Regional association study near ZmSWEET4c gene

Jinliang Yang

May 13th, 2015

Results

With the control of the population structure and polygenic effects of background QTLs, we conducted the regional association scanning around the ZmSWEET4c gene. As show in Figure 1, two SNPs near the ZmSWEET4c gene were significantly (FDR < 0.01) associated with the traits of ten kernel weight and total kernel weight, which are two important yield indexes. The most significant SNPs could explained 4.3% and 4.3% of the total phenotypic variations for ten kernel weight and total kernel weight, respectively. And about 7.8% and 10.8% of the heritability of the traits could be accounted by the most significant markers for ten kernel weight and total kernel weight. Importantly, non-B73 like genotype has the magtitude of effects of 0.2g (9.1%) and 4.4g (12%) for ten kernel weight and total kernel weight, respectively.

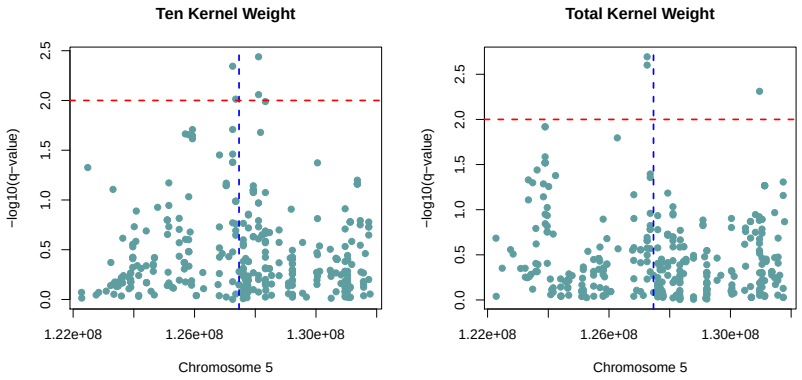

Figure 1: Regional association scanning results. Vertical dashed blue lines indicate the center of the ZmSWEET4c gene and horizontal red lines indicate the threshold of FDR=0.01.

Materials and Methods

Regional association study

A maize diversity panel composed of 282 inbred lines was employed for the regional association study. To conduct the analysis, we obtained Genotype-By-Sequencing (GBS) data from panzea (www.panzea.org) and obtained phenotypic data from Flint-Garcia et al., 2009 [1]. The SNPs data were filtered with minor allele frequency

[1] Flint-Garcia, Sherry A., et al. "Heterosis is prevalent for multiple traits in diverse maize germplasm." PloS one 4.10 (2009): e7433.

(MAF) > 0.05 and allele missing rate < 50%. After filtering, a total of 306,190 SNPs were remaining, including 79 SNPs in a 1-Mb region surrounding the ZmSWEET4c gene.

Association study with the mixed-model method was conducted using an R [2] add-on package "GenABEL" [3]. First of all, a kinship matrix was estimated from the genomic data to control population structure. Secondly, genome-wide polygenic effects were computed with the function "polygenic" for controlling background QTLs. Finally, the 79 SNPs near zmSWEET4c gene were tested one-by-one as the fixed effect and polygenic QTL effects derived from previous step were fitted as random effects using the function of "mmscore".

[2] Team, R. Core. "R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria, 2012." (2012).

[3] Aulchenko, Yurii, et al. "GenABEL: genome-wide SNP association analysis." R package version (2010): 1-6.

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/4.sweet/4.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 --nodelist=bigmem4
# run the shell
sh profiling/4.sweet/4.A.2_GBS_hdf2hmp.sh
```

A.3 Convert hapmap to BED5 format

```
source("../profiling/4.sweet/4.A.3_GBS_bed5format.R")
# Run the following python code
snpfrq -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/4.sweet/4.A.4_GBS_maf_mr.R")
```

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

```
source("../profiling/4.sweet/4.B.1_phenotype.R")
```

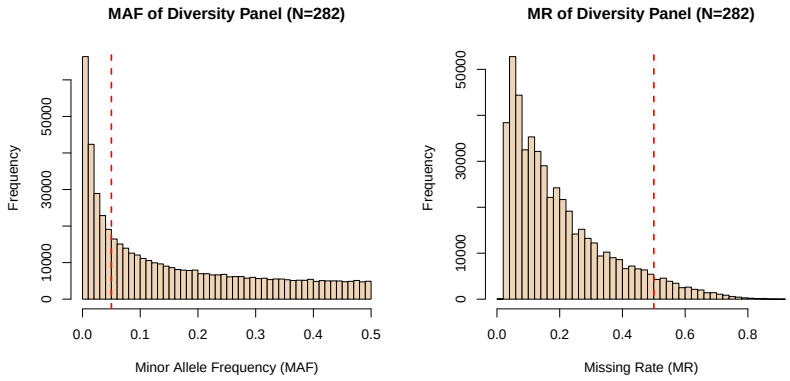

Figure 2: Minor allele frequency (MAF) and missing rate (MR) of the GBS SNPs of the maize diversity panel.

