Genetic dissection of complex traits, crop improvement through marker-assisted selection, and genomic selection

Awais Khan

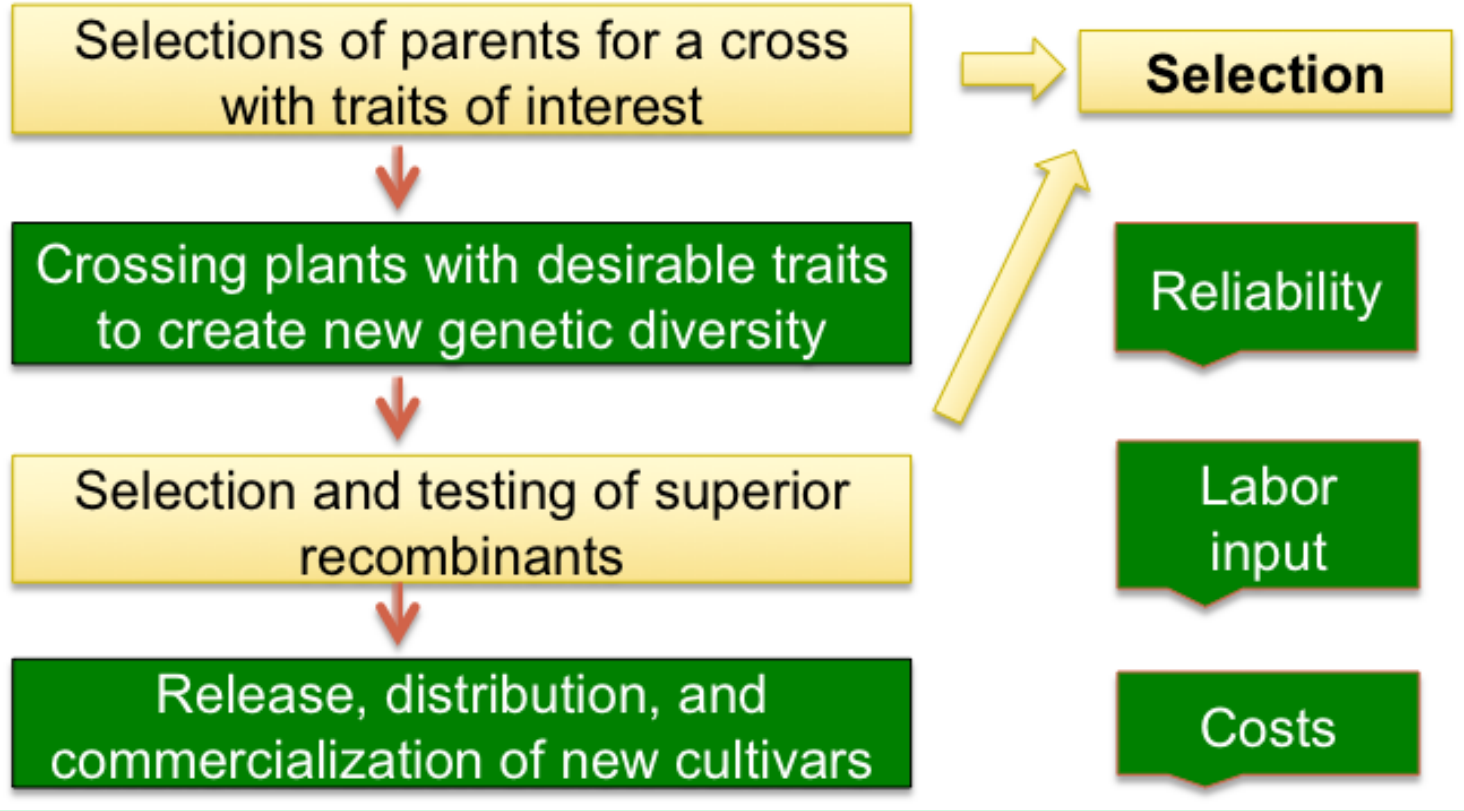
Adaptation and Abiotic Stress Genetics, Potato and sweetpotato

International Potato Center (CIP), Lima

June 19, 2014

Importance of selection in plant breeding

Systematic procedure for genetic improvement through **crossing** plants with desired traits and **selecting** progeny with improved performance and/or improved combinations of traits.



General steps in plant breeding (modified after Gepts 2002)

Phenotypic selection: Selection based on appearance and performance

Natural Populations

Selected individuals (phenotypic selection)

Progeny Testing

Elite (high EBV) individuals

Breeding Population

Repeat over generations

Improved individuals

I. Difficult to separate environmental & genetic contribution
II. Difficult to distinguish homozygous & heterozygous effects
III. Needs large space & labor input
IV. Slow & time consuming

DNA based selection methods

A. **Marker-assisted selection:** Selection for one or more (up to 8-10) alleles

B. **Marker-assisted backcrossing:** One or more (up to 6-8) donor alleles are transferred to an elite line

C. **Genome-wide selection:** Selection of several loci using genomic estimated breeding values (GEBVs) based on genome-wide marker profiling

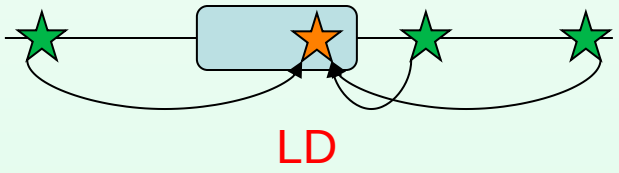
Trends Plant Sci 10: 621 630

Concept of Marker assisted selection

Molecular breeding

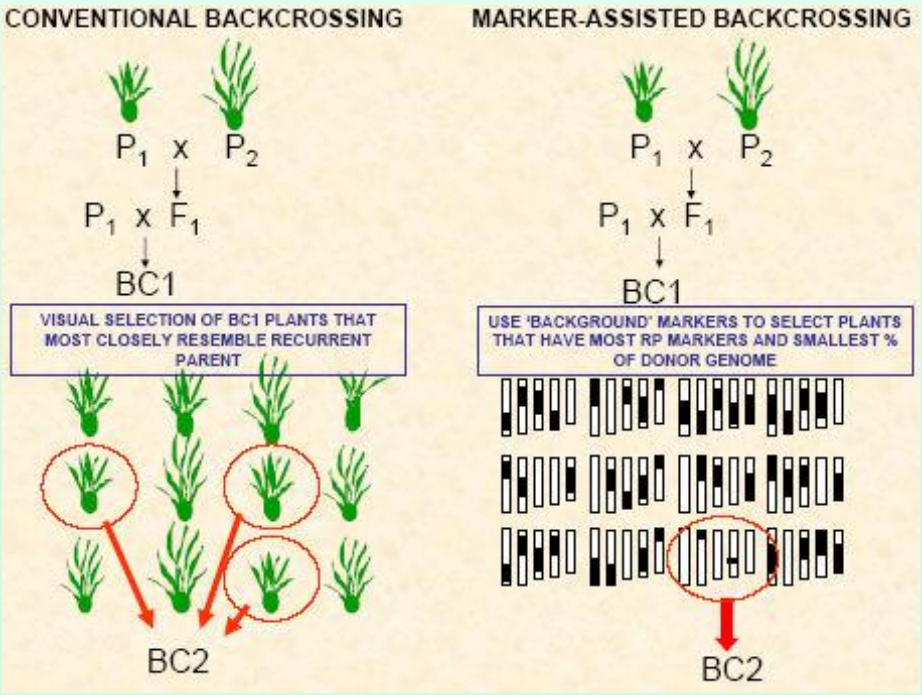
Association between molecular marker and causative gene

