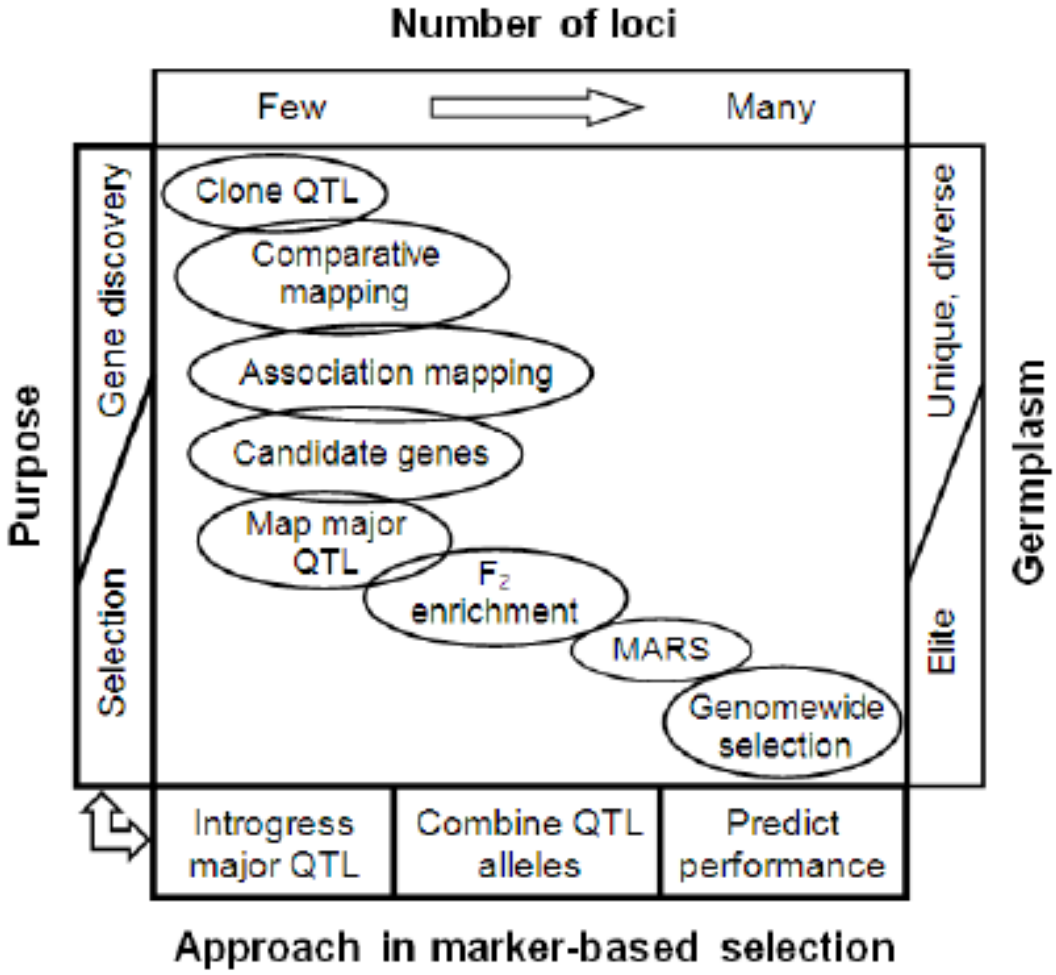
Lecture 7
Marker Assisted Selection and Genomic Selection

Lucia Gutierrez

Tucson Winter Institute

Molecular Breeding



Bernardo, 2008

Using Molecular Markers

GENES OF MAJOR EFFECT

In several species, genes of major effects have been identified and could be easily introgressed into elite cultivars. But why not using phenotypic selection?

- Recessive genes where progeny testing is needed
- Expressed after flowering time
- Expensive or difficult to evaluate

USING MOLECULAR MARKERS

Some of the advantages of using molecular markers instead of phenotypes to select are:

- Early selection (at seedling, or even for seeds)
- Reduced cost (fewer plants, shorter time)
- Reduced cycle time (if gene is recessive or measured after flowering)
- Screening more efficient (if it is a complex trait)

Moreaux, 2011

Genes vs. QTL for MAS

GENES

If the marker is on the gene or closely linked to it, selection on marker is straightforward (Bonnet et al., 2005). Flanking markers are advised if closely linked. Need to identify markers linked to the gene.

QTL

If the marker is associated to a QTL, there is a chance we will loose the gene by selecting on the marker. Flanking markers, and selecting on more than 2 markers is advised (Wang et al., 2007). Additional considerations should be taken into account: QTL effects small, inacurate positions, need to be validated.

Steps required:

- Identify potential sources of useful QTL alleles.
- Find markers closely linked to QTL of major effect.
- Confirm the effect of the QTL on different genetic backgrounds.
- Deploy the QTL in the breeding program.

