

Identifying a diamond in the rough: studying allelic variation for complex traits in maize landraces

Jorge Alberto Romero Navarro

Buckler Lab

Cornell University



**International Maize and Wheat Improvement Center
(CIMMYT)**

- Sarah Hearne
- Martha Wilcox
- Juan Burgueño
- Samuel Trachsel
- Ivan Ortiz-Monasterio
- Félix San Vicente
- Peter Wenzl
- Gary Atlin



Instituto Nacional de Investigaciones
Forestales, Agrícolas y Pecuarias

**Instituto Nacional de Investigaciones Forestales Agrícolas y
Pecuarias (INIFAP)**

- Ernesto Preciado
- Arturo Terron
- Victor Vidal
- Humberto Vallejo Delgado
- Alejandro Ortega
- Noel Orlando Gómez Montiel



Cornell University

Cornell University

- Edward Buckler
- Cinta Romay
- Eduardo Carrillo
- Kelly Swarts



Native varieties contain plenty of useful alleles...



each linked to hundreds of undesired alleles

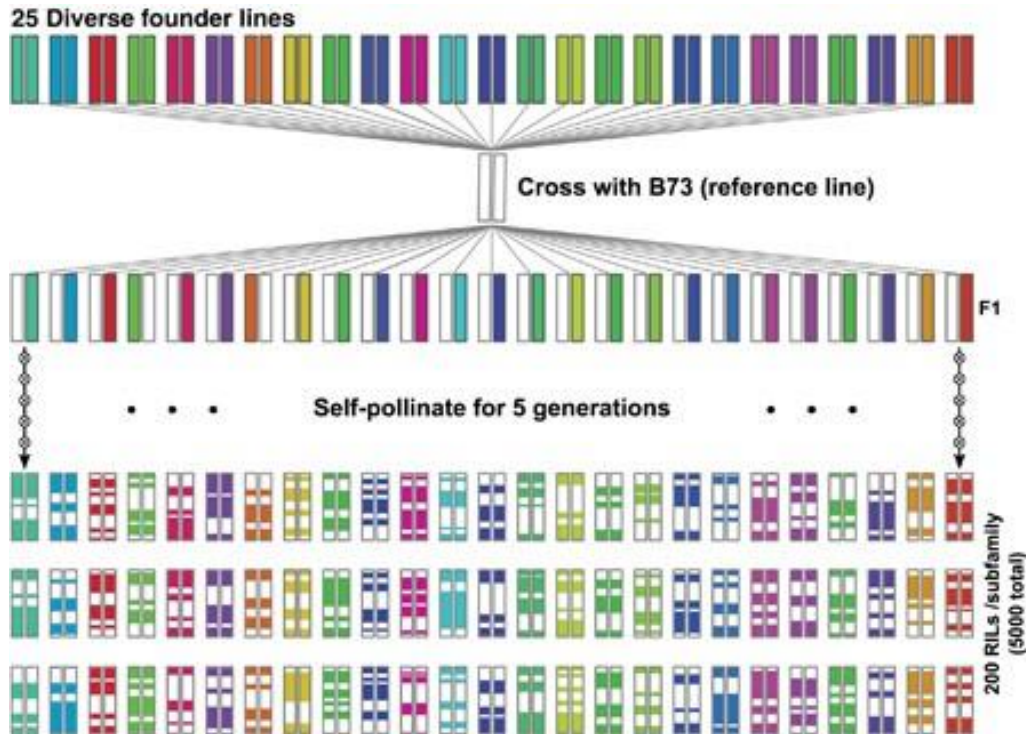


How can we identify the diamond in the rough?

- 1989-97 Latin American Maize Project (LAMP) 11 countries characterize CIMMYT's germplasm
 - ü Description of adaptation patterns
 - ü Definition of a Breeder's core collection
 - ✗ Genetic drag = limited impact
 - ✗ Expensive
 - ✗ Time consuming
 - ✗ Marker assisted selection unavailable



Association mapping helps identify and deploy useful variation



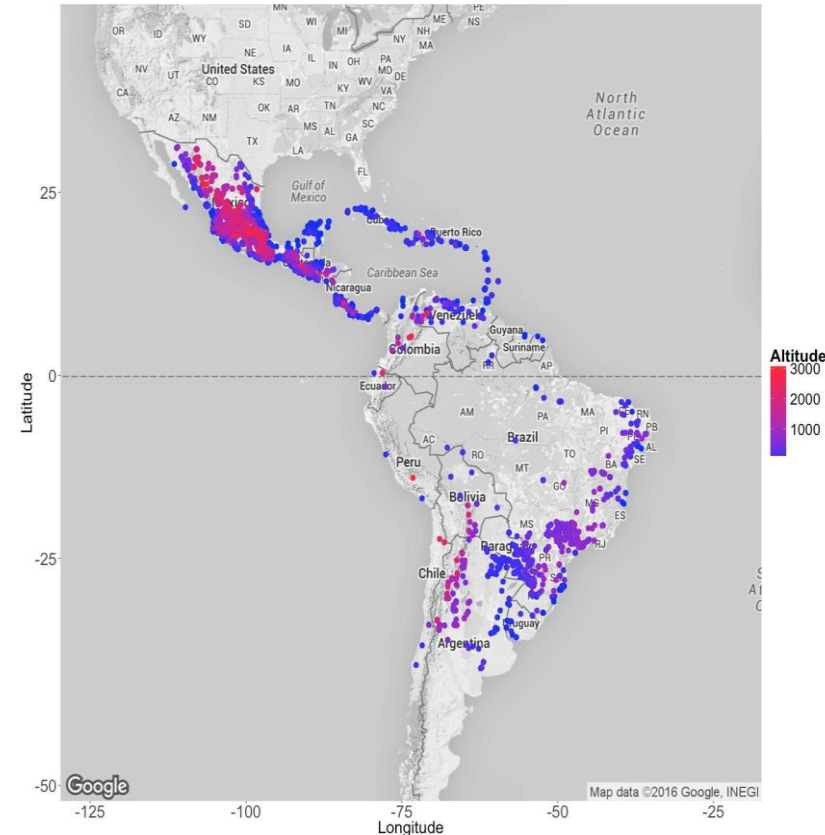
Buckler, et al, 2009



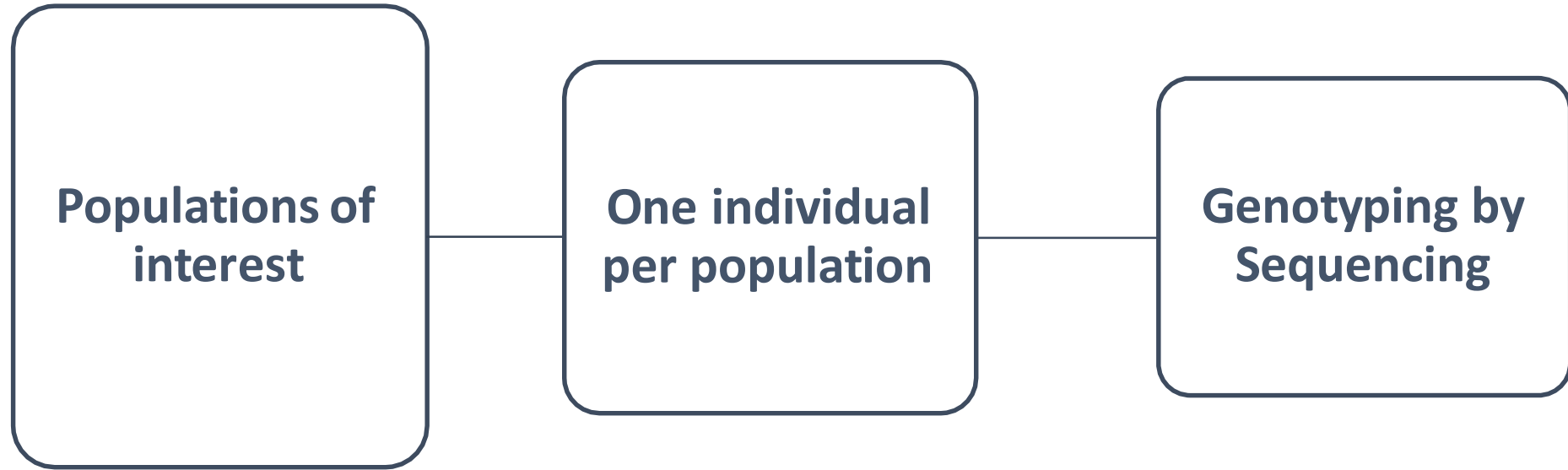
Harvestplus

How do you evaluate a large and diverse landrace panel?

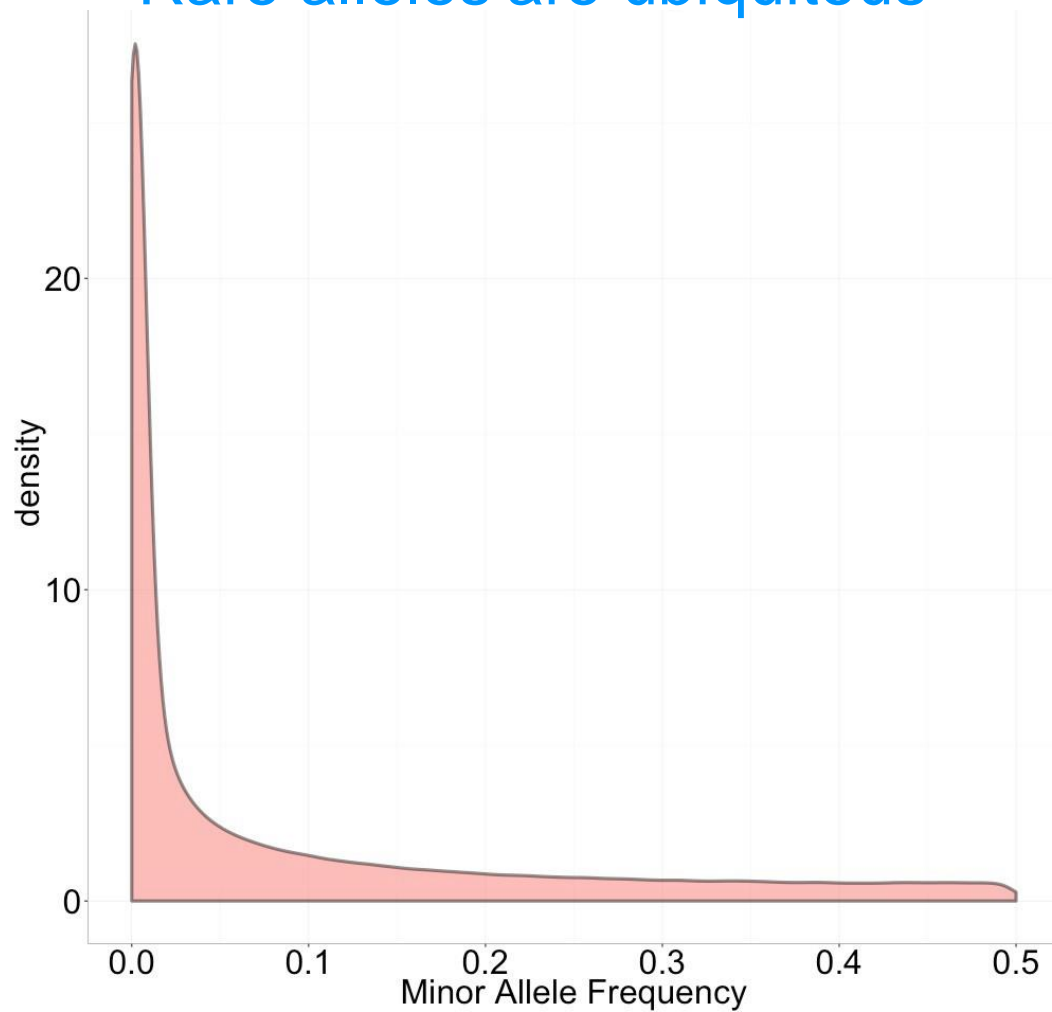
- Breeder's core collection:
4,500 CIMMYT landrace accessions
- Adaptation:
 - **Tropical** (low elevation)
 - **Subtropical** (mid-elevation)
 - **Highland**

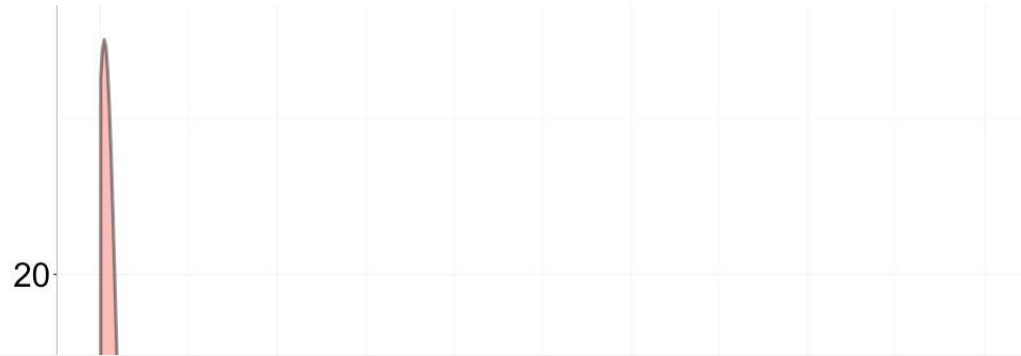


New approach for fast identification and deployment of useful alleles from landraces

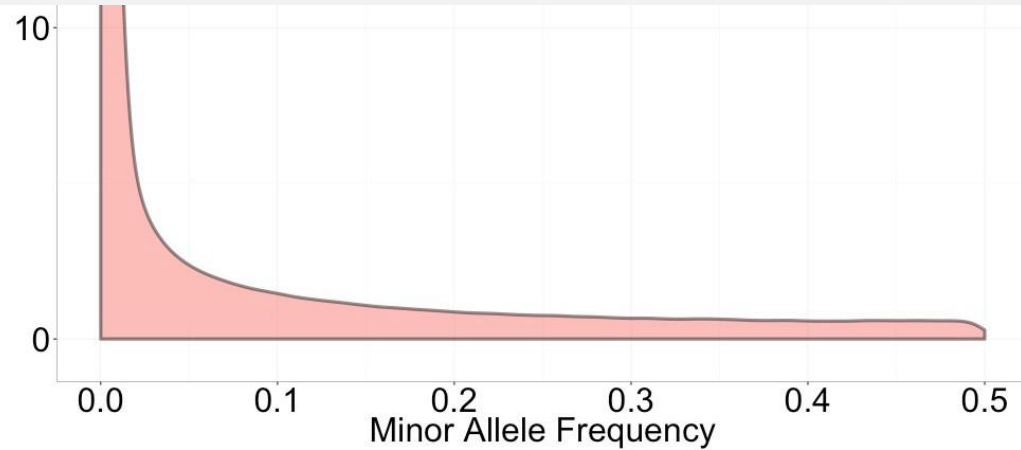


Rare alleles are ubiquitous





Which are the useful alleles?



