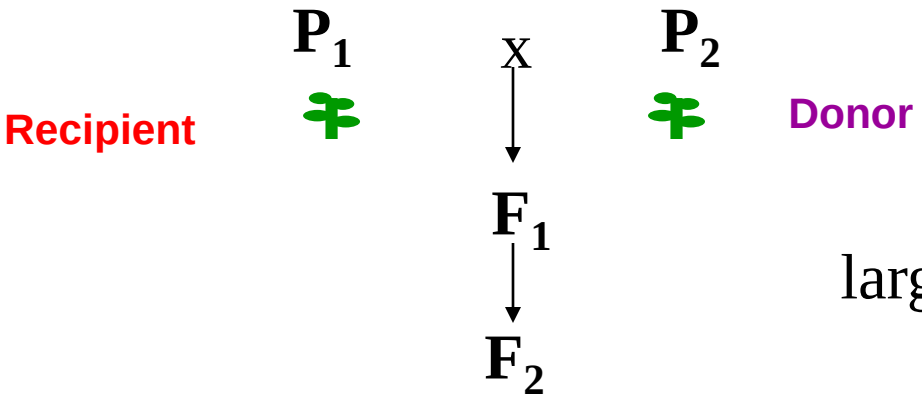
Marker assisted selection

CONVENTIONAL PLANT BREEDING



large populations consisting of thousands of plants

PHENOTYPIC SELECTION

Salinity screening in phytotron

Bacterial blight screening

Phosphorus deficiency plot

Glasshouse trials

Field trials

Marker assisted selection (MAS)

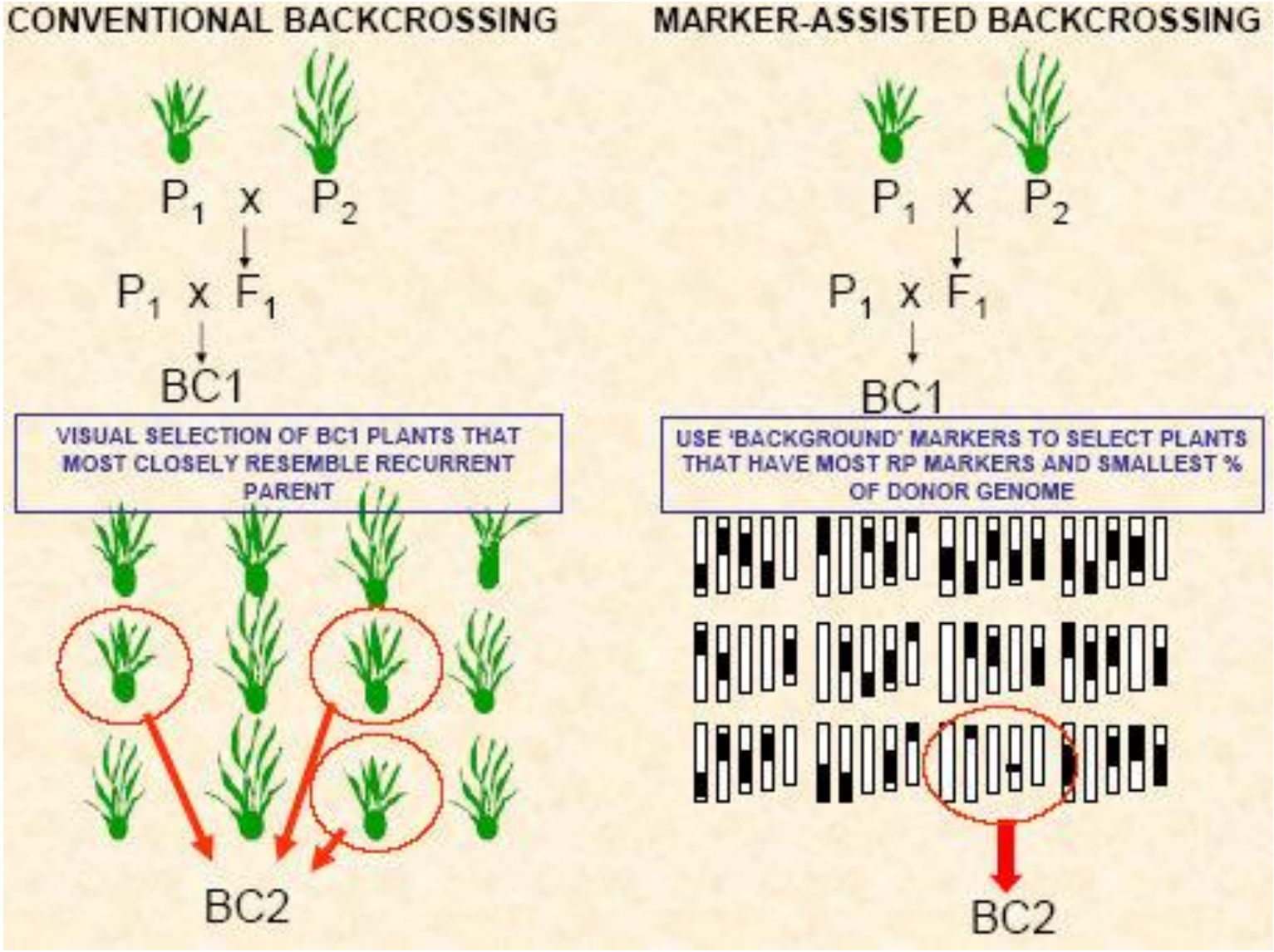
A method of selecting desirable individuals in a breeding scheme based on DNA molecular marker patterns instead of, or in addition to, their trait values.

A tool that can help plant breeders select more efficiently for desirable crop traits.

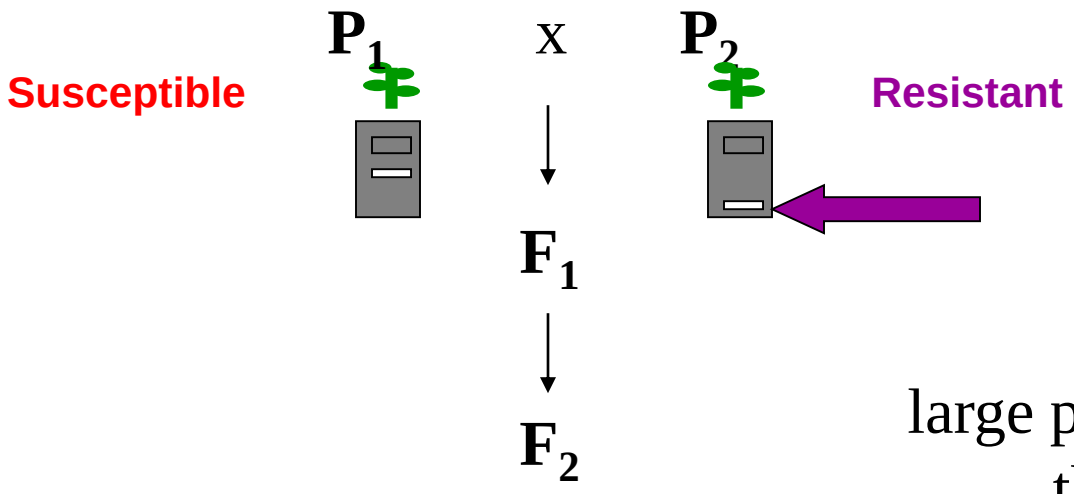
MAS is not always advantageous, so careful analysis of the costs and benefits relative to conventional breeding methods is necessary.