Direct association



Indirect association

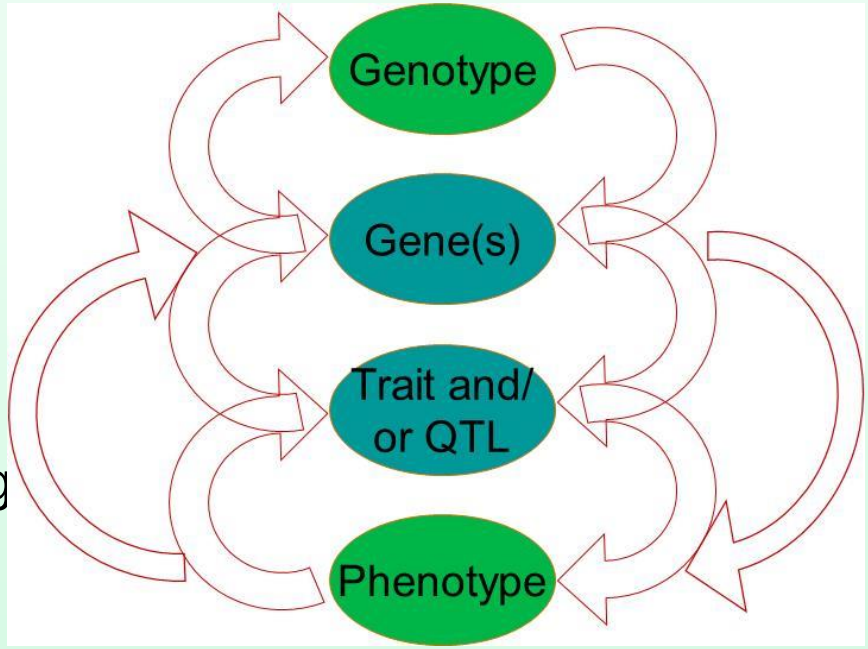
Hirschhorn & Daly, 2005



Identification of marker-trait associations for selection

Genetic mapping

Physical mapping



EST sequencing

Genome sequencing

Map-based cloning

Genetic mapping

Association mapping and QTL mapping

Trait correlations

Transcriptomics

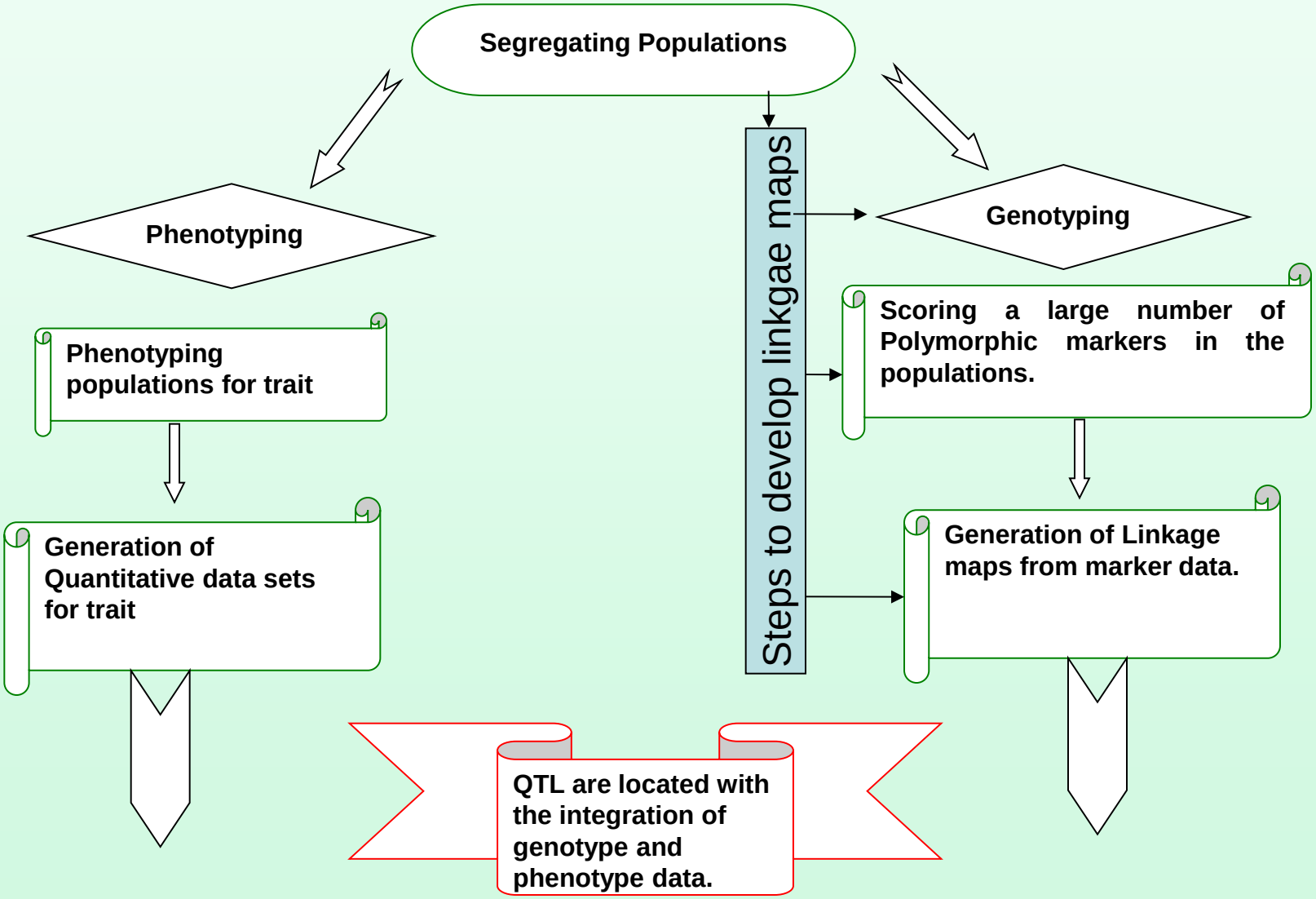
Proteomics

Metabolomics

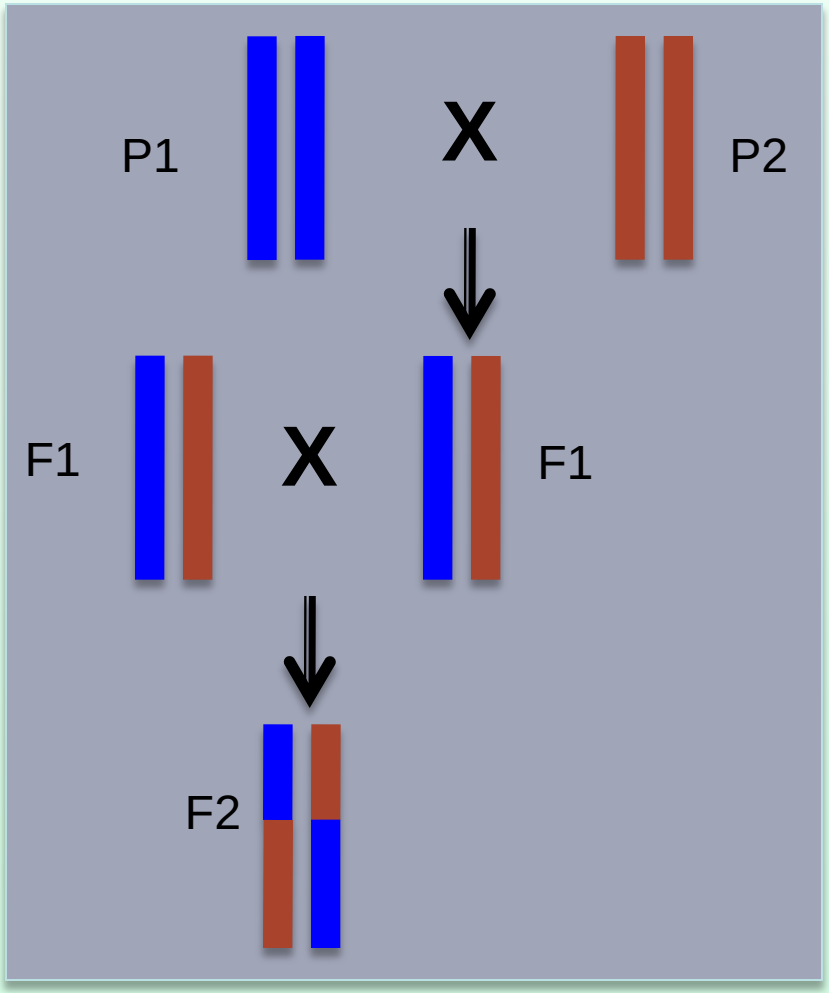
TILLING

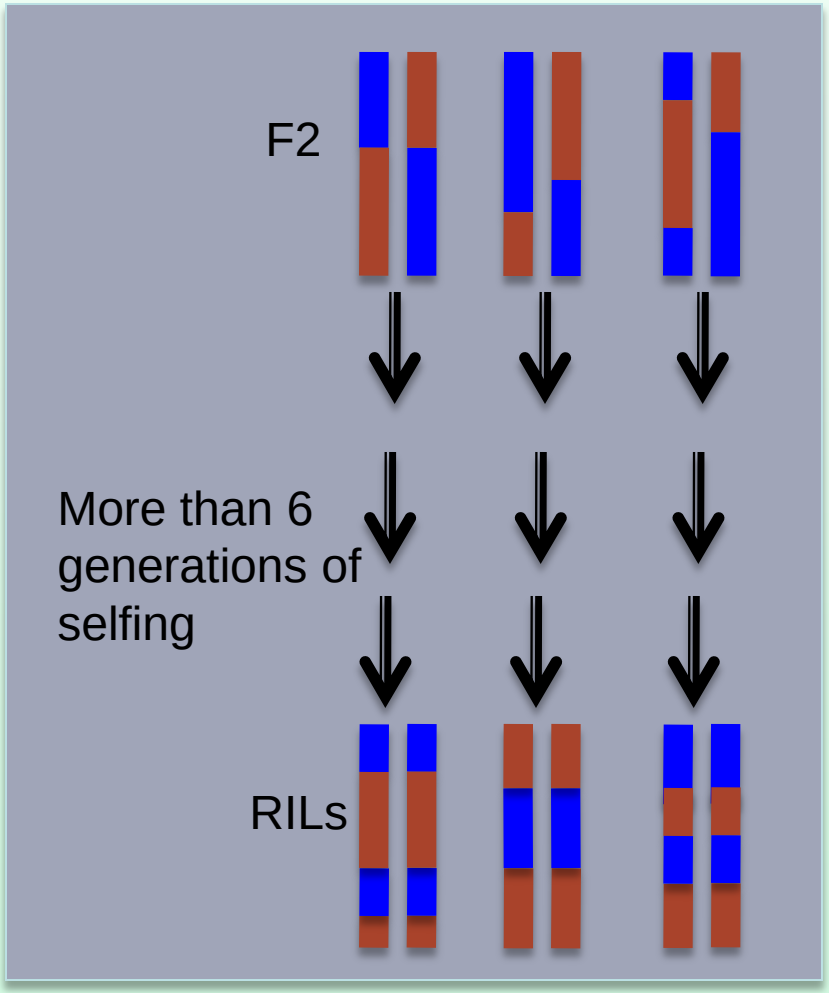
Adapted from Trends Biotech 24:490-499

Overview of marker-trait association via QTL mapping



