High” or “Low” based on the traits behaviour for identification of superior haplotypes e.g., for yield related traits “High” and disease scoring related traits “Low”.
v. Candidate genes and Superior haplotypes table will be displayed as follows.
vi. By clicking on “download” button, the results of candidate genes and superior haplotypes can be downloaded.

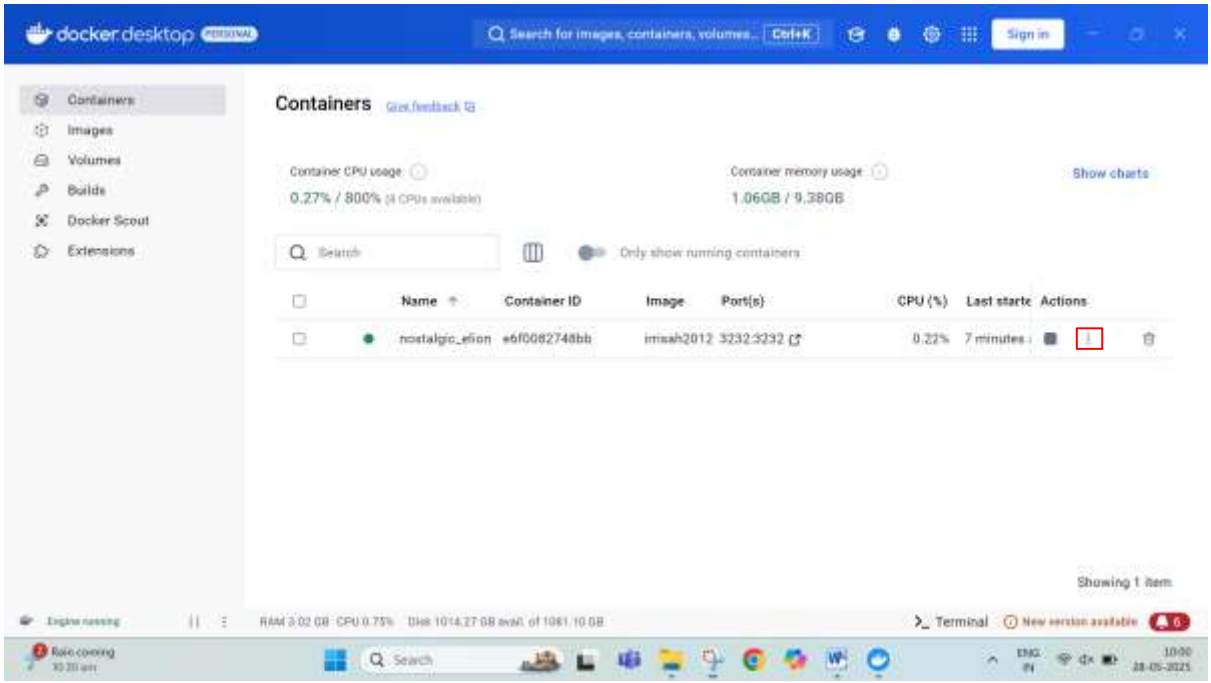




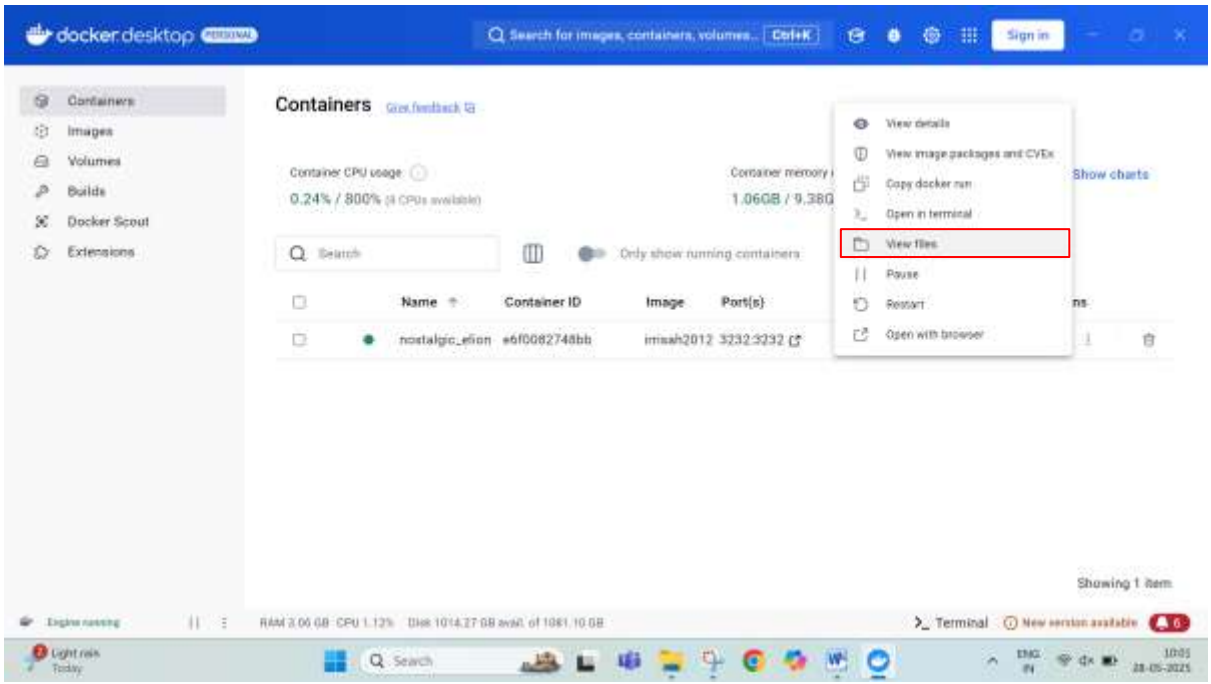

## Accessing/ downloading the results of GWAS, Et-GWAS and Haplopheno

i. After performing the analysis of GWAS, Et-GWAS and Haplopheno, the results can be downloaded by following the provided screenshots.

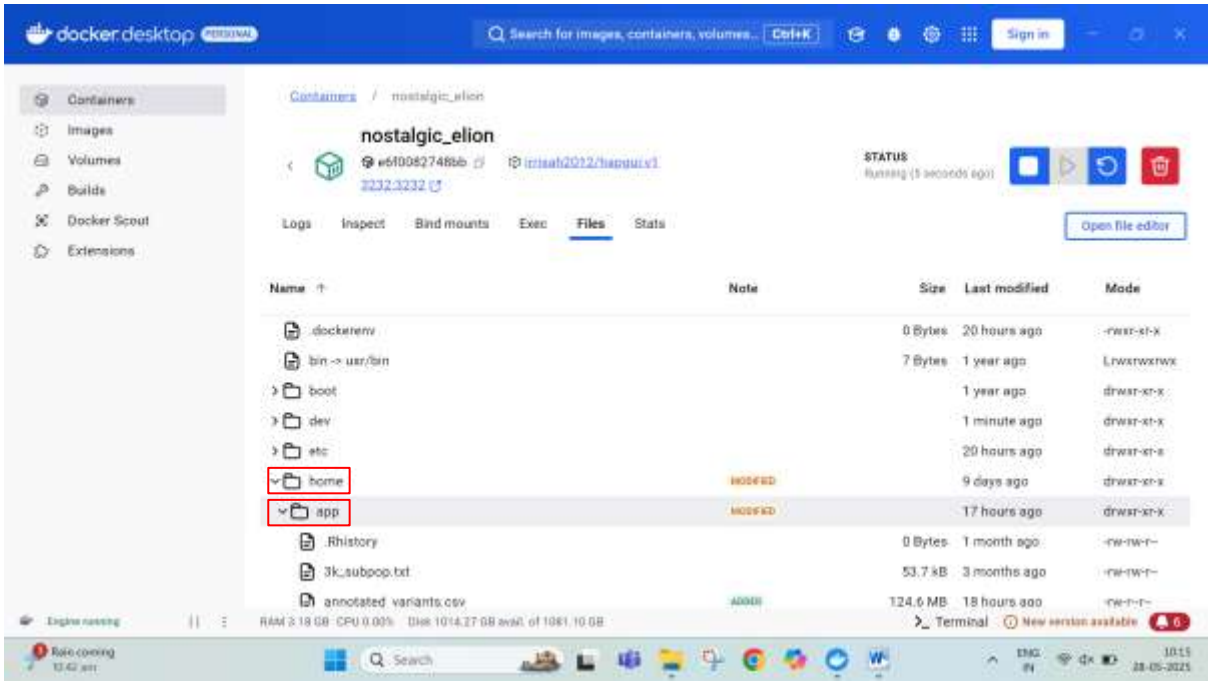
- Click on the symbol " ⋮ " as shown below.



