

Pipeline

Jinliang Yang

May 15, 2015

Analysis Pipeline and source codes

Note: currently the pipeline sits in a private repo on github: <https://github.com/yangjl/Misc>. I can share the complete repo upon request. And the SNP data sit on farm in the dir of /group/jrigrp4/AllZeaGBSv2.7impV5.

A.1 Obtain GBS and phenotypic data for maize diversity panel

```
source("profiling/A.1_GBS_diverse.R")
```

A.2 Run the following shell codes: convert HapMap to BED5 format

```
# open interactive srun on farm
srun.x11 -p bigmemh --ntasks=8 --odelist=bigmem4
# run the shell
sh profiling/4.sweet/4.A.2_GBS_hdf2hmp.sh
```

A.3 Convert hapmap to BED5 format

```
source("profiling/A.3_GBS_bed5format.R")
# Run the following python code
snprfq -p /group/jrigrp4/AllZeaGBSv2.7impV5 -i ZeaGBSv27_Ames282.bed5 \\\
-s 6 -m "0" -a 0 -b 1 -c 2 -o ZeaGBSv27_Ames282.frq
```

A.4 checking the SNP MAF and missing rate

```
source("profiling/A.4_GBS_maf_mr.R")
```

B.1 derive the BLUE values for the phenotypic data and plot the histogram distribution of the traits.

```
source("profiling/B.1_phenotype.R")

pheno <- read.table("data/pheno_ames282.txt",
  header = TRUE)
par(mfrow = c(1, 2))
traits <- c("10 kernel weight", "Total kernel weight")
hist(pheno[, 3], breaks = 30, col = "cadetblue",
  main = "10 kernel weight", xlab = "weight (g)")
hist(pheno[, 8], breaks = 30, col = "cadetblue",
  main = "total kernel weight", xlab = "weight (g)")
```

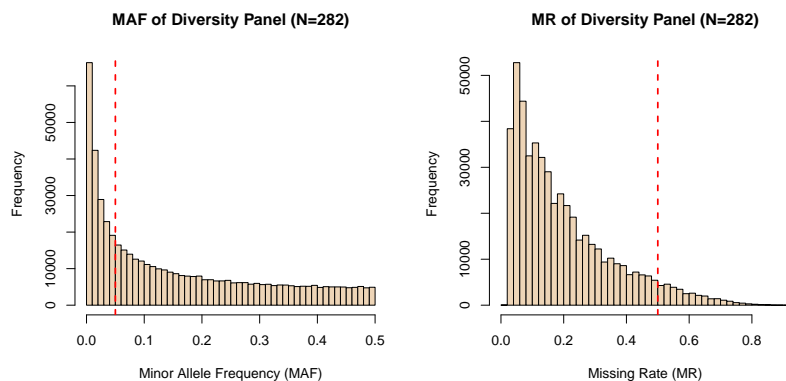


Figure 1: Minor allele frequency (MAF) and missing rate (MR) of the GBS SNPs of the maize diversity panel.

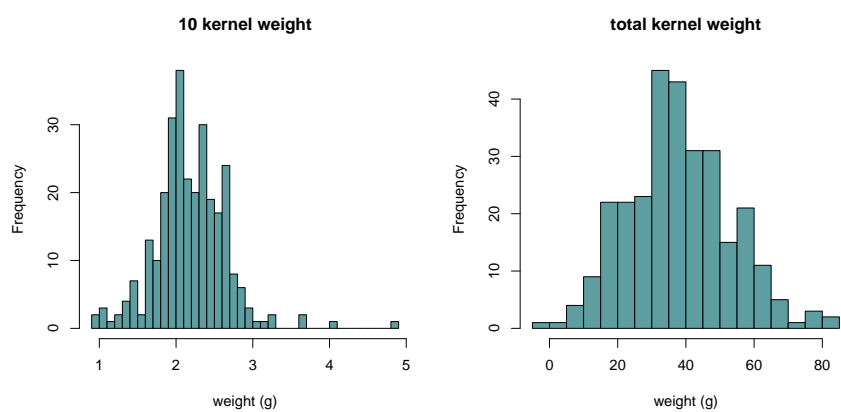


Figure 2: Histogram distribution of the phenotypic traits of 10 kernel weight and total kernel weight of the diversity panel.