We can use GWAS for adaptation to local environmental conditions

- Landraces have been selected for local adaptation for thousands of generations
- Passport information has been carefully collected and curated for decades for thousands of samples (and we get both amazing resources for free!)
- Example equation for environmental GWAS :
 - Average precipitation during growing season
 - **Precipitation = Altitude + Latitude + Longitude + SNP effect + ϵ**

We can use GWAS for adaptation to local environmental conditions

- Landraces have been selected for local adaptation for thousands of generations
- Passport information has been carefully collected and curated for decades for thousands of samples (and we get both amazing resources for free!)
- Example equation for environmental GWAS :
 - Average precipitation during growing season
 - **Precipitation = Altitude + Latitude + Longitude + SNP effect + ϵ**



Adaptation

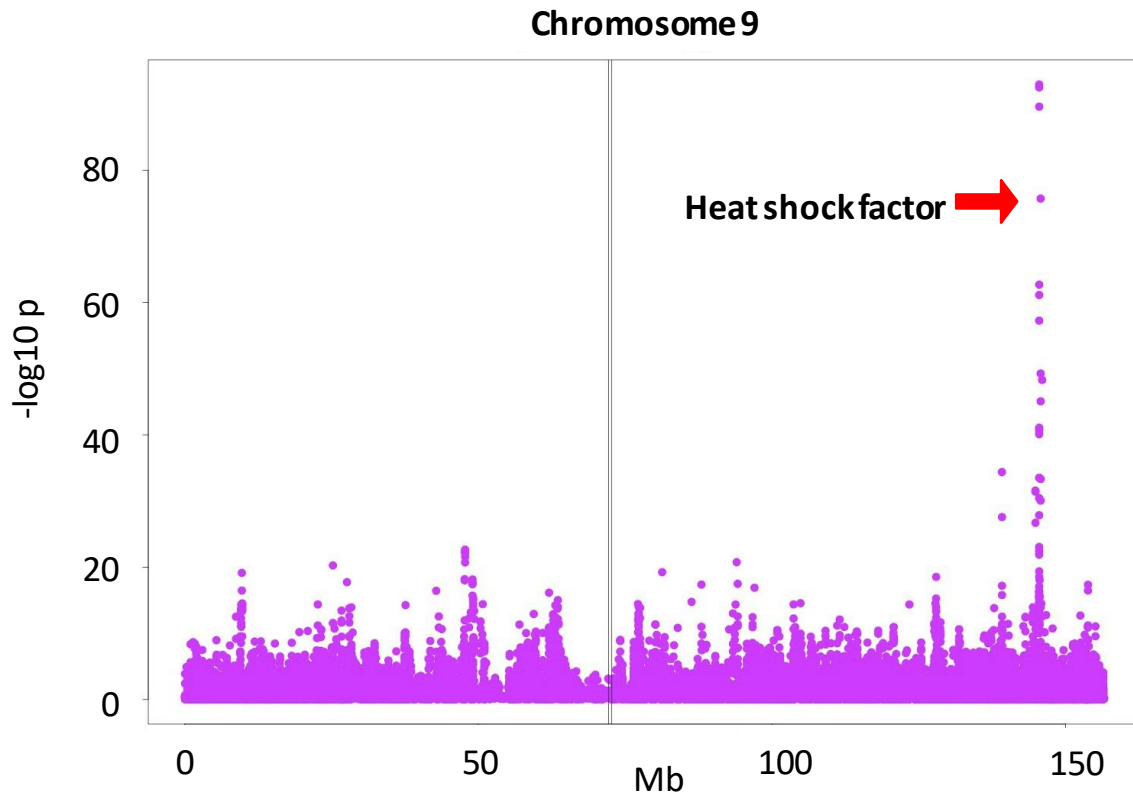
Account for isolation by distance

Estimate genotype effect

Using passport data we can perform high resolution GWAS for adaptation to climate

- Environmental traits

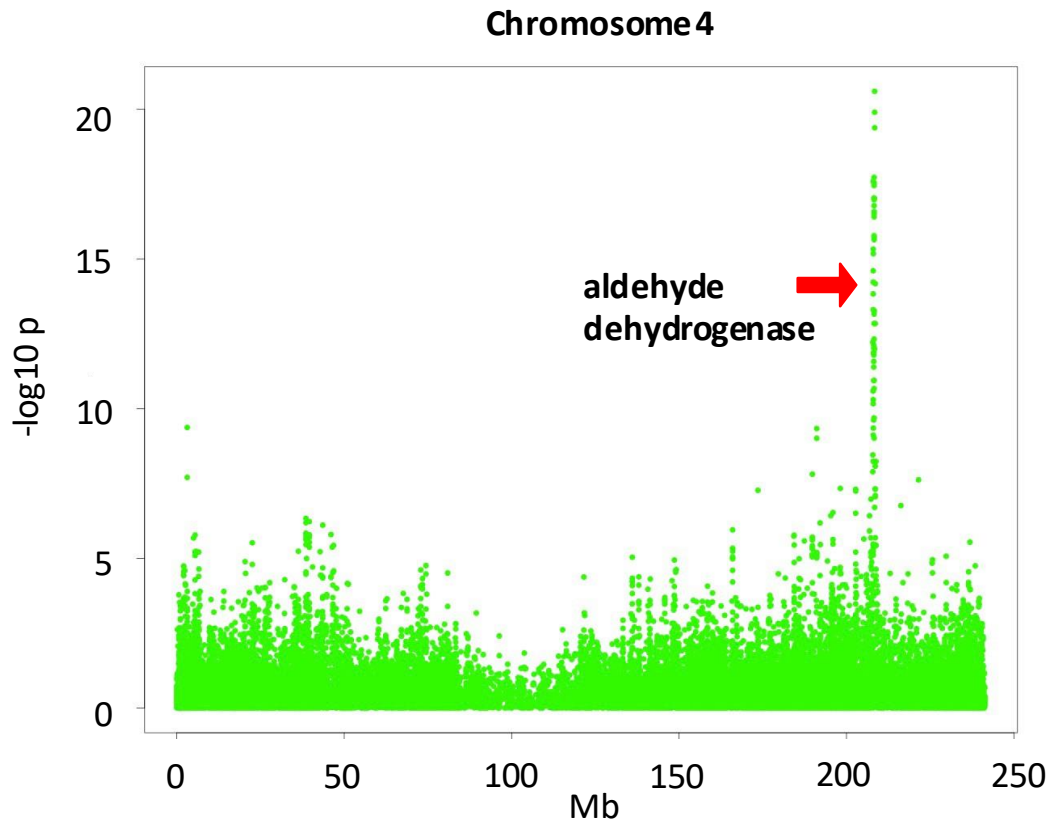
- Altitude
- Latitude/Longitude
- **Precipitation**
- Temperature
- Cloud cover
- Frost frequency
- Vapor pressure
- Aridity
- Waterlog
- pH



Using passport data we can perform high resolution GWAS for adaptation to soils

- Environmental traits

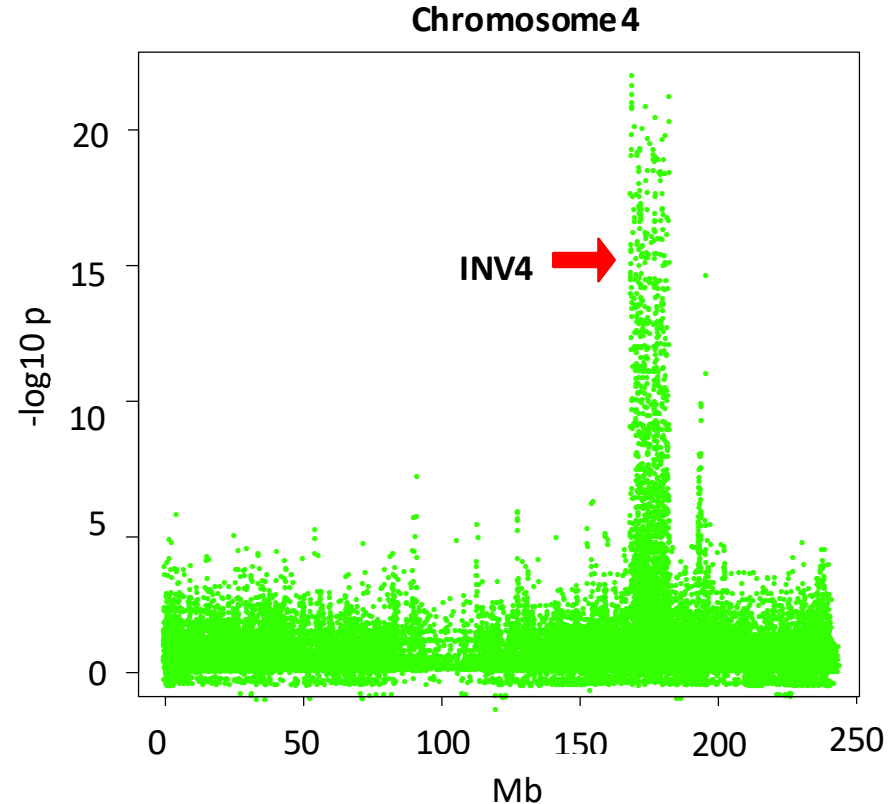
- Altitude
- Latitude/Longitude
- Precipitation
- Temperature
- Cloud cover
- Frost frequency
- Vapor pressure
- Aridity
- Waterlog
- **pH**



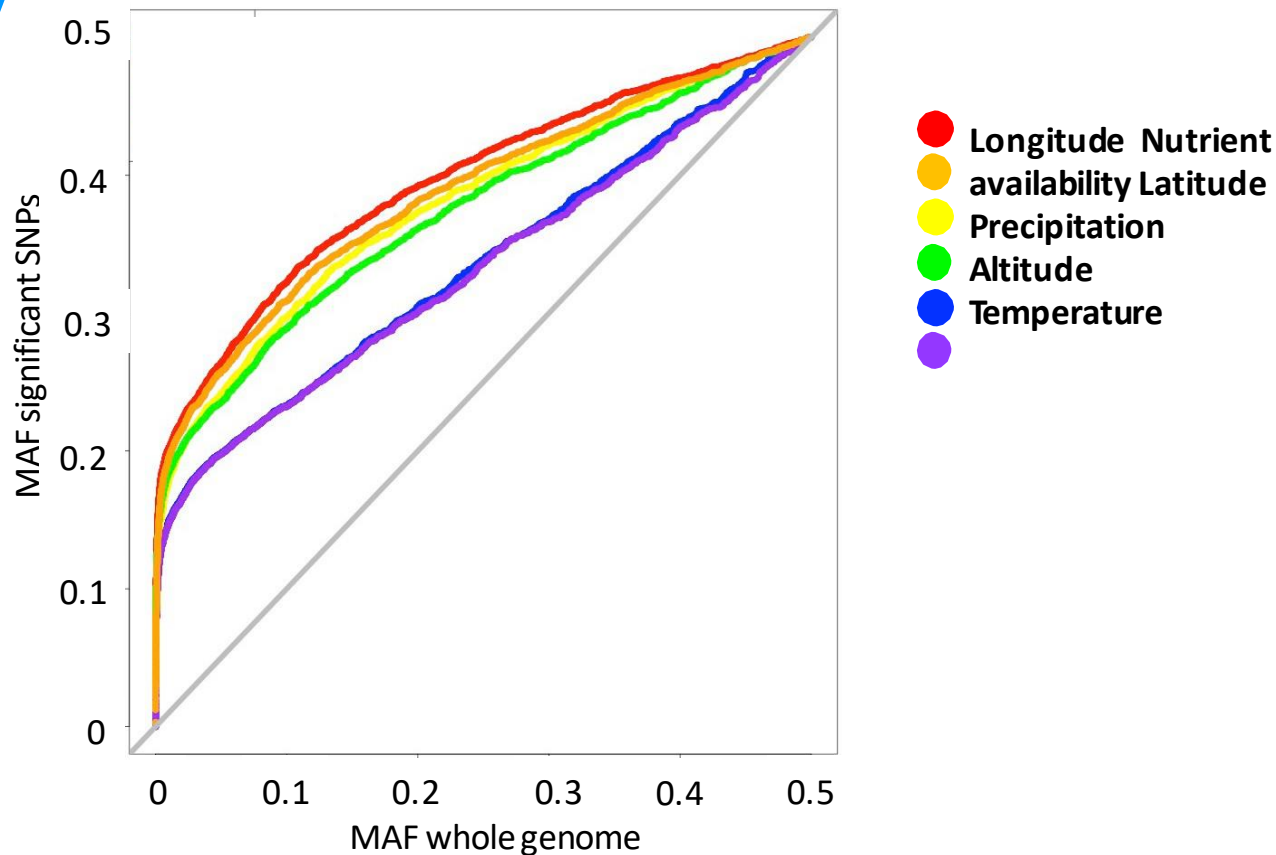
Using passport data we can perform high resolution GWAS for adaptation to geography