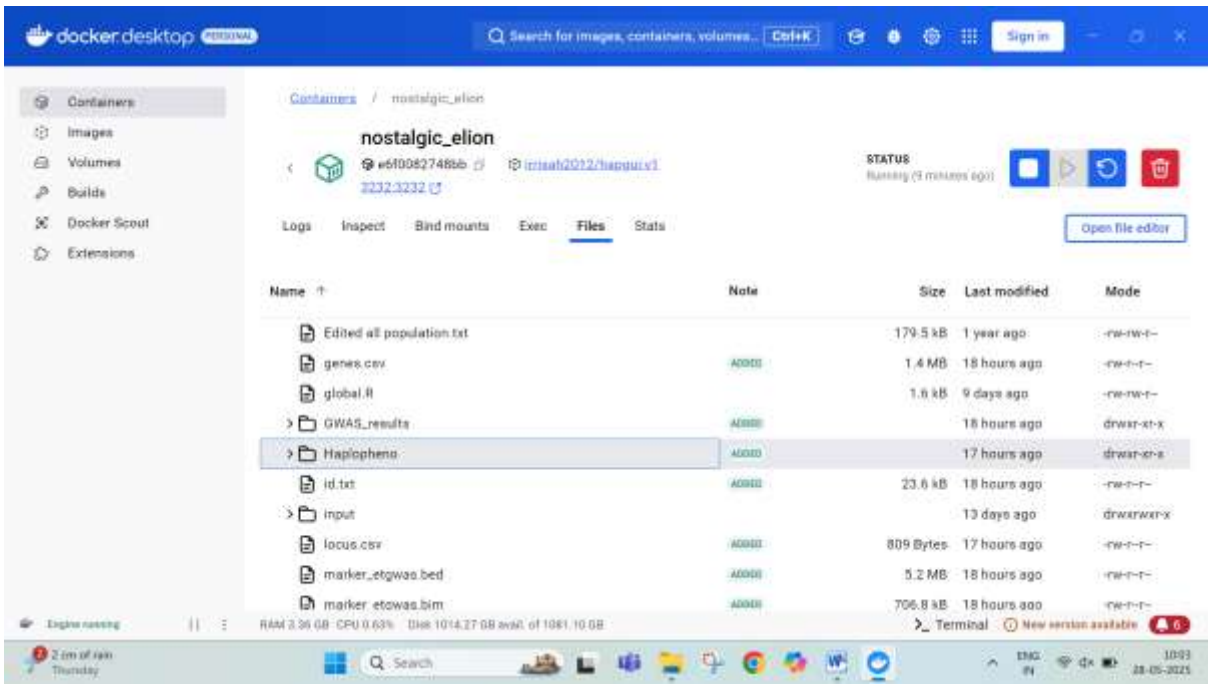

- Click on “view files”.



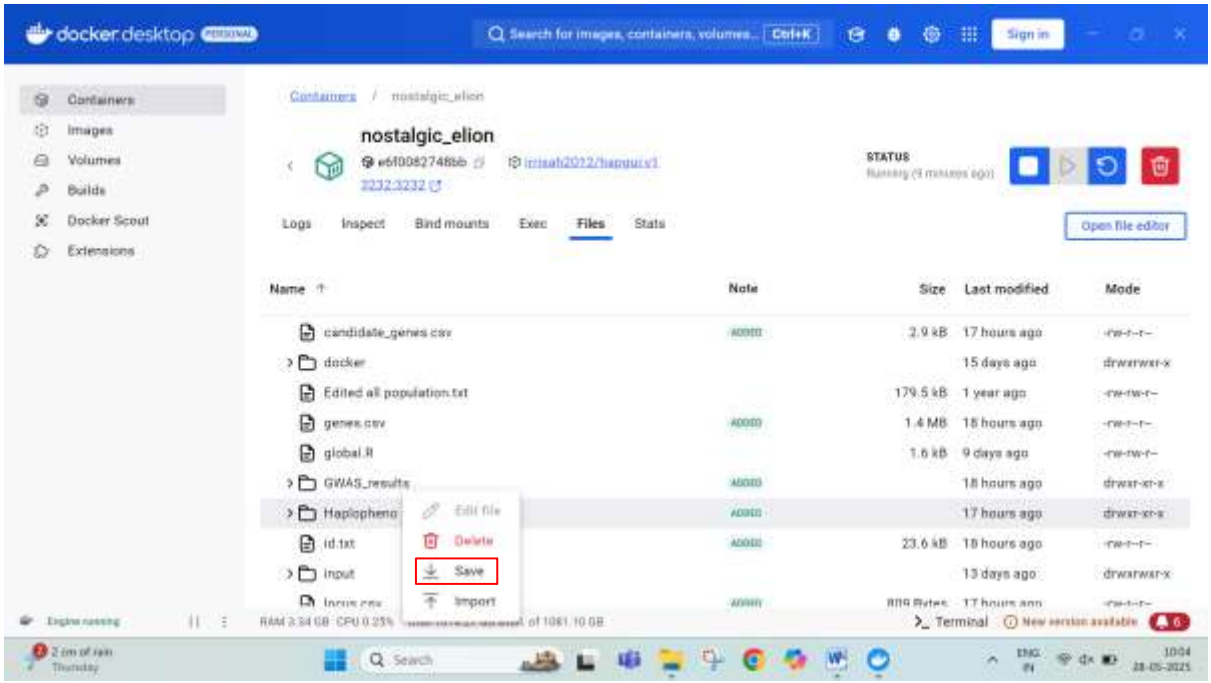

- Then Click on “home” and later on “app”.

