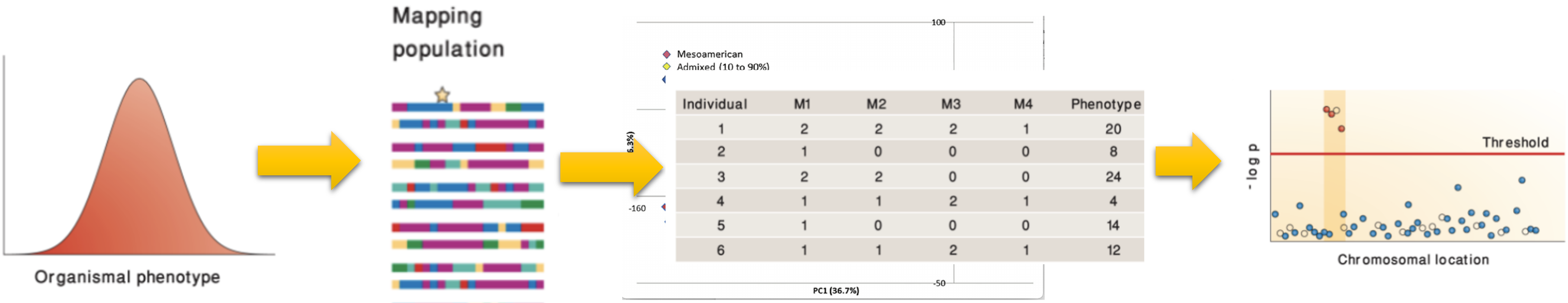
Association mapping

Association mapping

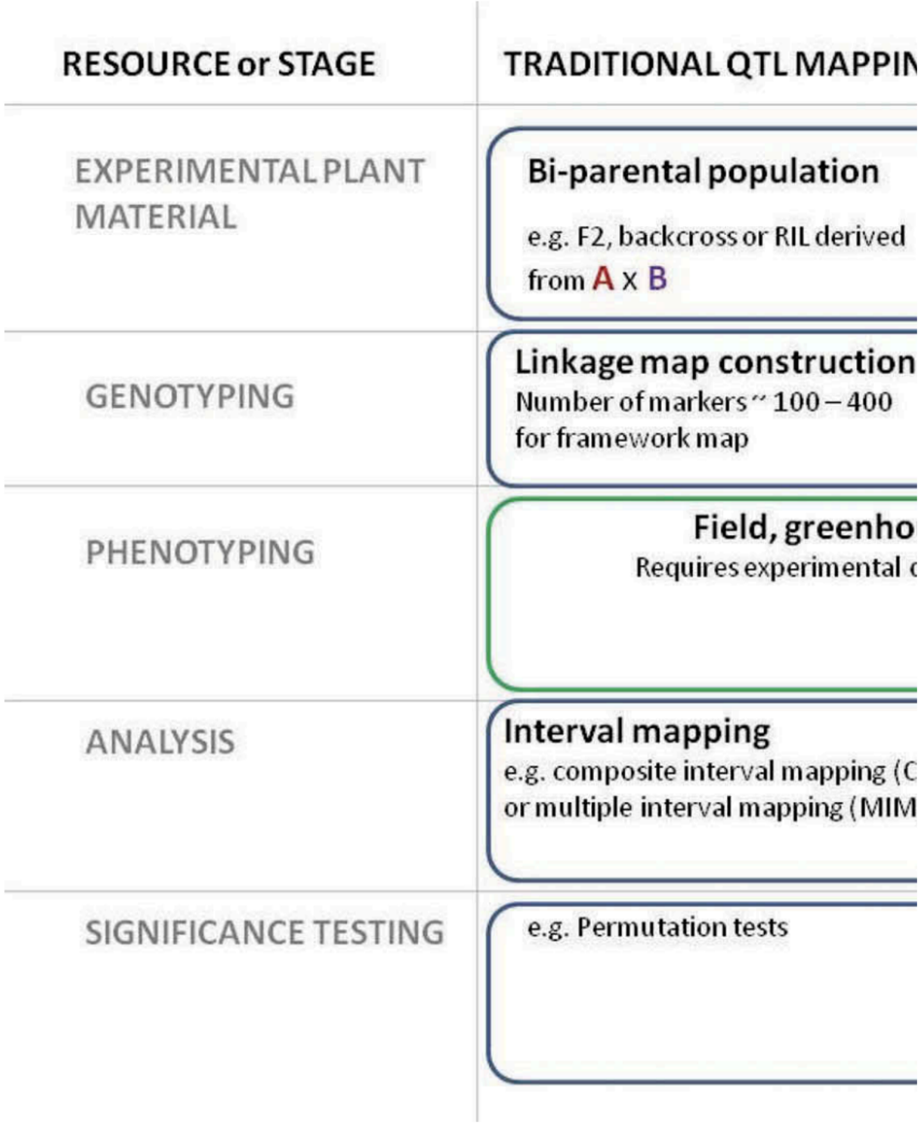
- Association analysis uses historical recombination events over many generations within a short interval surrounding a trait locus
- Association analysis is advantageous for the identification of relatively small genomic regions, in which only few genes may reside.



Individual	M1	M2	M3	M4	Phenotype
1	2	2	2	1	20
2	1	0	0	0	8
3	2	2	0	0	24
4	1	1	2	1	4
5	1	0	0	0	14
6	1	1	2	1	12

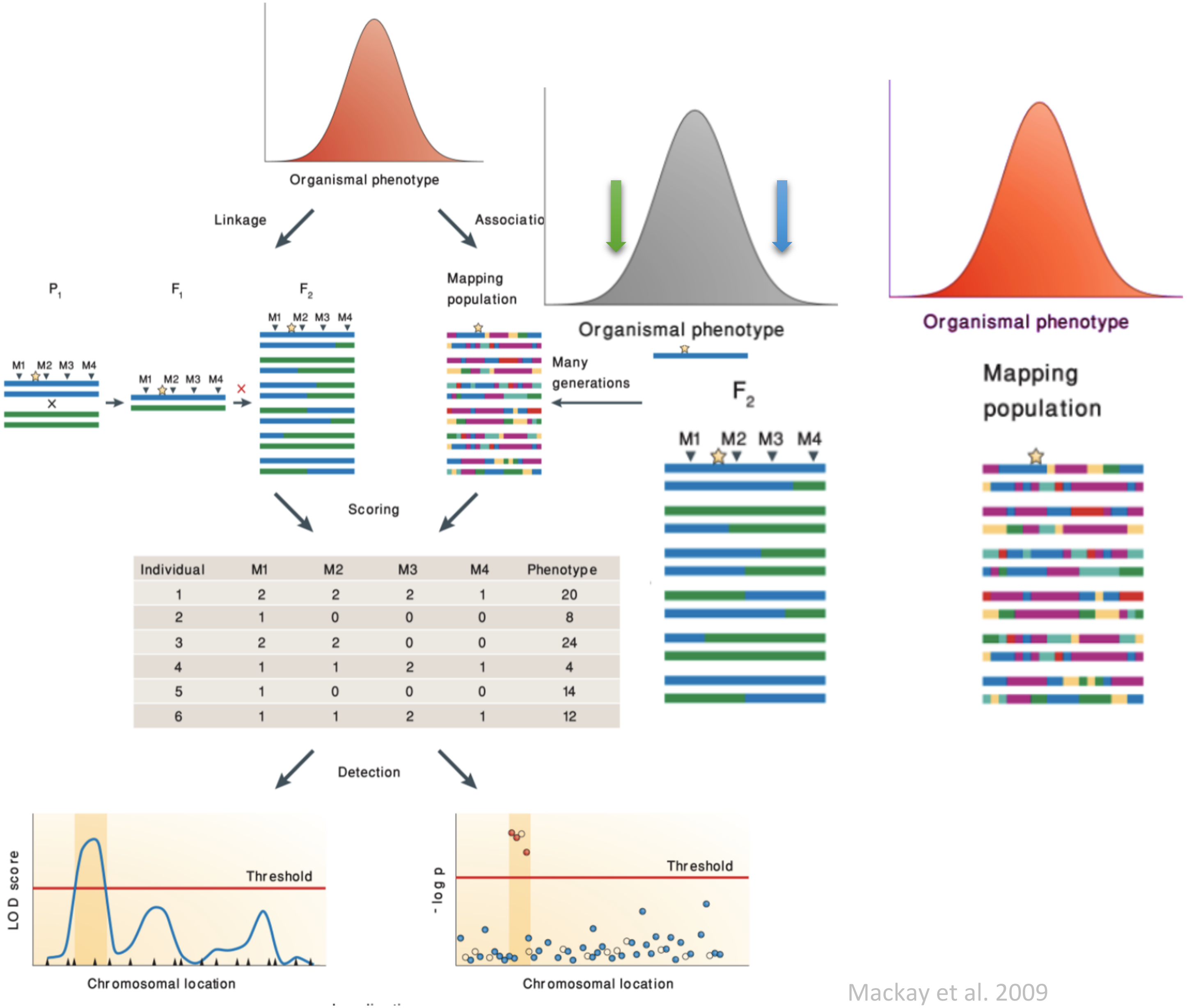
Cichy et al. 2015

Mapping population vs Association mapping



RESOURCE or STAGE	TRADITIONAL QTL MAPPIN
EXPERIMENTAL PLANT MATERIAL	**Bi-parental population** e.g. F2, backcross or RIL derived from A x B
GENOTYPING	**Linkage map construction** Number of markers ~ 100 – 400 for framework map
PHENOTYPING	**Field, greenho** Requires experimental c
ANALYSIS	**Interval mapping** e.g. composite interval mapping (C or multiple interval mapping (MIM
SIGNIFICANCE TESTING	e.g. Permutation tests