- Environmental traits

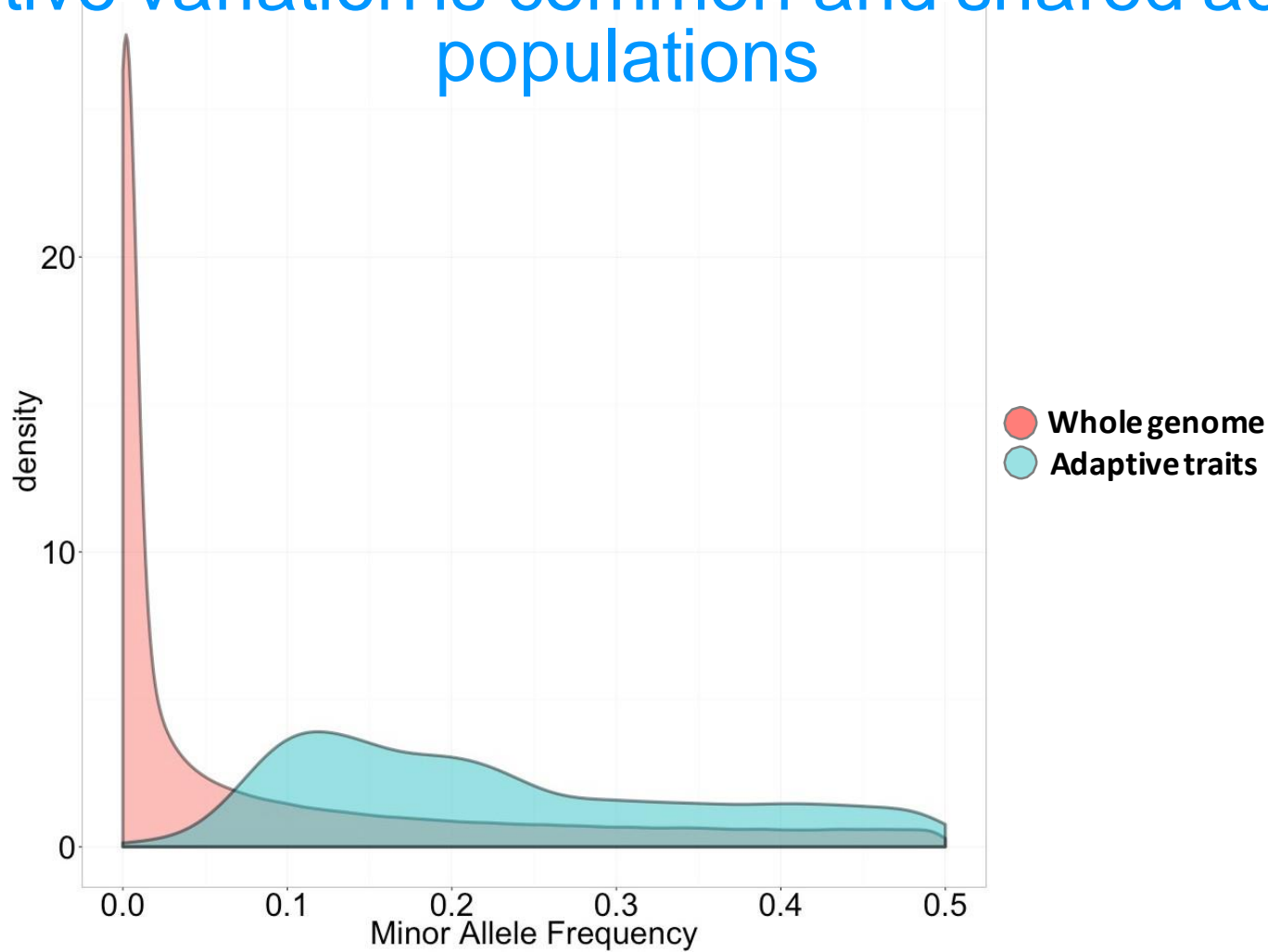
- **Altitude**
- Latitude/Longitude
- Precipitation
- Temperature
- Cloud cover
- Frost frequency
- Vapor pressure
- Aridity
- Waterlog
- pH

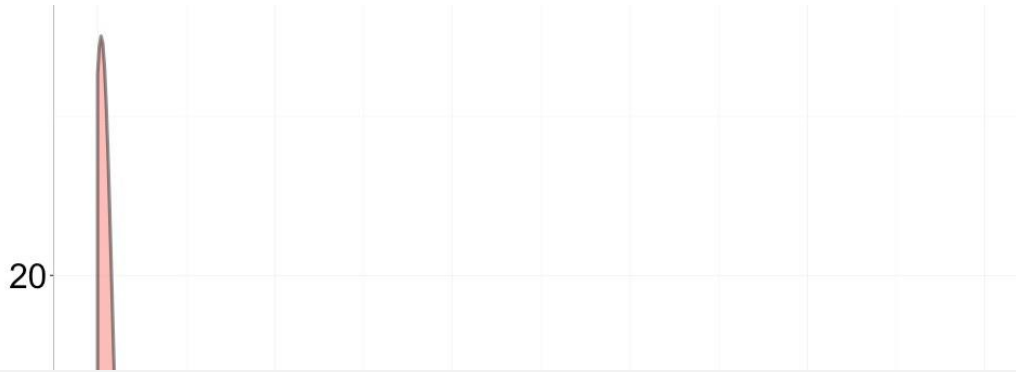


SNPs associating with adaptation are in high frequency

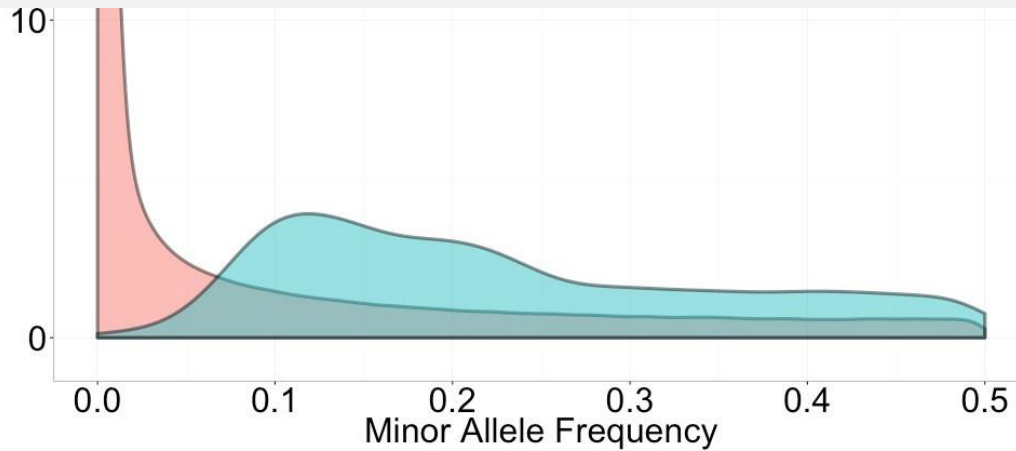


Adaptive variation is common and shared across populations

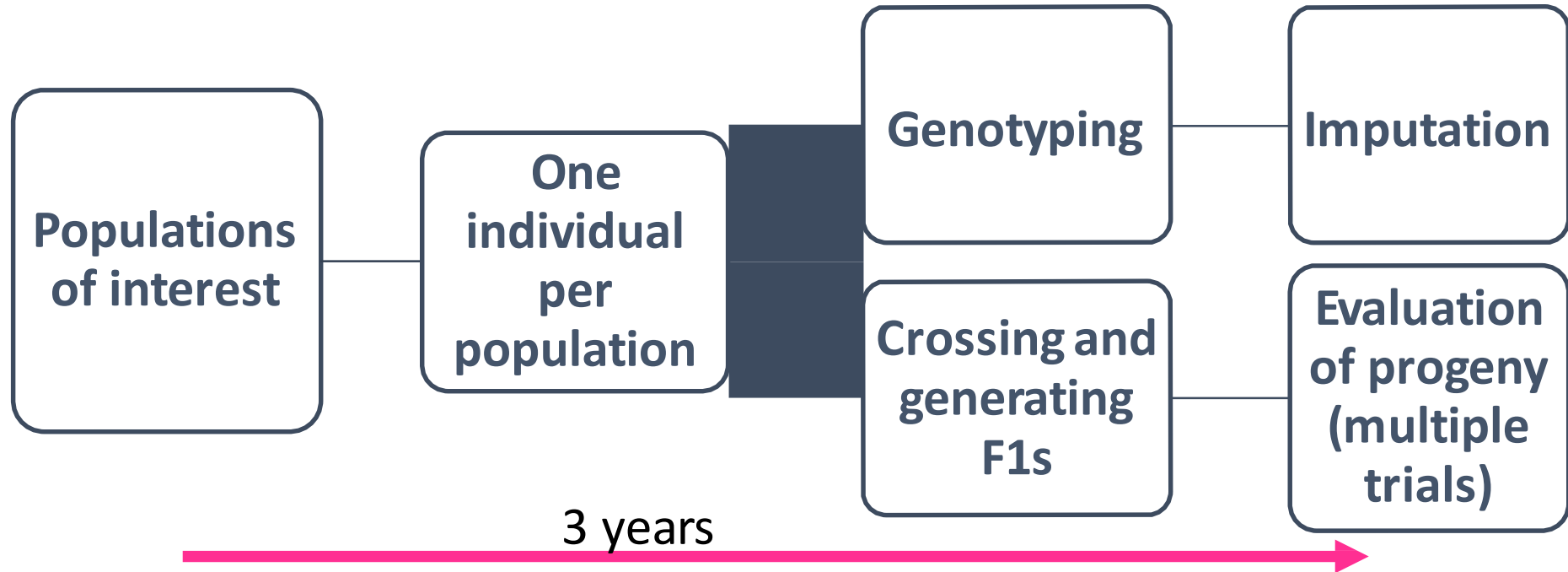




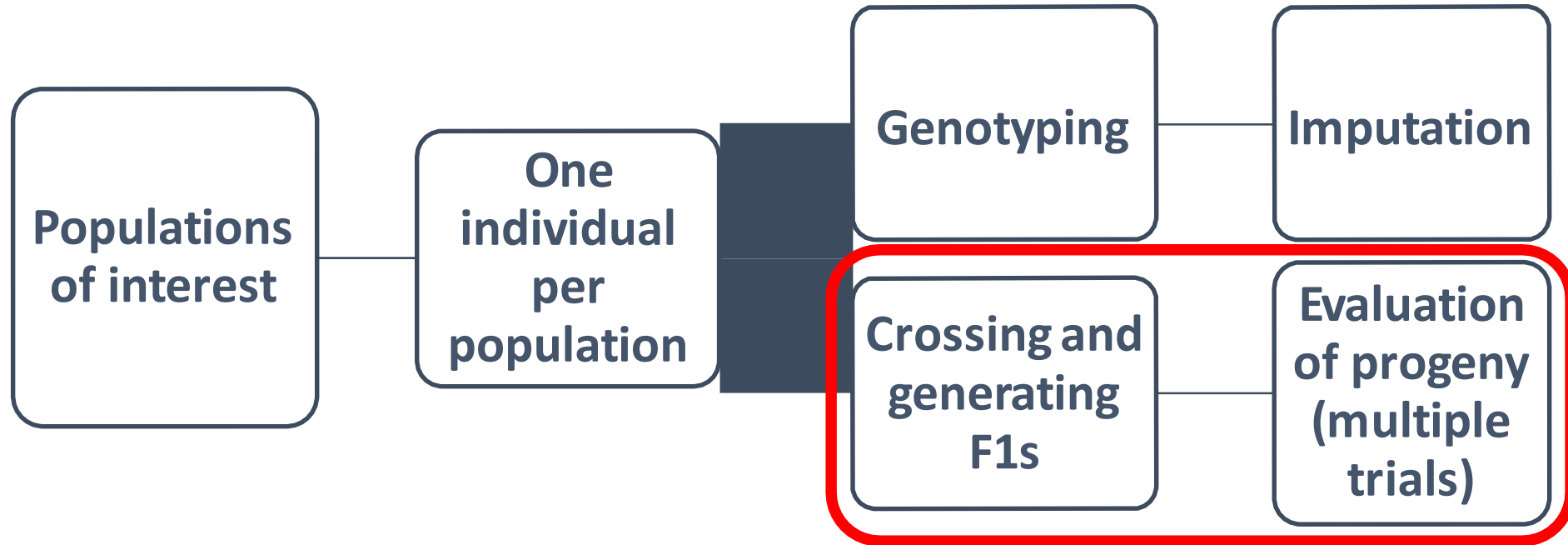
Can we do genome wide association
for complex traits in landraces?



