

Module 2: Linkage Disequilibrium in R

Fundamentals of Genomic Prediction and Data-Drive Crop Breeding

(November 24-28, 2025)



Waseem Hussain

Senior Scientist-I

International Rice Research Institute
Rice Breeding Innovations Platform

waseem.hussain@cgiar.org

whussain2.github.io

Mahender Anumalla

Scientist-I

International Rice Research Institute
South-Asian Hub, Hyderabad

m.anumalla@cgiar.org

Margaret Catolos

Associate Scientist

International Rice Research Institute
South-Asian Hub, Hyderabad

m.catolos@cgiar.org

November 20, 2025

Contents

Install and load the packages	1
Load the Data	1
Estimate LD	2
Heat map to Visualize the LD	2
Additional Read and Literature	3

Install and load the packages

- Here in this section we will install and load the required packages

```
> rm(list=ls()) # remove the previous history
> # Install
> #install.packages("BGLR")
> #install.packages("genetics")
> # Installing snpStats package from Bioconductor
> #if (!requireNamespace("BiocManager", quietly = TRUE))
>   #install.packages("BiocManager")
> #BiocManager::install("snpStats")
> # Load the packages
> library(BGLR)
> library(genetics)
> library(pheatmap)
> #library(LDheatmap)
```

Load the Data

Here we will use R package [genetics](#) to measure LD. We will use **mice marker data** given with the package and subset only first 20 markers to estimate LD. The data sets are also available in the folder.

```
> # Read the mice marker data
> #mice.X<-read.csv(file="mice.X.csv", header = TRUE)
> # Load the mice data
> data(mice)
> # Subset the mice data, first 20 markers
> mice.20<- mice.X[, 1:20] # use the first 10 markers
> # Visualize first 5 rows and columns
> mice.20[1:5, 1:5] # Data is coded 0, 1 and 2
```

	rs3683945_G	rs3707673_G	rs6269442_G	rs6336442_G	rs13475700_A
A048005080	1	1	1	1	0
A048006063	1	1	2	1	1
A048006555	2	0	2	2	0
A048007096	1	1	1	1	1
A048010273	2	0	2	2	0

Estimate LD

First we will convert the allele counts into genotypes or haplotypes using *makeGenotypes* function of genetics package. Then we will measure LD using function *LD*. Function *LD* return you list of 8 outputs including D and r^2 values.

```
> # Make genotypes
> mice.20.G<- makeGenotypes(mice.20, convert=c(colnames(mice.20)), method=as.genotype.allele.count)
> # Visualize first 5 rows and columns
> mice.20.G[1:5, 1:5] # Data is coded 0, 1 and 2
```

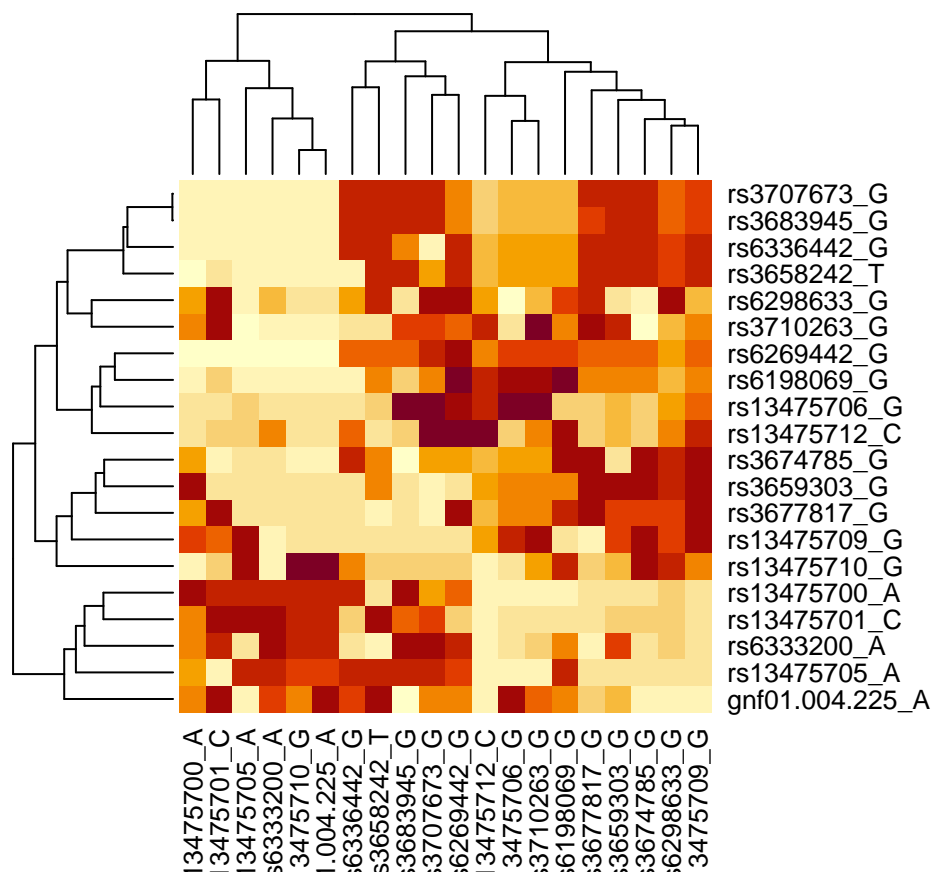
	rs3683945_G	rs3707673_G	rs6269442_G	rs6336442_G	rs13475700_A
A048005080	A/B	B/A	A/B	A/B	B/B
A048006063	A/B	B/A	A/A	A/B	B/A
A048006555	A/A	B/B	A/A	A/A	B/B
A048007096	A/B	B/A	A/B	A/B	B/A
A048010273	A/A	B/B	A/A	A/A	B/B

```
> # Now calculate the LD
> LD.20<- LD(mice.20.G) # This will return the list
> names (LD.20)
[1] "call" "D" "D'" "r" "R^2" "n" "X^2"
[8] "P-value"
> # Extract r2 ( Hill and Robertson (1968)
> r2<-LD.20$`R^2`
> # Copy upper part of matrix to lower for visualizations
> lowerTriangle(r2) <- upperTriangle(r2)
> # Convert Diagonal to 1
> diag(r2)<-1
```

Heat map to Visualize the LD

- Here we will visualize the r^2 matrix as heatmap.

```
> ld.map<-heatmap(r2)
```



```
> ld.map
$rowInd
[1] 19 12 15 7 5 18 16 11 9 10 20 13 8 3 14 17 6 4 1 2

$colInd
[1] 5 7 12 15 18 19 4 6 1 2 3 20 13 14 8 11 9 10 17 16

$Rowv
NULL

$Colv
NULL
```

Additional Read and Literature

- Genome-wide association studies
- Next-generation genetics in plants
- Association study designs for complex diseases
- Linkage Disequilibrium and the Search for Complex Disease Genes
- Linkage disequilibrium: what history has to tell us
- Methods for linkage disequilibrium mapping in crops
- Structure of linkage disequilibrium in plants

For any suggestions or comments, please feel to reach at waseem.hussain@cgiar.org; m.anumalla@cgiar.org; m.catolos@cgiar.org