Verdep



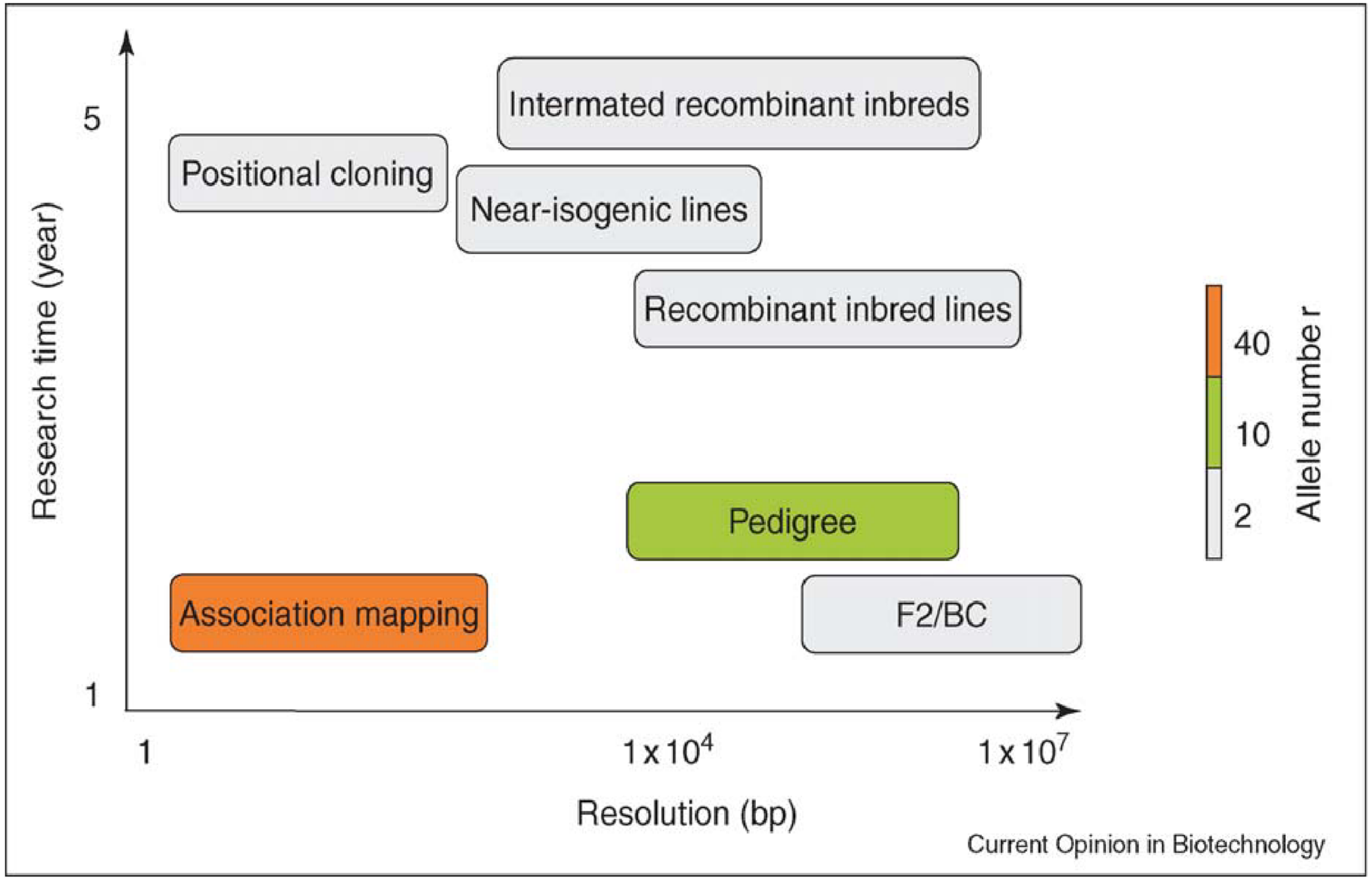
Individual	M1	M2	M3	M4	Phenotype
1	2	2	2	1	20
2	1	0	0	0	8
3	2	2	0	0	24
4	1	1	2	1	4
5	1	0	0	0	14
6	1	1	2	1	12

Mackay et al. 2009

Mapping population vs Association mapping

Association mapping offers three advantages over linkage analysis:

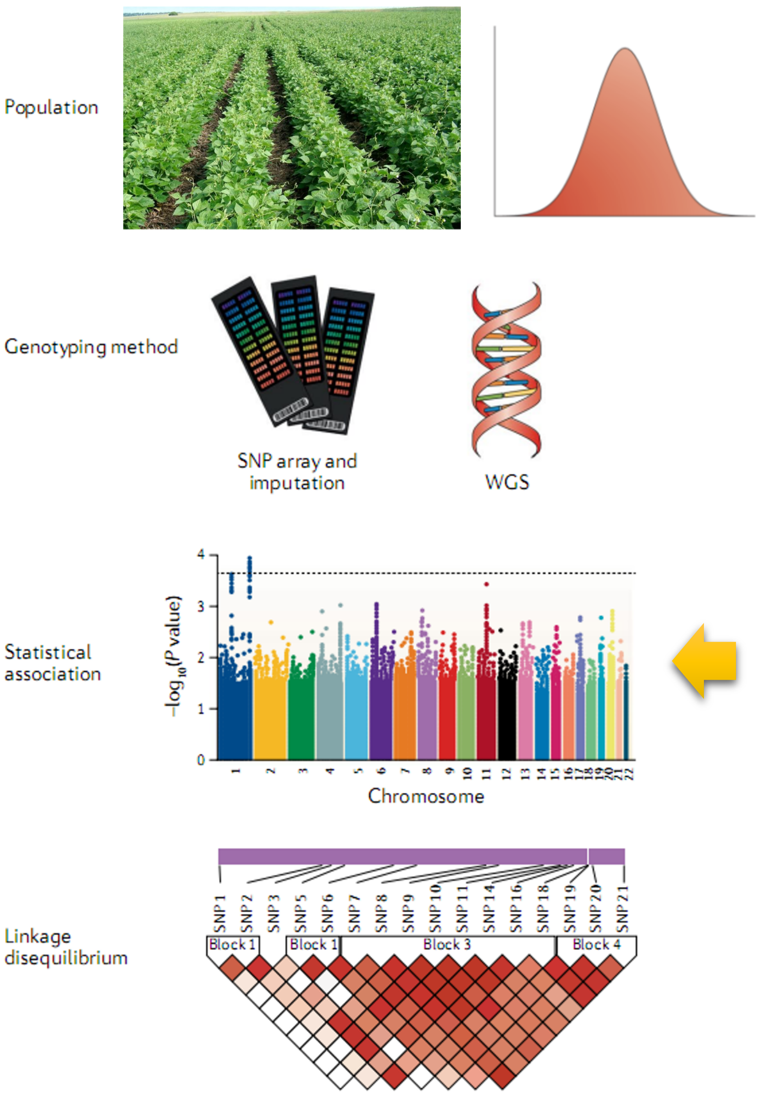
- Higher mapping resolution
- Greater allele number
- Broader reference population
- Less research time in establishing an association



