

Run GAPIT on the ICCR server

To perform the GWAS analysis using GAPIT on the server, it is necessary to have an account granting access. If you don't have one, please reach out to an administrator. GAPIT can be executed with the R program, and in this case, we will utilize an implementation of RStudio available on the ICCR server.

GWAS analysis is used to search for associations between a phenotype and a genomic region, for that you need to input phenotypic and genomic data.

The phenotype data needs to be organized in a table, the format should be a tab separated value (TSV) with 2 columns, the first column is the accession notation (e.g. AEG_1006_4) and the second column is the phenotypic value.

Example of first 5 lines of a table:

accession	phenotype
AE_121	3
AE_122	3
AE_124	9
AE_125	9

You can create the table using Excel by saving the table as 'Text (Tab delimited)'.

IMPORTANT: do not insert accessions that are not in the genomic data (use only accession that have been sequenced and there is SNP data for them).

To log in to the server you can use a program like [MobaXterm](#) (or any other you prefer), the server address is: icci-2.tau.ac.il

To be able to access the table within the server you need to upload it to a folder of choice on the server.

Run GAPIT3:

open a chrome internet browser and go to:

icci-2.tau.ac.il:9000

log in with your ICCR server account credentials.

Open the file “run_gapit.R” with: ‘file -> open File’.

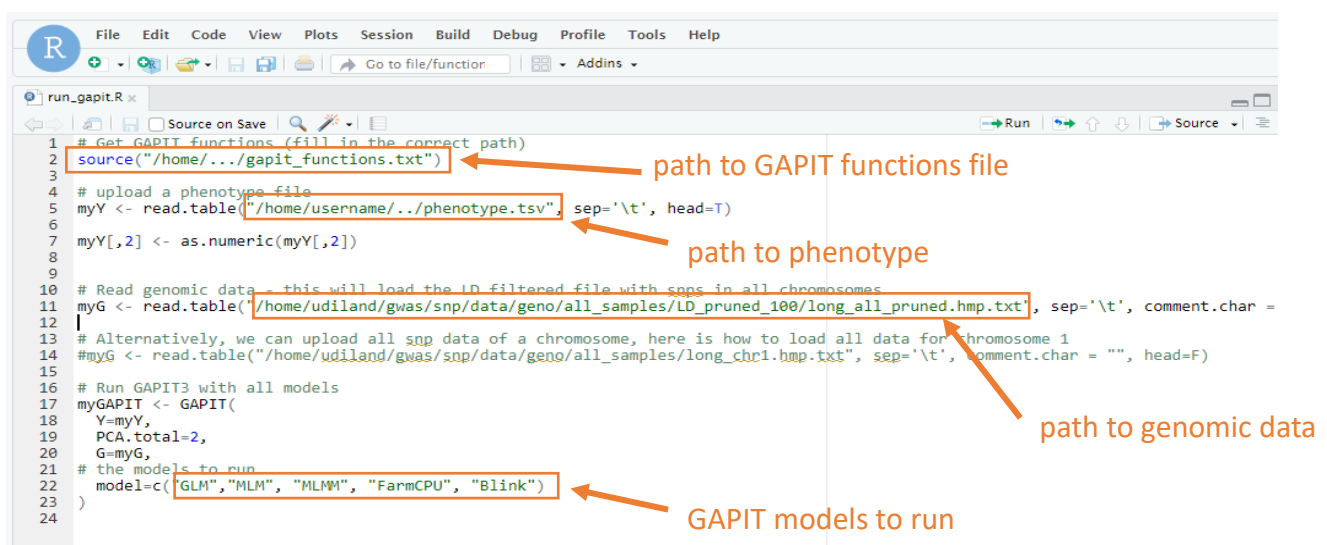
This file has a template of how to run GWAS analysis using GAPIT with multiple models, you need to edit it to your needs.

First, edit the command that loads the GAPIT functions to read the ‘gapit_functions.txt’ file in your file system, for example: ‘/home/udiland/gwas/snp/gapit_functions.txt’ (see example below).

Then, read the phenotype table, write the path to your phenotypic data table (you need to upload it first), example of path to a phenotype table:

“/home/tomerparpar/phenotypes.pheno” ,the path will always start with “/home/<your username>/”.

