

BIS I 80L

# Genome-Wide Association Mapping

May 2, 2017

Professor MALOOF

[jnmalooof@ucdavis.edu](mailto:jnmalooof@ucdavis.edu)

# Association Mapping

- Goals
  - find genes associated with diseases or other traits
  - assess relative risk based on genotype (personalized medicine)
- Association mapping
  - look at associations at a few loci (you have *a priori* candidates)
- Genome Wide Association Mapping (GWAS)
  - scan the whole genome for associations
- Takes advantage of historical recombination

# Many GWAS publications

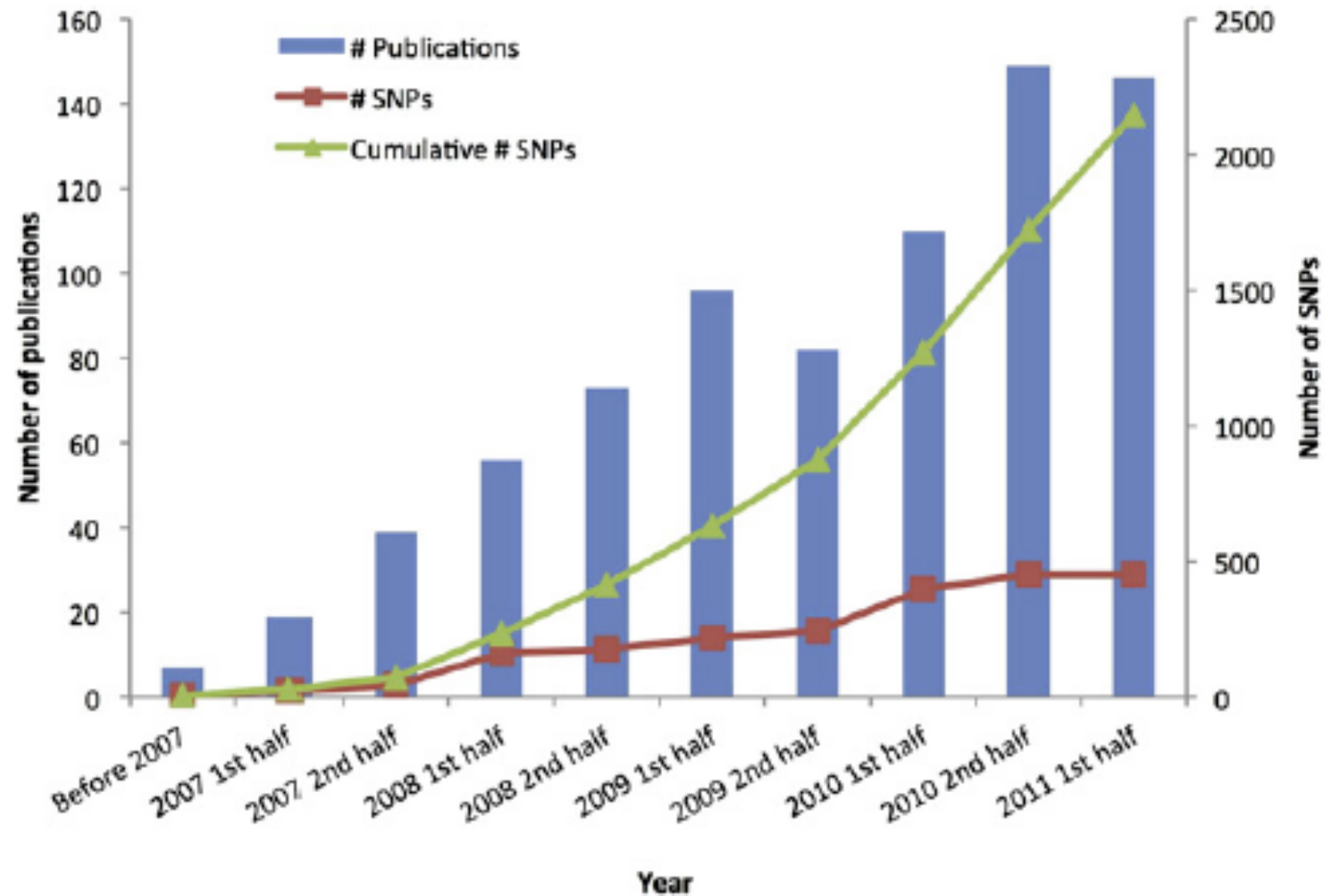


Figure 1. GWAS Discoveries over Time

## Lecture Outline

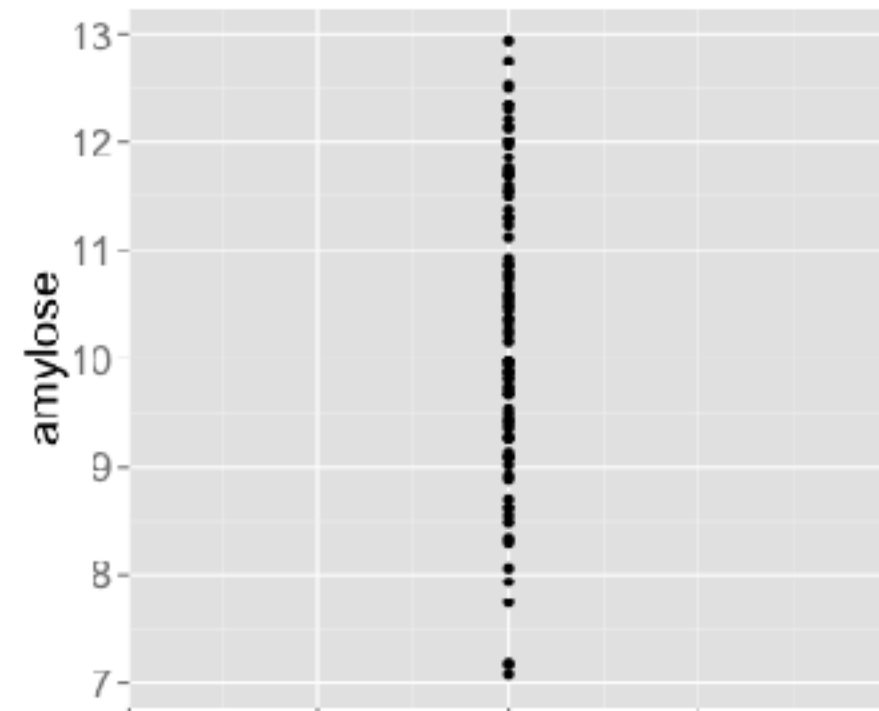
- Association Mapping
- Genome-Wide Association Mapping (GWAS)
- The problem of population structure

## Association Mapping

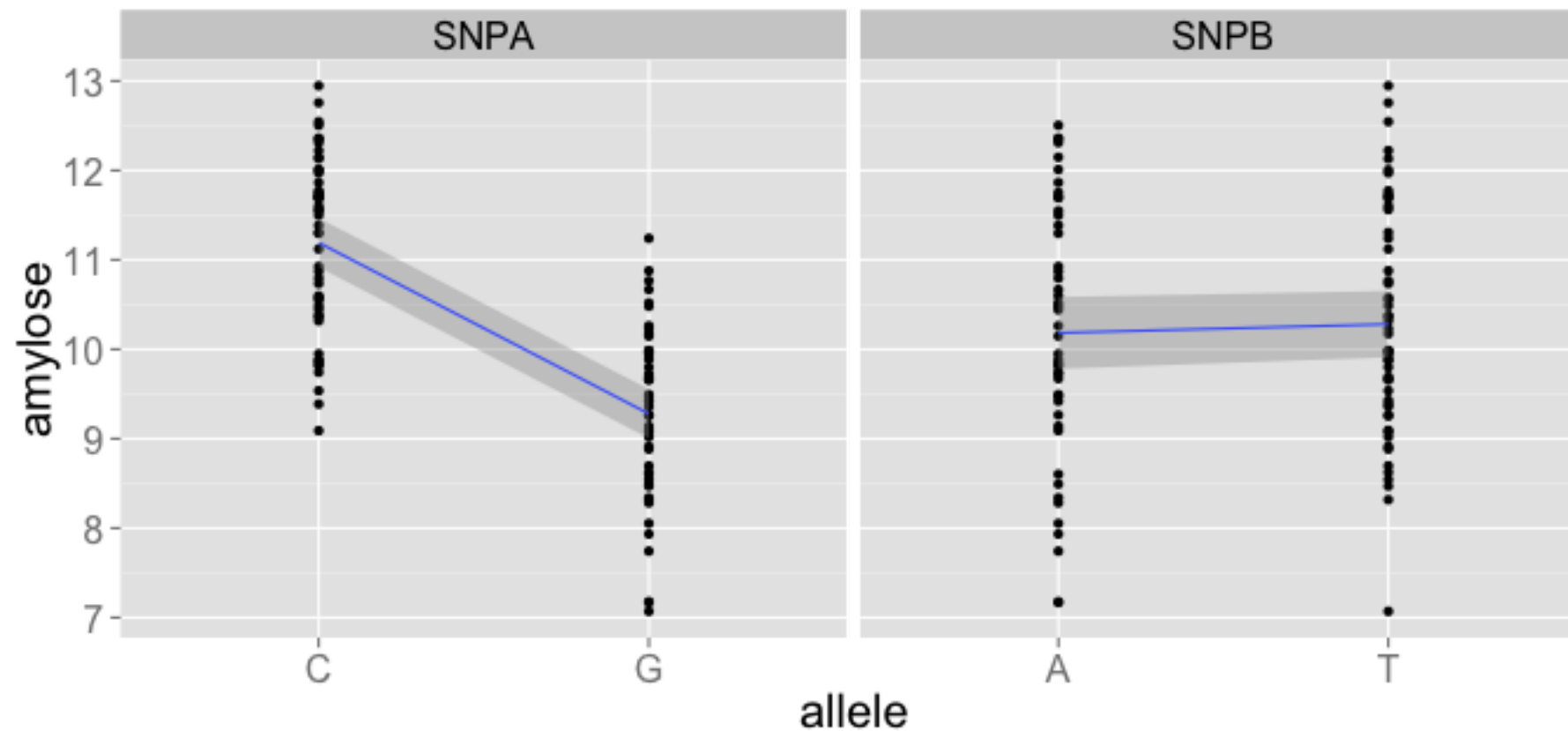
- Trying to find genetic basis for a trait or disease
- Look for statistical association between a SNP allele state and a phenotype