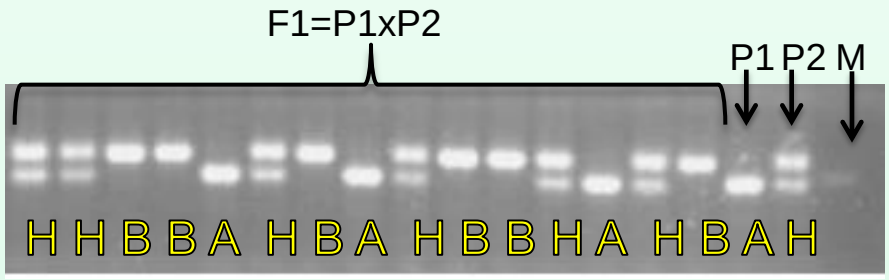
Segregating population





Genotyping and phenotyping

- Testing a large number of robust, high-throughput genetic markers on a segregating population (genotyping)
- Phenotyping segregating population for a large number of traits of multiple years and locations



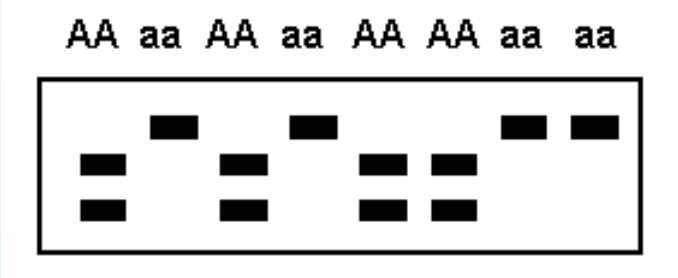
Marker data

Key:
A=Homozygous for allele P1
B=Homozygous for allele P2
H=Heterozygous
M=Ladder

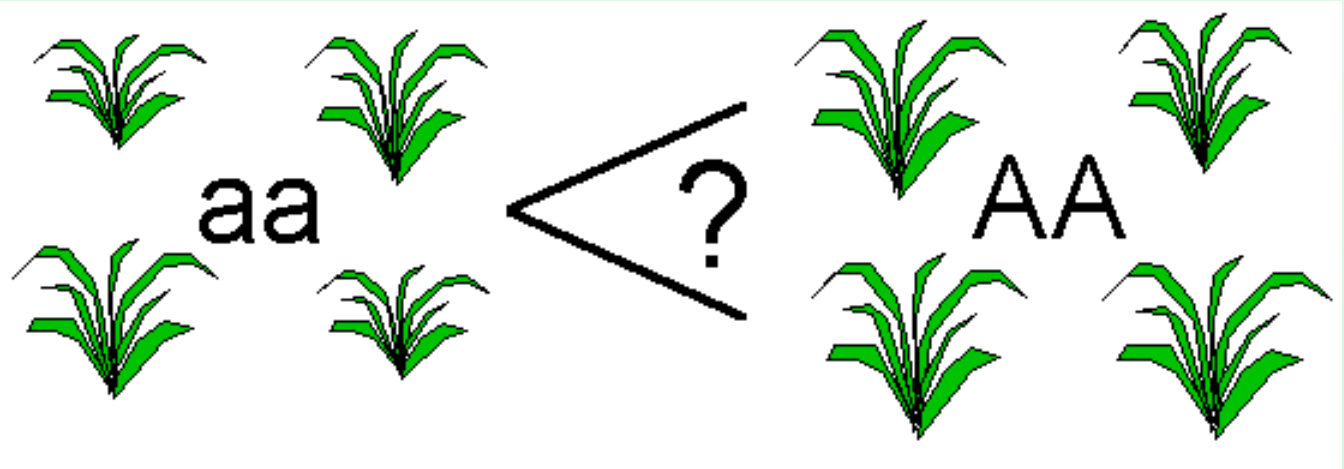
QTL analysis

Is there a significant link between genetic makeup (genotype) and trait phenotype?