B.2 change the genotype format to GenABEL

```
source("../profiling/4.sweet/4.B.2_GBS_2GenABEL.R")
```

B.3 and B.4 Regional association study

```
source("profiling/B.3_GenABEL_step1.R")
```

```
source("profiling/B.4_GenABEL_step2.R")
```

```
load("cache/gwas_res.RData")
```

```
library(GenABEL)
```

```
## Loading required package: MASS
```

```
## Loading required package: GenABEL.data
```

```
par(mfrow = c(1, 2))
```

```
plot(res1, main = "10 kernel weight", pch = 16,
     col = "cadetblue")
```

```
abline(v = 127466000, lwd = 2, col = "red", lty = 2)
```

```
plot(res2, main = "total kernel weight", pch = 16,
     col = "cadetblue")
```

```
abline(v = 127466000, lwd = 2, col = "red", lty = 2)
```

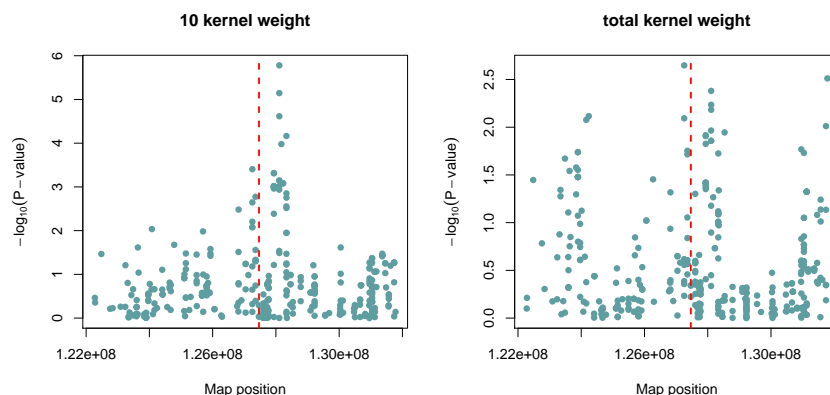


Figure 3: Regional GWAS with the simplest linear model.

```
# load("../cache/gwas_res.RData")
```

```
par(mfrow = c(1, 2))
```

```
plot(res1.eg, main = "10 kernel weight", pch = 16,
     col = "cadetblue")
```

```
abline(v = 127466000, lwd = 2, col = "red", lty = 2)
```

```
plot(res2.eg, main = "total kernel weight", pch = 16,
     col = "cadetblue")
```

```
abline(v = 127466000, lwd = 2, col = "red", lty = 2)
```

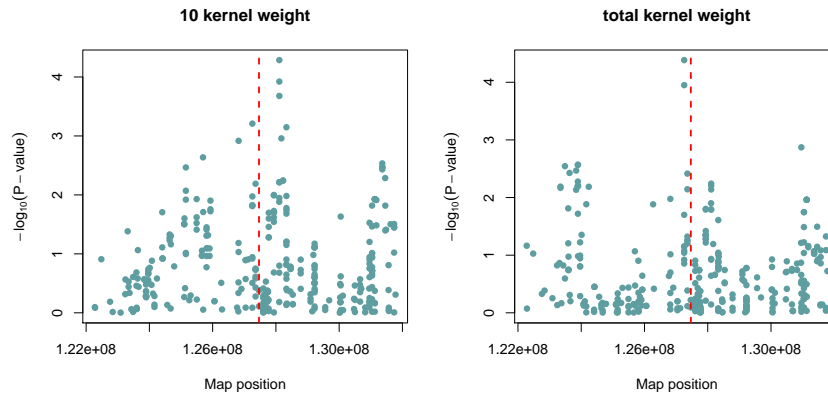


Figure 4: Regional GWAS with kinship matrix calculated from genome-wide marker to control the population structure.

```
# load('.../cache/gwas_res.RData')
par(mfrow = c(1, 2))
plot(res1.mm, main = "10 kernel weight", pch = 16,
      col = "cadetblue")
abline(v = 127466000, lwd = 2, col = "red", lty = 2)
plot(res2.mm, main = "total kernel weight", pch = 16,
      col = "cadetblue")
abline(v = 127466000, lwd = 2, col = "red", lty = 2)
```

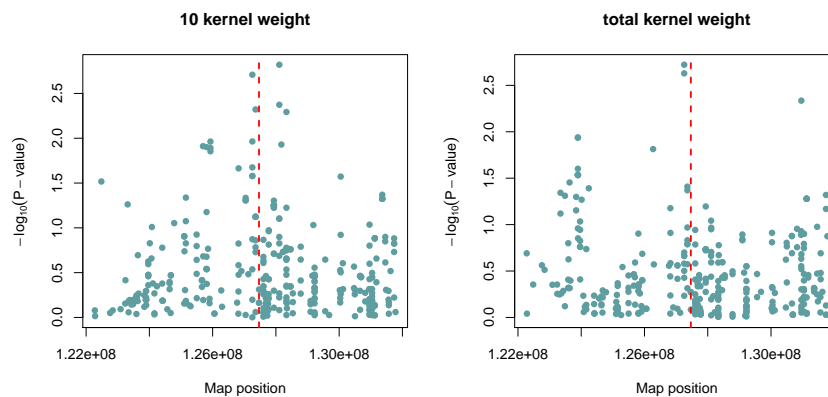


Figure 5: Regional GWAS with kinship matrix calculated from genome-wide marker to control the population structure and with the background QTL control.