The genomic data is already on the server, you can choose to use the SNP data for all the chromosomes which is a subset of all SNPs identified, the filtering was done using $LD R^2 > 0.6$ in blocks of 100bp (this is recommended when first running the GWAS), alternatively, you can use the complete SNPs data per chromosome (recommended to use for more detailed identification of a loci), see examples of the running file below.



```
1 # Get GAPIT functions (fill in the correct path)
2 source("/home/.../gapit_functions.txt")
3
4 # upload a phenotype file
5 myY <- read.table("/home/username/./phenotype.tsv", sep='\t', head=T)
6
7 myY[,2] <- as.numeric(myY[,2])
8
9
10 # Read genomic data - this will load the LD filtered file with snps in all chromosomes
11 myG <- read.table("/home/udiland/gwas/snp/data/geno/all_samples/LD_pruned_100/long_all_pruned.hmp.txt", sep='\t', comment.char =
12 |
13 # Alternatively, we can upload all snp data of a chromosome, here is how to load all data for chromosome 1
14 #myG <- read.table("/home/udiland/gwas/snp/data/geno/all_samples/long_chr1.hmp.txt", sep='\t', comment.char = "", head=F)
15
16 # Run GAPIT3 with all models
17 myGAPIT <- GAPIT(
18   Y=myY,
19   PCA.total=2,
20   G=myG,
21 # the models to run
22   model=c("GLM", "MLM", "MLMM", "FarmCPU", "Blink")
23 )
24
```

Annotations in the image:

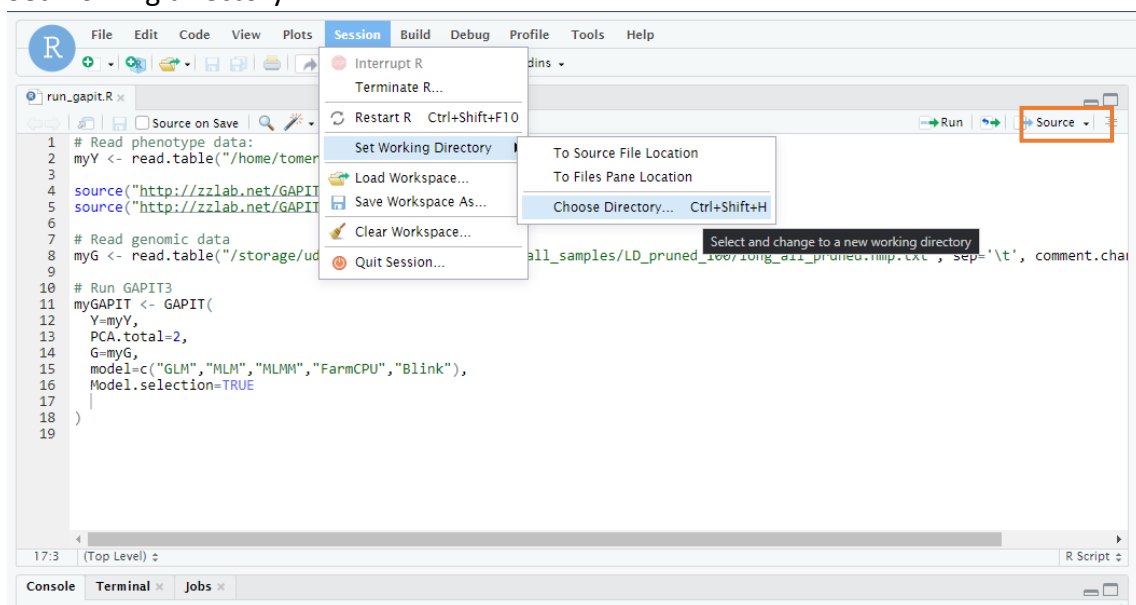
- path to GAPIT functions file (points to line 2)
- path to phenotype (points to line 5)
- path to genomic data (points to line 11)
- GAPIT models to run (points to line 22)

Select output directory:

You need to set the working directory before running the program, all GAPIT output files will be written to this folder.

Choose the directory of output (working directory) by selecting the 'Session' tab on the menu bar, then go to 'Set Working Directory' and 'Choose Directory' to choose the directory of the GWAS output files, it is recommended to save the 'run_gapit.R' file in the same location so you will have it for future reference.

