

Module 5: Predict Performance of Cross Combinations 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](https://github.com/whussain2)

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

| | |
|--|----------|
| Introduction | 1 |
| Load the Libraries | 1 |
| Data Sets | 1 |
| Construct GRM Matrix | 2 |
| Single Trait Usefulness Criterion | 3 |
| 1. Assigning the parents | 3 |
| 2. Creating the crossing plan | 3 |
| 3. Calculate the UC | 3 |
| Creating the mating plan | 4 |

Introduction

For this section we will use the R Package **SimpleMating** to identify the optimal crosses with high UC and crosses that have high variance and minimum inbreeding. To get more details check the manuscript on [SimpleMating](#) published by authors and [GitHub Page](#)

Recall that **Usefulness Criterion (UC)** of cross is the expected mean performance plus the expected selection gain ([Reference](#)) and is given as:

$U = \mu + \sigma^2 P$; where μ is mean of cross and $\sigma^2 P$ is the variance of cross.

SimpleMating gives us parental average, cross total genetic value, and/or usefulness (using additive and/or non-additive effects), allowing for multi-trait scenarios. Furthermore, it uses a mate allocation algorithm to create a mating plan aiming to maximize a target criterion (mean parental average, total genetic value, or usefulness) and constrains the next generation inbreeding levels (Copied from authors tutorial on GitHub).

Load the Libraries

```
> # Install Library
> #library(devtools)
> #install_github('Resende-Lab/SimpleMating')
> library(SimpleMating)
> library(AGHmatrix)
```

Data Sets

Here we will use data sets that are given in package. I have saved these data sets in the **Data** folder and we can directly upload it from there. We need four data files as shown here:

```
> # 1. Marker Data
> geno<-read.csv(file="./Data/Geno_Marker.csv", header=TRUE,
+               row.names = 1)
> geno<-as.matrix(geno)
> dim(geno)
```

```
[1] 100 1230
```

```
> #kable(head(geno))
>
> # 2. Markers effects for target trait
```

```

> MarEff<-read.csv(file="./Data/Marker_effects.csv", header=TRUE,
+                 row.names = 1)
> #kable(head(MarEff))
> # 3. Genetic Map or LD Matrix
> Map_info<-read.csv(file="./Data/Map_info.csv", header = TRUE,
+                  row.names = 1)
> #kable(head(Map_info))

```

Construct GRM Matrix

We will use AHGmatrix R package to build the GRM.

```

> #GRM<- (lines_Geno %*% t(lines_Geno)) / ncol(lines_Geno)
> GRM<- Gmatrix(geno, missingValue=NA,
+              maf=0.05, method="VanRaden")

```

Initial data:

```

Number of Individuals: 100
Number of Markers: 1230

```

Missing data check:

```

Total SNPs: 1230
0 SNPs dropped due to missing data threshold of 0.5
Total of: 1230 SNPs

```

MAF check:

```

524 SNPs dropped with MAF below 0.05
Total: 706 SNPs

```

Heterozygosity data check:

```

No SNPs with heterozygosity, missing threshold of = 0

```

Summary check:

```

Initial: 1230 SNPs
Final: 706 SNPs ( 524 SNPs removed)

```

Completed! Time = 0.034 seconds

```

> dim(GRM)

```

```

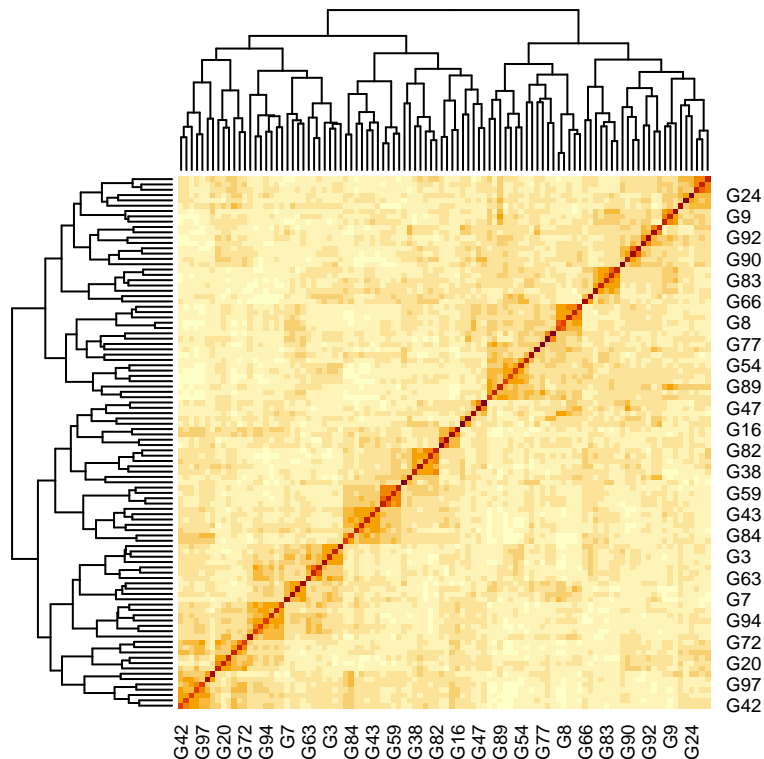
[1] 100 100

```

```

> heatmap(GRM)

```



Single Trait Usefulness Criterion

Here we will use steps as guided in manual to calculate the UC.