Marker Assisted Selection (MAS)

FOREGROUND SELECTION

Use markers to transfer genes or QTL of major effects. One or multiple genes may be transferred. Markers should be closely linked to the gene of interest to avoid loosing them by recombination.

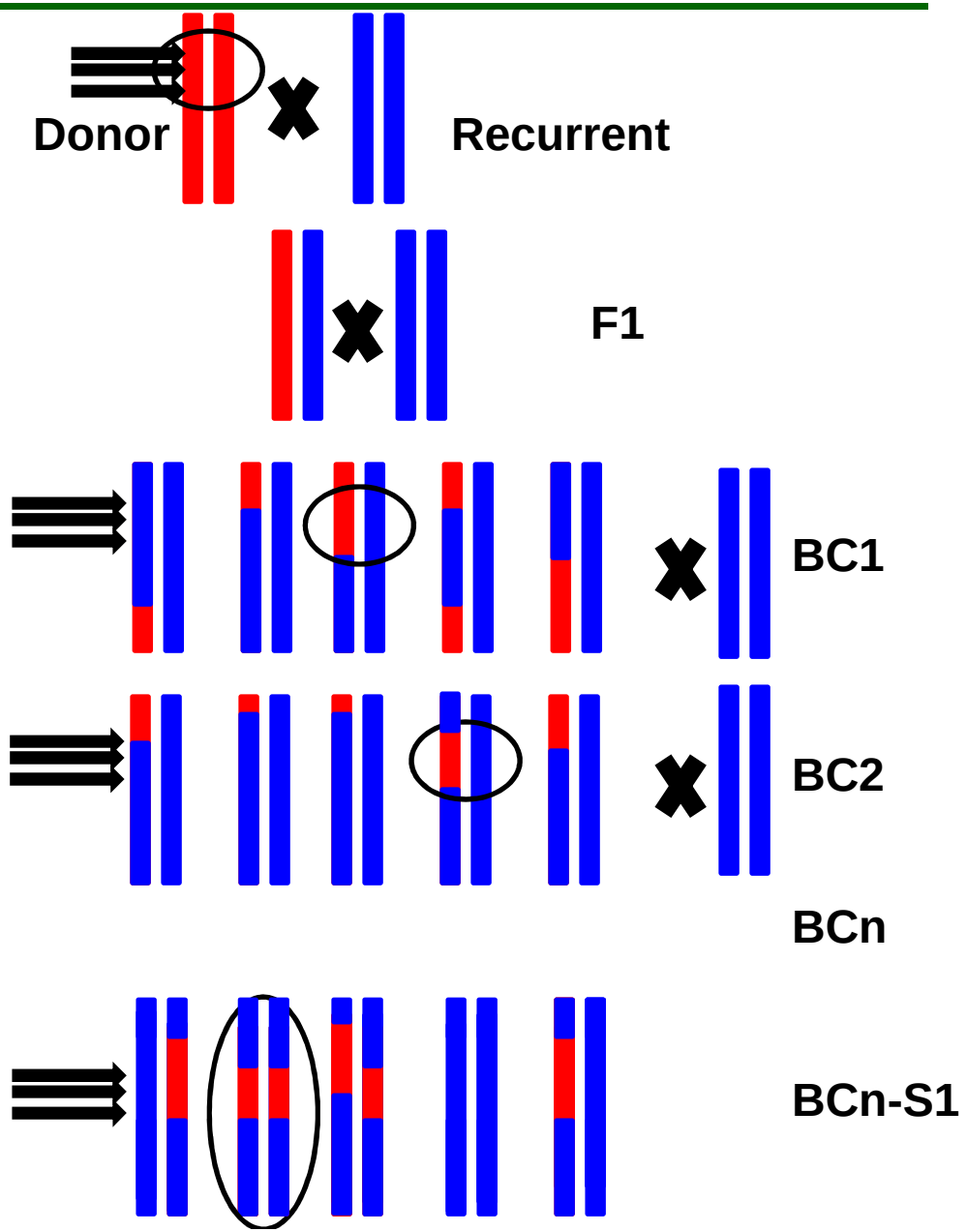
BACKGROUND SELECTION

Use markers to control for genetic background in a BC cycle. To speed the process of recovery of the elite germplasm, markers may be used along the genome.