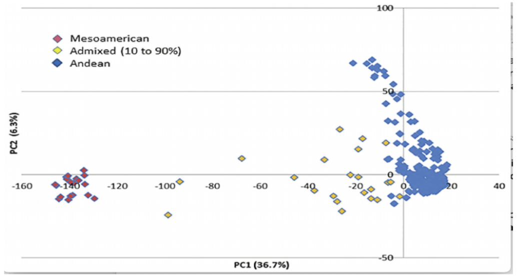
Yu & Buckler 2005

Association mapping

- Linkage disequilibrium
- Population structure
- Statistical approaches
- QTL vs Association mapping



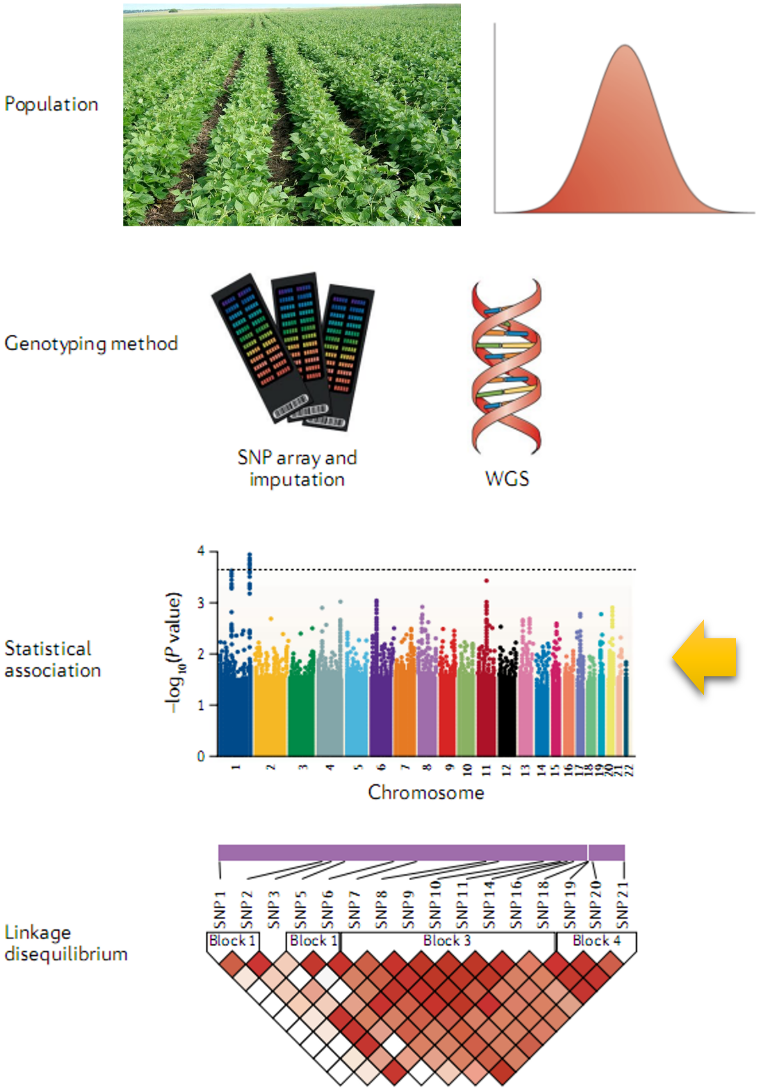
Tam et al 2019



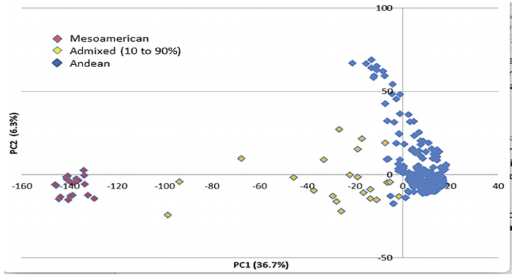
Cichy et al. 2015

Association mapping

- Linkage disequilibrium
- Population structure
- Statistical approaches
- QTL vs Association mapping



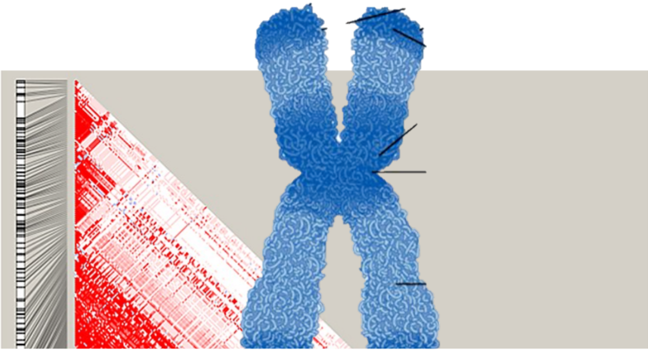
Tam et al 2019

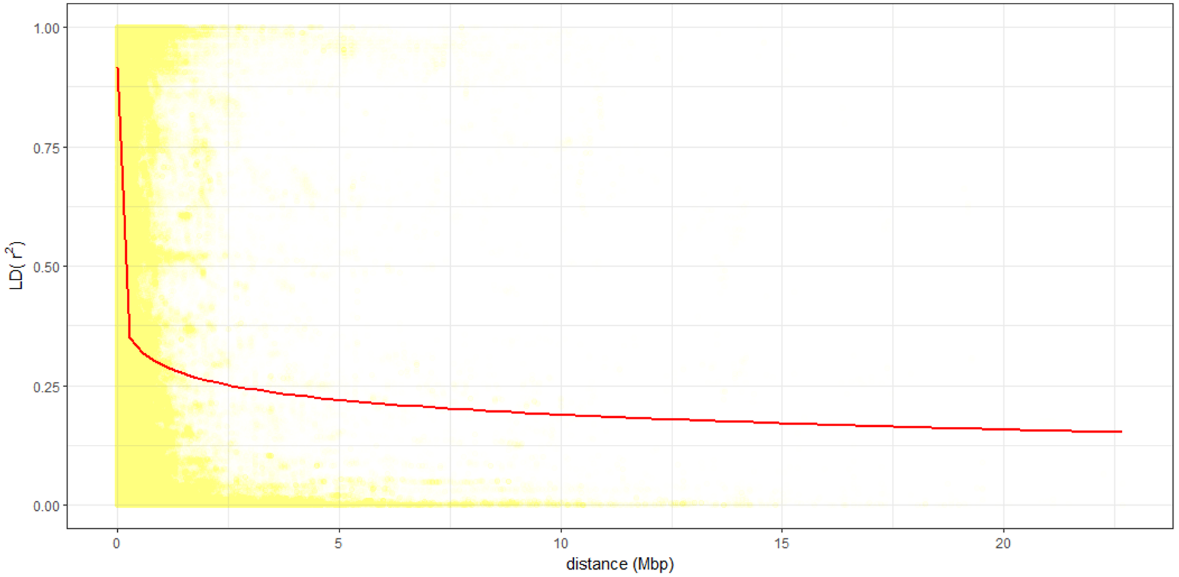


Cichy et al. 2015

Linkage disequilibrium

- LS is defined as the non-random association of allele between loci.
- The resolution is dependent of the LD.
- LD determines the density of markers needed.





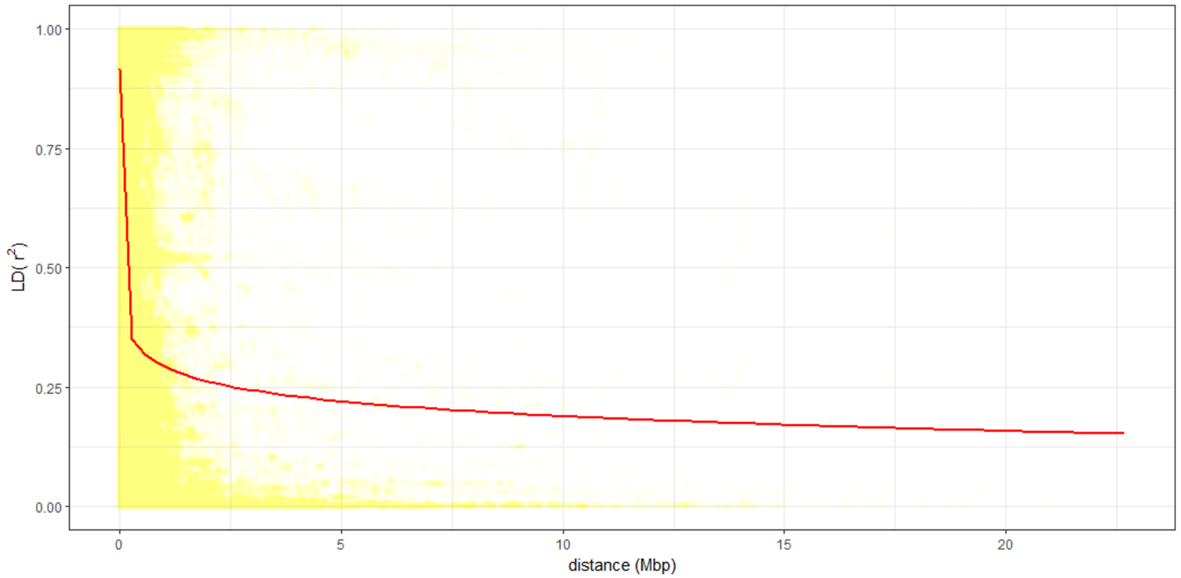
Supplementary Table 7. Physical (Kb) and genetic (cM) position of the last marker mapped in each chromosome and recombination rate (Kb/cM) per chromosome and comparison between pericentromeric regions and euchromatic arms.