ASSISTED NEGATIVE SELECTION:
against undesired features from one of the parental lines -> multiple markers (position of genes responsible for the traits are unknown)

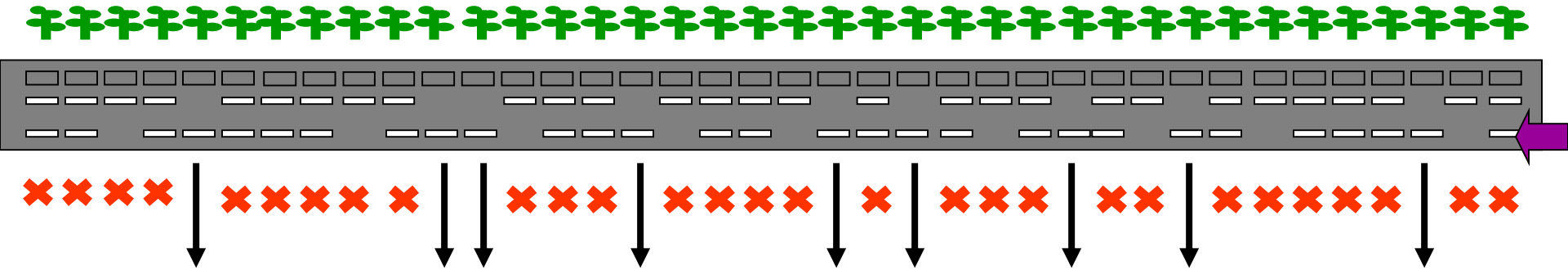
ASSISTED POSITIVE SELECTION:
selection of plants that received the trait of interest (few markers; map position is well established)


CONVENTIONAL BACKCROSSING
P1 x P2
P1 x F1
BC1
VISUAL SELECTION OF BC1 PLANTS THAT MOST CLOSELY RESEMBLE RECURRENT PARENT
BC2
MARKER-ASSISTED BACKCROSSING
P1 x P2
P1 x F1
BC1
USE 'BACKGROUND' MARKERS TO SELECT PLANTS THAT HAVE MOST RP MARKERS AND SMALLEST % OF DONOR GENOME
BC2

MARKER-ASSISTED BREEDING



large populations consisting of thousands of plants



MARKER-ASSISTED SELECTION (MAS)

Method whereby phenotypic selection is based on DNA markers

Advantages of MAS

- **Simpler method compared to phenotypic screening**
  - Especially for traits with laborious screening
  - May save time and resources
- **Selection at seedling stage**
  - Important for traits such as grain quality
  - Can select before transplanting
- **Increased reliability**
  - No environmental effects
  - Can discriminate between homozygotes and heterozygotes and select single plants

Potential benefits from MAS

- more accurate and efficient selection of specific genotypes
  - May lead to accelerated variety development
- more efficient use of resources
  - Especially field trials

Crossing house

Backcross nursery

Overview of ‘marker genotyping’

(1) LEAF TISSUE SAMPLING

↓

(2) DNA EXTRACTION

↓

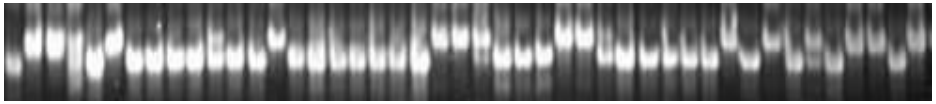
(3) PCR

↓

(4) GEL ELECTROPHORESIS

↓

(5) MARKER ANALYSIS



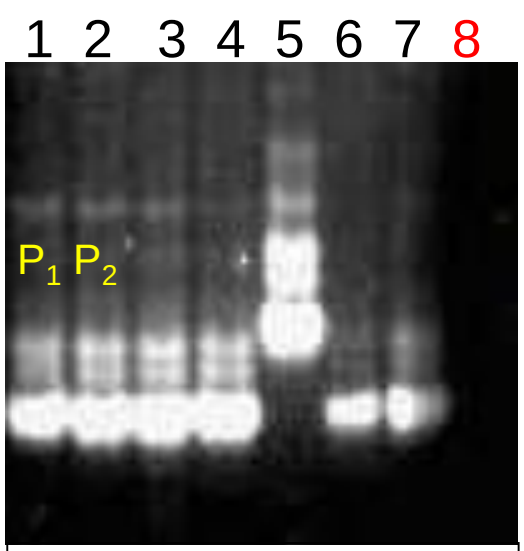
Markers

- What makes a good marker:
  - co-dominant (so homozygotes and heterozygotes can be distinguished)
  - many alleles at each locus (so most individuals will be heterozygous and different from each other)
  - many loci well distributed throughout the genome
  - easy to detect, especially with automated machinery
- No system is perfect

Feature	RFLPs	RAPDs	AFLPs	SSRs	SNPs
DNA required (g)	10	0.02	0.5-1.0	0.05	0.05
DNA quality	High	High	Moderate	Moderate	High
PCR-based	No	Yes	Yes	Yes	Yes
Number of polymorph loci analyzed	1.0-3.0	1.5-50	20-100	1.0-3.0	1.0
Ease of use	Not easy	Easy	Easy	Easy	Easy
Amenable to automation	Low	Moderate	Moderate	High	High
Reproducibility	High	Unreliable	High	High	High
Development cost	Low	Low	Moderate	High	High
Cost per analysis	High	Low	Moderate	Low	Low

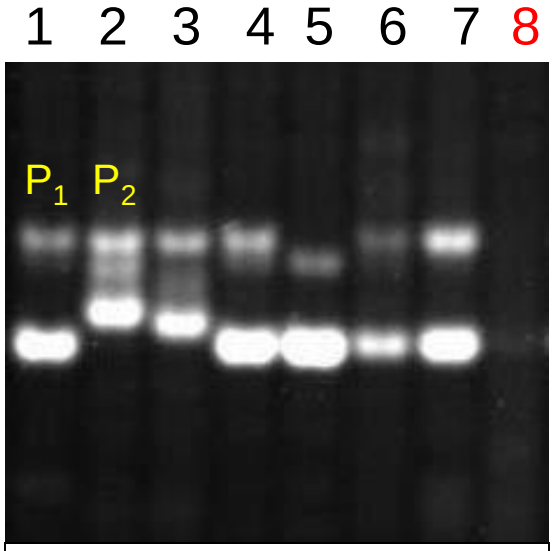
Markers *must* be polymorphic

RM84



Not polymorphic

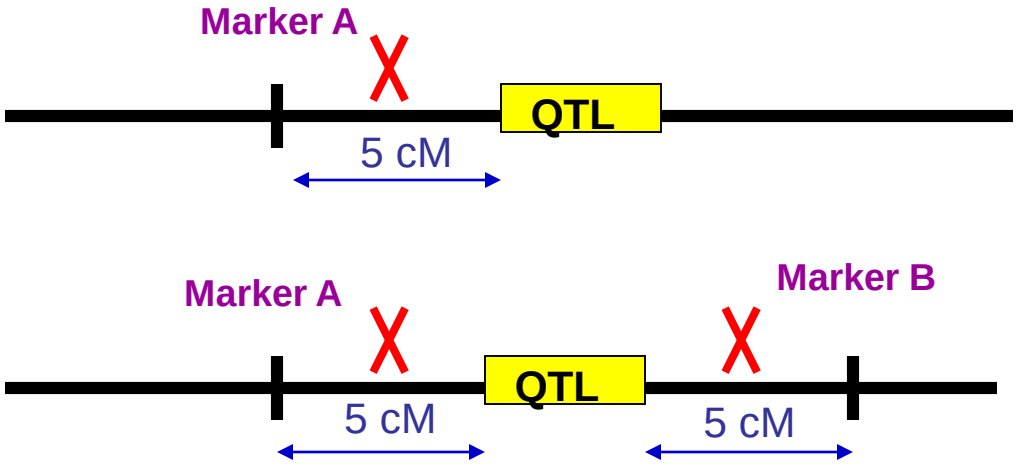
RM296



Polymorphic!

Markers must be tightly-linked to target loci!