Foreground Selection

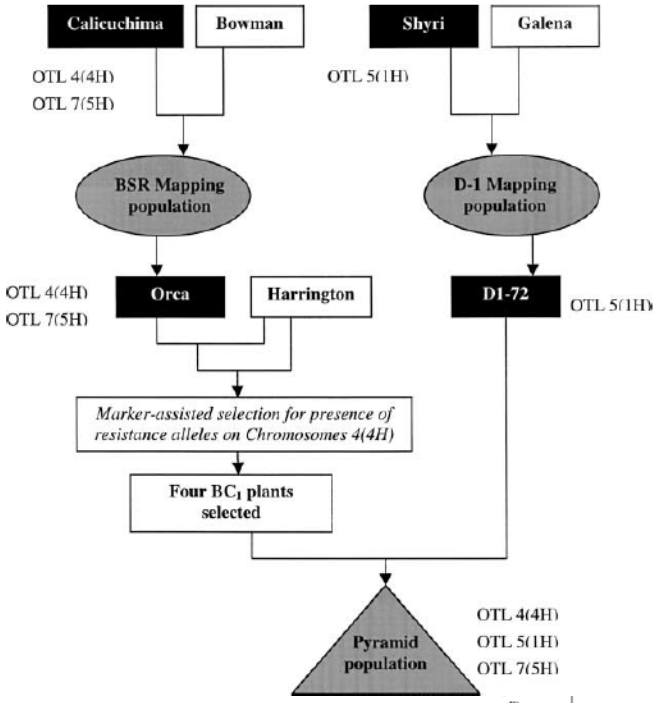
MARKER ASSISTED BACK-CROSS (MABC)

On each cycle, selection of individuals with favorable alleles at the loci of interest is performed. In the final step, after selfing, selection of homozygote individuals at the loci of interest is performed.



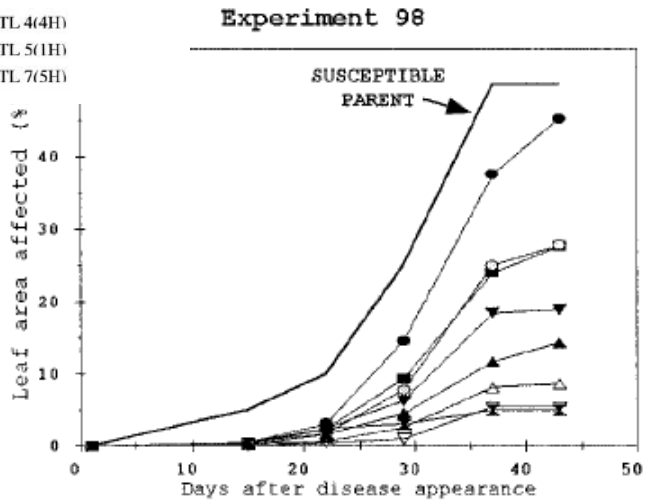
Foreground Selection

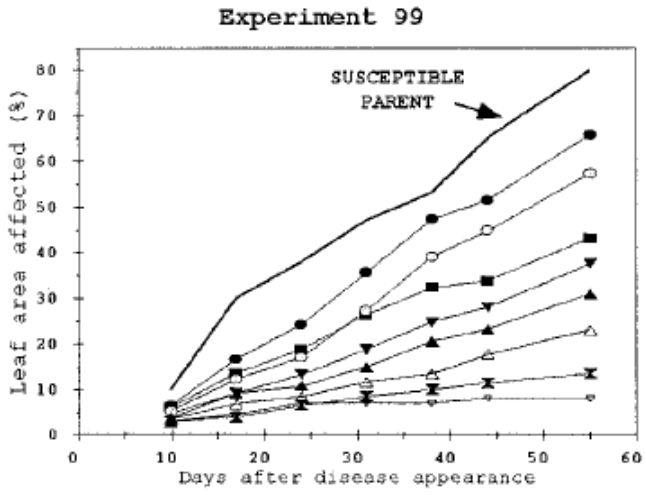
GENE PYRAMIDING (MAJOR QTL)



QTL identification and validation is necessary before conducting a MAS. Castro et al., 2003 detected 3 QTL for quantitative resistance to barley stripe rust and used a MAS to introgress them into a desirable line.

● No QTL ▼ 4(4H) △ 4(4H) + 5(1H) ▲ 5(1H) + 7(5H)
○ 7(5H) ■ 5(1H) ⧗ 4(4H) + 7(5H) ▽ 4(4H) + 5(1H) + 7(5H)