## Association Mapping, simulated example for amylose

- Measure amylose in many rice varieties



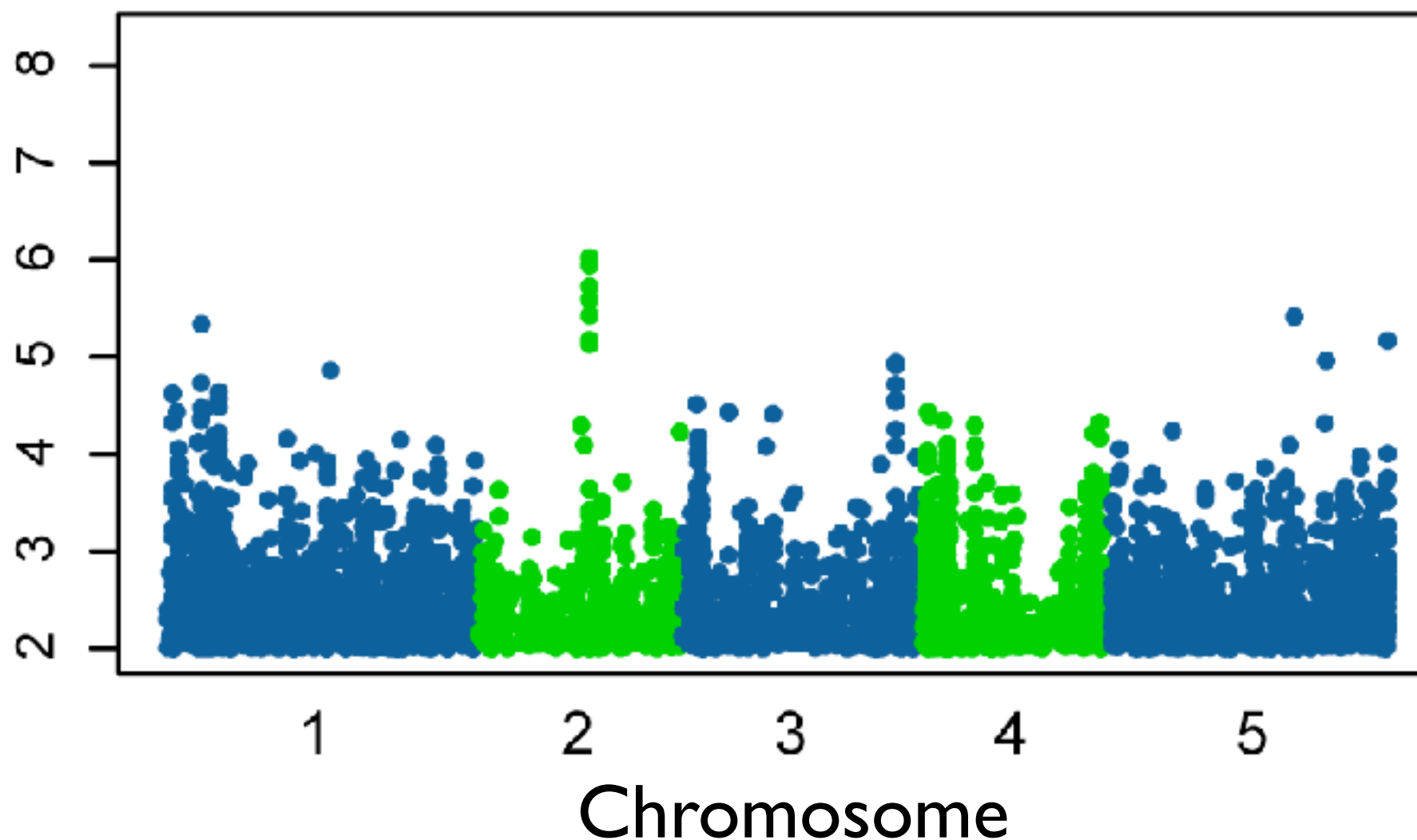
- Separate measurements according to SNP allele



- Test for association. Slope not equal to 0 = association.