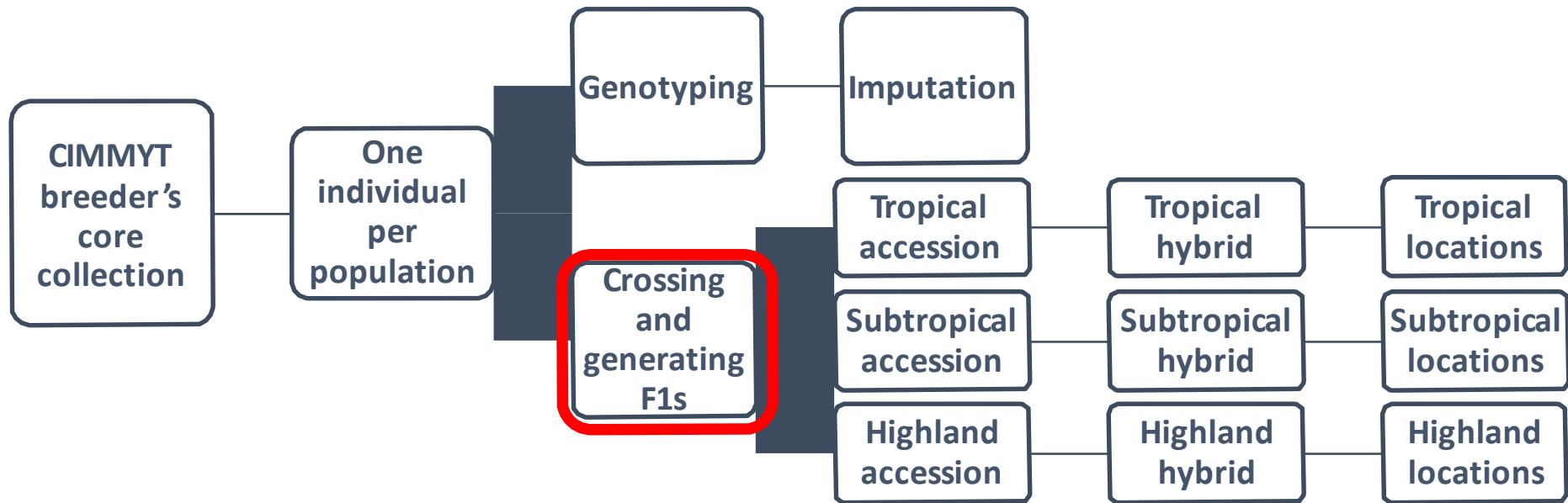
F-One Association Mapping (FOAM) design allows fast evaluation



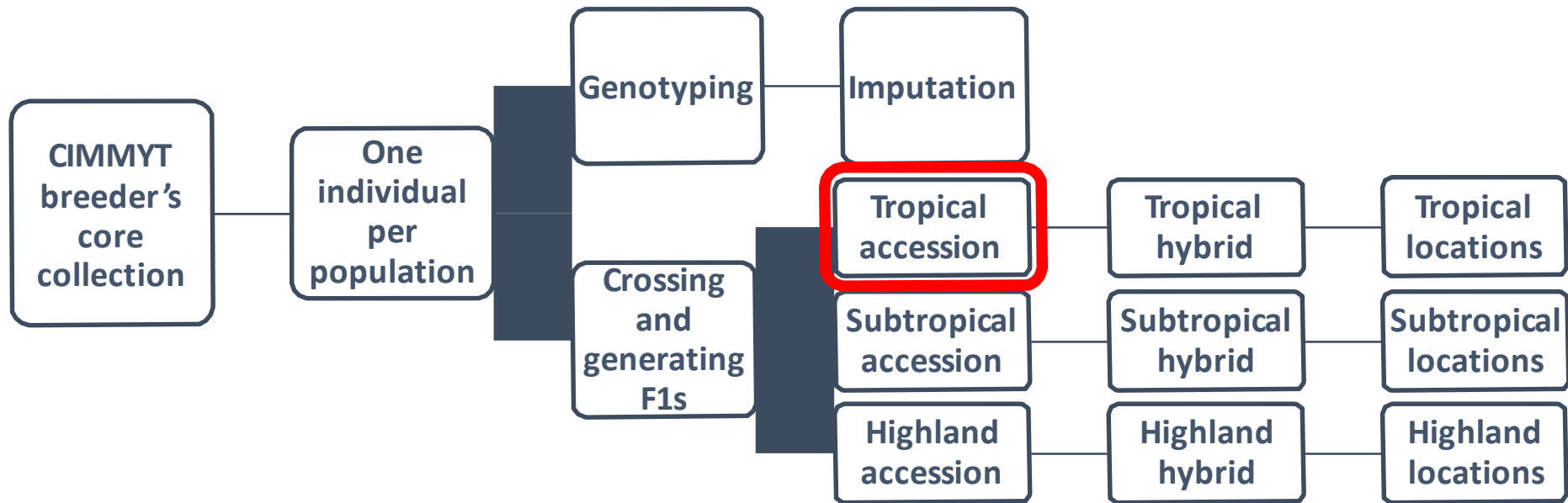
Custom FOAM design nests adaptation class
in order to do accurate evaluation



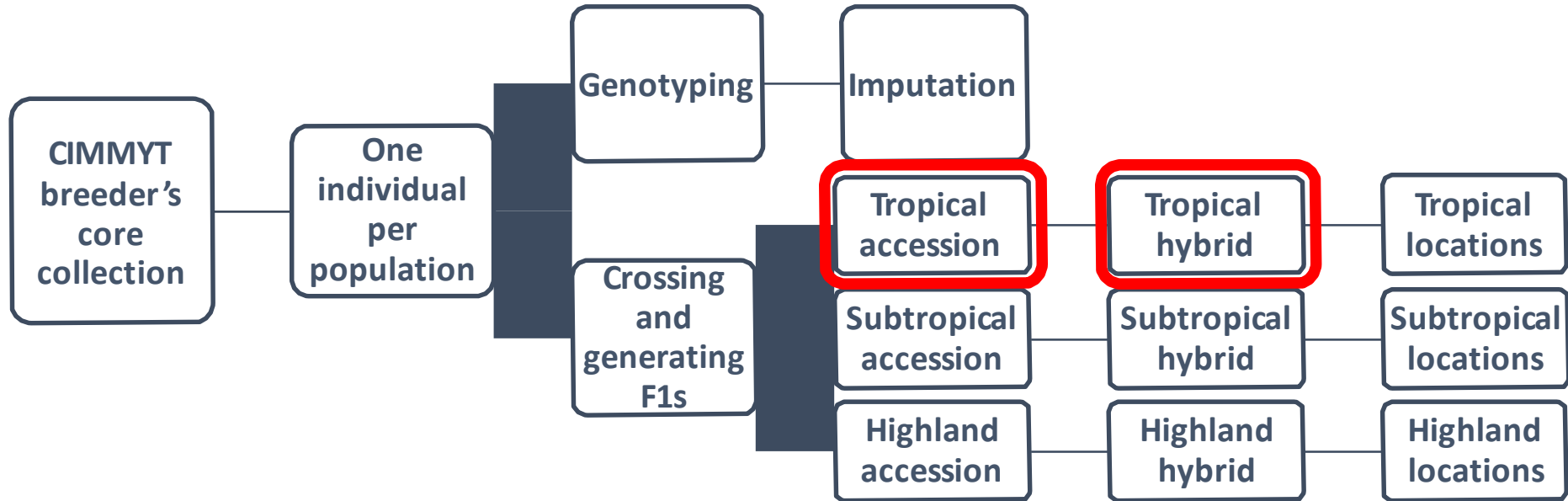
Custom FOAM design nests adaptation class



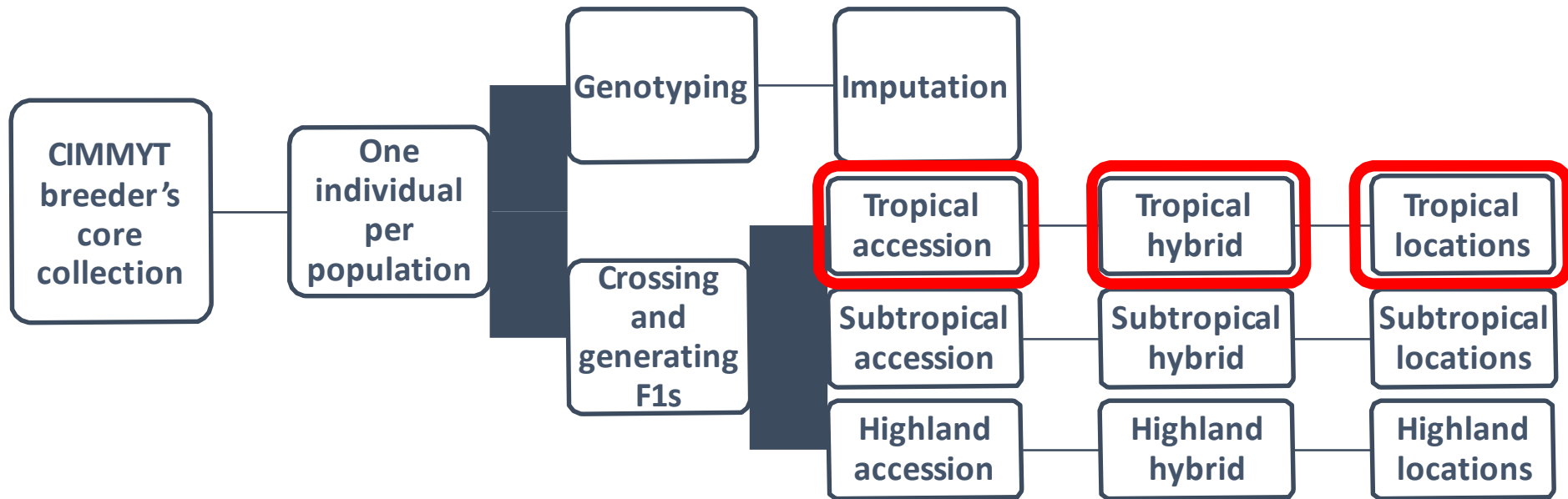
Custom FOAM design nests adaptation class



Custom FOAM design nests adaptation class



Custom FOAM design nests adaptation class





30 trials used to estimate ~19,000 breeding values



Martha Wilcox

Juan Burgueño

- Trial design allowed for a large number of accessions per location
 - Extended row-column
- Each trial contains a different subset of the entire collection
 - Lack of balanced replication sacrifices GxE effect estimation



GWAS mixed model for complex traits in FOAM design

$$\text{Precipitation} = \text{Altitude} + \text{Latitude} + \text{Longitude} + \text{SNP effect} + \epsilon$$

Adaptation

Account for isolation by distance

Estimate genotype
effect

GWAS mixed model for complex traits in FOAM design

$$\text{Precipitation} = \text{Altitude} + \text{Latitude} + \text{Longitude} + \text{SNP effect} + \epsilon$$



Adaptation



Account for isolation by distance



Estimate genotype effect

$$\text{Flowering time} = \text{trial} + \text{hybrid parent} + \text{PCs} + \text{kinship matrix} + \text{SNP effect}_{[\text{trial}]} + \epsilon$$