- Ideally markers should be <5 cM from a gene or QTL



RELIABILITY FOR SELECTION

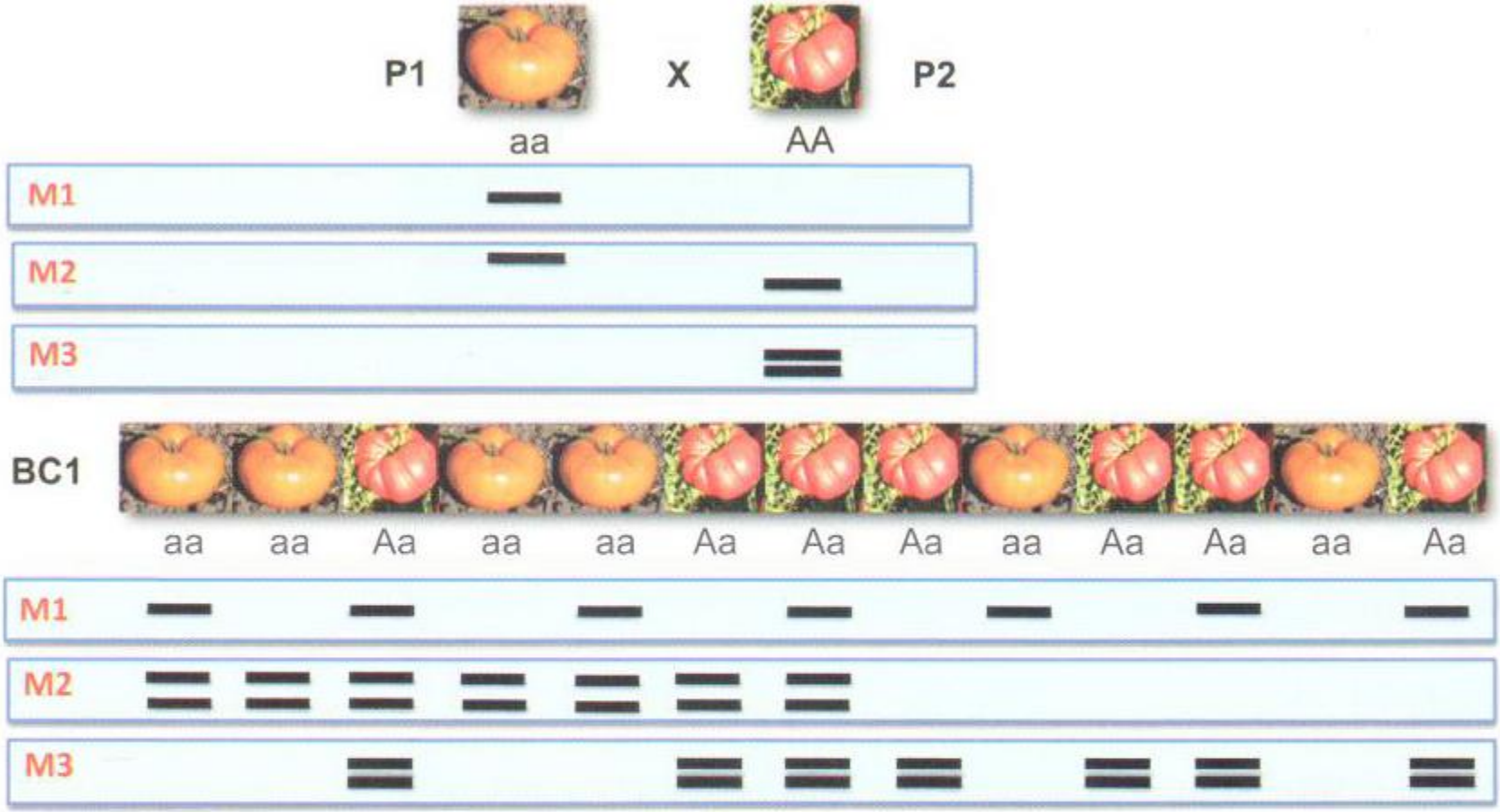
Using marker A only:

$1 - r_A$ = ~**95%**

Using markers A and B:

$1 - 2\, r_A r_B$ = ~**99.5%**

- Using a pair of flanking markers can greatly improve reliability but increases time and cost



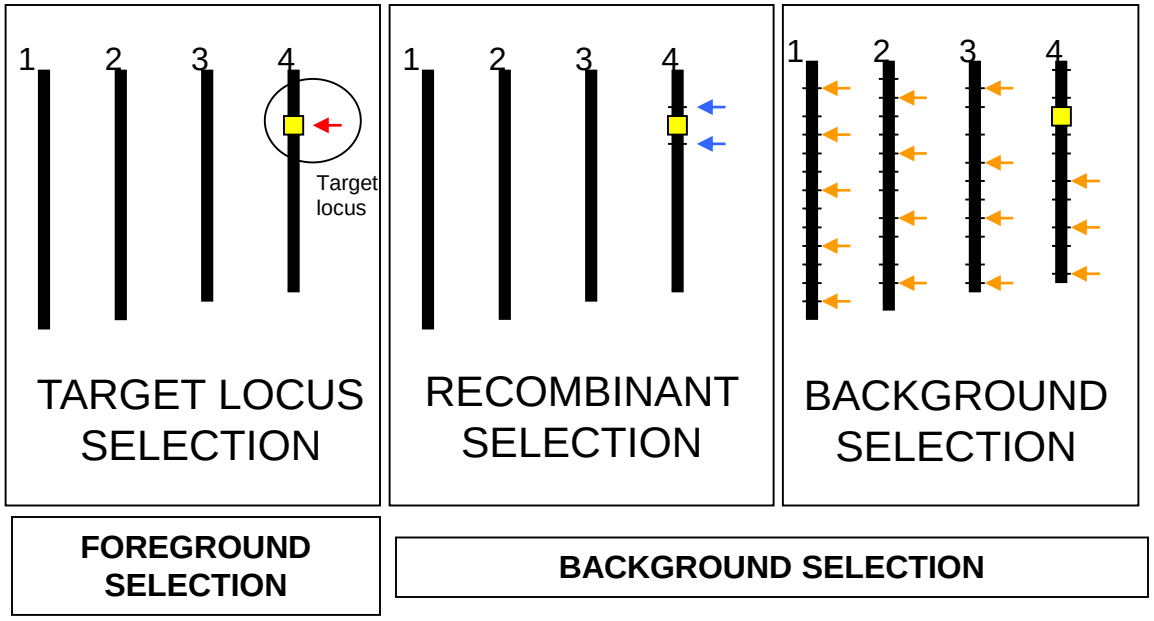
	Rosso Banda +	Rosso Banda-	Giallo Banda +	Giallo Banda -
M1	4	3	3	3
M2	3	4	4	2
M3	7	0	0	6

MAS BREEDING SCHEMES

1. Marker-assisted backcrossing
2. Pyramiding
3. Early generation selection
4. ‘Combined’ approaches

2.1 Marker-assisted backcrossing (MAB)

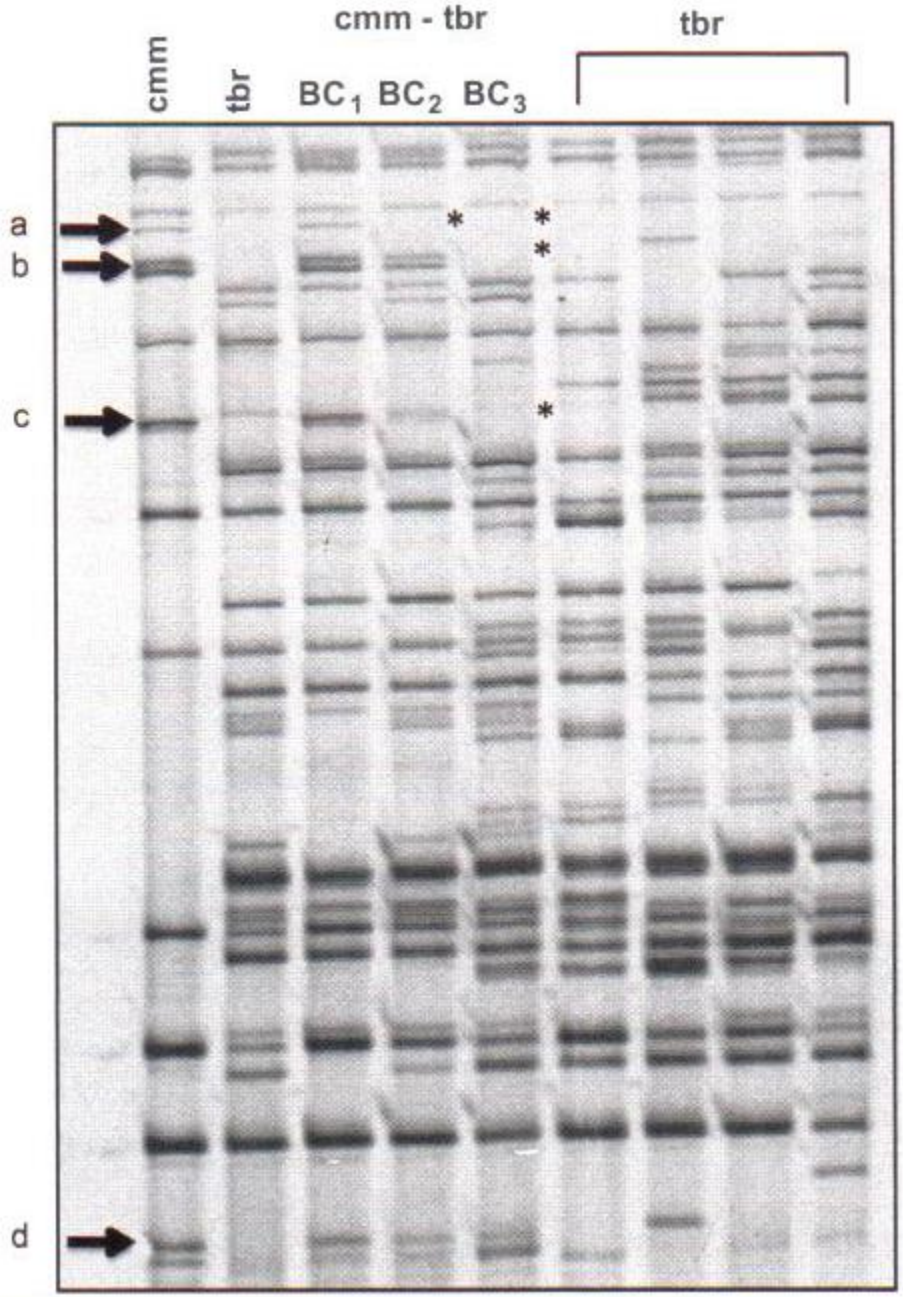
- MAB has several advantages over conventional backcrossing:
  - Effective selection of target loci
  - Minimize linkage drag
  - Accelerated recovery of recurrent parent