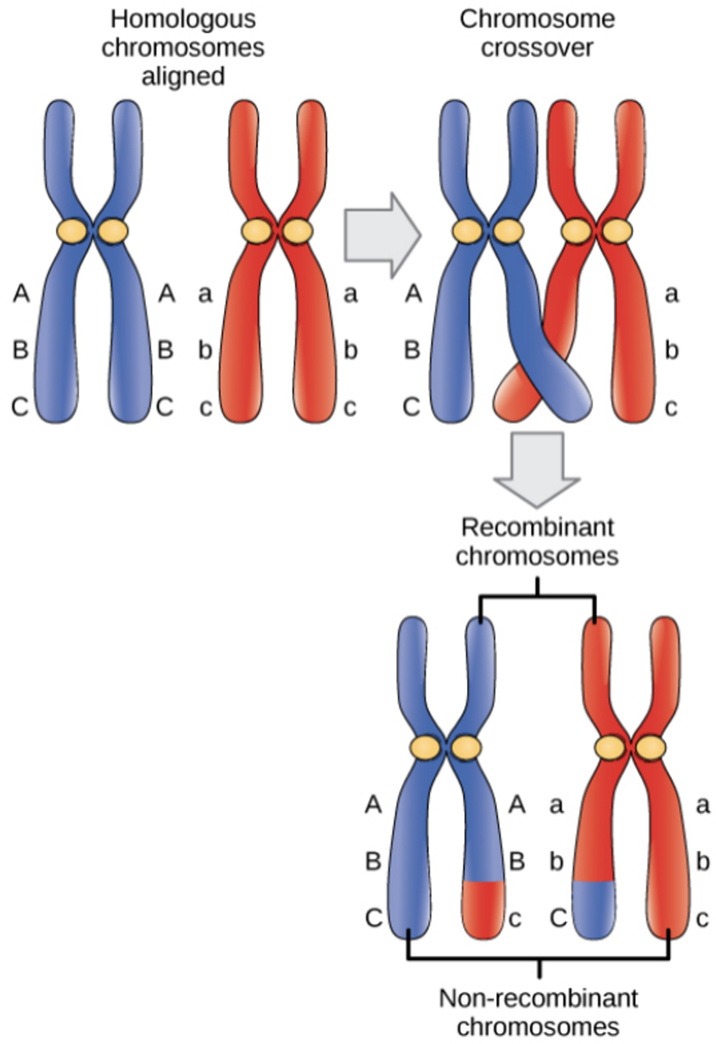
Schmutz et al 2014

Pistis et al. 2013

Linkage disequilibrium

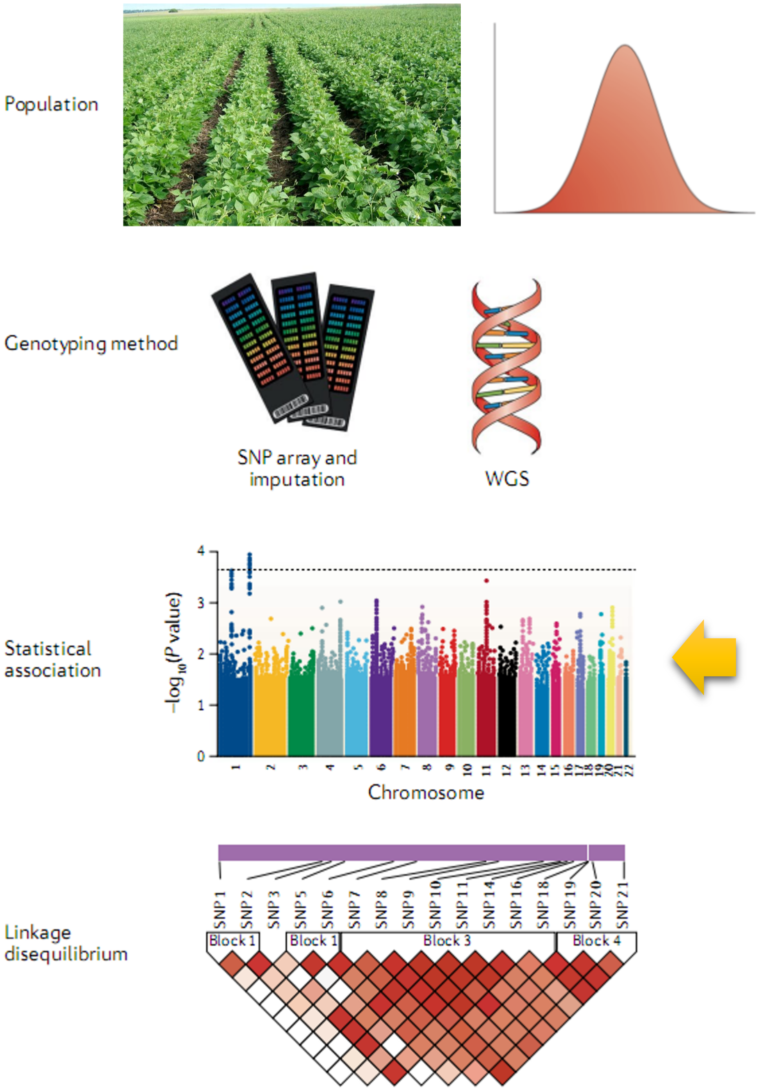
$$r^2 = \frac{p(AB) - p(A)p(B)}{p(A)(a)(B)(b)}$$



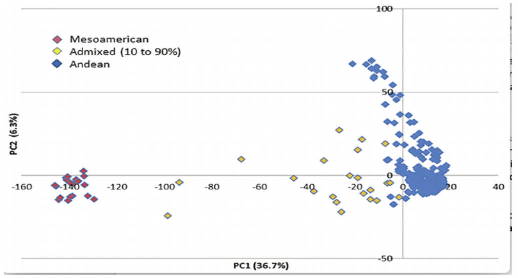


Association mapping

- Linkage disequilibrium
- Population structure
- Statistical approaches
- QTL vs Association mapping



Tam et al 2019



Cichy et al. 2015

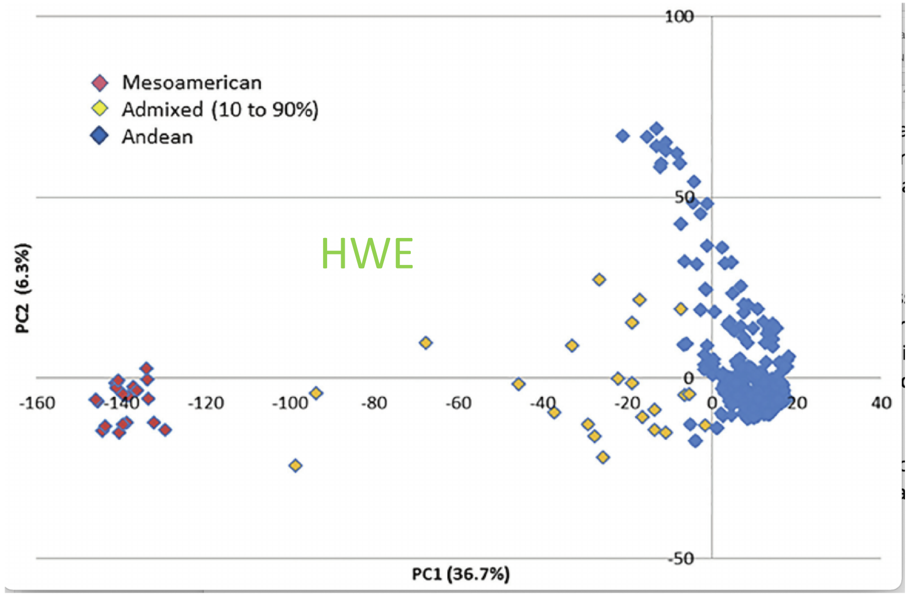
Population structure

- Population structure can cause false-positive results.

