Introgression Analysis

Jinliang Yang

Jan. 5th, 2015

\begin{abstract} Test whether Totontepec maize lines have admixture with *mexicana* using SNP50k data. \end{abstract}

Phasing parental reference populations using **fastPHASE**

Run the following R script to format the SNP50k data and the formatted data were used as the `fastPHASE`[1] input. The formatted data included: 1. Mexicana (N=120) 2. Parviglumis (N=130) 3. Landraces (N=94)

```
source("profiling/1.Introgression/1.A.1_fastPHASE_input.R")
```

And then run `fastPHASE` to do the haplotype phasing with HMM.

```
cd /largedata/fphase
fastPHASE -oMex mex_120.fp
fastPHASE -oParv parv_
fastPHASE -oLand land_
```

[1] Scheet, P., & Stephens, M. (2006). A Fast and Flexible Statistical Model for Large-Scale Population Genotype Data : Applications to Inferring Missing Genotypes and Haplotypic Phase, 78(April), 629–644.

The phenotypic data *per se*, GCA and SCA was reformatted from Sofiane's files and transformed them to a data matrix. The following chunk of code could not be re-ran because data from external folders were loaded. The original phenotypic data could be found here `/group/jrigrp2/DiallelSofiane/GeneticValues/` on farm.

12 Ex-PVP lines are B73, Mo17 and 10 others (LH1, LH123HT, LH82, PH207, 4676A, PHG39, PHG47, PHG84, PHJ40, PHZ51). And the 7 traits of interest are: ASI: anthesis-silk Interval; DTP: days to pollen; DTS: days to silking; EHT: ear height; GY: adjusted grain yield; GY: adjusted grain yield; TW: test weight.