1. Assigning the parents

```
> # Select parents from geno
> Parents <- rownames(geno)
```

2. Creating the crossing plan

```
> # Create a Half sib plan
> Cross_plan <- planCross(TargetPop = Parents,
+                         MateDesign = "half") # see other plans in arguments
```

Number of crosses generated: 4950

3. Calculate the UC

```
> # Usefulness of trait number 1
> UC_crosses<- getUsefA(MatePlan = Cross_plan,
+                       Markers = geno,
+                       addEff = MarEff[, 1], # Trait with Marker effects
+                       Map.In = Map_info,
+                       K = GRM,
```

```
+ propSel = 0.05, # Proportion selected
+ Type = "RIL", # Type DH or RILs
+ Generation = 6) # Which generation
```

Usefulness predicted for 4950 crosses.

```
> kable(head(UC_crosses[[1]], 10))
```

| Cross.ID | Parent1 | Parent2 | Mean | Variance | sd | Usefulness |
|----------|---------|---------|---------|-----------|-----------|------------|
| G1_G2 | G1 | G2 | 4.51880 | 0.1236724 | 0.3516708 | 5.24420 |
| G2_G4 | G2 | G4 | 4.18298 | 0.2053613 | 0.4531681 | 5.11774 |
| G2_G3 | G2 | G3 | 4.01124 | 0.2676103 | 0.5173106 | 5.07830 |
| G1_G3 | G1 | G3 | 4.10644 | 0.2065431 | 0.4544701 | 5.04388 |
| G1_G4 | G1 | G4 | 4.27818 | 0.1282092 | 0.3580631 | 5.01676 |
| G3_G4 | G3 | G4 | 3.77061 | 0.2908438 | 0.5392994 | 4.88303 |
| G2_G5 | G2 | G5 | 3.38943 | 0.2994511 | 0.5472213 | 4.51819 |
| G1_G5 | G1 | G5 | 3.48462 | 0.2342698 | 0.4840142 | 4.48300 |
| G1_G30 | G1 | G30 | 3.40499 | 0.2518281 | 0.5018247 | 4.44011 |
| G1_G24 | G1 | G24 | 3.09747 | 0.4016060 | 0.6337239 | 4.40466 |

Creating the mating plan

Here we will create a mating plan to restrict the levels of inbreeding, by means of the covariance in the individual-cross level, and maximize the criterion used. Please see manual from authors on [GitHub Page](#)

```
> # Use Usefulness Data above
> maxGainPlan <- selectCrosses(data = UC_crosses[[2]],
+ n.cross = 20, # Set Target crosses
+ max.cross = 2, # Maximum crosses parent can be part
+ min.cross = 1, # Minimum crosses parent can be part
+ culling.pairwise.k = 1) # covariance between cross, restriction
> # Mating Plan
> maxGainPlan[[2]]
```

| Parent1 | Parent2 | Y | K |
|---------|---------|---------|------------|
| G1 | G2 | 5.24420 | 0.9581274 |
| G2 | G3 | 5.07830 | 0.0295735 |
| G1 | G3 | 5.04388 | 0.2682648 |
| G4 | G24 | 4.15885 | -0.1444431 |
| G4 | G30 | 4.15516 | 0.1623349 |
| G5 | G30 | 3.32222 | -0.0417702 |
| G5 | G12 | 3.14401 | -0.1664666 |
| G12 | G24 | 2.89113 | -0.1196279 |
| G20 | G26 | 2.81435 | -0.2547157 |
| G23 | G26 | 2.77752 | -0.1427371 |
| G23 | G25 | 2.71351 | -0.1079958 |
| G19 | G25 | 2.57126 | -0.0751156 |
| G20 | G56 | 2.55632 | -0.3173741 |
| G9 | G56 | 2.55347 | -0.1002410 |
| G9 | G19 | 2.37980 | 0.4241353 |
| G10 | G13 | 2.32380 | -0.1419616 |
| G7 | G13 | 2.29550 | -0.1034980 |

| Parent1 | Parent2 | Y | K |
|---------|---------|---------|------------|
| G8 | G10 | 2.28721 | -0.4882888 |
| G7 | G54 | 2.28316 | -0.2419979 |
| G22 | G54 | 2.21373 | -0.1221094 |

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