Castro et al., 2003

Foreground Selection

POPULATION SIZE

Population size required in a MAS program is a function of the desired probability of not having at least one targeted genotype in the population and the frequency of targeted genotypes in the population is (Wang et al., 2007):

$$N = \frac{\log \alpha}{\log(1 - f)}$$

Population size needed for transferring multiple genes simultaneously is large. The minimum population size required in BC1 population to recover desirable genotype for different number of QTL (q) and one (m=1) or two (m=2) flanking marker is shown below (Hospital and Charcosset, 1997):

q	N(m=1)	N(m=2)
2	17	19
3	35	44
4	72	95
5	146	207
6	293	445

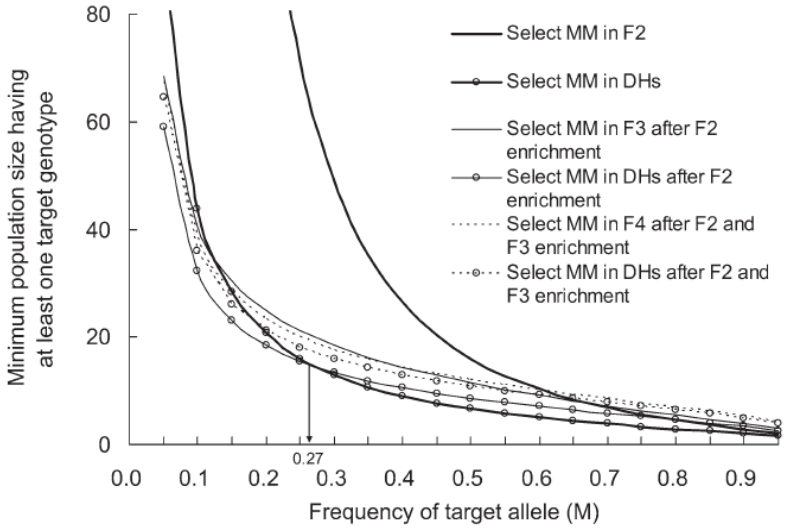
Foreground Selection

F2 ENRICHMENT

An alternative to increase proportion of favorable alleles before selfing is to “enrich” early generations by selecting F2 individuals with the targeted alleles in homozygosis and heterozygosis for all QTL. This reduces the population size needed increasing the frequency of favorable alleles (Bonnet et al., 2005).

This is true for several bi-parental crossing schemes (Wang et al., 2007).

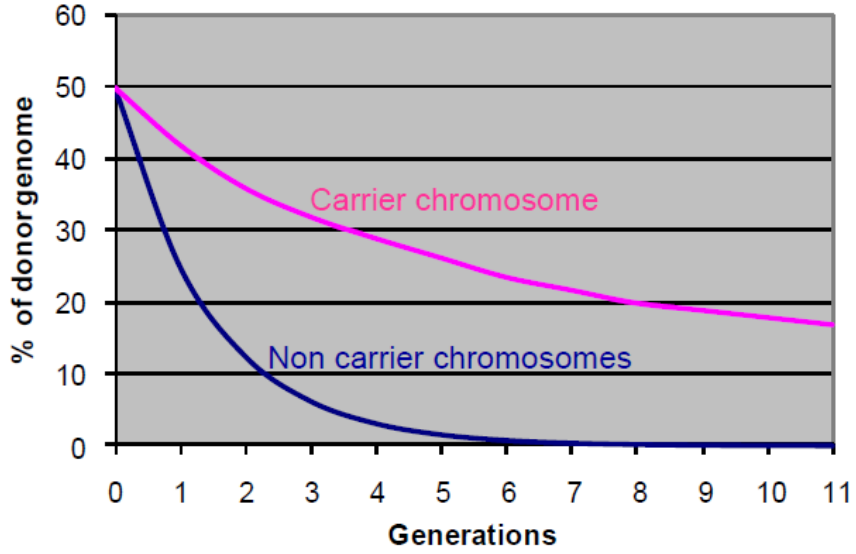
Loci:	Fix	Enrich
1	11	3
2	47	4
3	191	6
4	766	8
5	3067	11
6	12270	16
7	49081	21
8	196327	29
9	785312	39
10	3141252	52



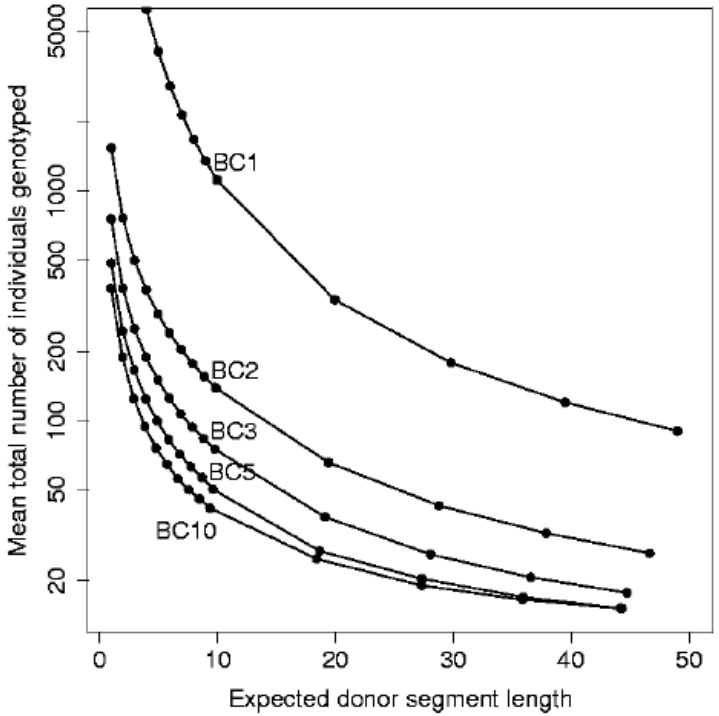
Background selection

MABC BACKGROUND SELECTION

When selecting for markers on a region, linkage drag will determine low recovery of the carrier chromosome.