Breeding values



Main effects

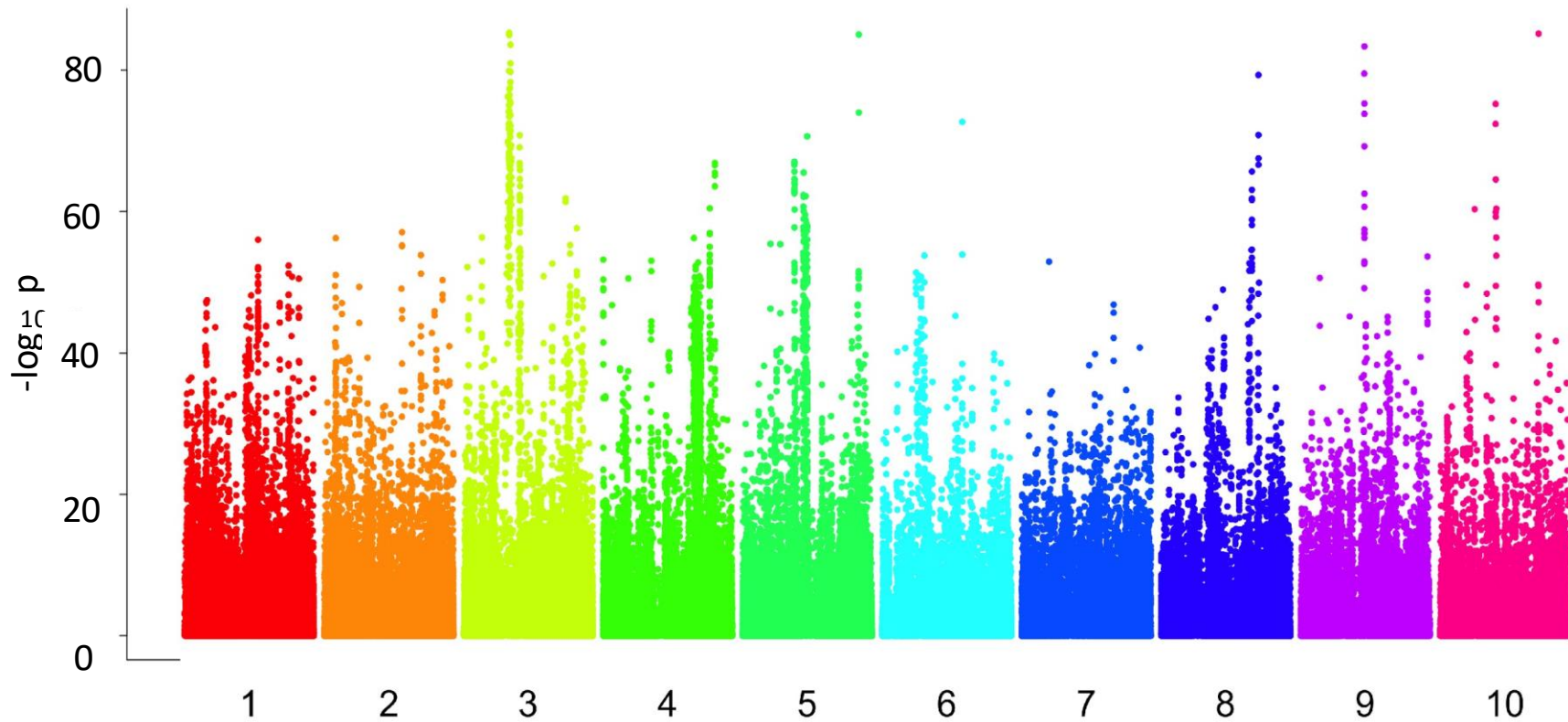


Fixed effect: Population structure
Random effect: relatedness



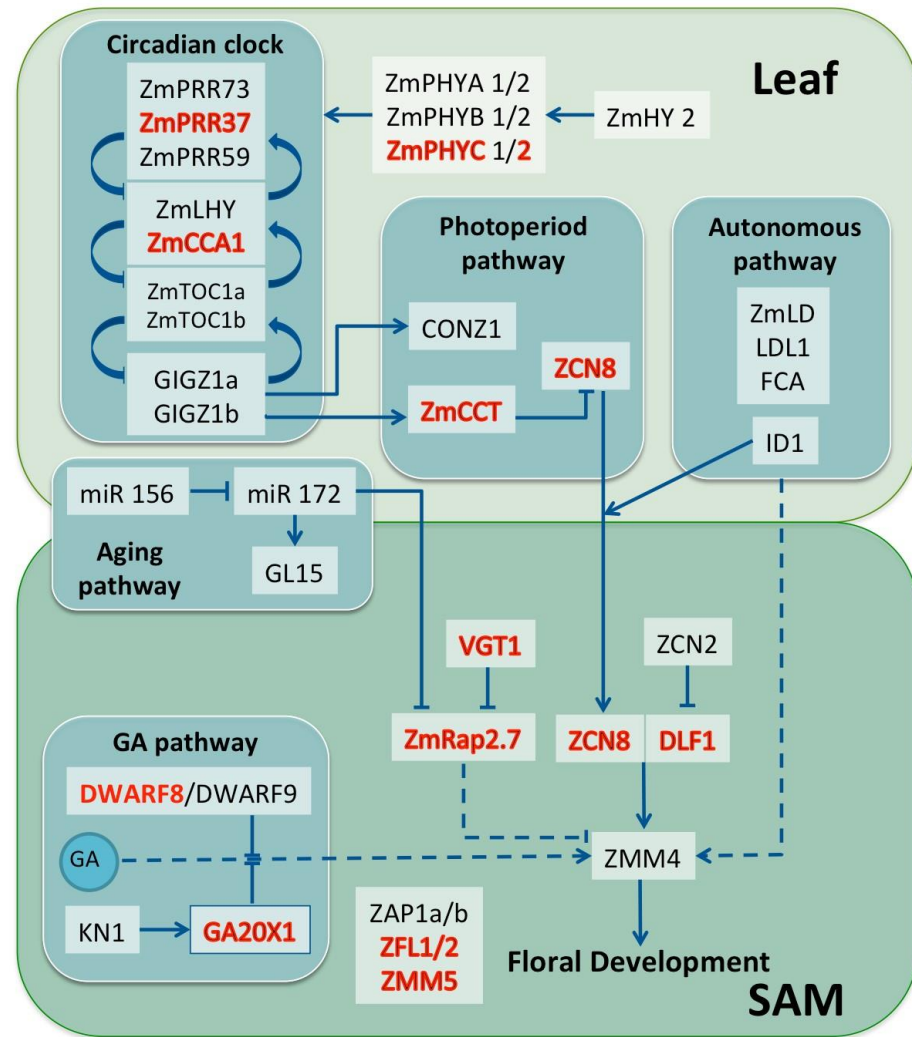
Estimate nested genotype effect

Flowering time GWAS hits 883 genes



Enrichment at candidate desi

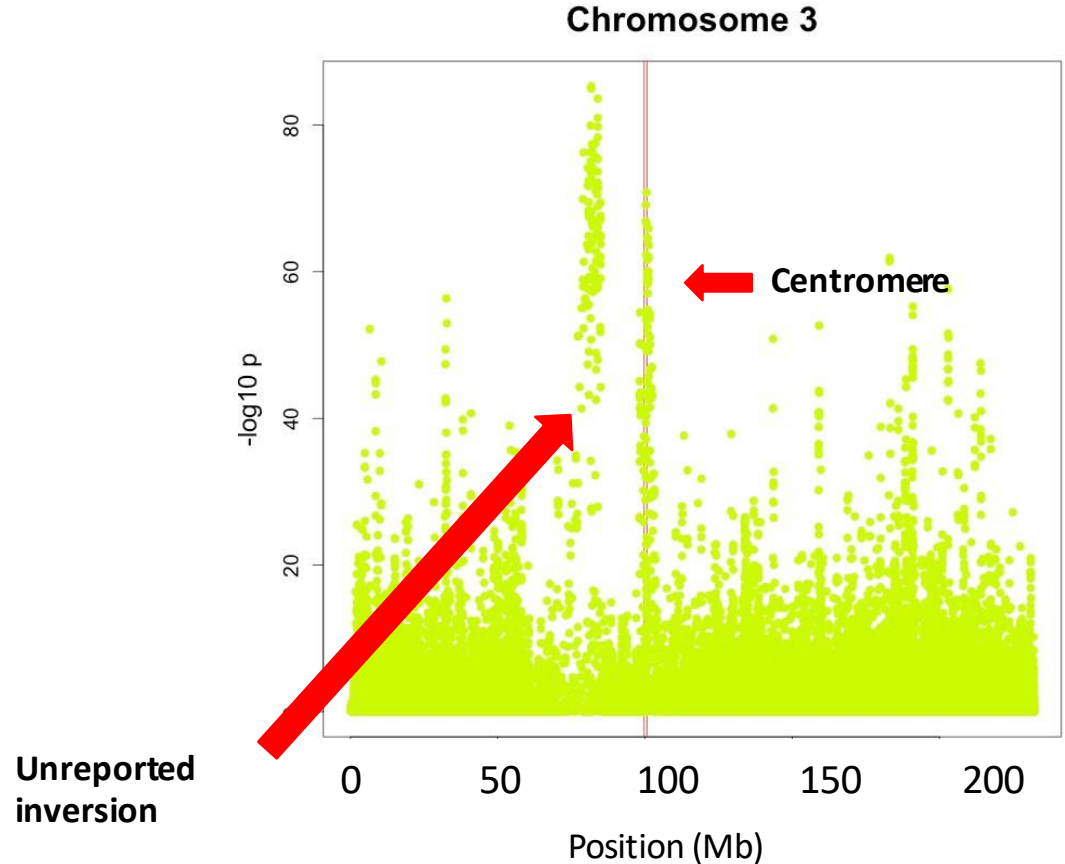
- High overlap between male and female flowering time control
- Results include association at ZCN8, VGT1, and *dwarf8* region



Modified from Dong *et al*, 2012

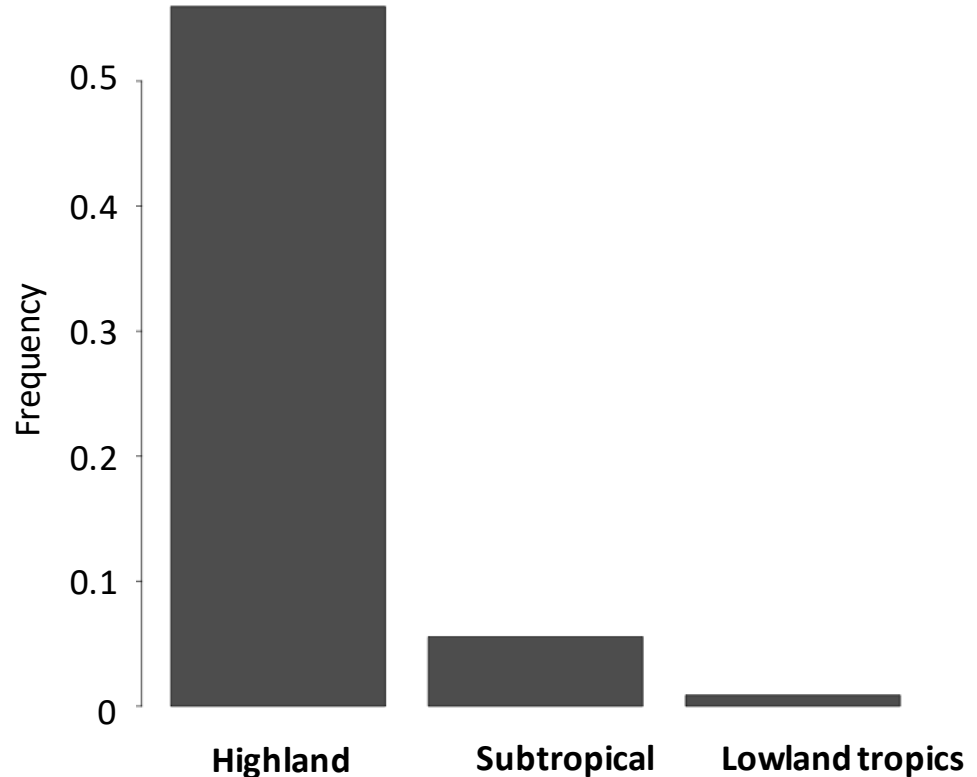
Structural variants account for 30% of explainable variance

- Centromeres 3,5 and 6
- Putative unreported inversion on chromosome 3



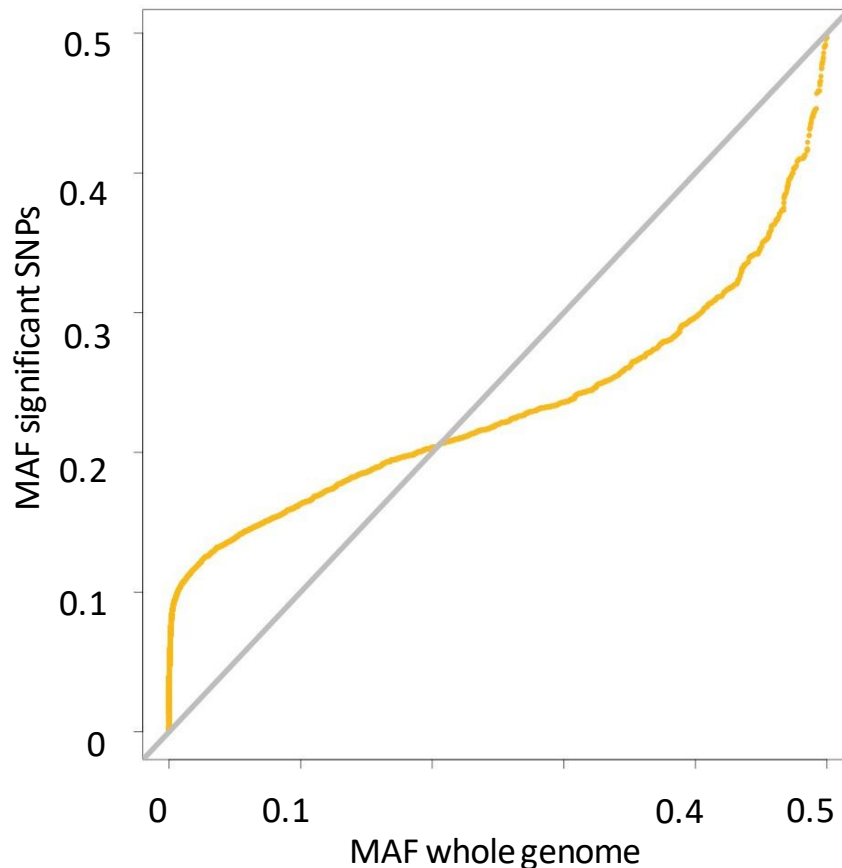
Structural variants account for 30% of explainable variance

- Centromeres 3,5 and 6
- Putative unreported inversion on chromosome 3
- INV4
 - -3 days effect
 - Adaptation dependent frequency
 - Fixed in several tropical highland improved lines