## Association Mapping vs Genome-Wide Association Mapping

- For GWAS repeat the analysis for SNPs across the whole genome.
- Can plot the results as a manhattan plot:
  - each point is a SNP
  - X-axis is position in the genome. In this case there are 5 chromosomes
  - Y-axis is  $-\log_{10}(P)$  for association with the trait. Higher values are more significant.

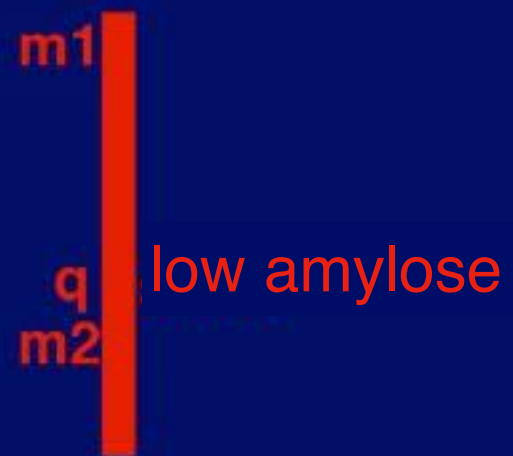
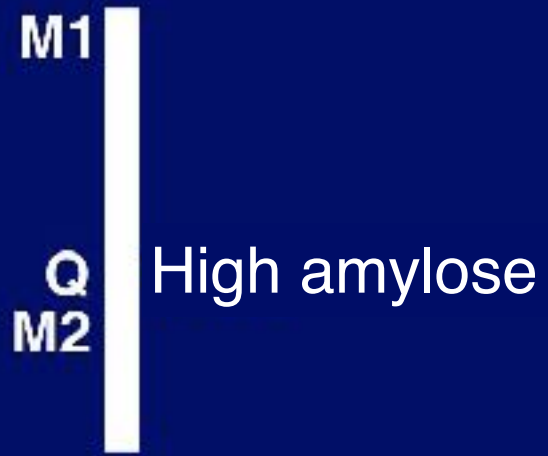