Set working directory:

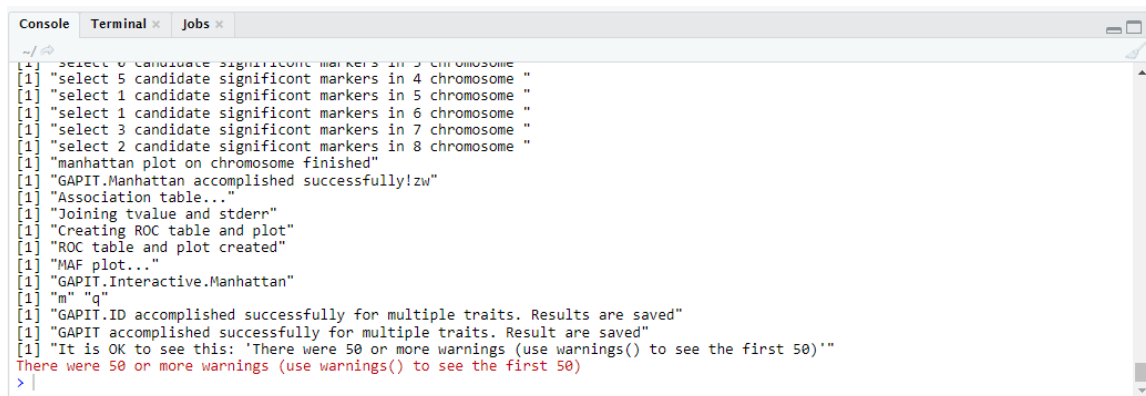


To run the analysis, click on 'Source' button (top right corner)

The analysis will take some time to run, you can see the log printing to the console.

When the analysis is finished you will see the console cursor ('>') indicating the program is ready for other command to run (it is ok to get a line painted in red that says there are warnings).

GAPIT has finished:

A screenshot of a terminal window with tabs for 'Console', 'Terminal', and 'Jobs'. The 'Terminal' tab is active, showing a series of status messages from the GAPIT software. The messages include instructions to select candidate significant markers on various chromosomes (4, 5, 6, 7, 8), completion of a Manhattan plot, successful completion of the GAPIT Manhattan process, creation of an association table, joining of t-values and standard errors, creation of ROC tables and plots, creation of a MAF plot, and successful completion of GAPIT ID and GAPIT Interactive Manhattan processes. It also shows the saving of results for multiple traits and a warning about 50 or more warnings being generated.

```
~/ |  
[1] "select 6 candidate significant markers in 3 chromosome "  
[1] "select 5 candidate significant markers in 4 chromosome "  
[1] "select 1 candidate significant markers in 5 chromosome "  
[1] "select 1 candidate significant markers in 6 chromosome "  
[1] "select 3 candidate significant markers in 7 chromosome "  
[1] "select 2 candidate significant markers in 8 chromosome "  
[1] "manhattan plot on chromosome finished"  
[1] "GAPIT.Manhattan accomplished successfully!zw"  
[1] "Association table..."  
[1] "Joining tvalue and stderr"  
[1] "Creating ROC table and plot"  
[1] "ROC table and plot created"  
[1] "MAF plot..."  
[1] "GAPIT.Interactive.Manhattan"  
[1] "m" "q"  
[1] "GAPIT.ID accomplished successfully for multiple traits. Results are saved"  
[1] "GAPIT accomplished successfully for multiple traits. Result are saved"  
[1] "It is OK to see this: 'There were 50 or more warnings (use warnings() to see the first 50)'"  
There were 50 or more warnings (use warnings() to see the first 50)  
> |
```

Get the results:

The results are on the server in the working directory you choose.

Go to the result folder and **download the files you need to your computer and delete what you don't need from the server** (In order to have enough storage space for others)

For each model there are few files and there are also some general outputs. You need to go over and know them.

The results of the associated SNPs are in the files with the structure:

GAPIT.<name of model>.phenotype_value.GWAS.Results.csv

What next?

In order to interpret the results a several applications where created:

Plot Manhattan: <http://icci-2:9100/manhattan/>

Compare models: <http://icci-2:9100/compare/>

Contact me for any questions/suggestions

udiland@tauex.tau.ac.il