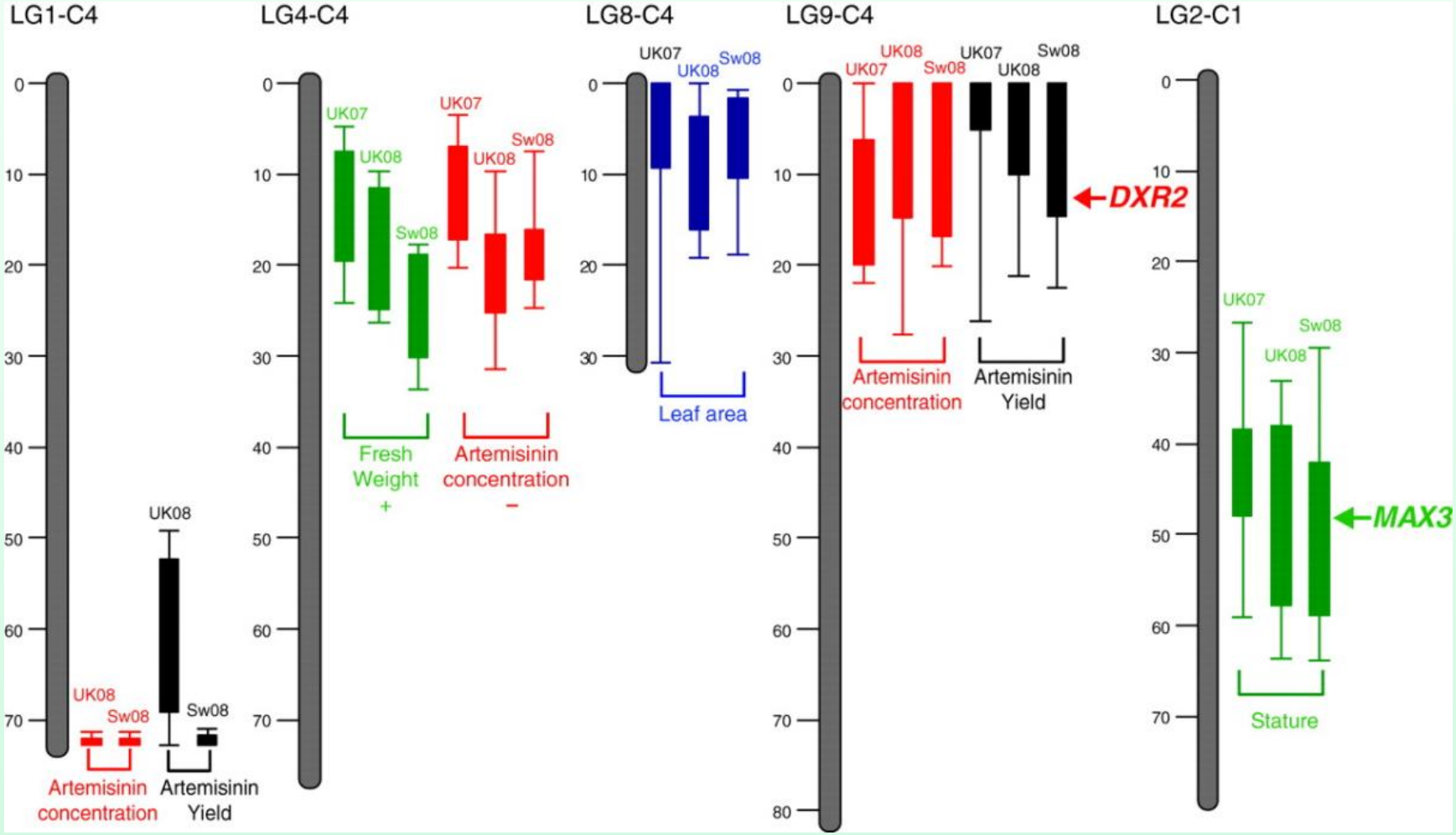
Single marker Analysis



Marker genotype

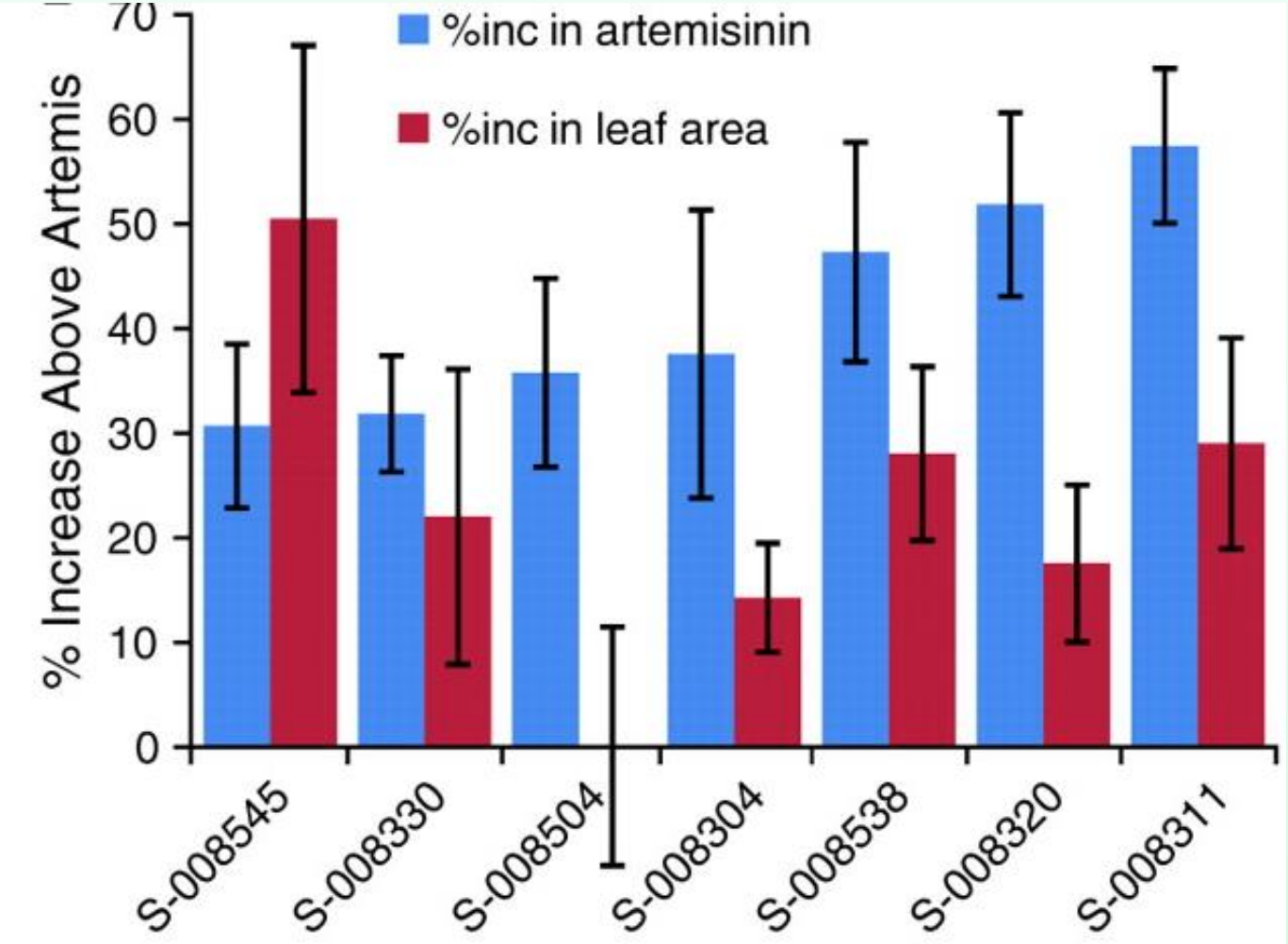


QTL mapping



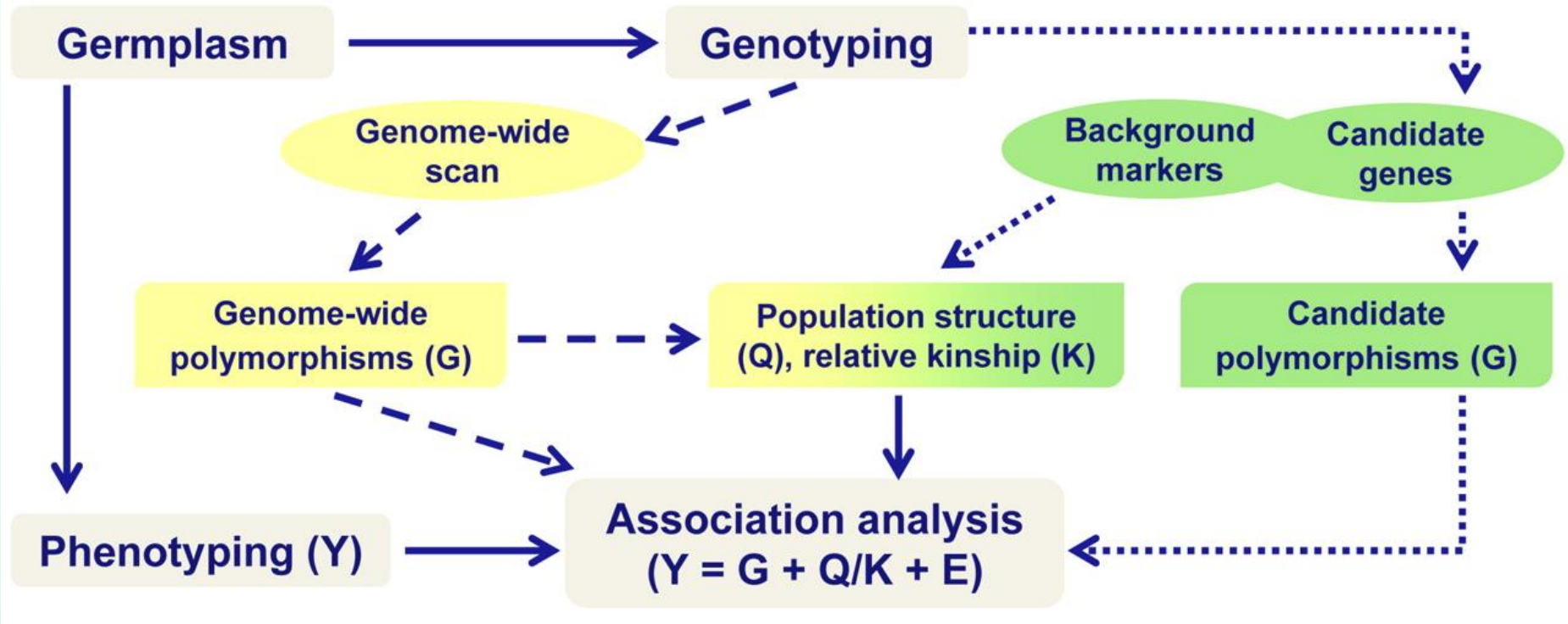
A selection of QTLs for key traits. QTLs are shown to the right and distances in centimorgans to the left of each linkage group.

Results:
Use of MAS in hybrid production



The increase (%) in artemisinin concentration (in blue) and leaf area (in red), over Artemis F1 for seven hybrids produced from crosses of selected high-yielding individuals.

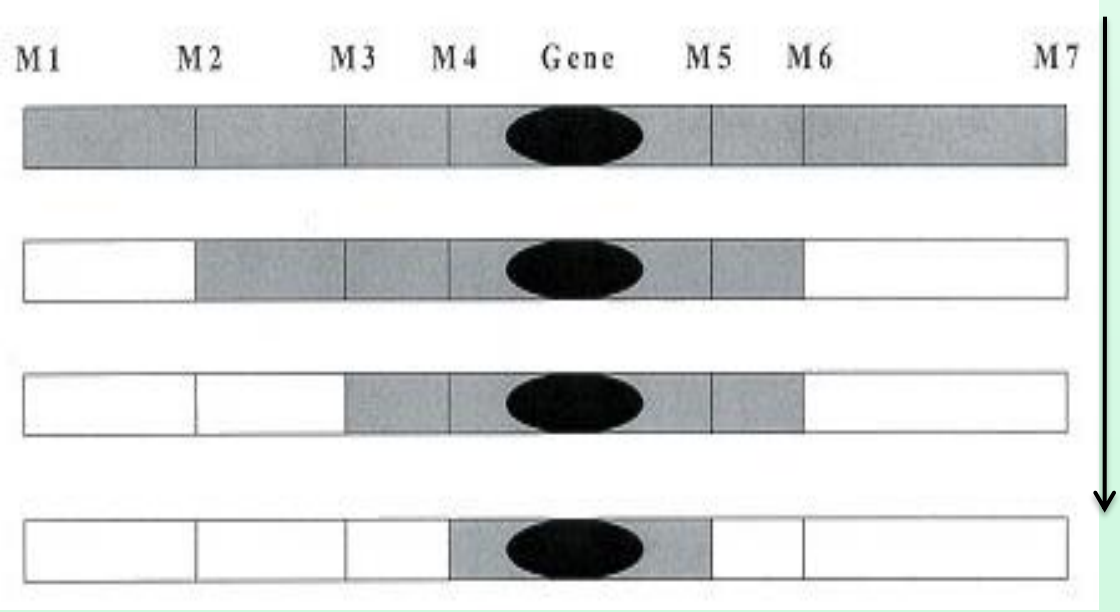
Overview: Association mapping analysis



Zhu *et al.* 2008