The number of plants required for small donor fragments is large (Hospital et al., 2001)

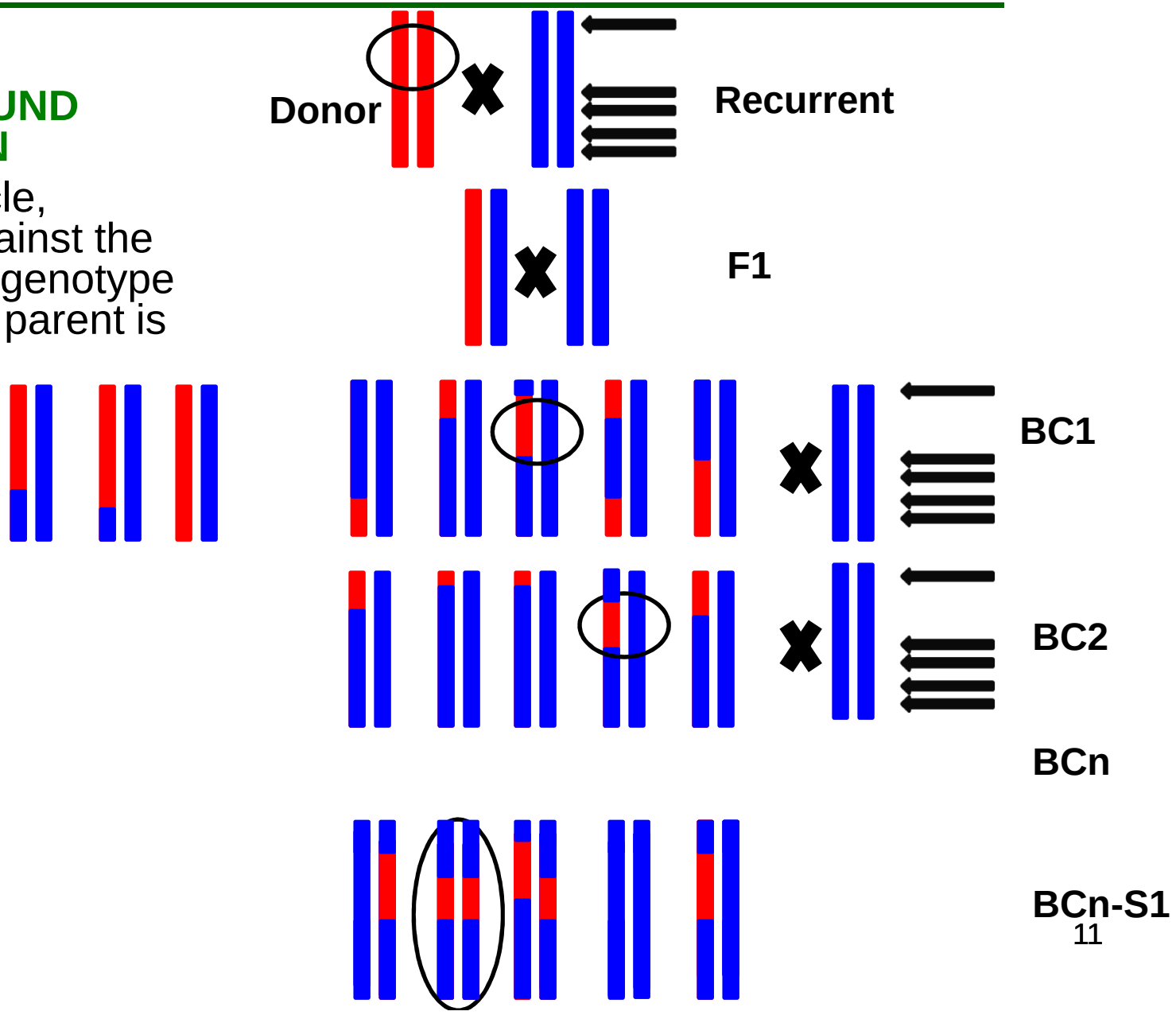


Moreaux, 2011

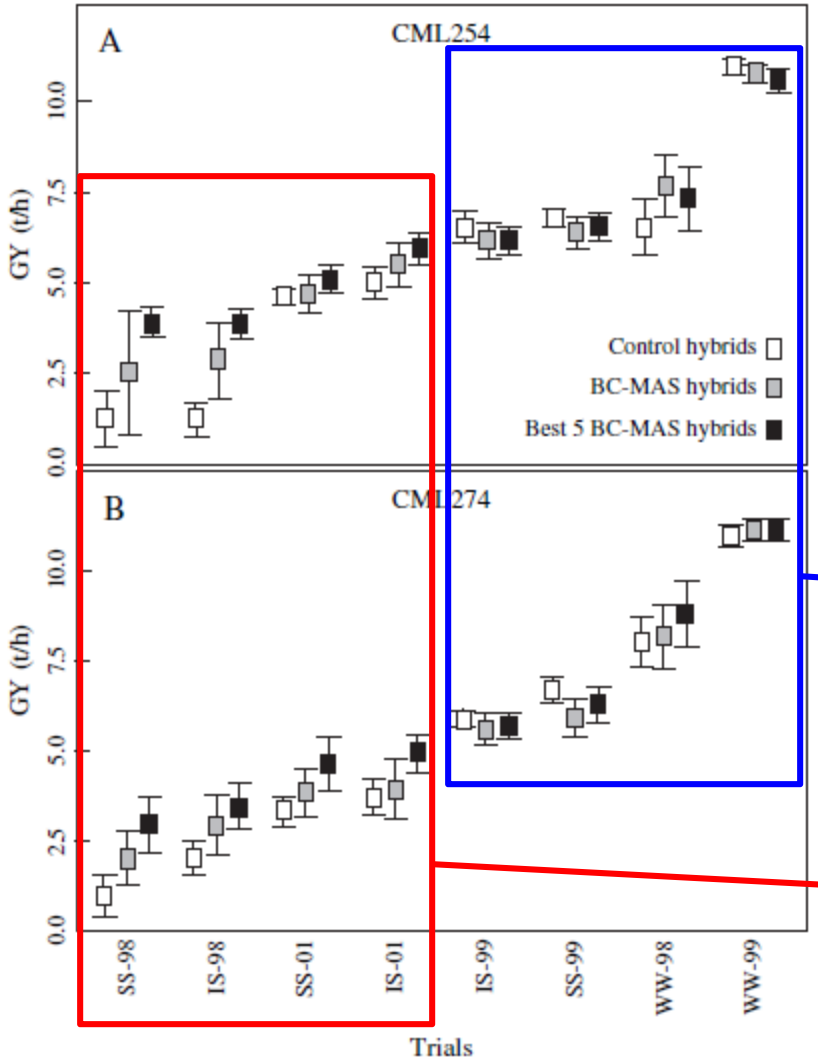
Background selection

MABC BACKGROUND SELECTION

On each cycle, selection against the background genotype of the donor parent is performed.



MABC



Ribaut and Ragot (2007) showed success for grain yield (5 grain yield component QTL) under stress in maize.

However, in general, MABC works really good for few genes or QTL of major effect. It does not work very good for complex traits.

MABC under mild water stress was not different than control

MABC under severe water stress performed better than control

Marker Assisted Recurrent Selection

Markers could be used to assist selection at times where phenotypic selection cannot be performed. Marker-assisted recurrent selection (MARS) is a recurrent selection that utilizes markers (Eathington et al., 2007) to increase gain per unit time.

Steps in a MARS in Maize:

1. MAS in Cycle 0.
    1. Create an F2 (Cycle 0)
    2. Test-cross the F2
    3. Evaluate progeny in multiple environments (with high h^2)