$$r^2 = \frac{p(AB) - p(A)p(B)}{p(A)(a)(B)(b)}$$

A = 0.2
a = 1 – p = 0.8

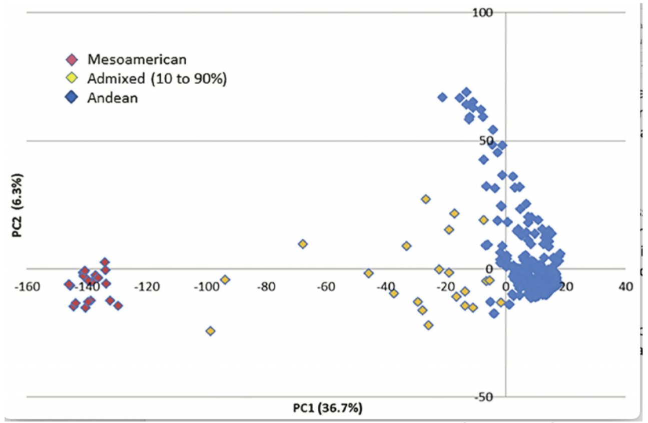
A = 0.6
a = 1 – p = 0.4



Cichy et al. 2015

- Statistically controlling for population structure can result in false-negatives, particularly for small sample sizes.

Population structure

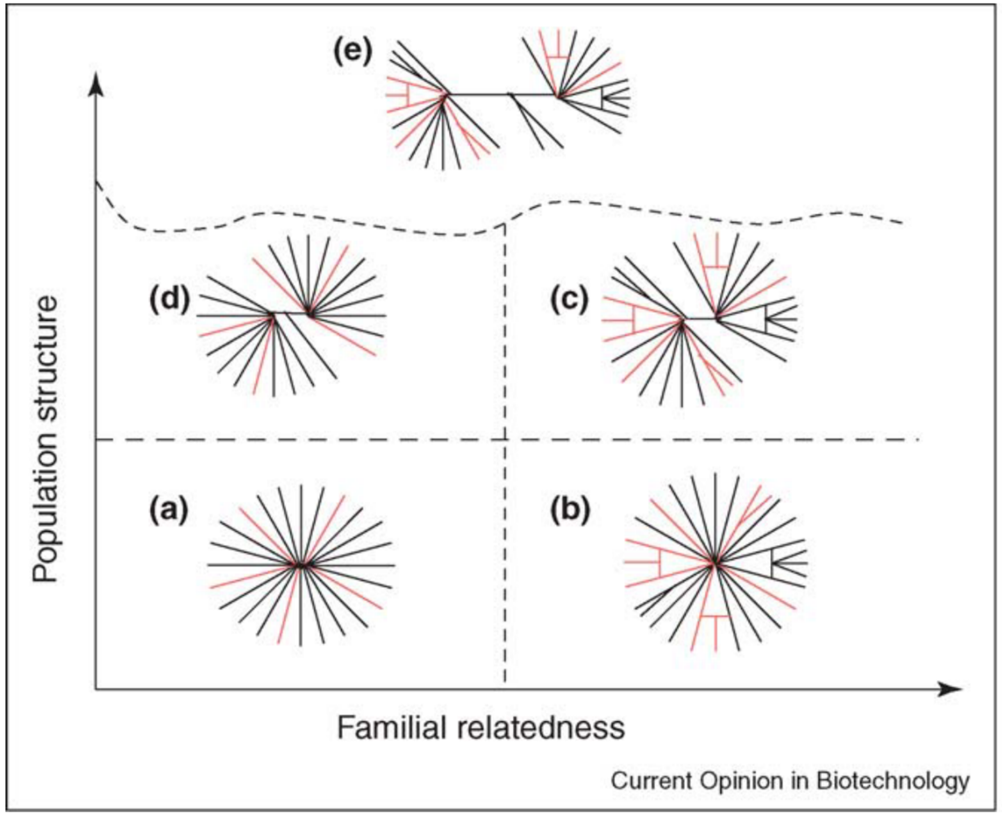


Cichy et al. 2015

- Population structure can cause false-positive results.

$$r^2 = \frac{p(AB) - p(A)p(B)}{p(A)(a)(B)(b)}$$

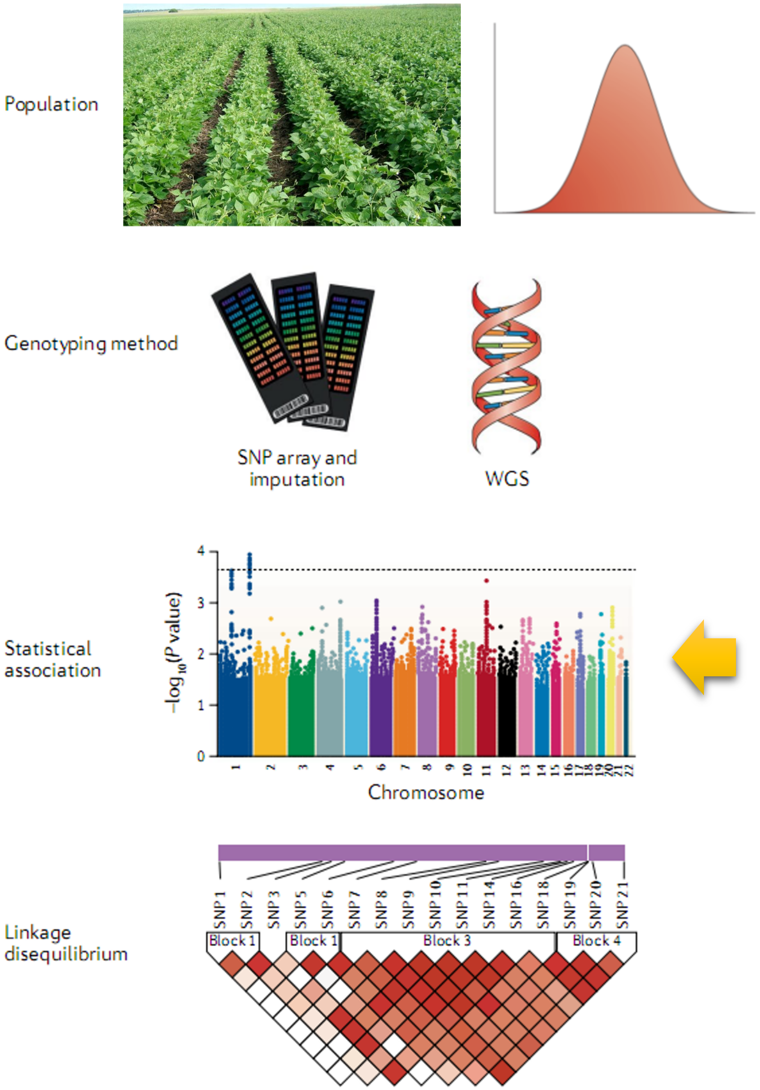
- Statistically controlling for population structure can result in false-negatives, particularly for small sample sizes.
- Association studies, therefore, are best carried out in independent populations with a large sample size.



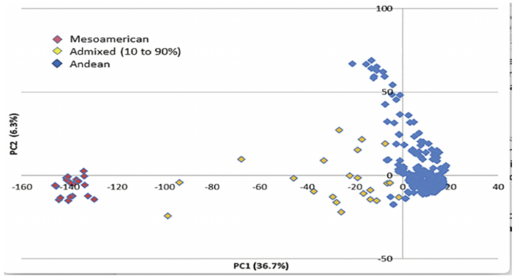
Yu & Buckler 2006

Association mapping

- Linkage disequilibrium
- Population structure
- Statistical approaches
- QTL vs Association mapping