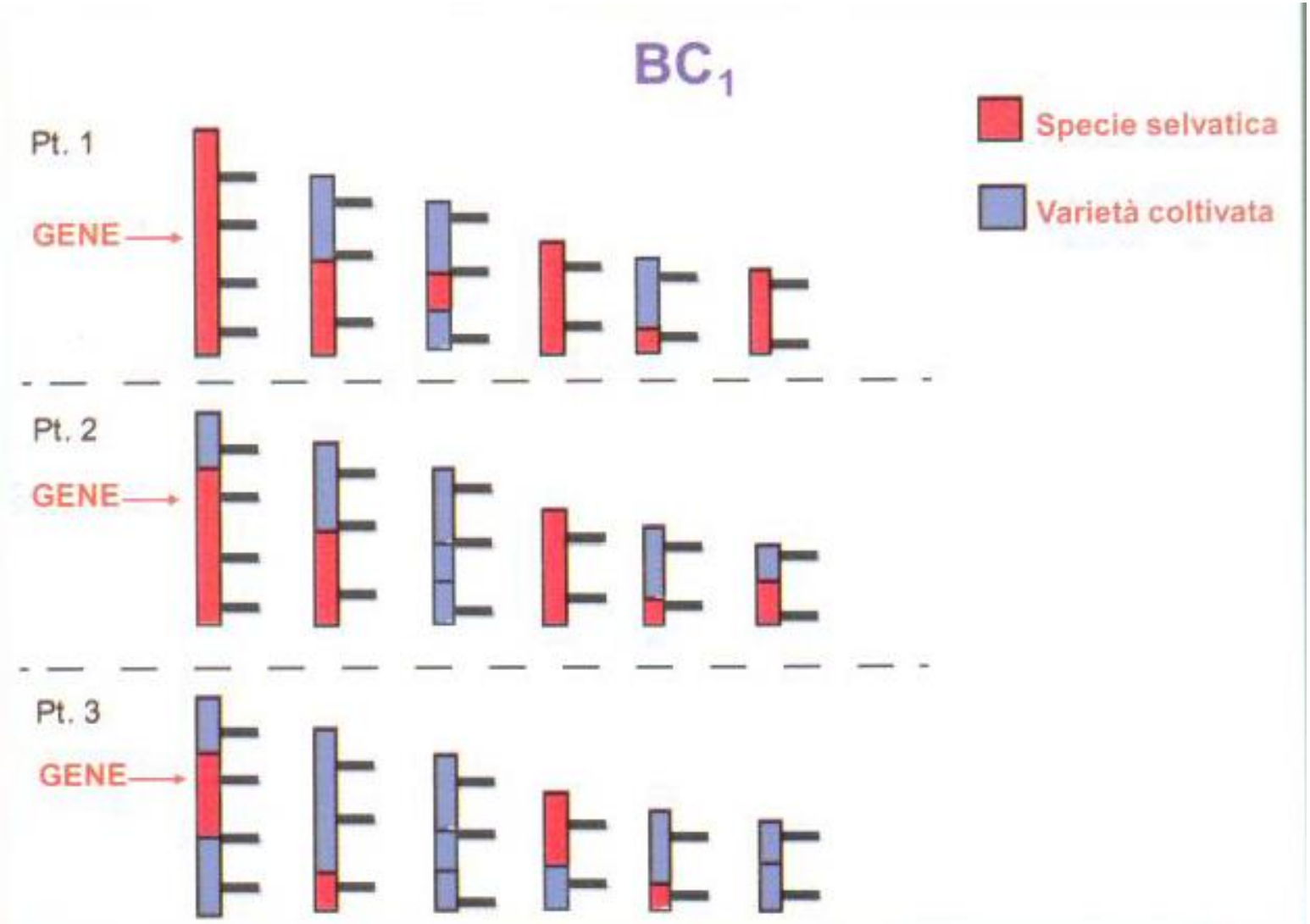


Negative selection

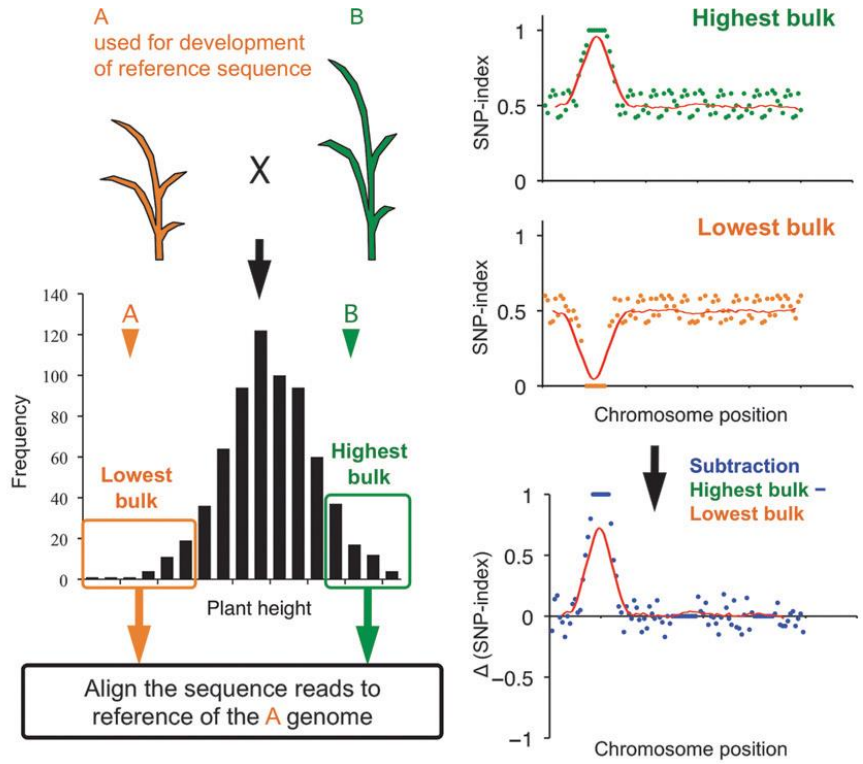
Autogamic species: 99% genome of one parental (recurrent parental genome) recovered after 6 generations of selfing

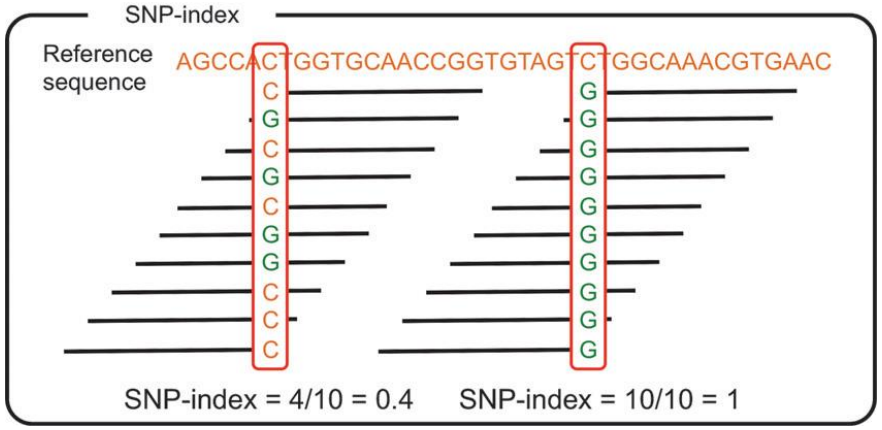
Using MAS, the same % of genome can be recovered in 3 generations (using markers widely and homogenously distributed in the genome)


cmm - tbr
tbr
cmm
tbr
BC1
BC2
BC3
a
b
c
d
*
*
*
*


BC1
Specie selvatica
Varietà coltivata
Pt. 1
GENE
Pt. 2
GENE
Pt. 3
GENE

Bulked segregant analysis for QTLs





2.2 Pyramiding

- Widely used for combining multiple disease resistance genes for specific races of a pathogen
- Pyramiding is extremely difficult to achieve using conventional methods
  - Consider: phenotyping a single plant for multiple forms of seedling resistance – almost impossible
- Important to develop ‘durable’ disease resistance against different races