## Association Mapping: historical recombination

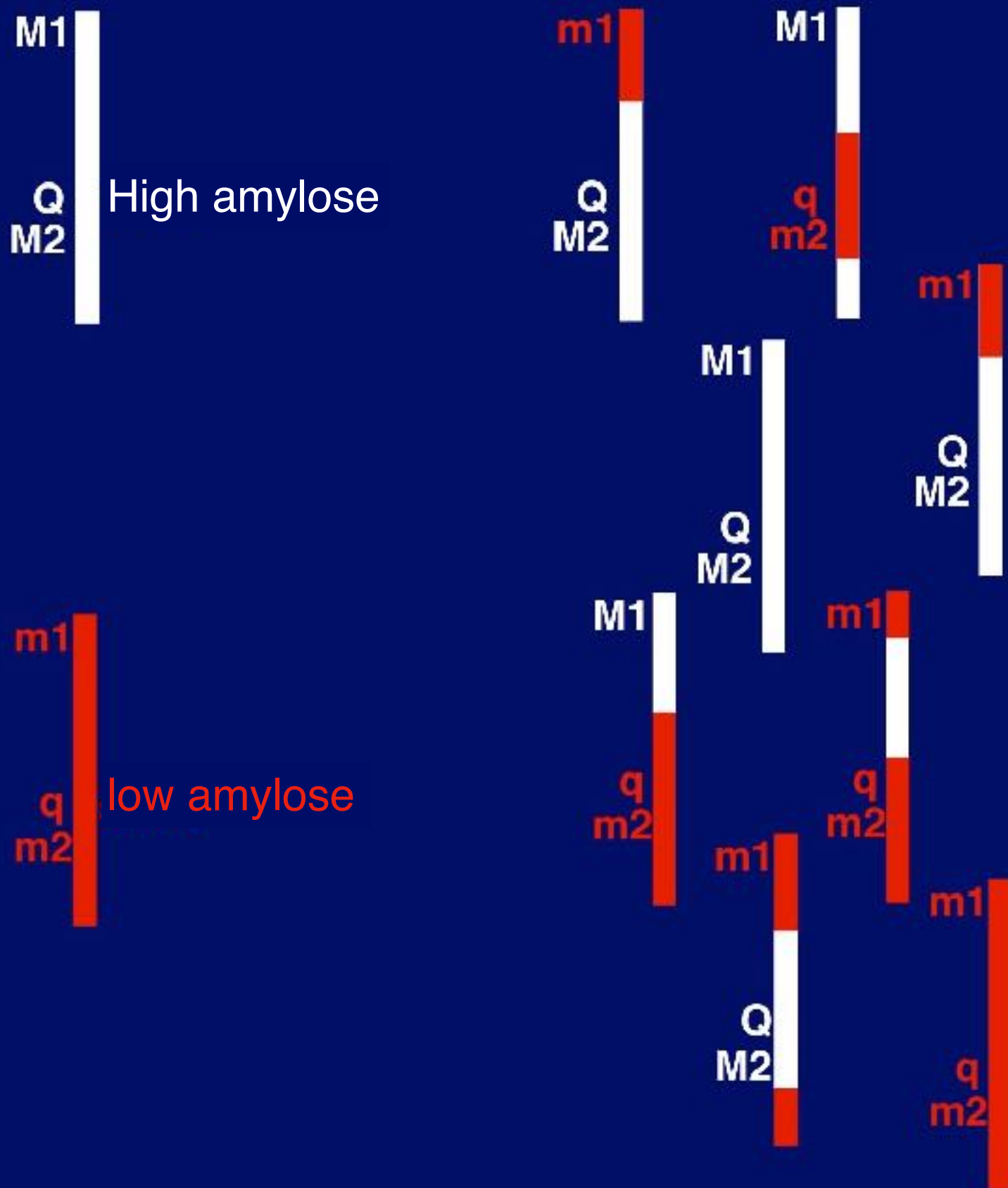
- Why might some SNPs be associated with a trait and not others?
- Historical Recombination!



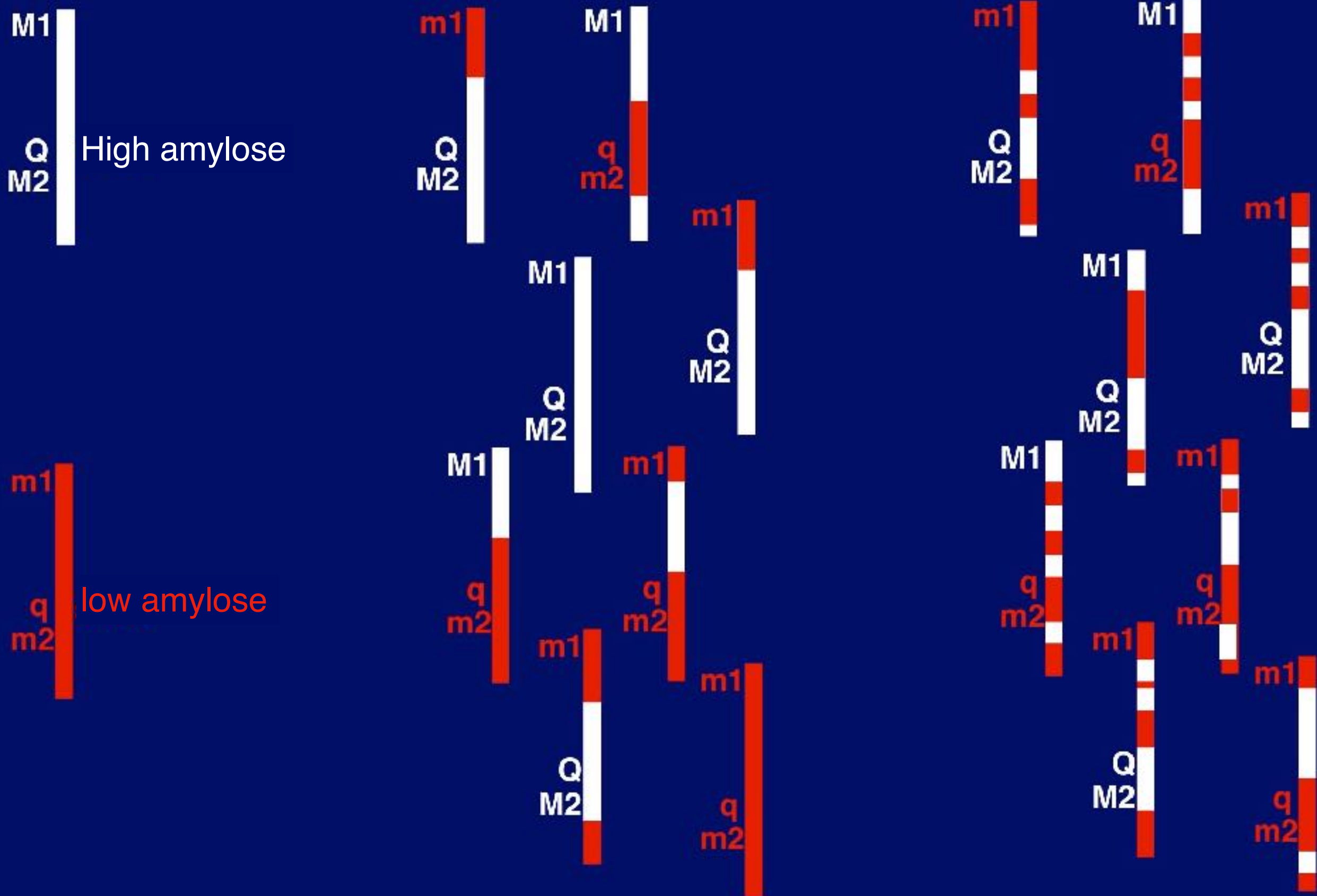
# Association mapping and historical recombination