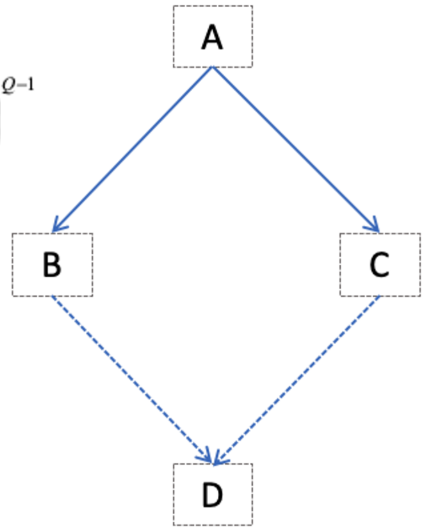
Tam et al 2019



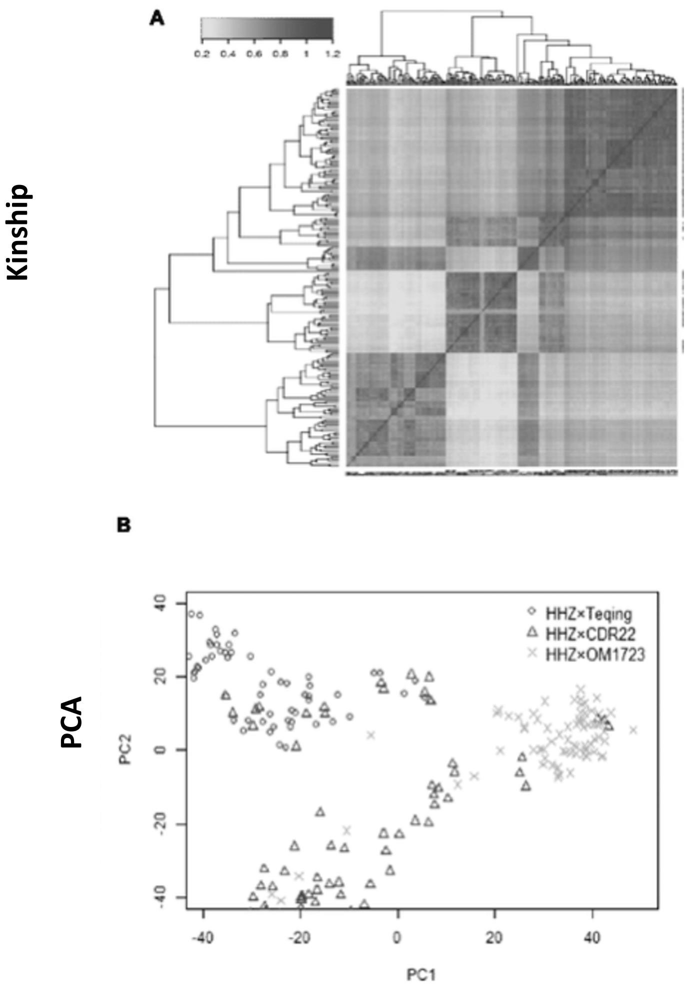
Cichy et al. 2015

Statistical approaches

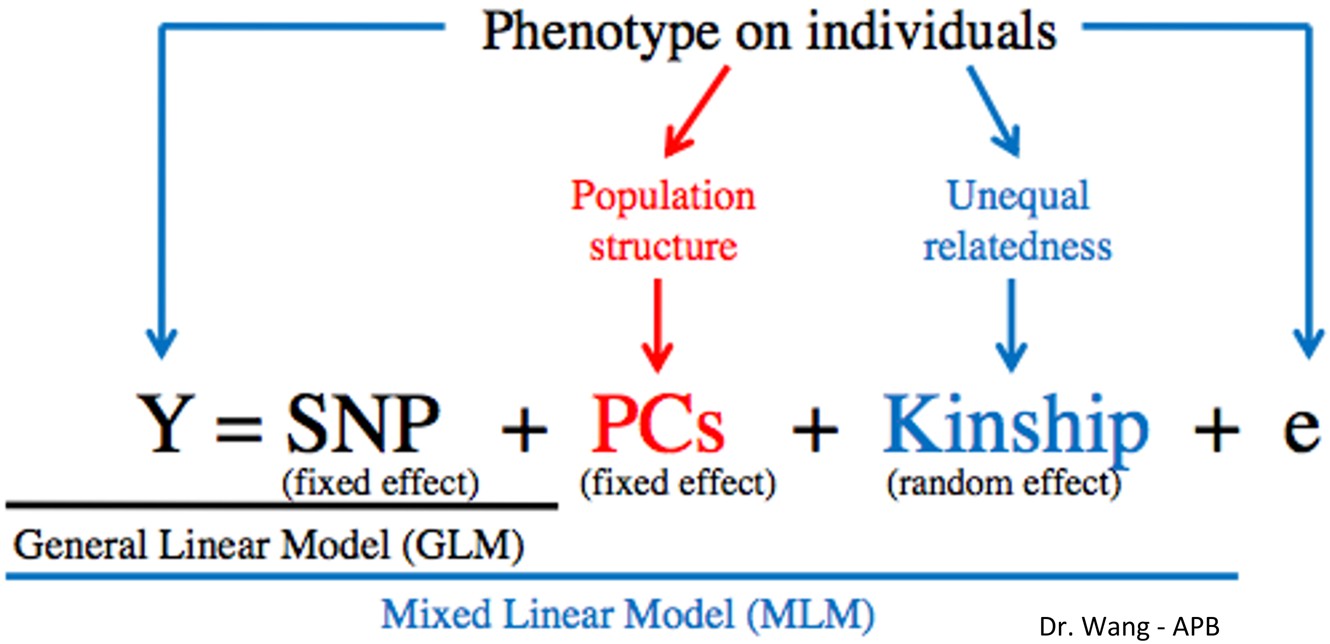
$$\Theta_{BC} = F_D = 2\left(\frac{1}{2}\right)^{Q} = \left(\frac{1}{2}\right)^{Q-1}$$



Kinship : a measure of genetic similarity between two individuals which is equal to the inbreeding coefficient of a hypothetical offspring of the two individuals.



Feng *et al* 2018



Phenotype on individuals

Population structure

Unequal relatedness

$$\underbrace{Y = \text{SNP}}_{\text{(fixed effect)}} + \underbrace{\text{PCs}}_{\text{(fixed effect)}} + \underbrace{\text{Kinship}}_{\text{(random effect)}} + e$$

General Linear Model (GLM)

Mixed Linear Model (MLM)

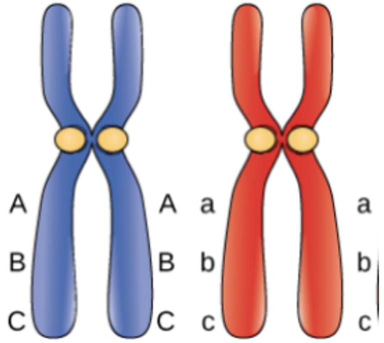
Statistical approaches

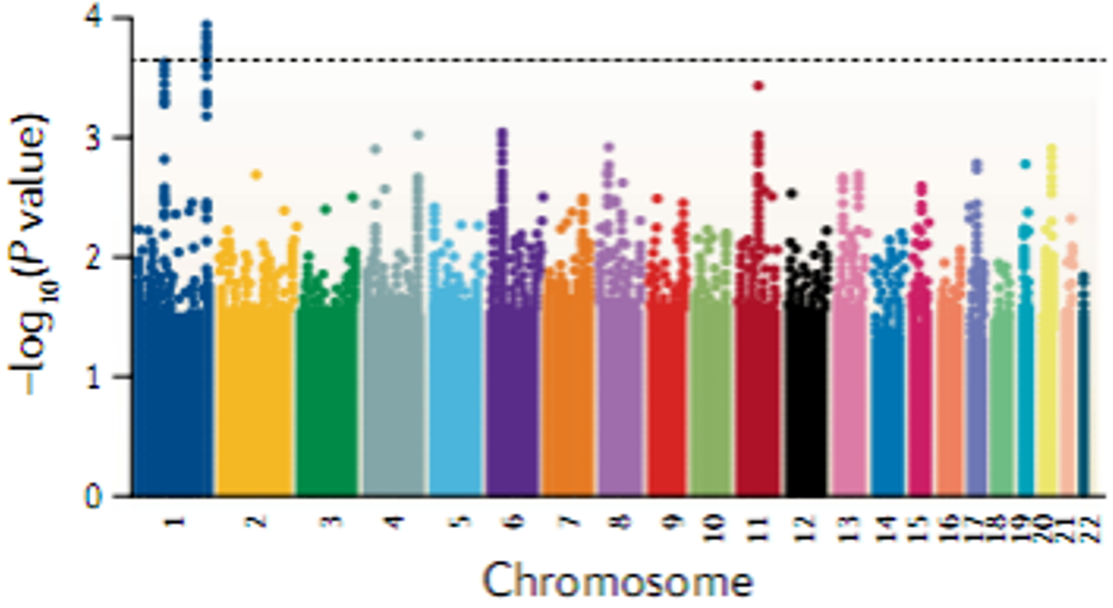
TASSEL 5.0

TASSEL Version 5.0 *(Getting Started!)*
(Build: February 10, 2021 Requires: Java 1.8)

$$F = \frac{\text{variance between treatments}}{\text{variance within treatments}}$$