```
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)")
```

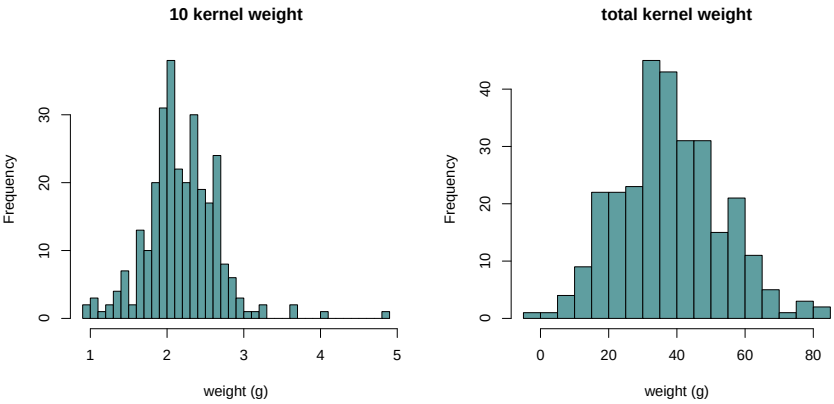

Figure 3: Histogram distribution of the phenotypic traits of 10 kernel weight and total kernel weight of the diversity panel.

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/4.sweet/4.B.3_GenABEL_step1.R")
source("../profiling/4.sweet/4.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)
```

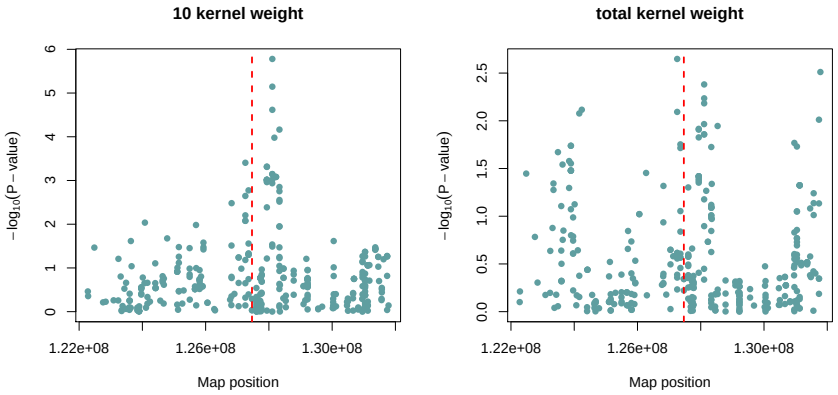

Figure 4: Regional GWAS with the simplest linear model.

```
# 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)
```

```
# 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)
```

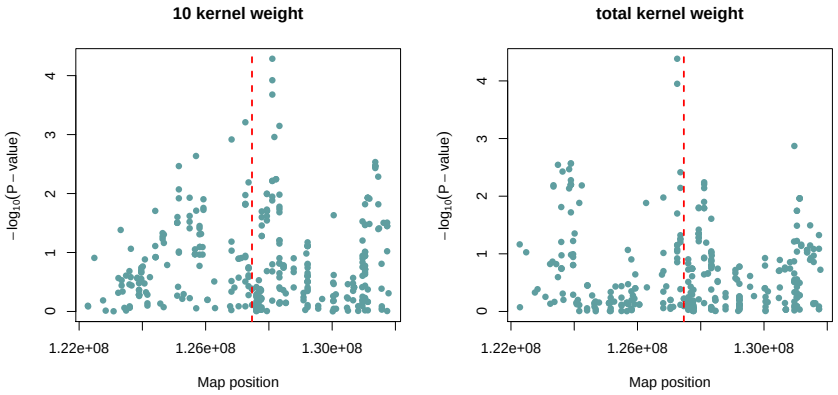

Figure 5: Regional GWAS with kinship matrix calculated from genome-wide marker to control the population structure.

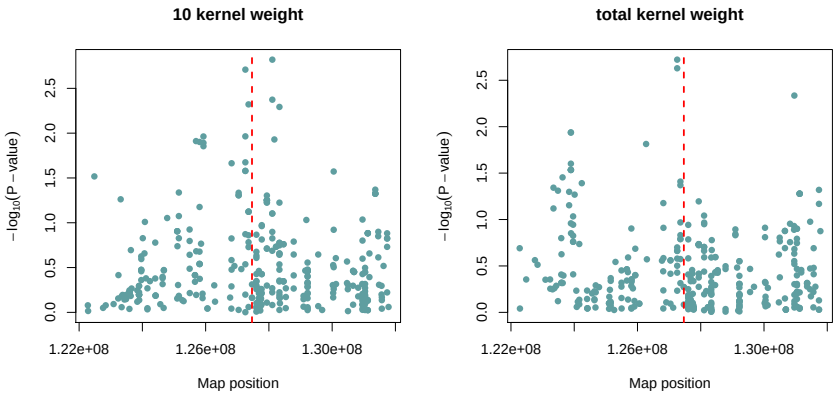

Figure 6: Regional GWAS with kinship matrix calculated from genome-wide marker to control the population structure and with the background QTL control.