    4. Identify markers associated with trait of interest
    5. Create an index weighting significant markers by their effect using multiple linear regression (Lande and Thompson 1990).
    6. Recombine best progeny (best individuals from Cycle 0)

2. Select in greenhouse or off-season nursery (up to 3 cycles in low h^2 environments).

Bernardo 2010

MARS

Lande and Thompson 1990 equations:

$$G = \sum_q a_q \theta_q$$

G = predicted genetic value

a_q : additive effect at marker q

θ_q : indicator (dummy) variable for marker q

	QTL1	QTL2	QTL3	QTL4	
Indiv.	+4	-3	+2	+1	G
1	1	0	-1	-1	1
2	1	0	1	1	7
3	-1	-1	0	1	2
.					
n	-1	-1	1	0	3

$$G = b_p P + b_m M$$

G : Predicted genetic value

P : Phenotype

M : Marker score

b : weighted coefficients

$$b_p = \frac{h^2 - R_p^2}{1 - R_p^2}$$

$$b_m = \frac{1 - h^2}{1 - R_p^2}$$

$$h^2 = \frac{\sigma_G^2}{\sigma_P^2}$$

R_p^2 : percentage of phenotypic variance explained by QTL

MARS

Relative Efficiency:

$$RE_M = \sqrt{\frac{R_G^2}{h^2}}$$

$$RE_C = \sqrt{\frac{R_G^2}{h^2} + \frac{\left(1 - R_G^2\right)}{1 - h^2 R_G^2}}$$

MARS

Eathington et al., 2007 showed that MARS outperformed conventional methods.