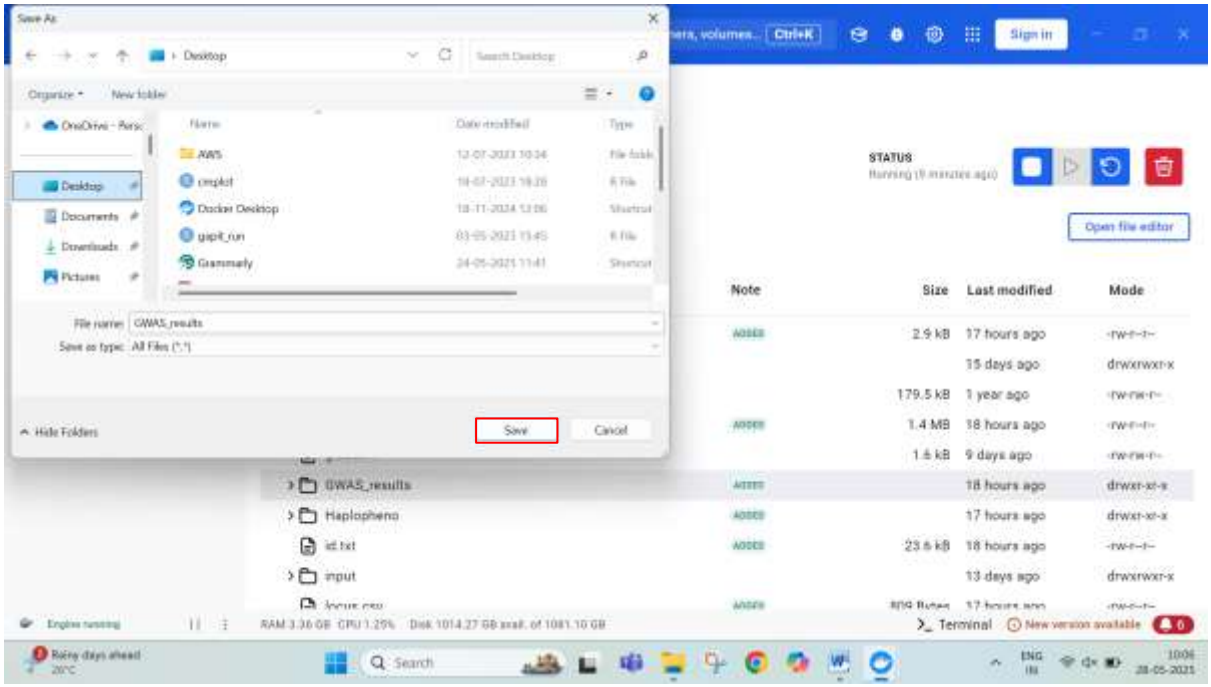



- Now you'll find the results namely "GWAS_esults", "EtGWAS_results" and "Haplopheno" folders where you can find respective results in respective folders.




- All these results can be downloaded by "right clicking". Then the results can be downloaded by clicking on "save" button, Delete by clicking on the "Delete" button.
- **Important Note:** In the similar way like save**,** please delete after saving the results, before running another analysis.


docker desktop
Search for images, containers, volumes... Ctrl+K
Sign in
Containers
Images
Volumes
Builds
Docker Scout
Extensions
Containers / nostalgic_elion
nostalgic_elion
STATUS
Logs
Inspect
Bind mounts
Exec
Files
Stats
Open file editor
Name
Note
Size
Last modified
Mode
candidate_genes.csv
2.9 kB
17 hours ago
docker
15 days ago
Edited all population.txt
179.5 kB
1 year ago
genes.csv
1.4 MB
18 hours ago
global.R
1.6 kB
9 days ago
GWAS_results
18 hours ago
Haplopheno
17 hours ago
id.txt
23.6 kB
input
13 days ago
Edit file
Delete
Save
Import
Terminal
New version available
Search


Save As
Desktop
Organize
New folder
Desktop
Documents
Downloads
Pictures
File name: GWAS_results
Save as type: All Files (*.*)
Hide Folders
Save
Cancel
STATUS
Open file editor
Note
Size
Last modified
Mode
2.9 kB
17 hours ago
15 days ago
179.5 kB
1 year ago
1.4 MB
18 hours ago
1.6 kB
9 days ago
GWAS_results
18 hours ago
Haplopheno
17 hours ago
id.txt
23.6 kB
18 hours ago
input
13 days ago
Terminal
New version available
Search