

# Notes

*Jinliang Yang*

*July 23, 2015*

## Infer mom's genotype by JRI

We have obs. mom and obs. (selfed) kids. We want to know  $P(G|\theta)$ , and  $P(G|\theta) \propto P(\theta|G) \times P(G)$ , where  $\theta$  is observed data. This consists of observed genotypes ( $G'$ ) of both mom and kids. So:  $P(G|\theta) \propto \left( \prod_{i=1}^k P(G'_i|G) \right) \times P(G'_{mom}|G) \times P(G)$  This function is to impute mom's genotype from a progeny array of k kids at a single locus. inferred\_mom=1 -> 00, 2->01, 3->11

---

## Current issues

1. ~~errors should not be used for inferring~~
2. ~~p unknown, could only be estimated from data.~~
3. ~~missing data?~~
4. File IO
5. switching error.?
6. mom + dad

## To Do List:

1. ~~impute Mom v~~
  2. ~~phase Mom v~~
  3. ~~phase Kids x~~
  4. impute Kids: with certain Mb window; HMM
- 

Running with John's idea, I just did chromosome-level (40K SNPs, 1.5 crossovers) sims assuming we know mom's genotype without error (e.g. from WGS data). V1 is switch errors, V2 is progeny genotyping error. This is with window of 11 SNPs and 10 progeny. It is better than using an imputed mom, but not as much as I expected. This will make a big difference however for small families and for plants with no selfs.

-Jeff

```

> summary(x)
      V1      V2
Min.   : 0.00  Min.   :0.04522
1st Qu.: 0.00  1st Qu.:0.07948
Median : 0.00  Median :0.08783
Mean   : 2.14  Mean   :0.08827
3rd Qu.: 2.00  3rd Qu.:0.09876
Max.   :95.00  Max.   :0.12422

```