Identification of marker-trait association via Association mapping

- The identification of marker alleles involved in the inheritance of traits, also known as linkage disequilibrium (LD) mapping
- Utilizes ancestral recombinations for identification of marker and trait association

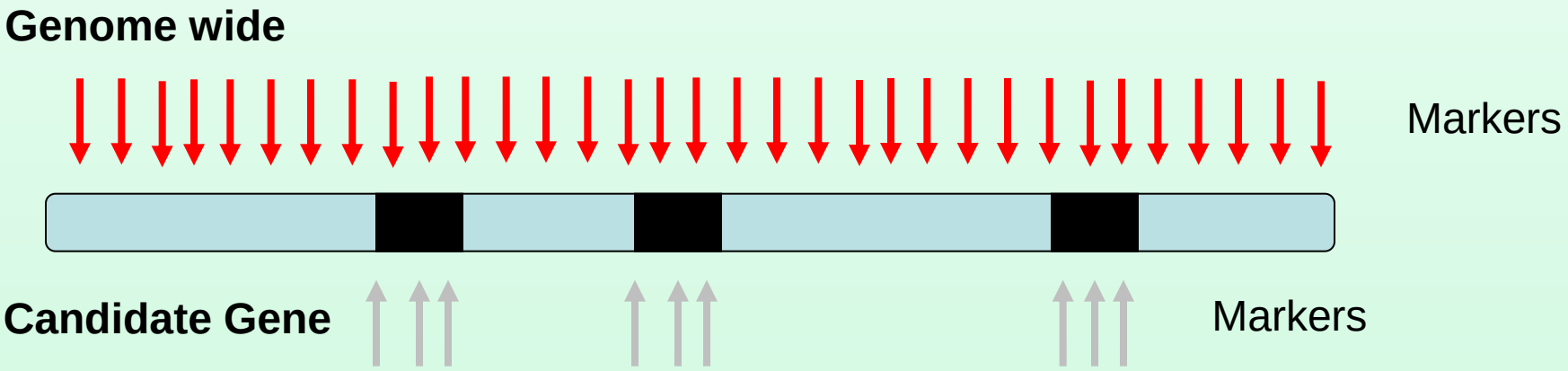


The traits we observe in a population are linked to the surrounding genetic sequence of the original evolutionary ancestor.

Approaches for Association mapping

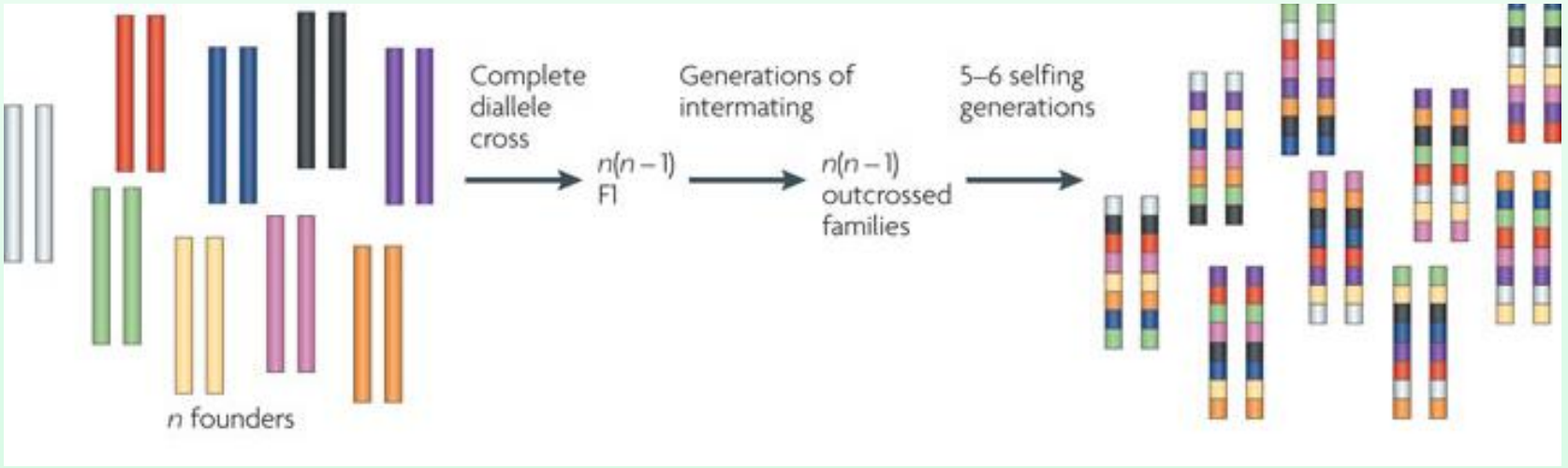
Candidate gene: Lower number of markers, based on prior knowledge: expert opinion, linkage mapping results