# Association mapping and historical recombination



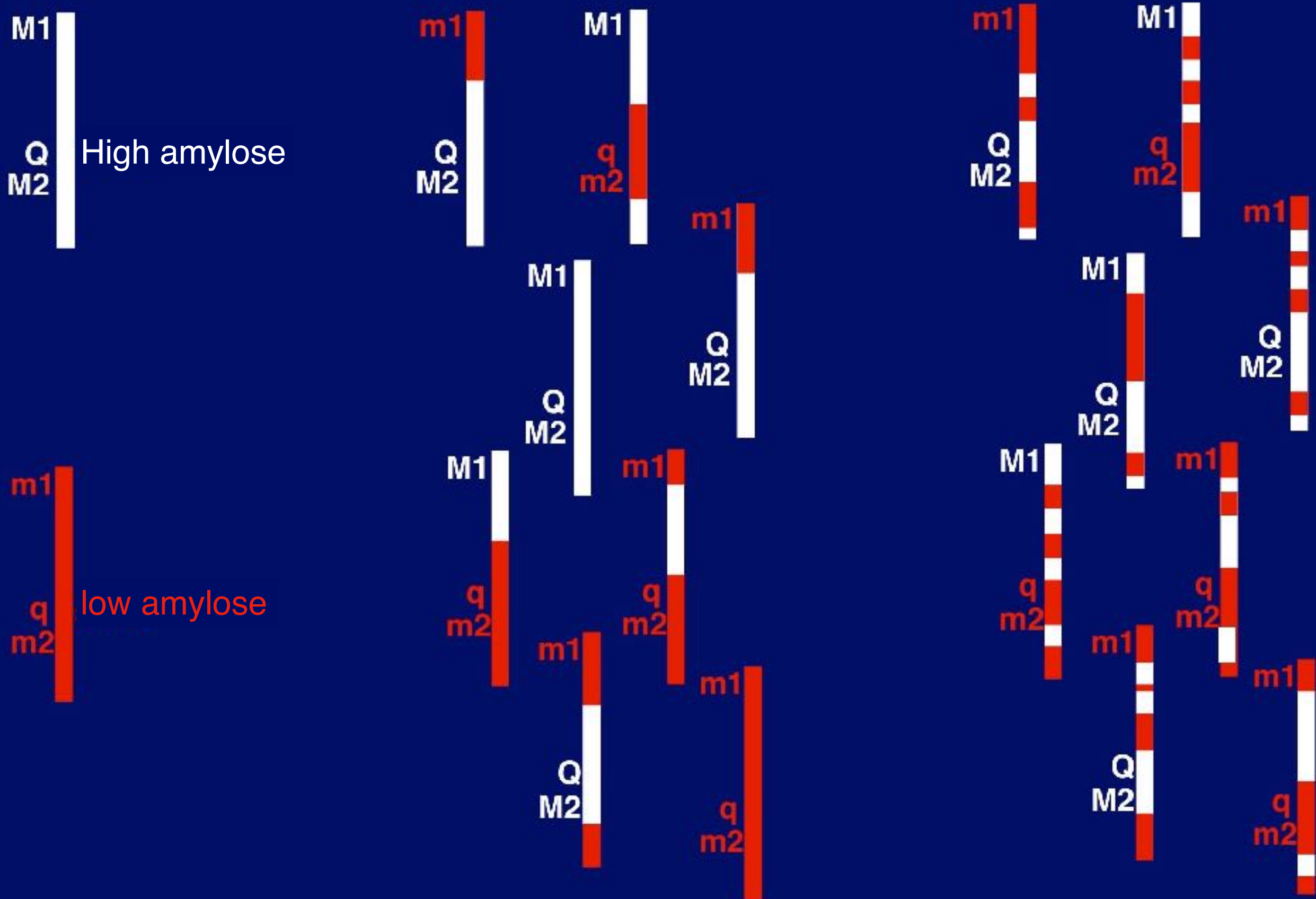
# Association mapping and historical recombination



