

Introgression Analysis

Jinliang Yang

Jan. 5th, 2015

SUMMARY

Test whether Totontepec maize lines have admixture with *mexicana* using SNP50k data.

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)

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.

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

And then run *fastPHASE* to do the haplotype phasing with HMM. Basically, I ran the code, for example

```
fastPHASE -oMex -S1234 mex_120.fp,
```

to do the hap phasing. To run on farm with *slurm*, I prepared the *slurm* codes with an R script.

```
source("profiling/1.Introgression/1.A.2_run_fastPHASE.R")
```

```
# And then run the following code on farm
```

```
cd /home/jolyang/Documents/Github/Introgression
sbatch -p bigmeml largedata/fphase/mex_fp_slurm.sh
sbatch -p bigmeml largedata/fphase/parv_fp_slurm.sh
sbatch -p bigmeml largedata/fphase/land_fp_slurm.sh
```