Genome wide: High density of molecular markers throughout genome



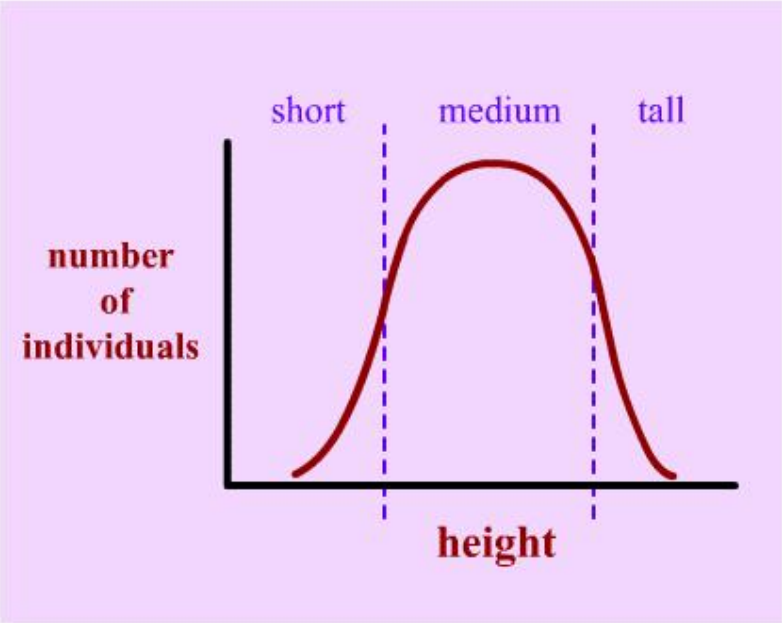
Choice of method depends mostly on how fast linkage disequilibrium decays in the crop

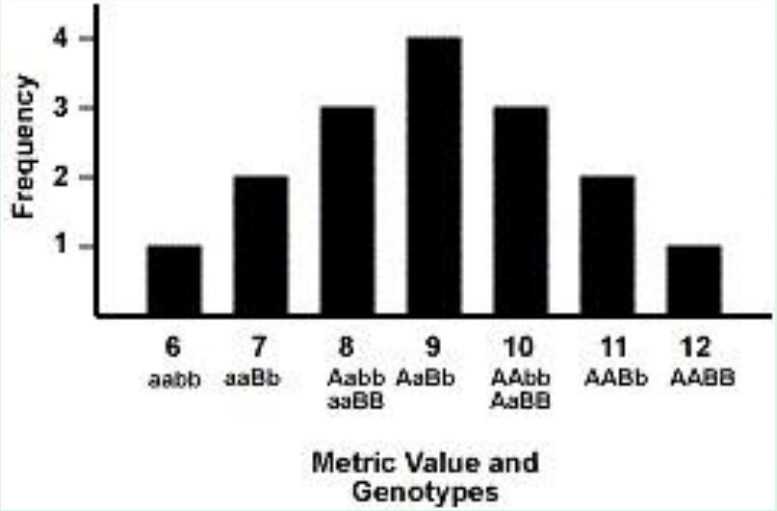
The Multi-parent Advanced Generation Inter-cross lines (MAGIC lines)



Joy Bergelson and Fabrice Roux, 2010

Quantitative traits are complex





1) Multiple loci

2) Pleiotropy (one gene has many effects)

3) Epistasis

4) Environment (produces a range of phenotypes)

Complex traits and QTL and association mapping

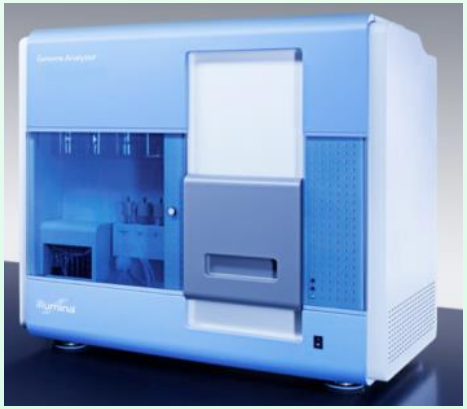
- Most of quantitative traits are controlled by several genes, QTL and association mapping will only allow identification of linked markers that **explain a small fraction of total genetic variance**

- Individual genes will have small effects and to accurately estimate small effects, **a large data set is needed (a large population to be genotyped and phenotyped)**

DNA marker technology coupled with Next-Generation Sequencing (NGS)