- Process of combining several genes, usually from 2 different parents, together into a single genotype

Breeding plan

P_1 x P_1

Gene A **Gene B**

↓

F_1

Gene A + B

↓

F_2

↓

MAS

↓

Select F2 plants that have

Gene A and Gene B

Genotypes

P_1: **AA**bb x P_2: aa**BB**

↓

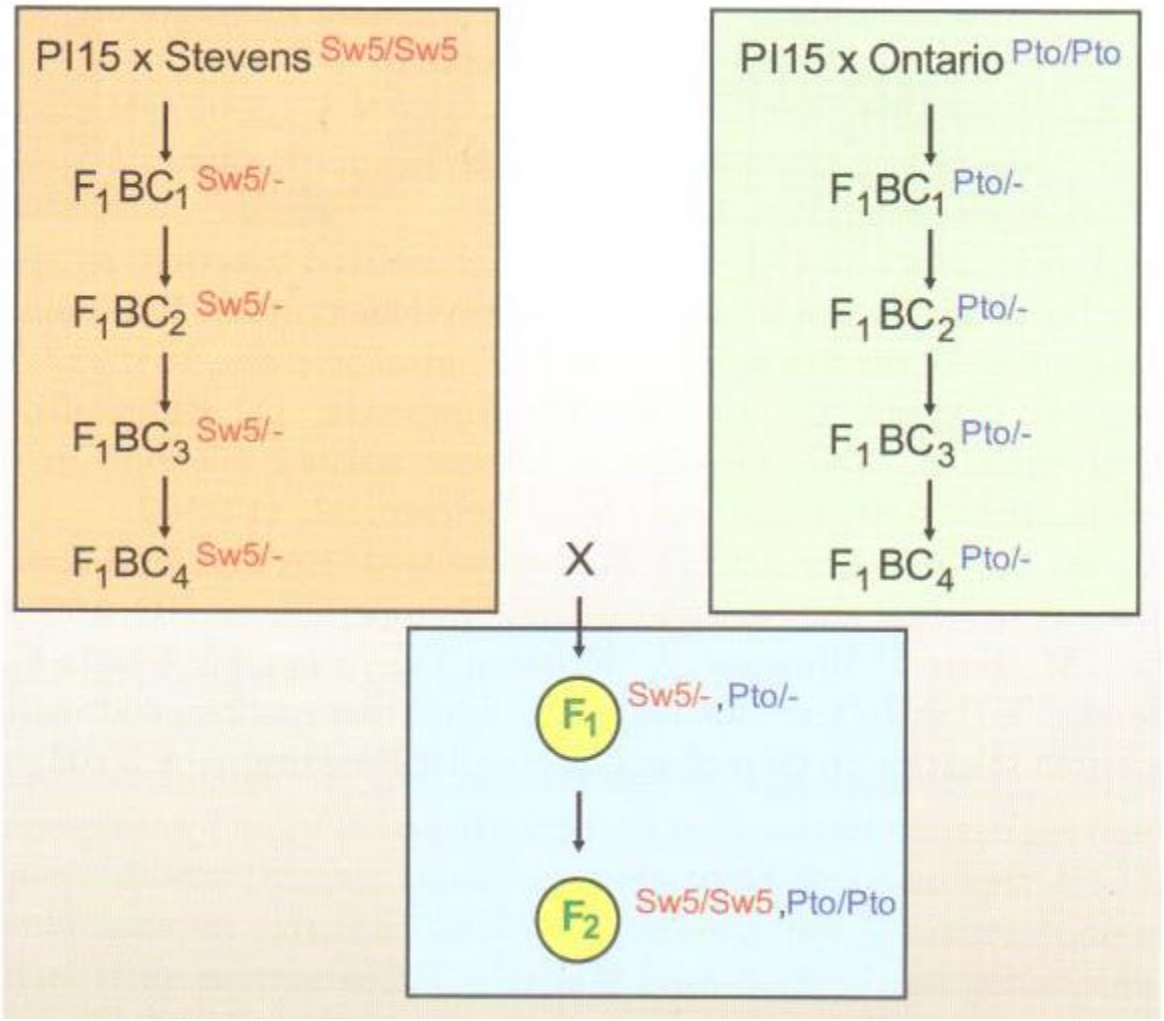
F_1: **Aa**B**b**

↓

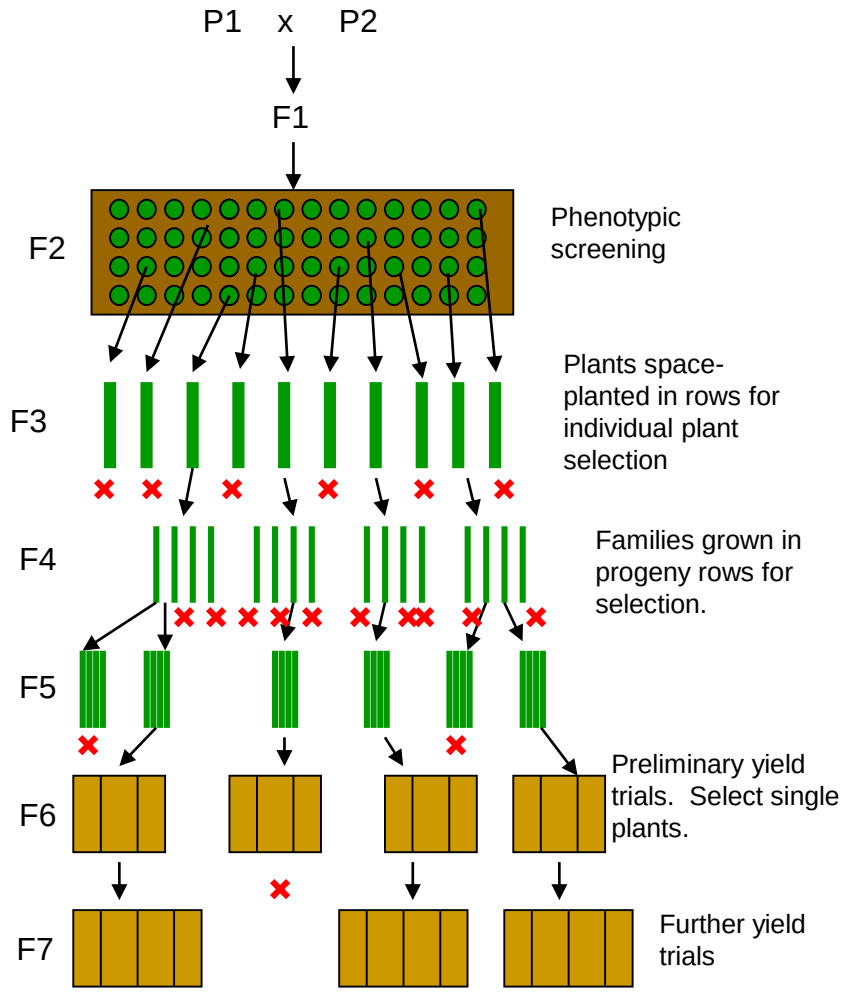
F_2	**AB**	**Ab**	**aB**	**ab**
AB	**AABB**	AABb	AaBB	AaBb
Ab	AABb	AAbb	AaBb	Aabb
aB	AaBB	AaBb	aaBB	aaBb
ab	AaBb	Aabb	aaBb	aabb

Hittalmani et al. (2000). Fine mapping and DNA marker-assisted pyramiding of the three major genes for blast resistance in riceTheor. Appl. Genet. 100: 1121-1128

Liu et al. (2000). Molecular marker-facilitated pyramiding of different genes for powdery mildew resistance in wheat. Plant Breeding 119: 21-24.