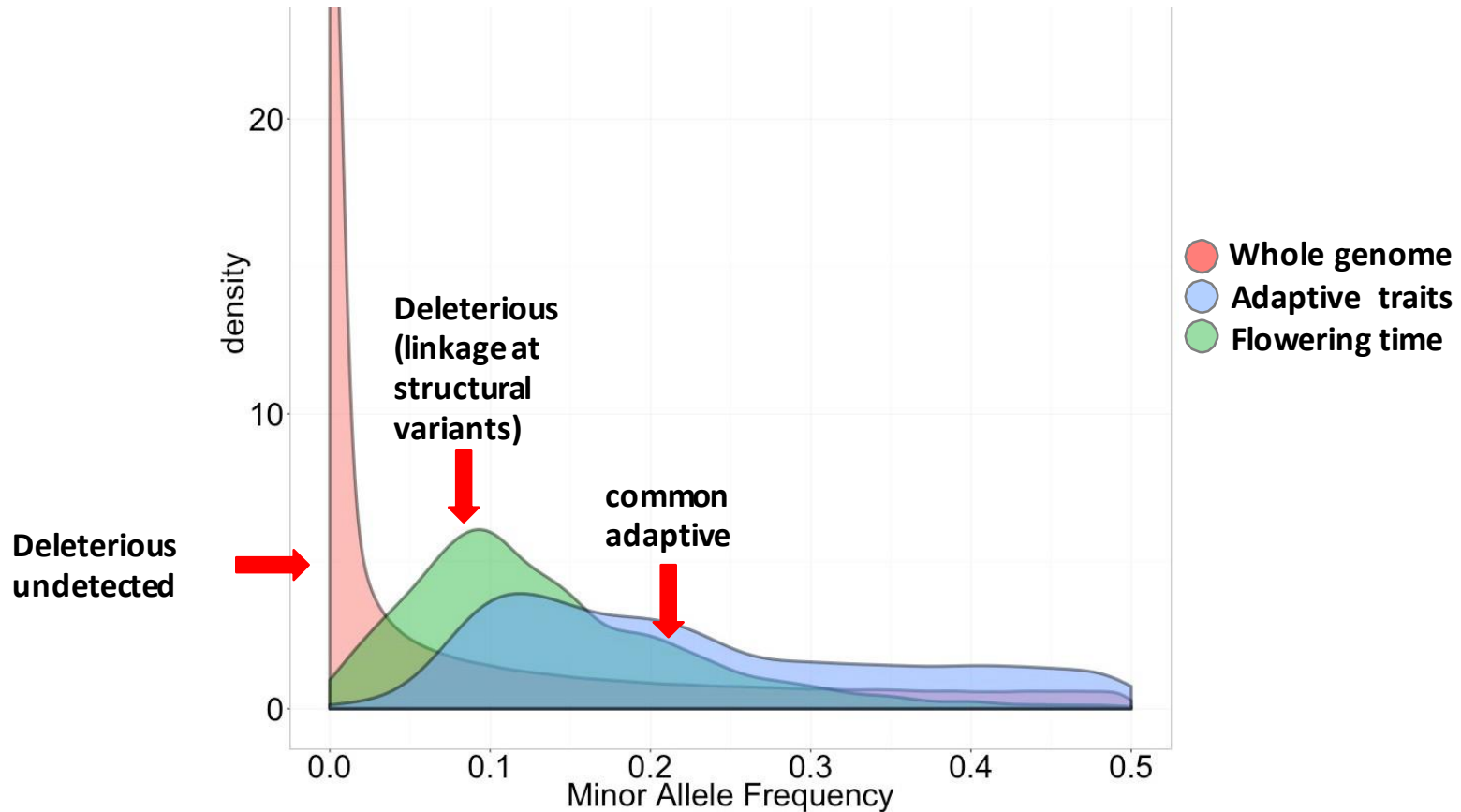


Most flowering time associating SNPs are in low frequency

- Common genes-uncommon variants
- Enrichment with environmental GWAS
 - 4x temperature and altitude
 - 3.5x precipitation

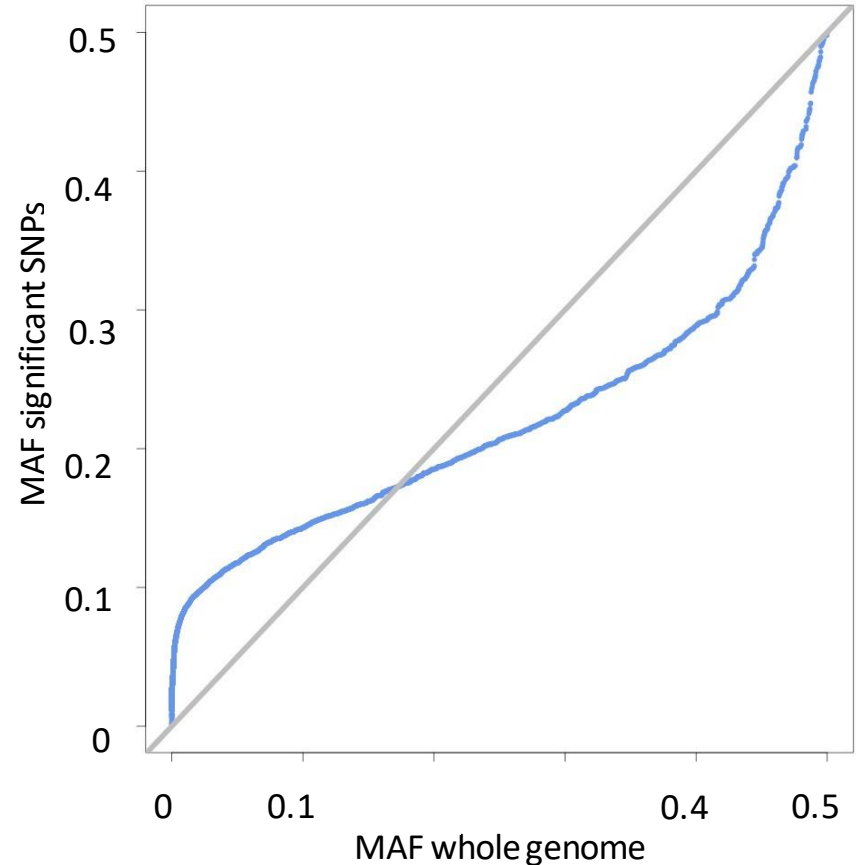


Flowering time control mixture of common adaptive and uncommon alleles

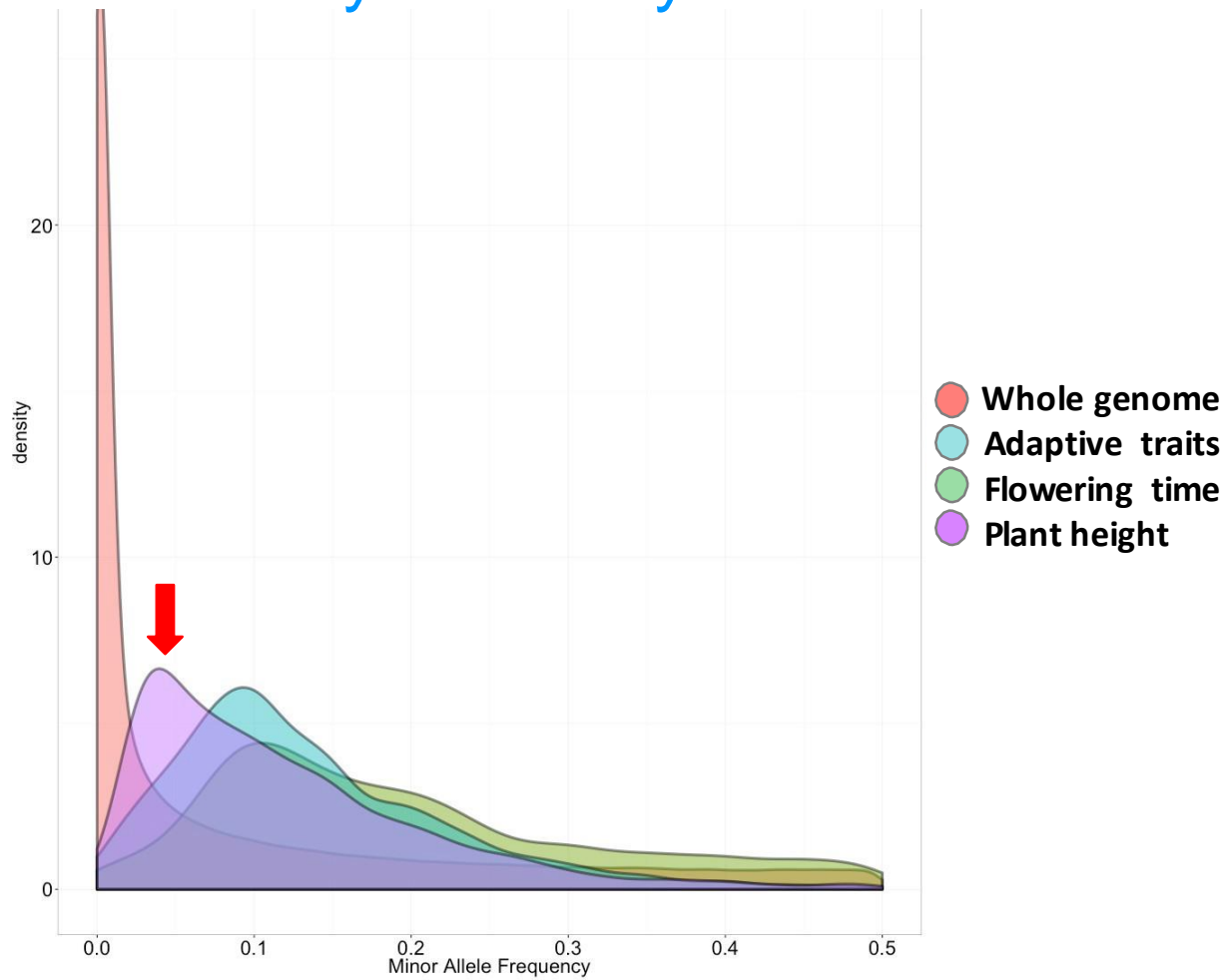


Plant height GWAS

- 1046 maize orthologs of *Arabidopsis* genes involved in hormonal pathways (auxin, gibberellin, brassinosteroids, abscisic acid, jasmonic acid, salicylic acid, ethylene, cytokinins)
 - 44% enrichment
- Genes associated with flowering time in Seed
 - Three-fold enrichment



Height control is mostly driven by uncommon alleles



Summary

- FOAM (F-One Association Mapping) population design
 - Advantages: Fast, comprehensive and cost effective
 - Considerations: Adaptation and lack of replication limit GxE estimation
- Structural variation has a significant effect on local adaptation and quantitative trait variation
- GWAS
 - Adaptive alleles are in high frequency and shared across populations
 - Deleterious alleles are rare and distinct across populations

FOAM GWAS allows fast identification of useful alleles...



we can break linkage with undesired alleles



**International Maize and Wheat Improvement Center
(CIMMYT)**

- Sarah Hearne
- Martha Wilcox
- Juan Burgueño
- Samuel Trachsel
- Ivan Ortiz-Monasterio
- Félix San Vicente
- Peter Wenzl
- Gary Atlin



Instituto Nacional de Investigaciones
Forestales, Agrícolas y Pecuarias

**Instituto Nacional de Investigaciones Forestales Agrícolas y
Pecuarias (INIFAP)**

- Ernesto Preciado
- Arturo Terron
- Victor Vidal
- Humberto Vallejo Delgado
- Alejandro Ortega
- Noel Orlando Gómez Montiel



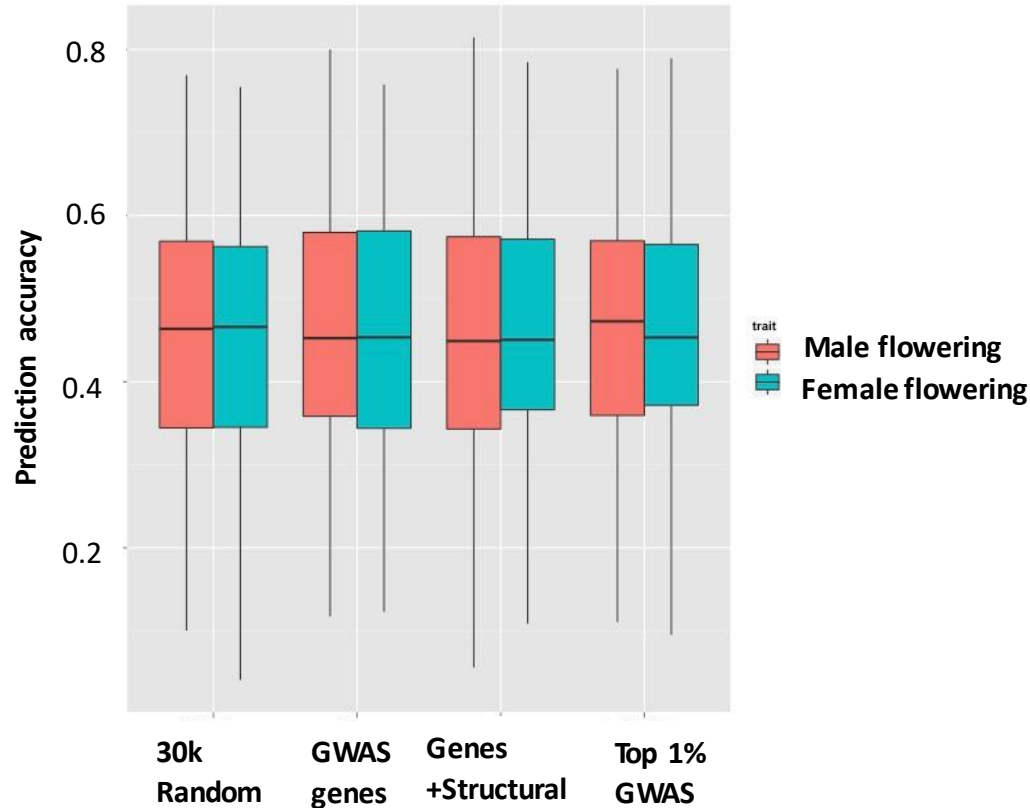
Cornell University

Cornell University

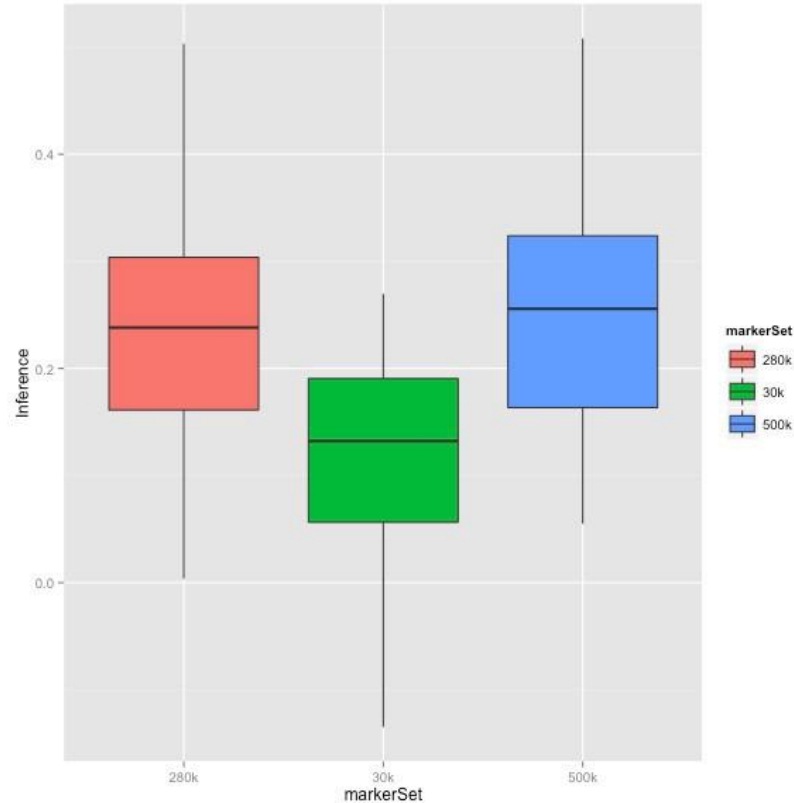
- Edward Buckler
- Cinta Romay
- Eduardo Carrillo
- Kelly Swarts