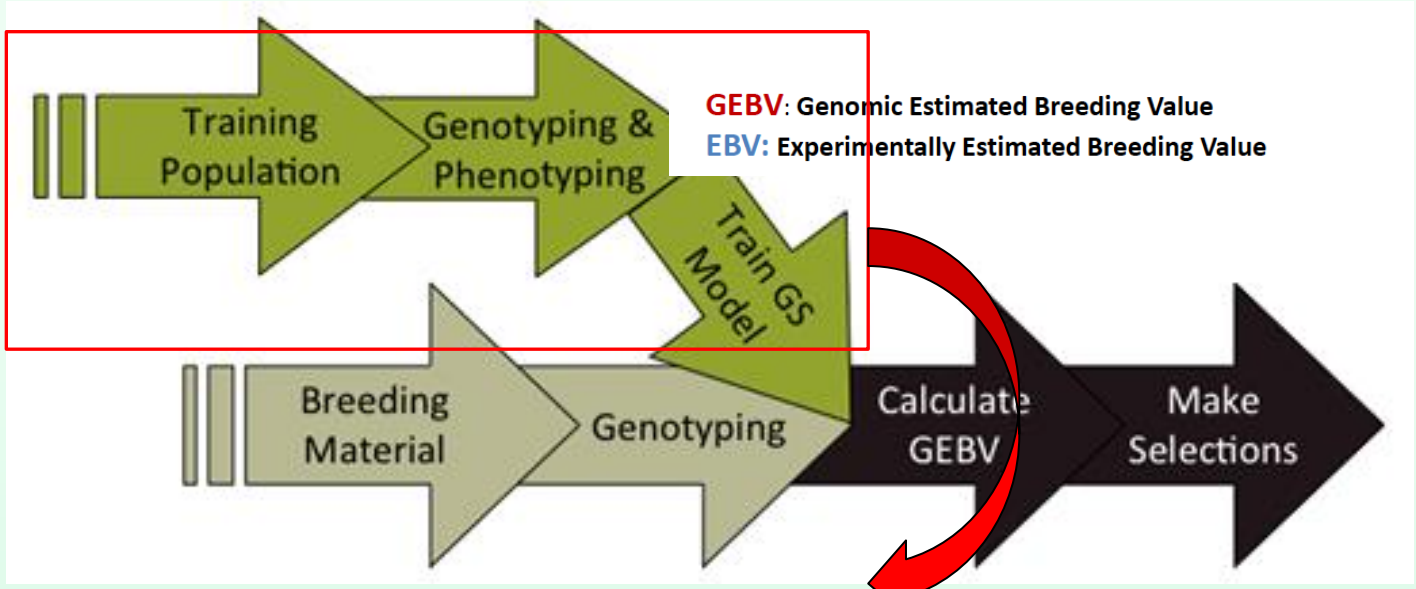
Cost and throughput comparisons

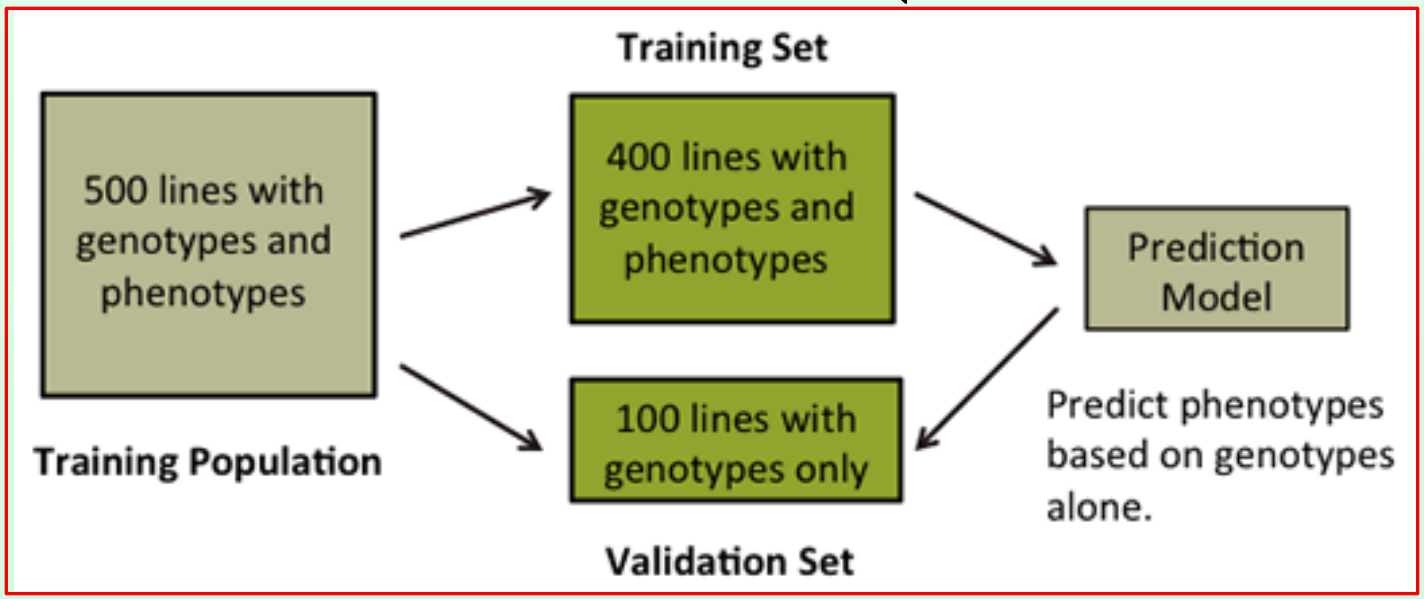
Sequencing of 3 Gbase genome to 18X coverage (54 Gbases)



	Sanger	454	Illumina
No. of plates:	756,000	120	3
Time:	48 years	6 months	2-3 weeks
Total cost:	$108 millions	1 million	$60k
Cost/Mbase:	$2,000	$18.5	$3

Predicting the phenotype: Genomic selection



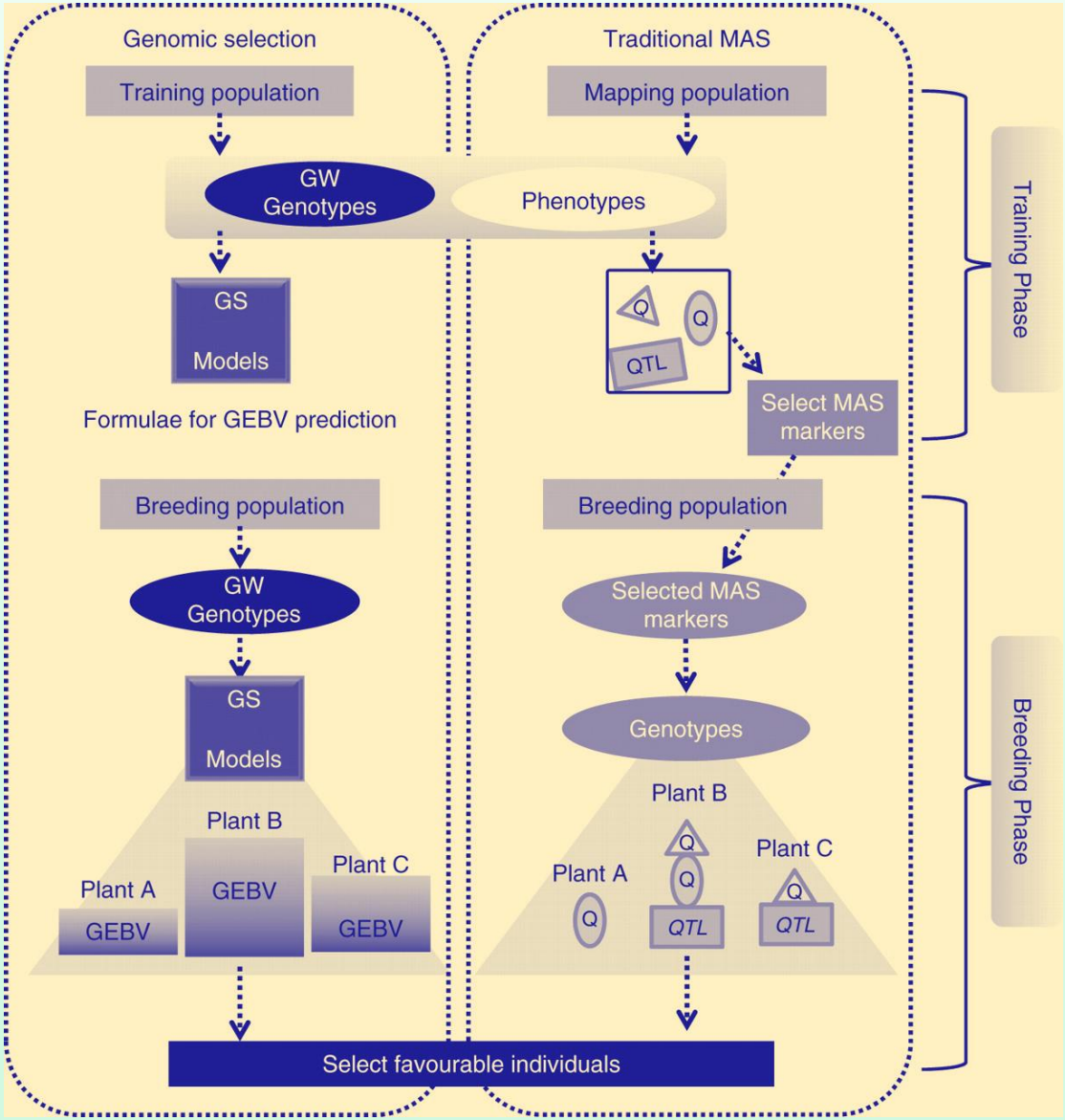


Predicting the phenotype: Genomic selection *vs* traditional MAS

Genomic Selection (GS): Selection of several loci genome-wide linked to traits of interest using Genomic Estimated Breeding Values (GEBVs) based on genome-wide markers