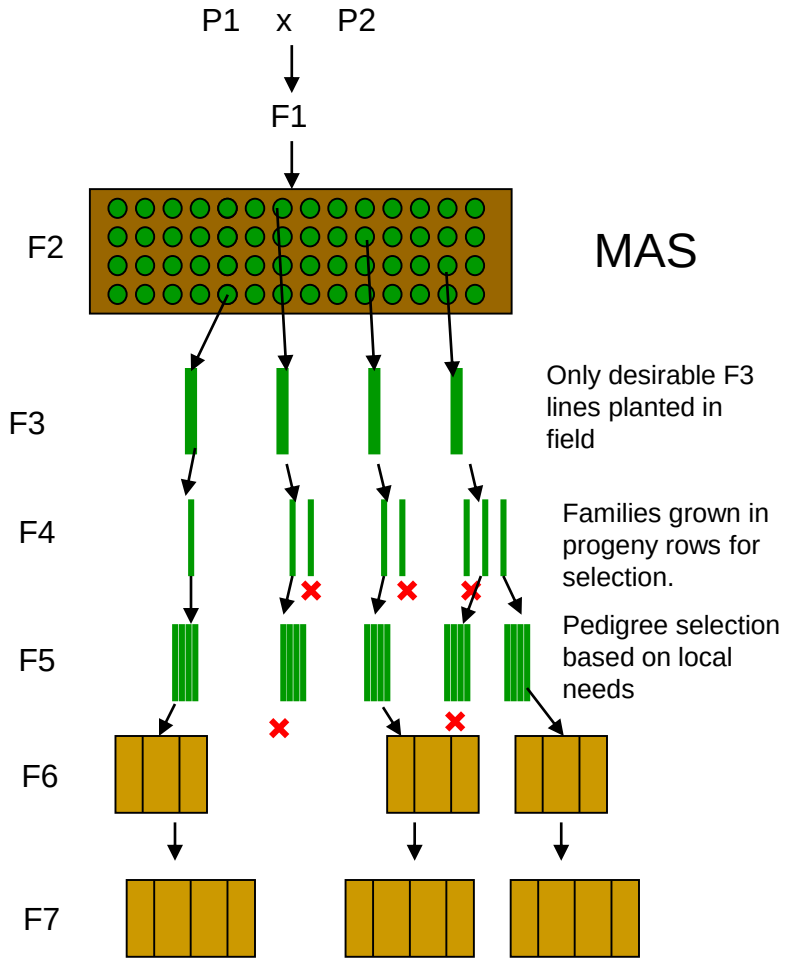

PI15 x Stevens Sw5/Sw5
F1BC1 Sw5/-
F1BC2 Sw5/-
F1BC3 Sw5/-
F1BC4 Sw5/-
PI15 x Ontario Pto/Pto
F1BC1 Pto/-
F1BC2 Pto/-
F1BC3 Pto/-
F1BC4 Pto/-
X
F1
Sw5/- , Pto/-
F2
Sw5/Sw5 , Pto/Pto

PEDIGREE METHOD



F8 – F12 Multi-location testing, licensing, seed increase and cultivar release

SINGLE-LARGE SCALE MARKER-ASSISTED SELECTION (SLS-MAS)



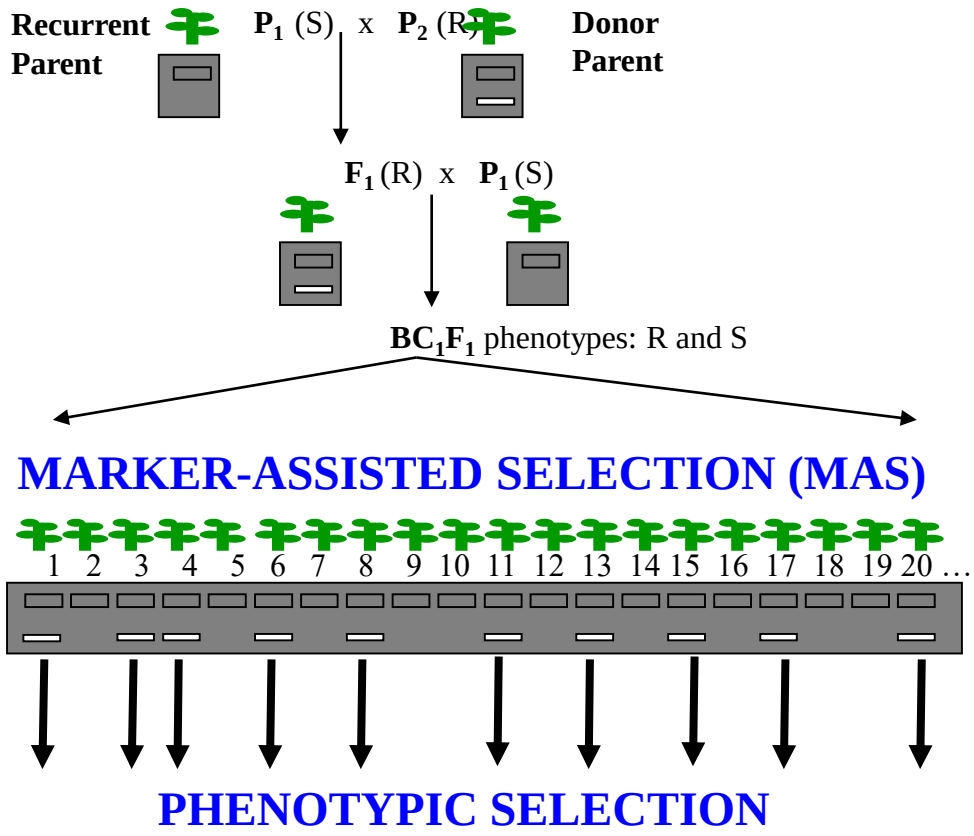
F8 – F12 Multi-location testing, licensing, seed increase and cultivar release

Benefits: breeding program can be efficiently scaled down to focus on fewer lines

- In some cases, a combination of phenotypic screening *and* MAS approach may be useful
  1. To maximize genetic gain (when some QTLs have been unidentified from QTL mapping)
  2. Level of recombination between marker and QTL (in other words marker is not 100% accurate)
  3. To reduce population sizes for traits where marker genotyping is cheaper or easier than phenotypic screening

‘Marker-directed’ phenotyping

(Also called ‘tandem selection’)



- Use when markers are not 100% accurate or when phenotypic screening is more expensive compared to marker genotyping