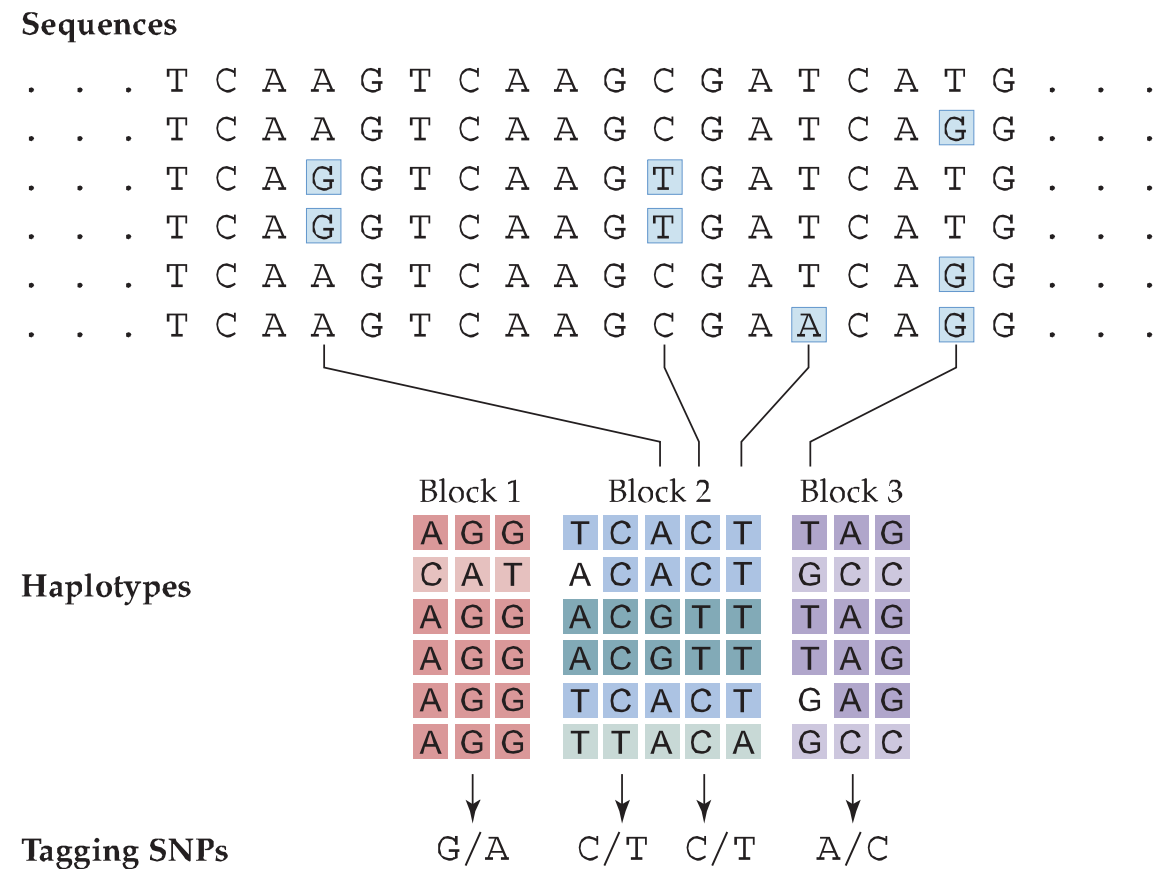
# The importance of tagSNPs

- Our rice SNP data set has ~36,000 SNPs.
- There are ~500,000 SNPs segregating among the rice varieties.
- Is it hopeless?
  - Do we have less than a 1 in 10 chance of finding an association because we are assaying less than 10% of the SNPs?
- Not hopeless
  - Because of linkage disequilibrium there is a strong correlation among closely linked SNPs