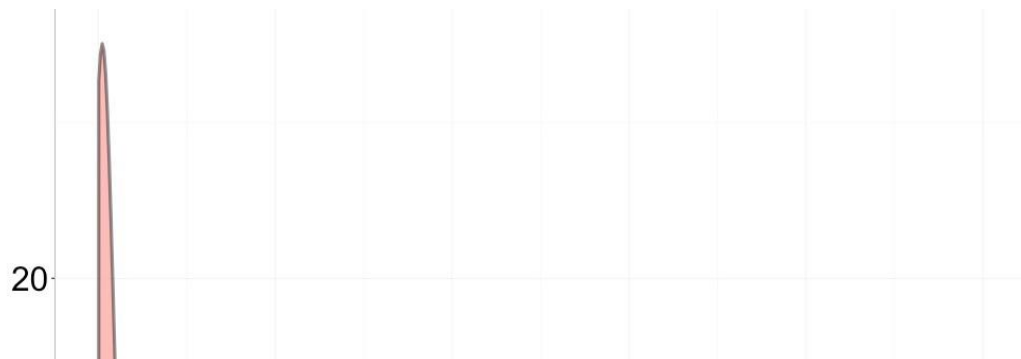


GWAS markers display prediction accuracy equivalent to 30,000 random genome wide SNPs

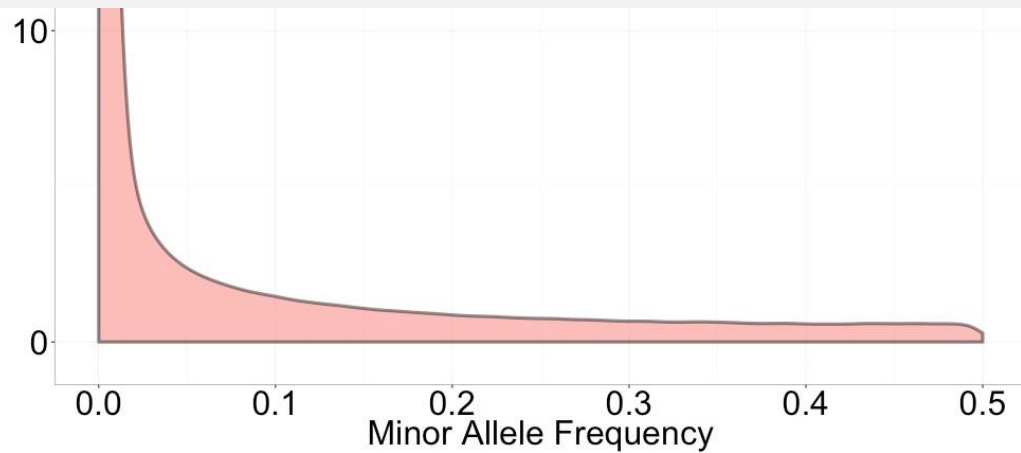


Plant height is highly polygenic and remains hard to predict even when using all markers

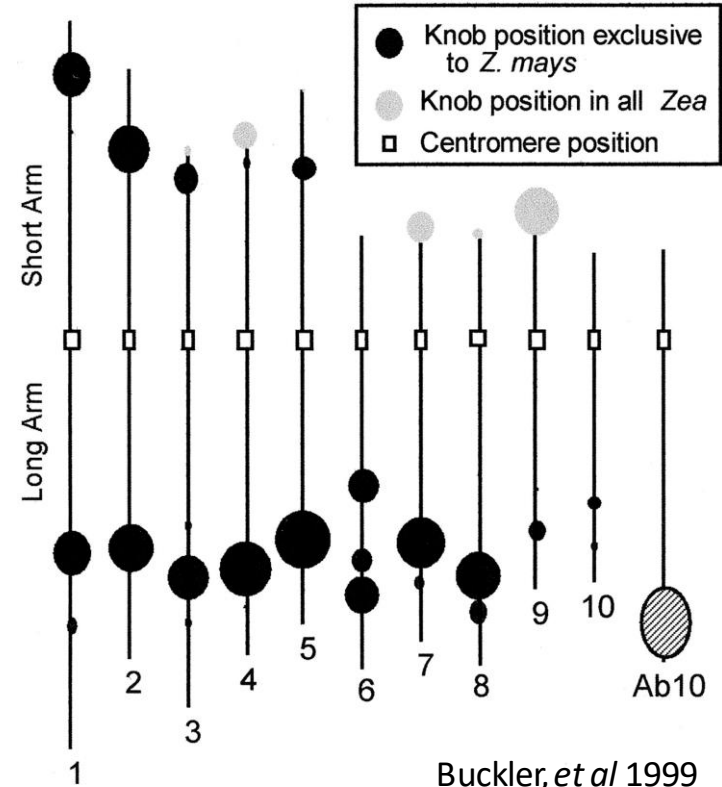




What shapes the underlying variation?



Landrace diversity has been studied extensively through cytological features

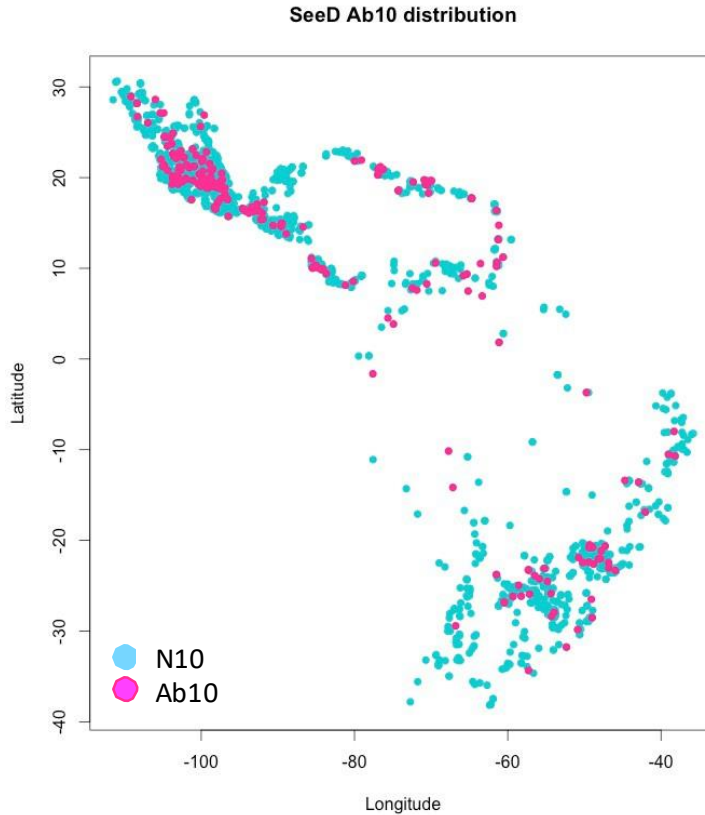


Buckler, *et al* 1999

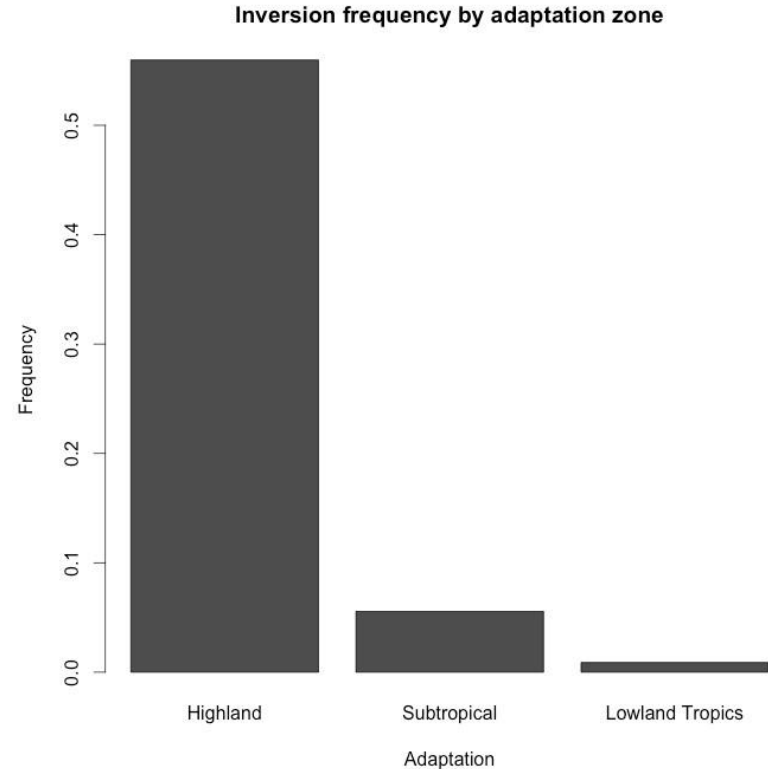
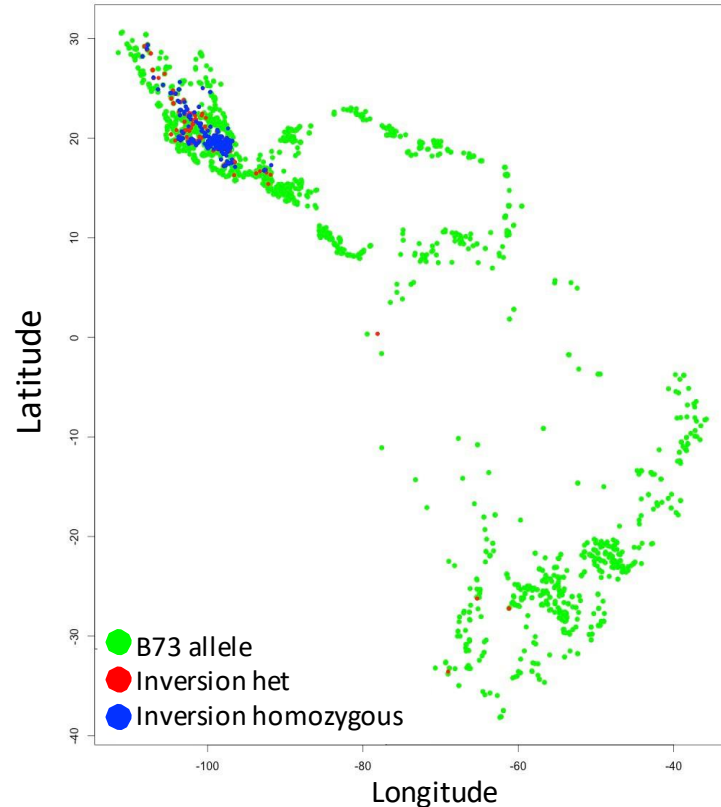
Ab10 also segregates at low frequency