SAVE TIME & REDUCE COSTS

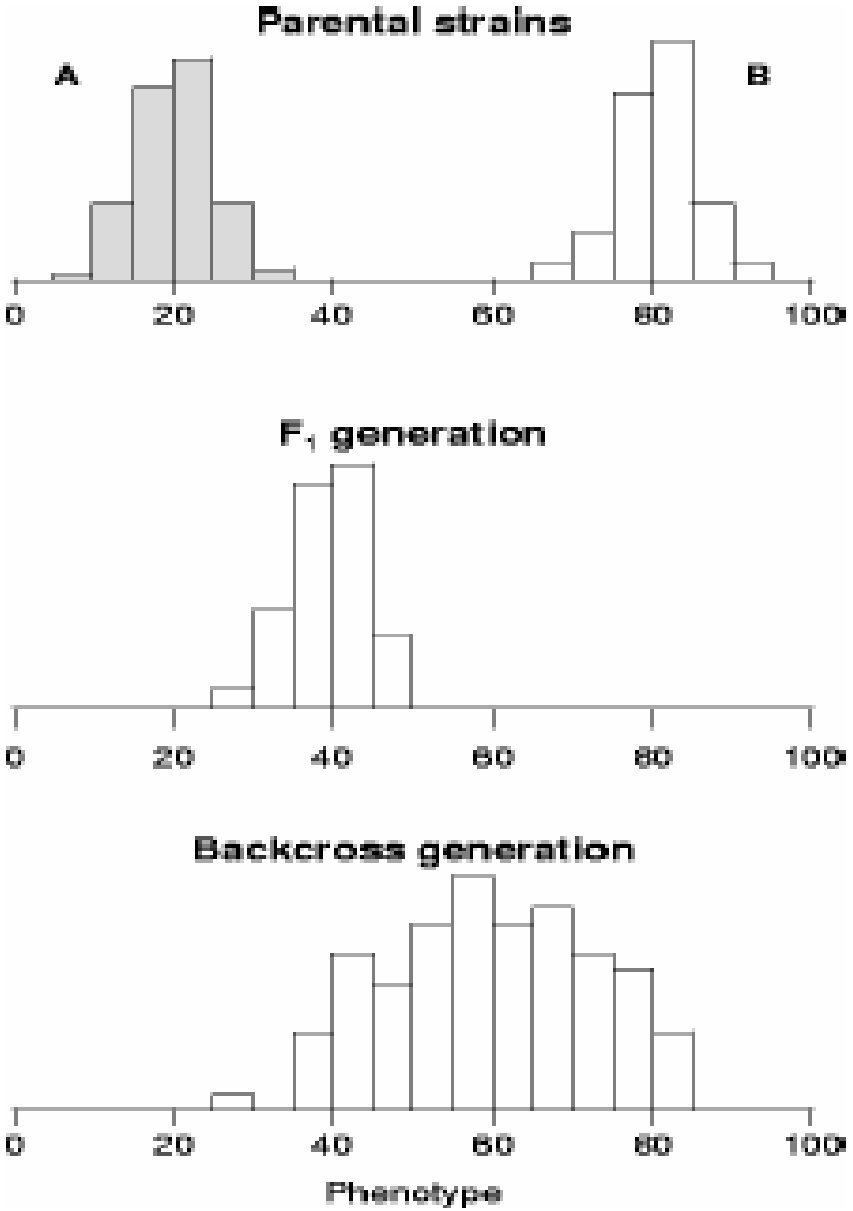
Especially for quality traits

References:

Han et al (1997). Molecular marker-assisted selection for malting quality traits in barley. Mol Breeding 6: 427-437.

Quantitative trait loci (QTLs)

- QTLs determine the genetic component of variation in quantitative traits.
- Quantitative traits are usually encoded by many genes (polygenes).

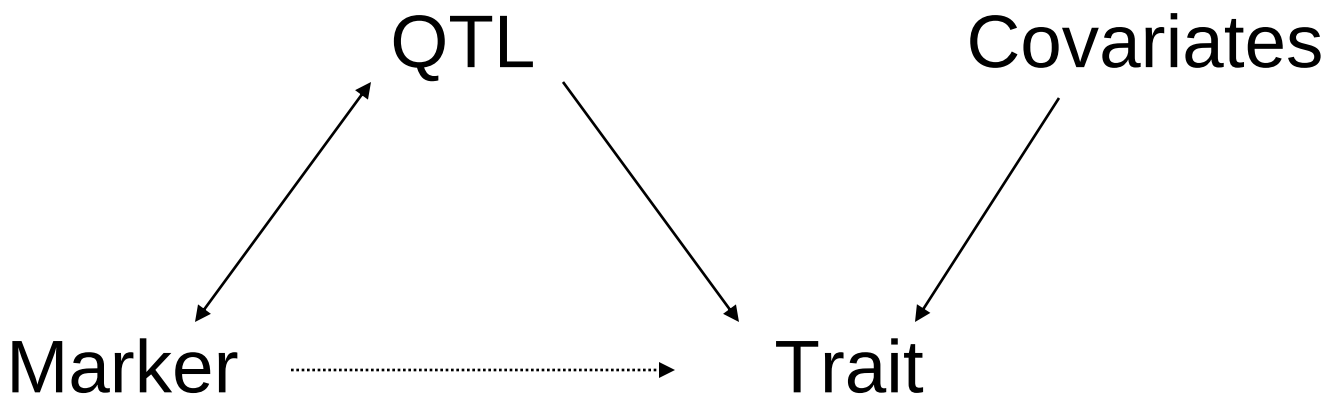

Parental strains
A
B
0
20
40
60
80
100
F1 generation
0
20
40
60
80
100
Backcross generation
0
20
40
60
80
100
Phenotype

Goals of QTL analysis

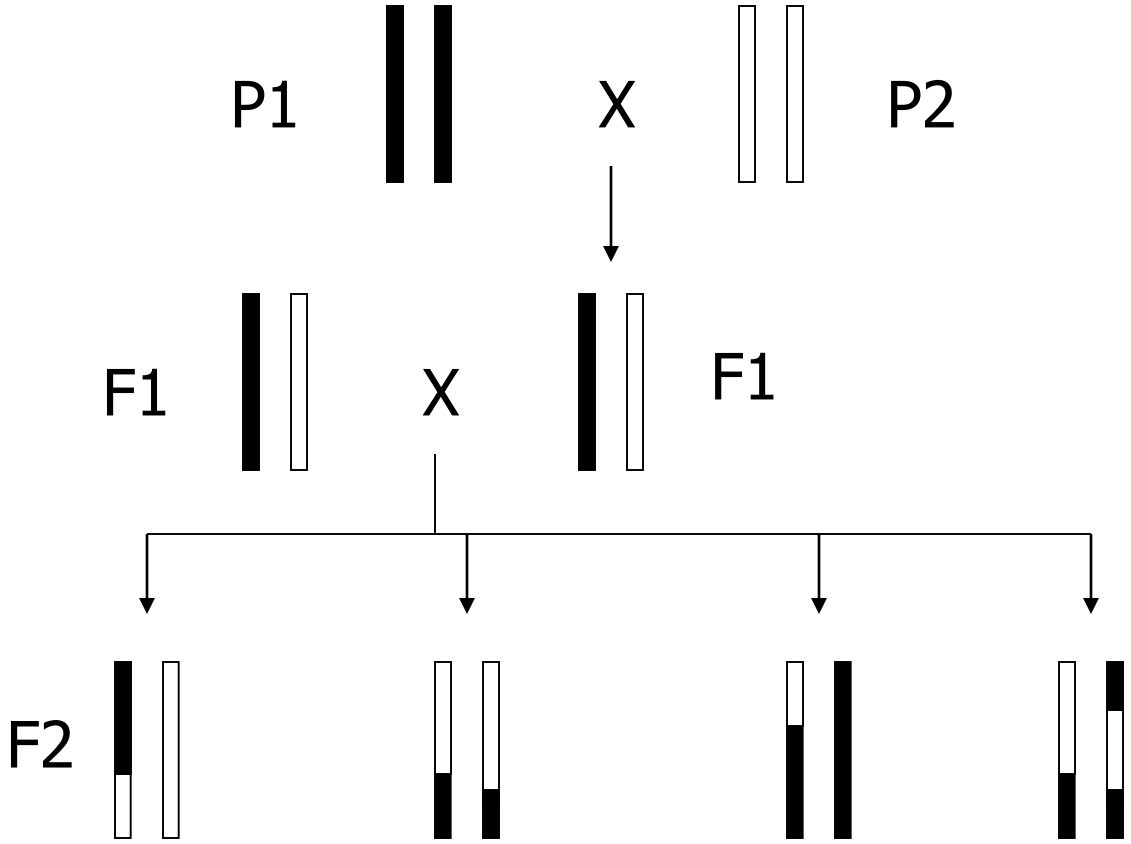
- Detect genetic effects
- QTL mapping: inference of the QTL location on chromosome

QTL mapping in experimental crosses

Experimental crossing creates associations between genetic marker loci and traits to allow localization of QTL.



Intercross



Data structure for a backcross experiment

- Phenotypes:

  y_i = quantitative measurement of trait

- Genotypes:

  x_{ij} = 0/1 coded for AA/AB at marker j

- Covariates:

  $\mathbf{Z}_i$ = environmental factors, demographics, etc.

where i = 1, …, n; j = 1, …, M.