Table 1. Comparison of multiple trait index (MTI) values following one year of marker assisted recurrent selection (MARS) (three cycles) and conventional selection (two cycles) in corn.

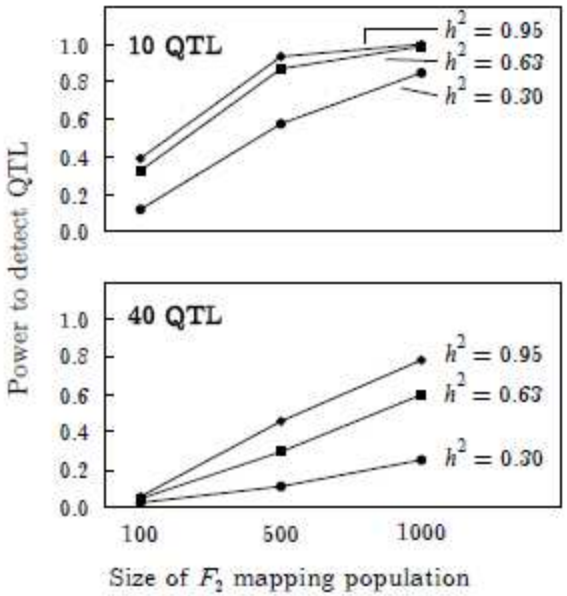
Year	No. of unique breeding populations	Multiple trait index†	
		Conventional selection	MARS
2002	79	0.63	1.10
2003	97	0.25	0.97
2004	72	0.76	1.62
All years	248	0.50	1.18

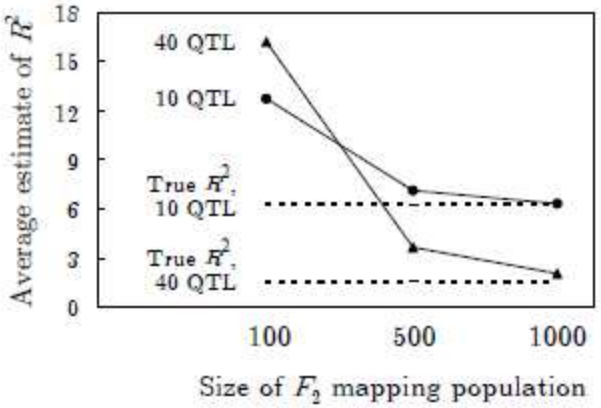
†Multiple trait index is scaled to the have the parental lines equal to zero. This index includes traits like grain yield, grain moisture, test weight, standability, etc.

Some Limitations

LIMITATIONS OF MAS AND MARS (BEAVIS EFFECT)

1. Underestimation of the number of QTL
2. Over-estimation of effects





Bernardo 2010

Genomic Selection

Meuwissen et al., 2001 proposed to skip the significance test of markers and use a large set of random markers directly to perform a marker-based selection using the predicted marker effects (instead of using only the markers significant above an arbitrary threshold). This is called Genomic Selection (GS) or Genomewide Selection (GWS).

Marker effects can be predicted through BLUP:

$$\mathbf{y} = \boldsymbol{\mu}\mathbf{1} + \mathbf{X}\mathbf{g} + \mathbf{e}$$

$\mathbf{y}$: vector of phenotypic menas (Nx1)

$\boldsymbol{\mu}1$: overall mean

$\mathbf{X}$: incidence matrix (NxN_M)

$\mathbf{g}$: vector of random effects ($N_M x1$)

$\mathbf{e}$: vector of residual effects (Nx1)