Large F ratio indicate that there are more variations between groups and likely a difference in the mean

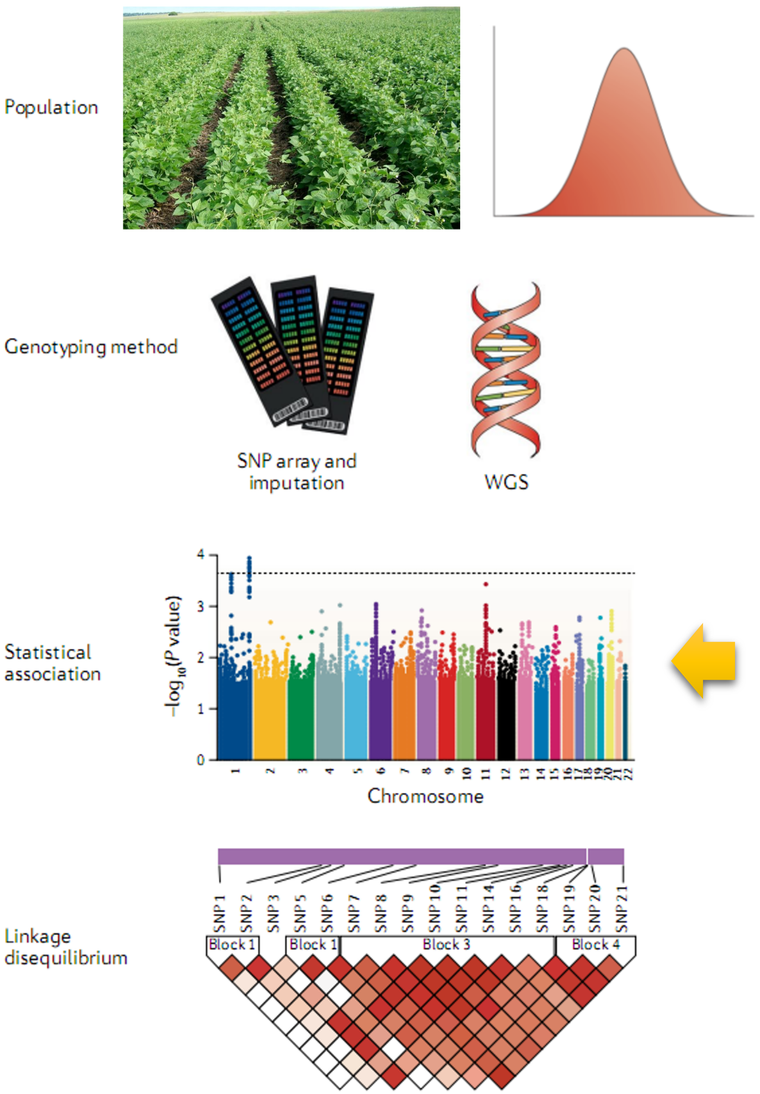




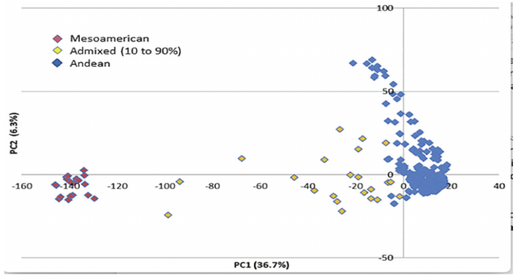
Gene

Association mapping

- Linkage disequilibrium
- Population structure
- Statistical approaches
- QTL vs Association mapping



Tam et al 2019

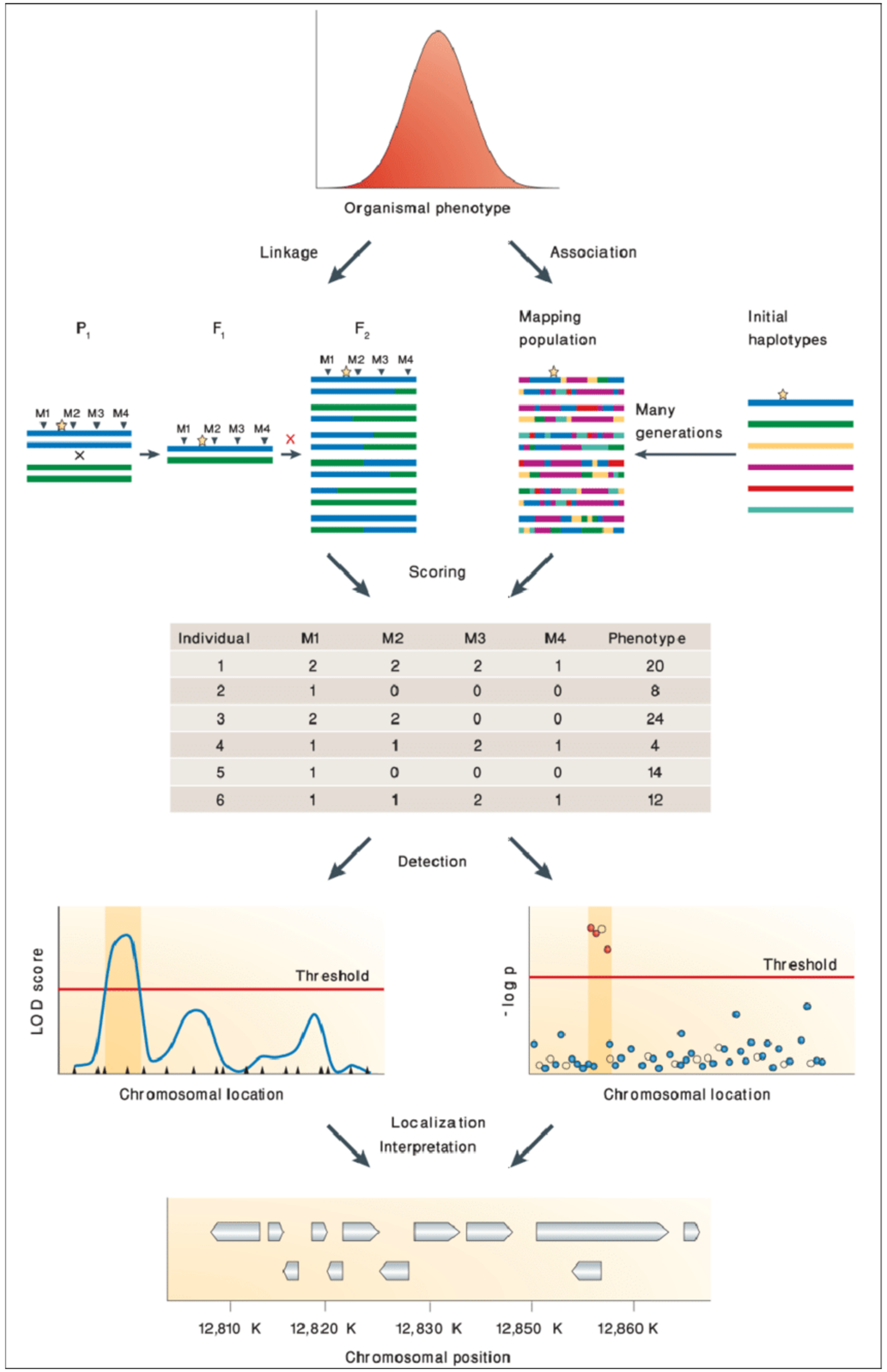


Cichy et al. 2015

Mapping population
vs
Association mapping

1. Cross required
2. Resolution
3. Alleles tested
4. Power

$$t = \frac{\overline{MM} - \overline{mm}}{\sqrt{\frac{\hat{V}(MM)}{N} + \frac{\hat{V}(mm)}{N}}}$$