# Not hopeless: SNPs near to one another are correlated...



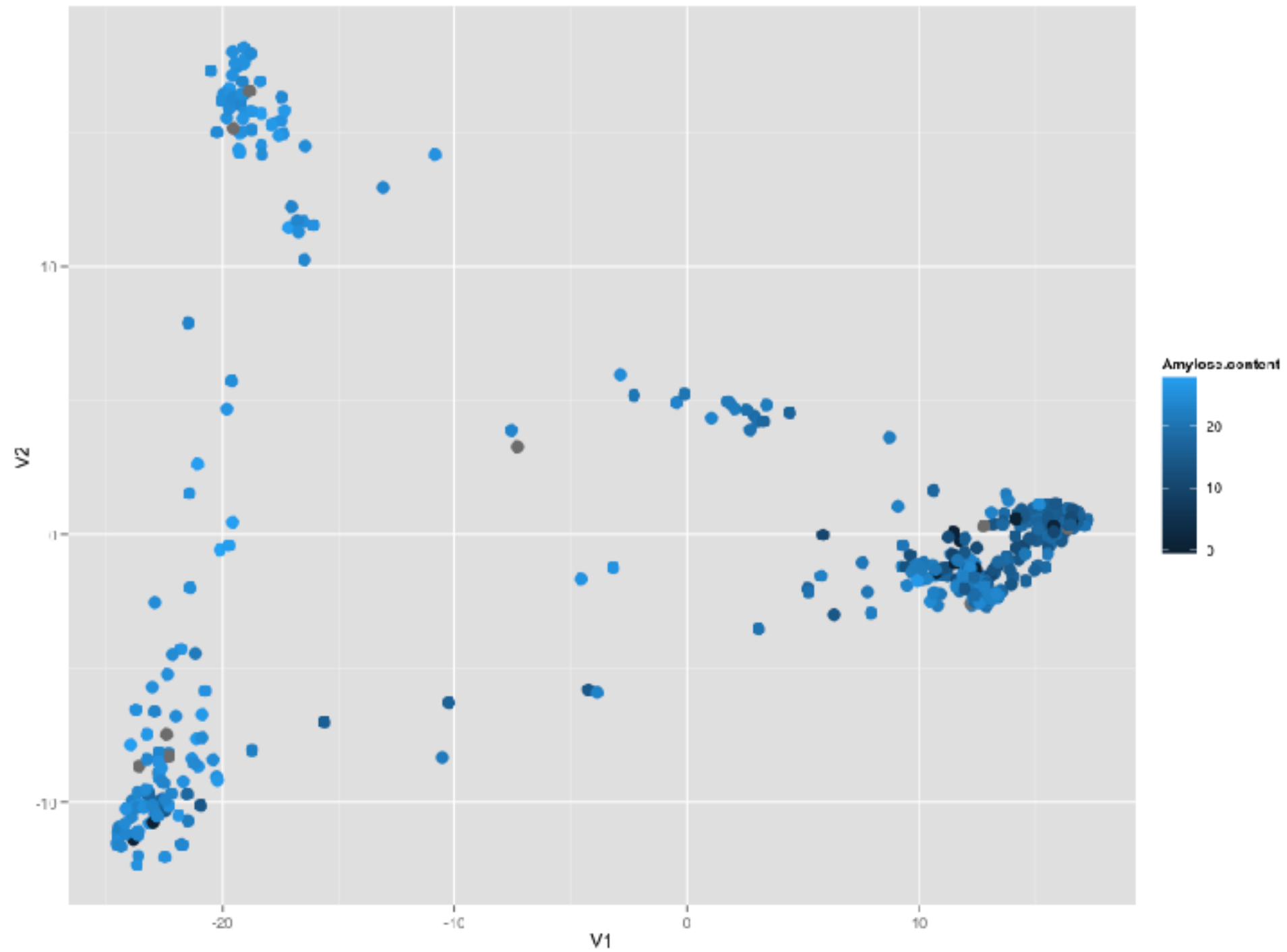
# Not hopeless: SNPs near to one another are correlated...



**Figure 3.5 Tagging SNPs and haplotype blocks.** Extraction of the polymorphisms from a set of sequences typically reveals a blocklike pattern of haplotypes. In this hypothetical example, Block 1 has two classes of haplotypes, one rare and one common; Block 2 has three classes of haplotypes; and Block 3 has two classes of haplotypes. Note that the boundaries between blocks are relatively sharp. The tagging SNPs can be used to define most of the variation in the sample.

# Population Structure can present a problem for GWAS

- What is the potential problem with a GWAS for amylose content?



## Population structure corrections

- Analyze within each population
- OR
- include structure information in the statistical model.
  - instead of:  $\text{amylose} \sim \text{SNPgenotype}$
  - use:  $\text{amylose} \sim \text{SNPgenotype} + \text{population\_membership}$
- Often it is best to include BOTH population membership and a kinship matrix (genetic relatedness). We will not use that method today (but checkout GAPIT)