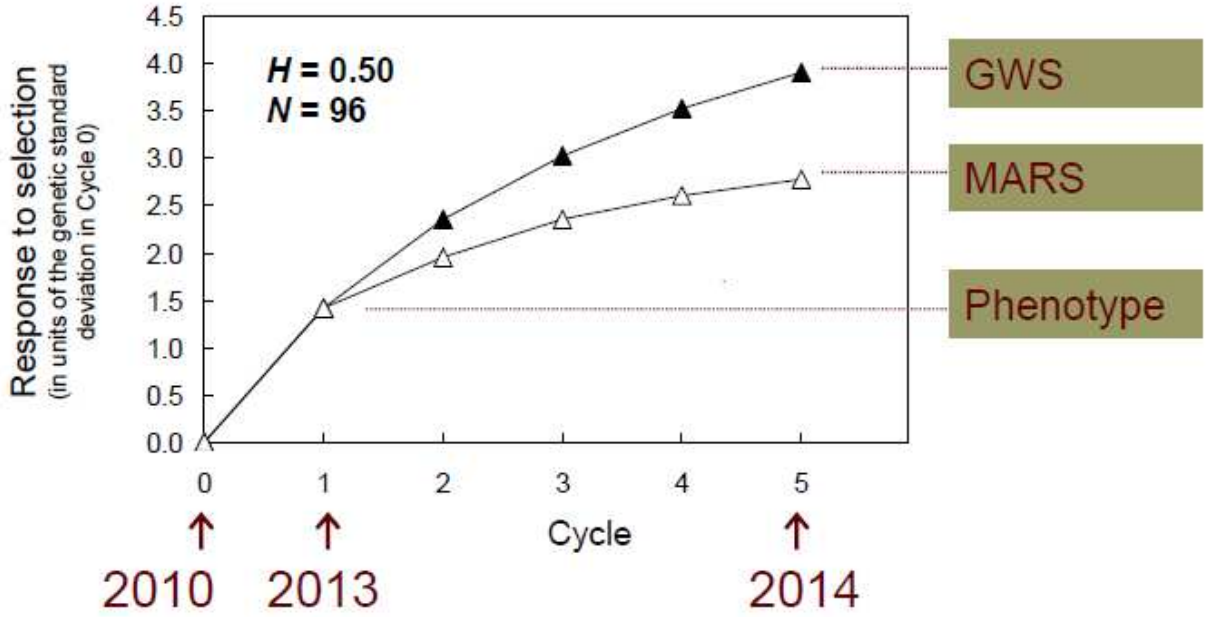
$$V(g) = \mathbf{I}V_{M1} = \mathbf{I}\left(V_G / N_M\right)$$

$$\begin{bmatrix} \hat{\mu} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} N & \mathbf{1'X} \\ \mathbf{X'1} & \mathbf{X'X} + \mathbf{I}\left(V_{\bar{Y}} / V_{Mi}\right) \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{1'y} \\ \mathbf{X'y} \end{bmatrix}$$

Genomic Selection

GS out-performes MARS: MLR of linear effects overestimates true effects of the QTL.

On the other hand, there is a shrinkage of predicted effects toward zero. Variance of marker is constant across all loci causing over-shrinkage of QTL of large effect and under-shrinkage of QTL of low effect.



Create the model when h^2 is high and use it when h^2 is zero or low

Bernardo 2012

Genomic Selection

OTHER APPROACHES

- BLUP methods assume that each marker explains the same amount of genetic variance, and that epistasis is absent.

- Bayesian methods do not require this assumptions (Meuwissen et al., 2001). For modeling the variance of the markers:
  - Bayes A uses the inverse of a chi-squared distribution with degrees of freedom and scale parameters such that the mean and variance of the distribution match the expected mean and variance of the marker.
  - Bayes B assumes a prior mass at zero allowing for markers with zero effect which is more realistic.
- Bayesian methods are supposed to work better due to the colinearity with large number of markers. However, in plant breeding this methods do not seem to work as good as in animals (Meuwissen et al., 2001, Lorenzana and Bernardo, 2009; Bernardo and Yu 2007).

- Epistasis did not seem to improve predictions either in plant context (Lorenzana and Bernardo, 2009).

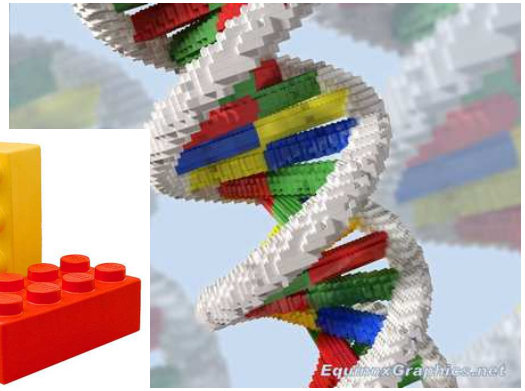
- .

Molecular Breeding

SELECTION RESPONSE VS. UNDERLING CAUSES

Trait	Genetic control	Population used for selection	Accuracy of gene/QTL position	Molecular Breeding (MB)method
Qualitative	Few major genes	Elite X Elite	**High accuracy** - Ideally marker = gene - Often, marker close to gene (stable association)	MAS
		Elite x Non-elite		
Quantitative	Few QTL of large effects	Elite x Non-elite	**Medium to low accuracy** (CI >10cM) - Markers linked to QTL - Associations marker-QTL depend on the genetic background	MABC
		Elite x Elite		
	Numerous QTL of small effects	Elite x Elite		MARS (or GWS)

Building blocks



Black box