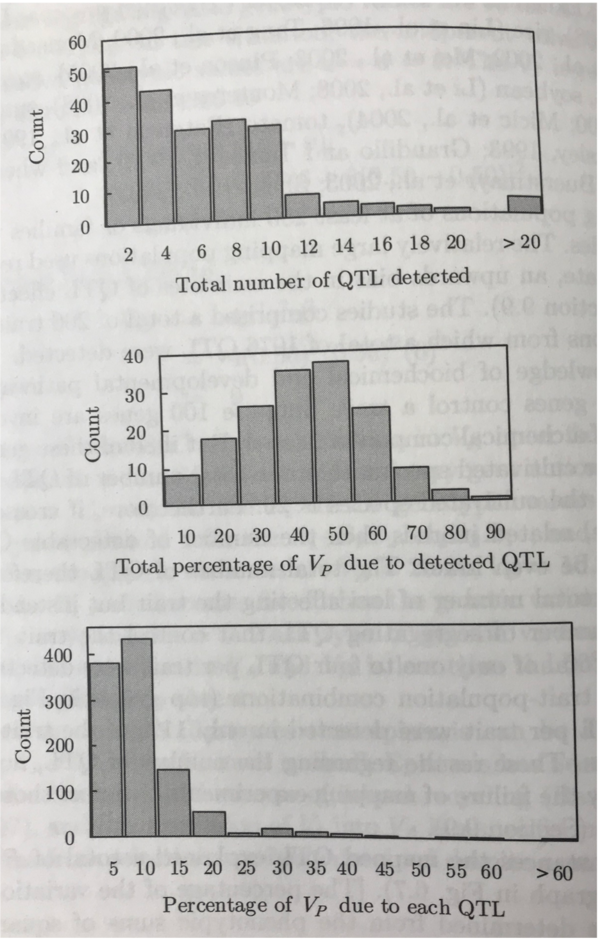


Individual	M1	M2	M3	M4	Phenotype
1	2	2	2	1	20
2	1	0	0	0	8
3	2	2	0	0	24
4	1	1	2	1	4
5	1	0	0	0	14
6	1	1	2	1	12

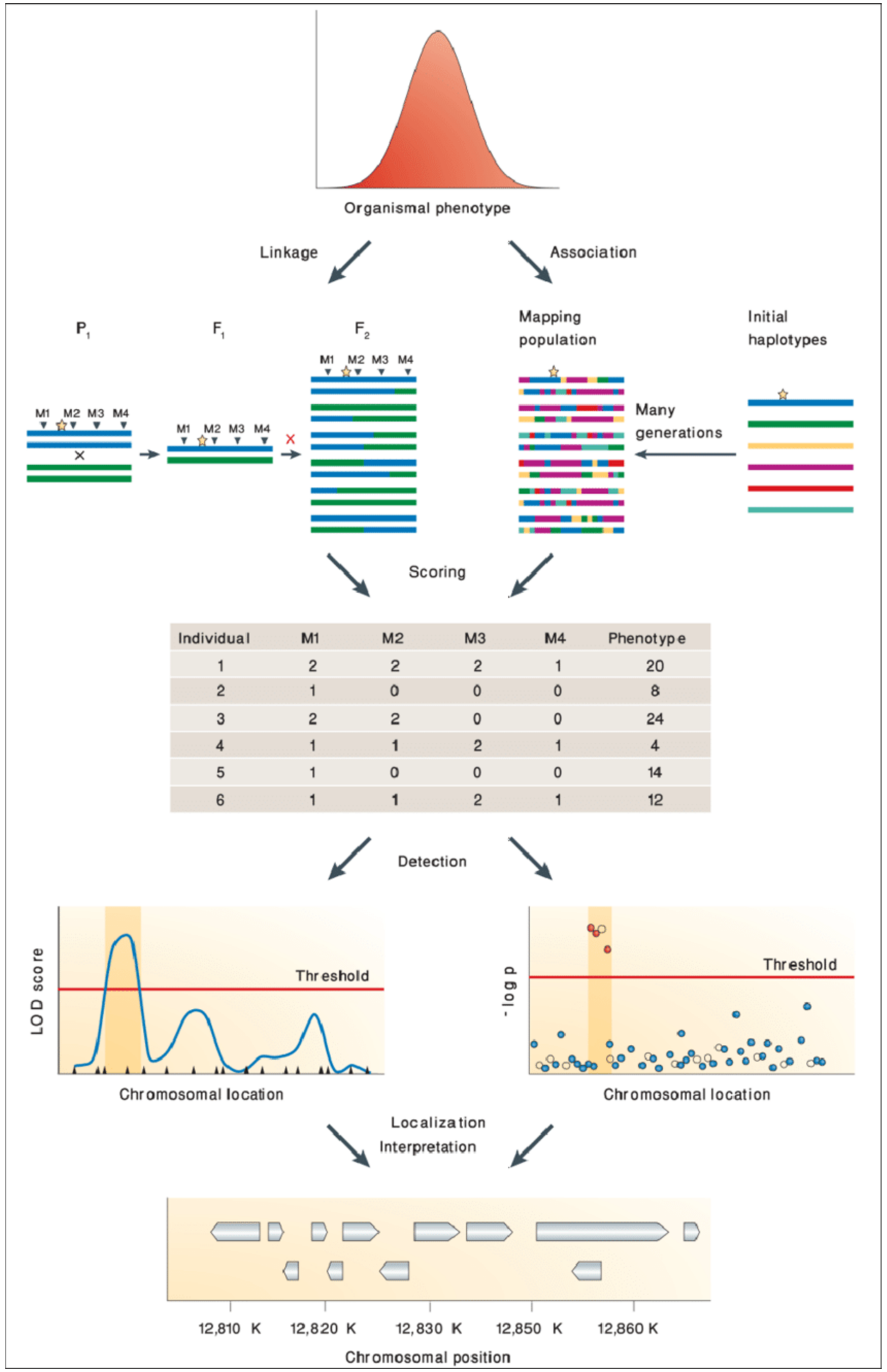
Mackay et al. 2009

Mapping population vs Association mapping

Number of QTL detected for a trait in mapping studies with >250 progeny

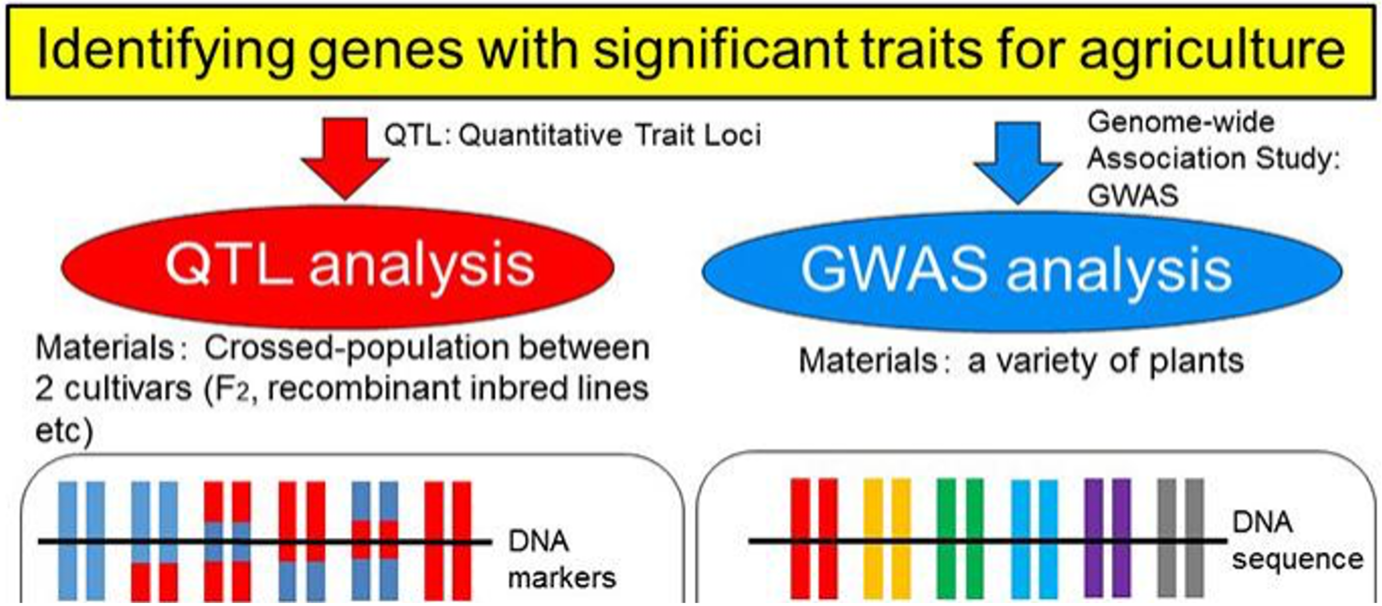


Bernardo 2020, p 152



Individual	M1	M2	M3	M4	Phenotype
1	2	2	2	1	20
2	1	0	0	0	8
3	2	2	0	0	24
4	1	1	2	1	4
5	1	0	0	0	14
6	1	1	2	1	12

Mackay et al. 2009



Kobe University

> "When the objective is to detect a rare variant, as is often the case in plant breeding, QTL mapping is therefore preferred over association mapping"

Bernardo 2020