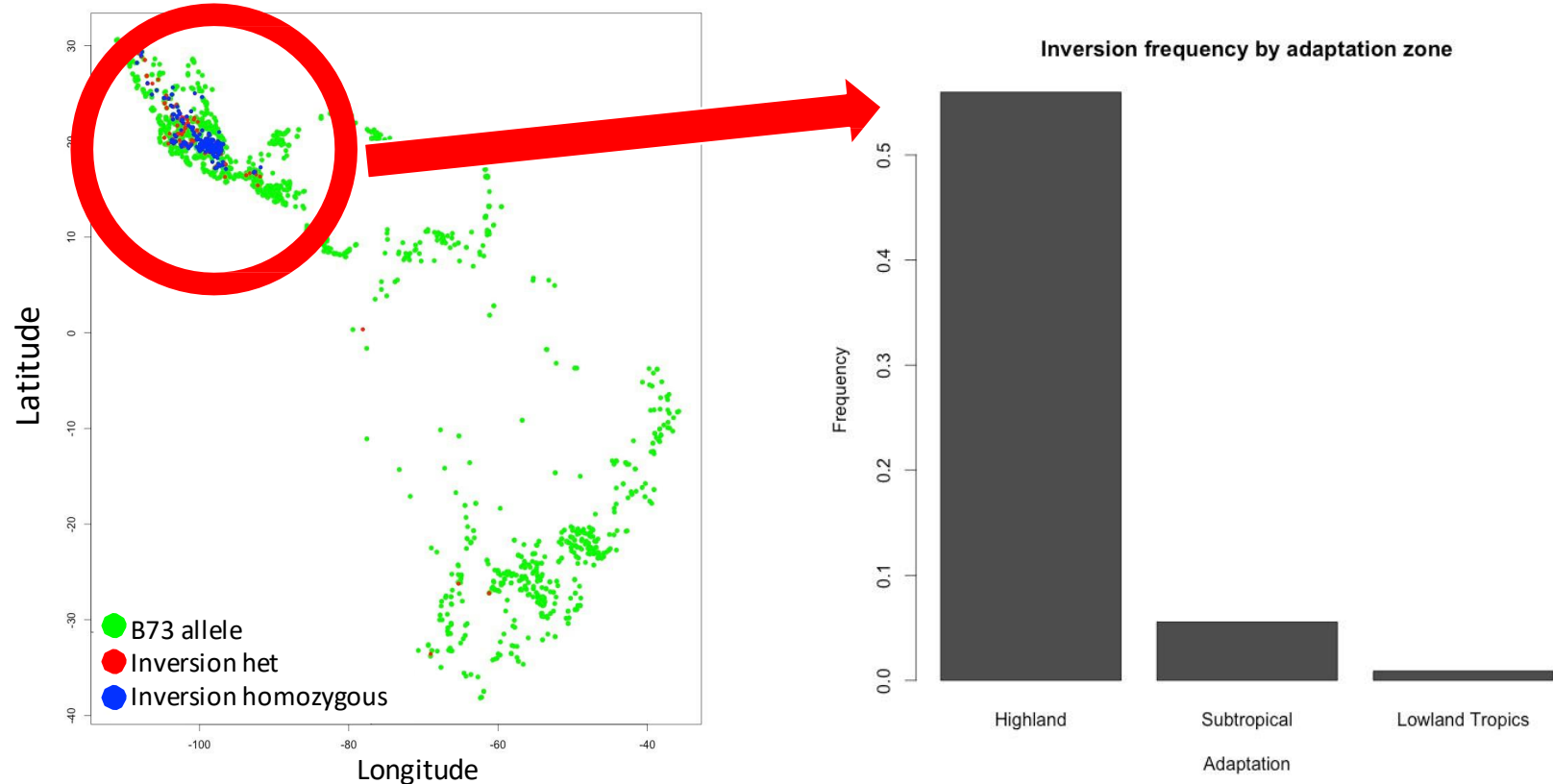
Kelly Dawe



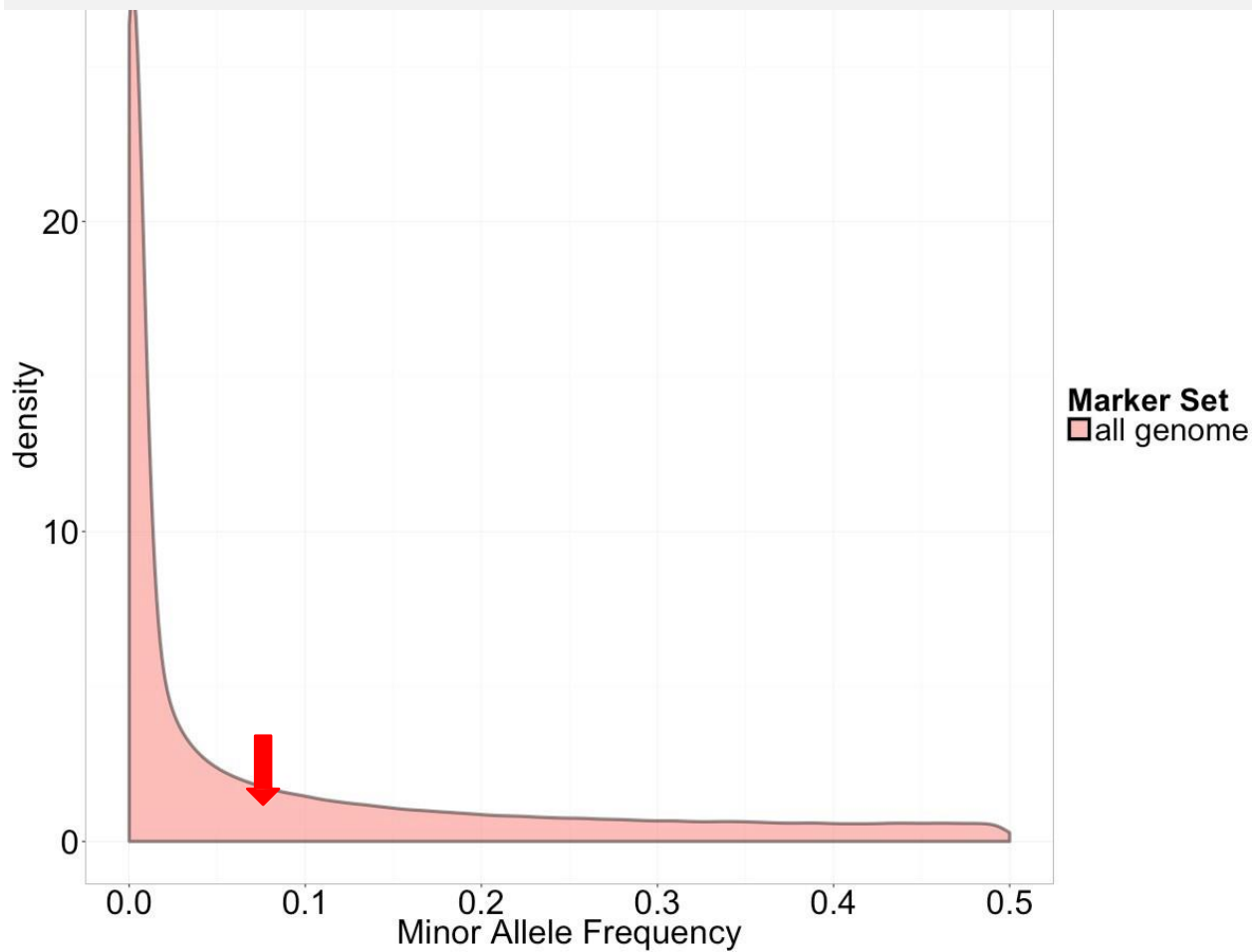
One reported inversion (INV4) is present at varying frequency across populations



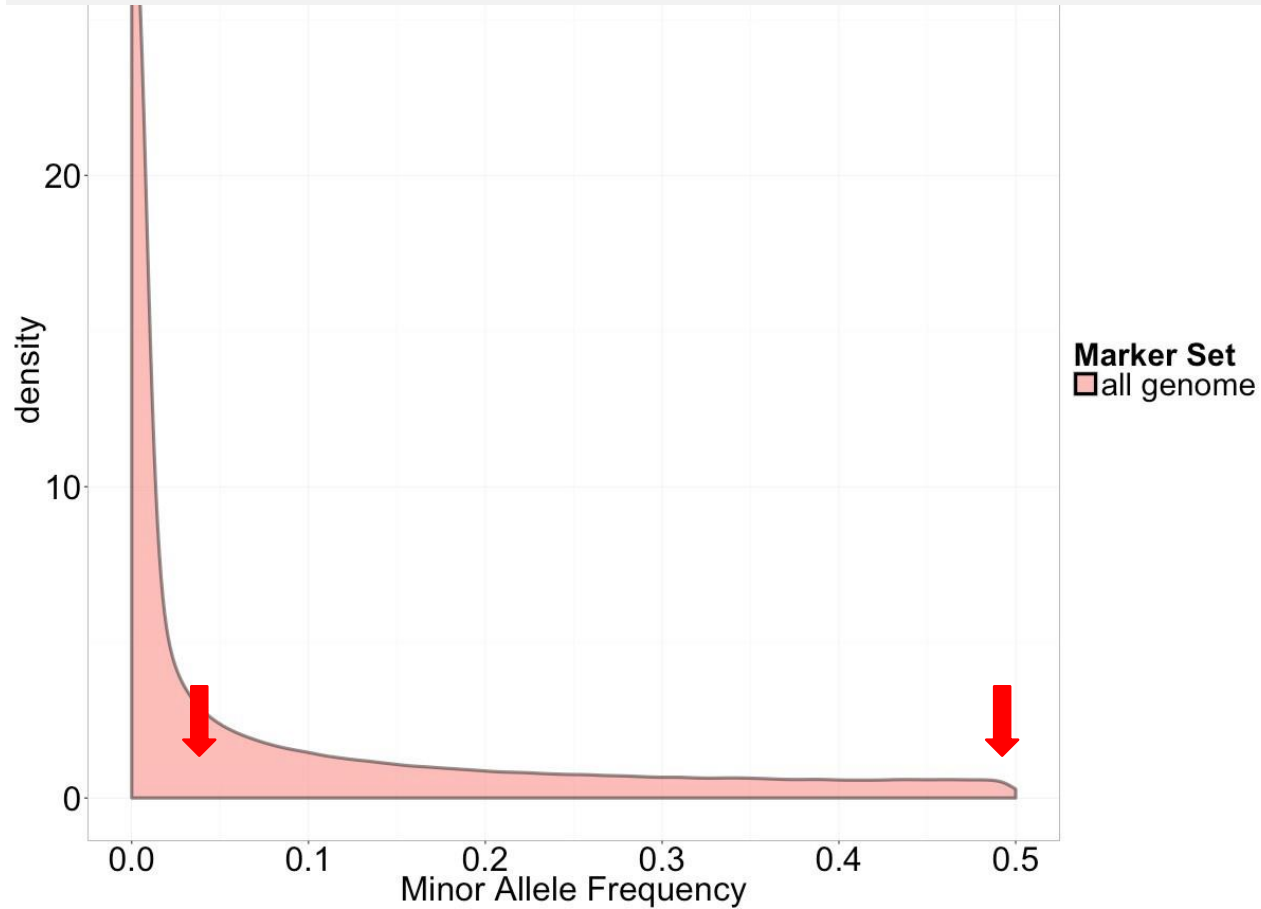
One reported inversion (INV4) is present at varying frequency across populations



Ab10 is rare and deleterious



INV4 frequency depends on environment

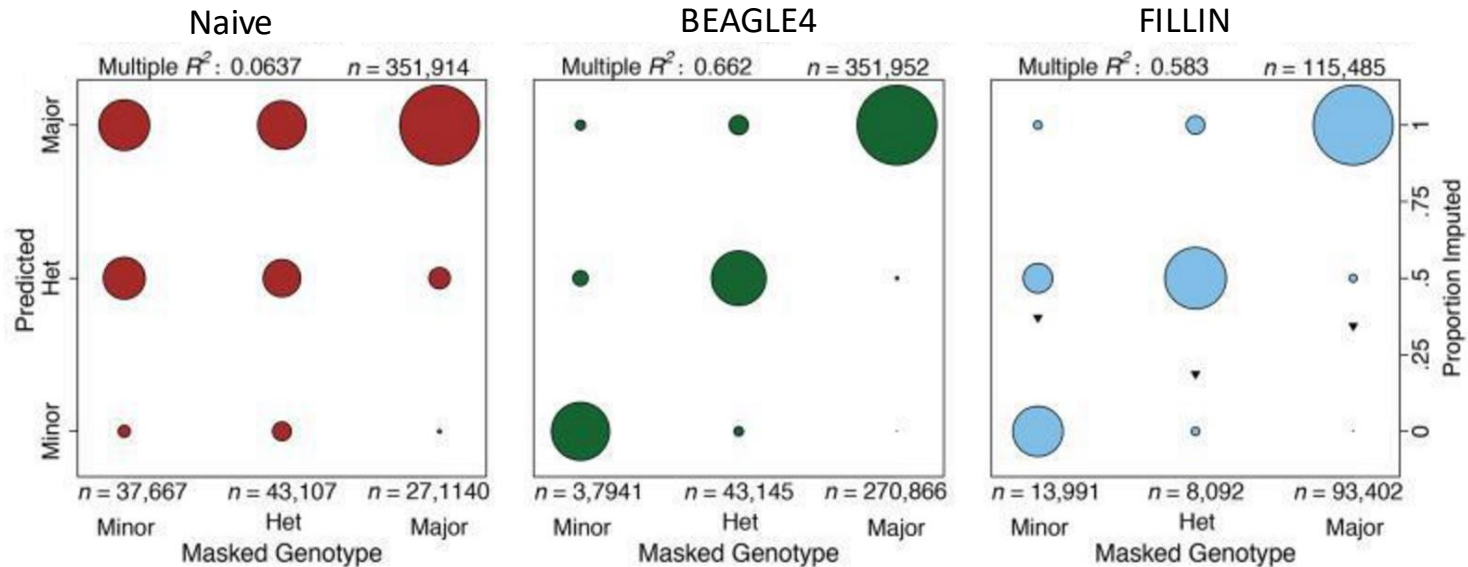


BEAGLE4 is the best performing imputation method for maize landraces



Kelly Swarts

- 366 diverse landraces



- SeeD: Multiple R^2 FILLIN 0.53, BEAGLE4 0.68