Module 5: Genomic Prediction with Main QTLs Fixed

Fundamentals of Genomic Prediction and Data-Drive Crop Breeding

(August 4-8, 2025)

Waseem Hussain
Senior Scientist-I
International Rice Research Institute
Rice Breeding Innovations Platfrom
waseem.hussain@irri.org
whussain2.github.io

Mahender Anumalla
Scientist-I
International Rice Research Institute
South-Asian Hub, Hyderabad
m.anumalla@irri.org

August 7, 2025

Contents

Background Information

In this section we will perform genomic selection but we will use **Major Effect or Know QTLs as Fixed Effects".

Treating them as fixed effect mean they will not be shrinked. And there effects will be maintained.

What is Our Goal

Fix the QTLs as fixed effect

Load the R Packages

```
>   library(AGHmatrix)
>   library(BGLR)
>   library(lme4)
>   library(ggplot2)
>   library(sommer)
```

Read the Data Sets

The Data has 5 environments and has yield data. The data comes from the different locations in Bangladesh and India. BLUEs alrady extracted. We will upload the file and use it for analysis.

```
>   rm(list=ls()) # remove History
> # Read the phenotypic data
>   BLUEs.all<-read.csv(file="./Data/BLUES.ALL.csv")
>   BLUEs.all<-subset(BLUEs.all, Environment=="ENV1")
```

Read Genotype Data

This marker data as 844 genotypes with 396511 SNP Markers, and the file is saved as **.rds**. We will subset 252 genotypes and use it estimate the **GEBVs**.

```
>   geno<-readRDS("./Data/GBS_datav2.rds")
>   dim(geno)
```

```
[1]    844 396511
```

```
> # Match genotype with Phenotype
>   Ids<-unique(BLUEs.all$Genotype)
>   length(Ids)
```

```
[1] 251
```

```
> # Now subet the genotype Data based on IDs
>   geno<-geno[row.names(geno)%in%Ids,]
>   dim(geno)
```

```
[1]    251 396511
```

Build the G matrix

- Here we will construct the **Genomic Relationship Matrix (GRM)** using marker data. The GRM will be based on **VanRanden (2008)**.
- The steps used to create this GRM is:
  - Create a center of marker data (X matrix)
  - Create a Cross Product (XX)
  - Divide the (XX) by number of markers

$$GRM = XX^t/m$$

- More on relationship matrix can be found here Source 1, Source2
- We will use the AGHmatrix package to build G matrix.

```
>   GM<- Gmatrix(SNPmatrix=geno, missingValue=NA,
+                            maf=0.05, method="VanRaden")
```

```
Initial data:
    Number of Individuals: 251
    Number of Markers: 396511

Missing data check:
    Total SNPs: 396511
     0 SNPs dropped due to missing data threshold of 0.5
    Total of: 396511  SNPs

MAF check:
     25572 SNPs dropped with MAF below 0.05
    Total: 370939  SNPs

Heterozigosity data check:
    No SNPs with heterozygosity, missing threshold of =  0

Summary check:
    Initial:  396511 SNPs
    Final:  370939  SNPs ( 25572  SNPs removed)

Completed! Time = 31.436  seconds
```

```
>   dim(GM)
```

```
[1] 251 251
```

Fit QTLs as Fixed Effects

Fit a multi-kernel model using BGLR to treat some large-effect QTL as fixed effects, and remaining QTL as random effects. QTL here were previously declared, significant using a GWAS analysis. SNP positions of QTL were 1926, 829, 683, 678.

```
> # Let us take position of markers based on columns
>    qtl <- c(1120, 1126, 1128, 1129)
>
> # Create a ETA list
>     ETA <- list(
+       list(X = geno[, qtl], model = 'FIXED', probIn = 0.10),
+       list(K = GM, X = geno[, -qtl], model = 'RKHS', probIn = 0.10)
+     )
>
> # Fit the model
>   model_fix<- BGLR(y=BLUEs.all$BLUEs, ETA=ETA, burnIn=500, nIter=2000,
+                        verbose=FALSE)
> # Extract the GBVs
>   GEBVs_fixed<-data.frame(GEBVs=model_fix$yHat)
```

Note: For questions specific to data analysis shown here contact waseem.hussain@cgiar.org

---

If your experiment needs a statistician, you need a better experiment - Ernest Rutherford

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