Model and assumptions

- No interference in the recombination process
- Independence
- Normal distribution

$$y_i|X \sim N(\mu_X, \sigma_X^2)$$

- Homoscedasticity (constant variance)

$$\sigma_X^2 = \sigma^2$$

LOD curve

- Likelihood profile (profilo di verosimiglianza)
- A clear peak is taken as the QTL
- 1.5-LOD support interval

Breeders' QTL mapping 'checklist'

- LOD & R^2 values will give us a good initial idea but probably more important factors include:

1. *What is the population size used for QTL mapping?*
2. *How reliable is the phenotypic data?*
   - *Heritability estimates will be useful*
   - *Level of replication*
3. *Any confirmation of QTL results?*
4. *Have effects of genetic background been tested?*
5. *Are markers polymorphic in breeders' material?*
6. *How useful are the markers for predicting phenotype? Has this been evaluated?*

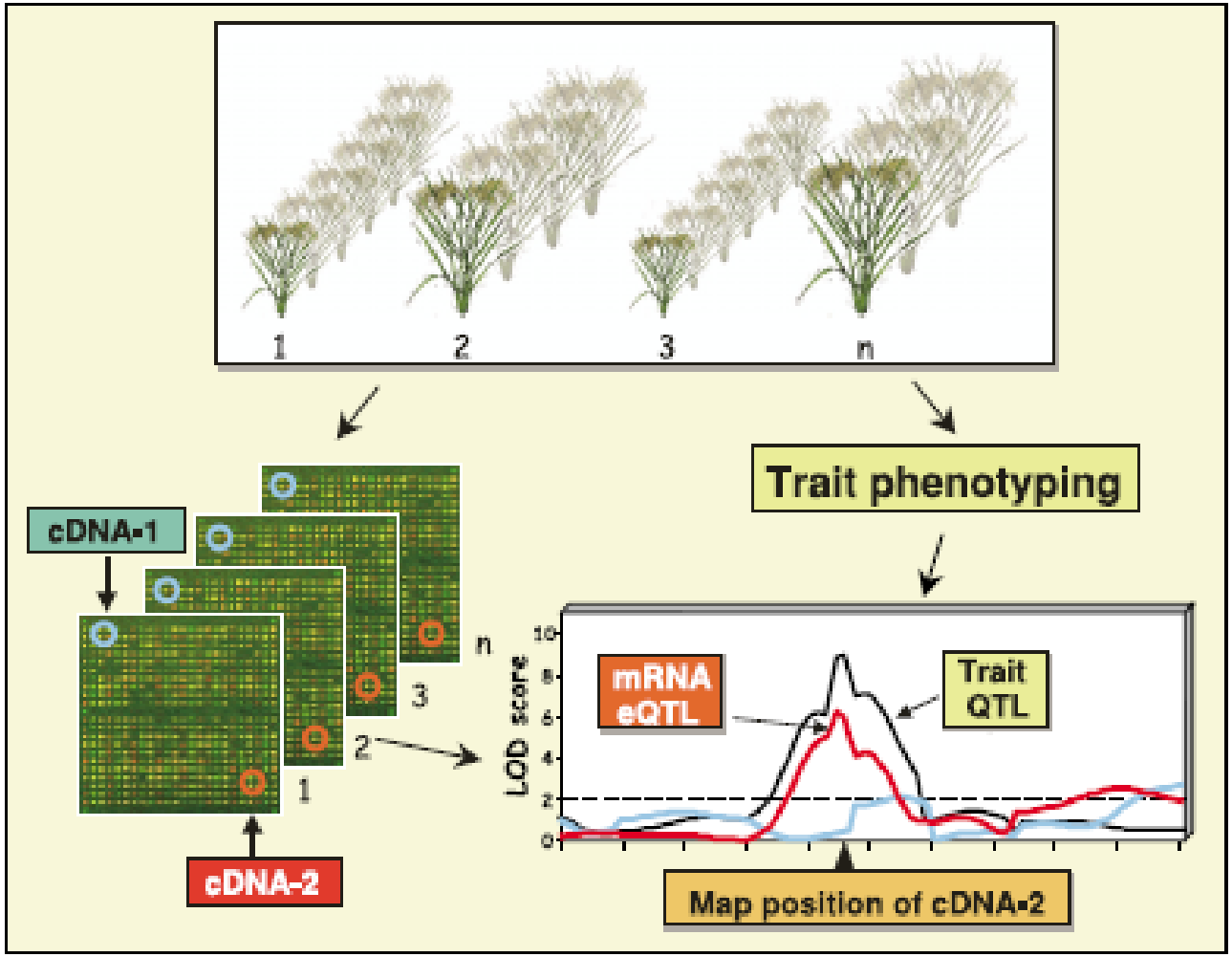


Figure 5. Expression profiling of a mapping population at the mRNA level via microarray analysis to identify expression QTLs (eQTLs) for specific cDNAs. Correspondence between an eQTL peak for a specific cDNA (e.g. cDNA-2) and a QTL peak for a trait causally linked to the function of the protein encoded by the cDNA provides circumstantial evidence supporting the role of the cDNA as a candidate gene for the target trait.

Current status of molecular breeding

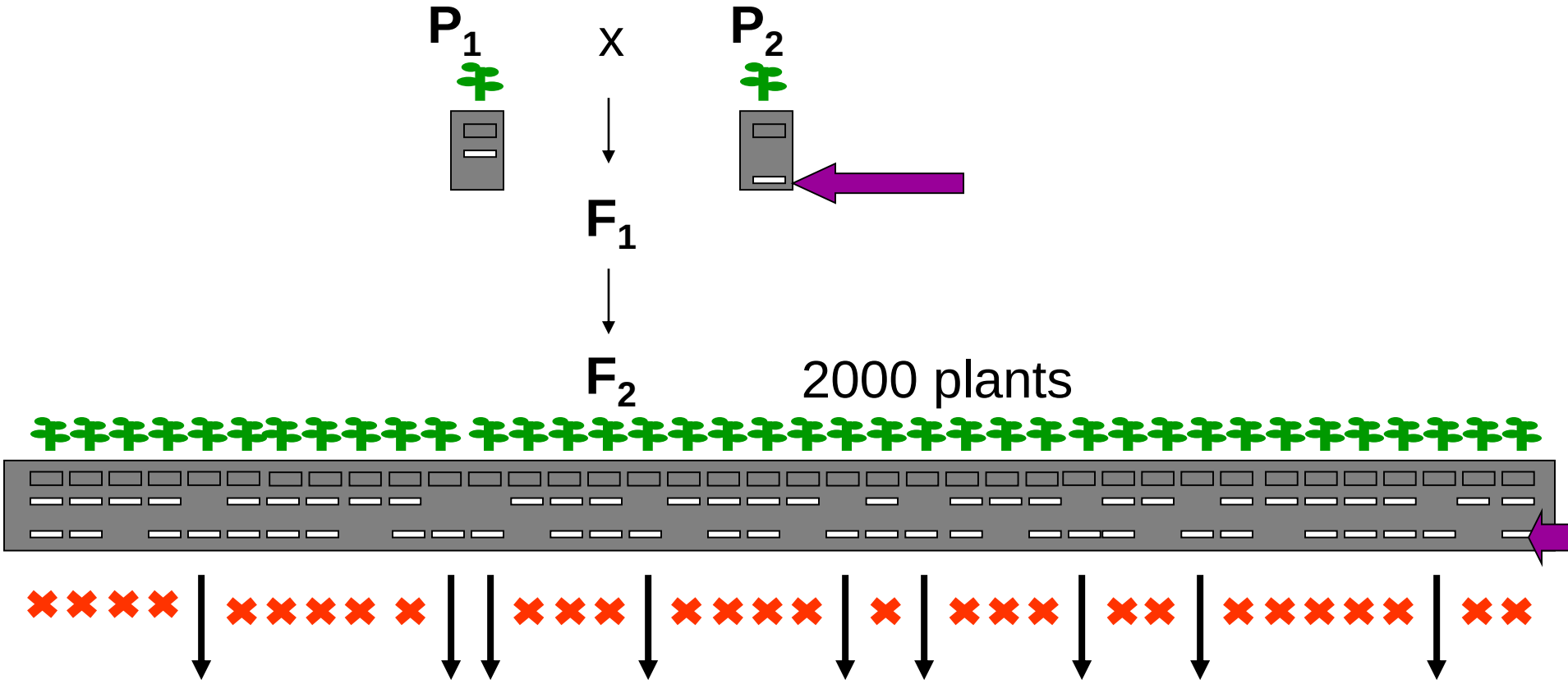
- A literature review indicates thousands of QTL mapping studies but not many *actual* reports of the application of MAS in breeding

- *Why is this the case?*

Some possible reasons to explain the low impact of MAS in crop improvement

- Resources (equipment) not available
- Markers may *not* be cost-effective
- Accuracy of QTL mapping studies
- QTL effects may depend on genetic background or be influenced by environmental conditions
- Lack of marker polymorphism in breeding material
- Poor integration of molecular genetics and conventional breeding

Cost of MAS in context: Example 1: Early generation MAS



USD **$640** to screen 2000 plants with a single marker for one population

Reliability of QTL mapping is critical to the success of MAS

- Reliable phenotypic data critical!
  - Multiple replications and environments
- Confirmation of QTL results in independent populations
- “Marker validation” must be performed
  - Testing reliability for markers to predict phenotype
  - Testing level of polymorphism of markers
- Effects of genetic background need to be determined