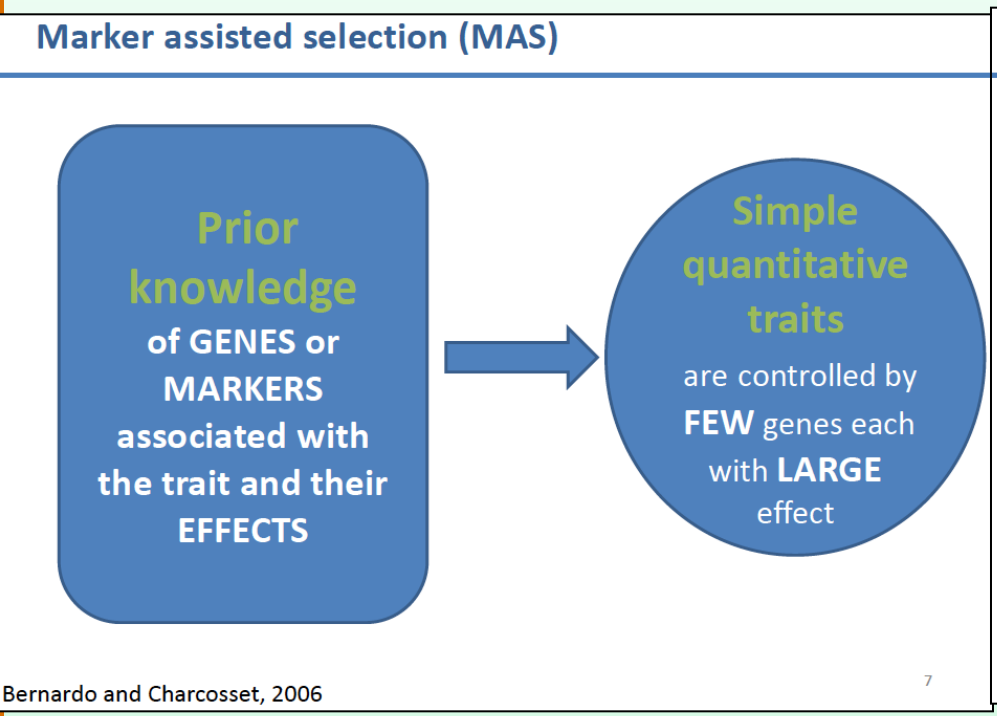
Traditional MAS:

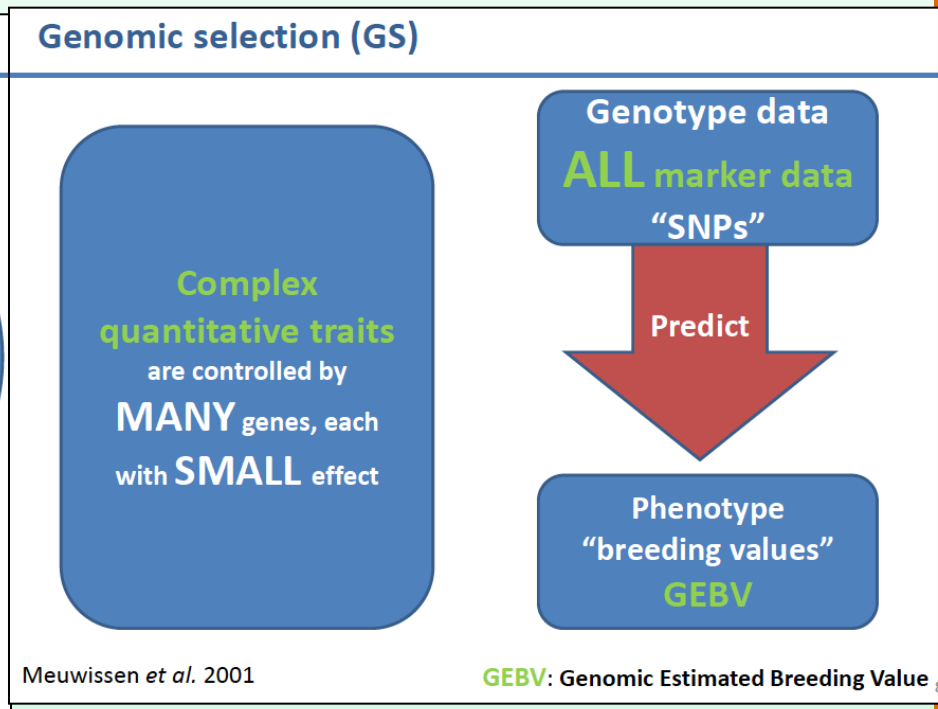
DNA markers that are tightly-linked to target loci are used to select genotypes with desirable combination of alleles. Usually allele of a DNA marker associated to trait of interest are identified through prior quantitative trait loci (QTLs) mapping.



Nakaya and Isobe, 2012

Comparison of MAS and GS





GS → Increased gains per unit of time

Lorenz et al. 2011

Genomic selection vs. Traditional breeding

	Traditional breeding	GS
Time To identify superior individuals	Individuals must mature to estimate BV	BV can be estimated earlier
Cost	Space requirements of trials and phenotype measurements are costly	Continuing decline in the cost of marker technologies “Genotyping”

Lorenz et al. 2011

Prediction Accuracy of Genomic Selection

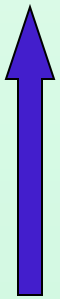
Correlation between

GEBV
(Genomic estimated breeding value)

EBV
(Experimentally estimated breeding value)

Affected by:

1. LD between markers and QTLs (↑LD)
2. Size of Training population (↑n)
3. Heritability of the trait in question (↑h^2)
4. Genetic structure of the trait (↓#QTLs)



Accuracy of GS

Hayes et al. 2009

Factors affecting accuracy of GS: Heritability of trait and population size

With greater heritability of trait, fewer records are required (population size) in training set for achieving high accuracy of GEBV in target breeding population.

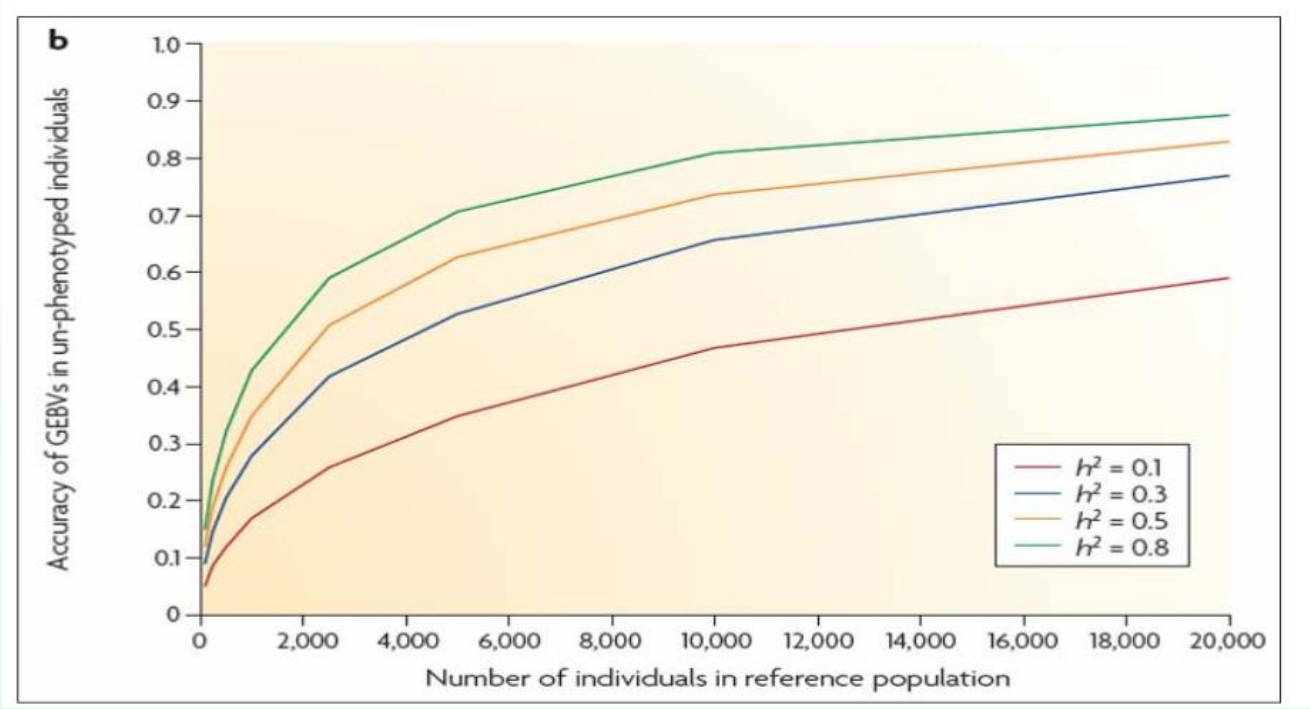
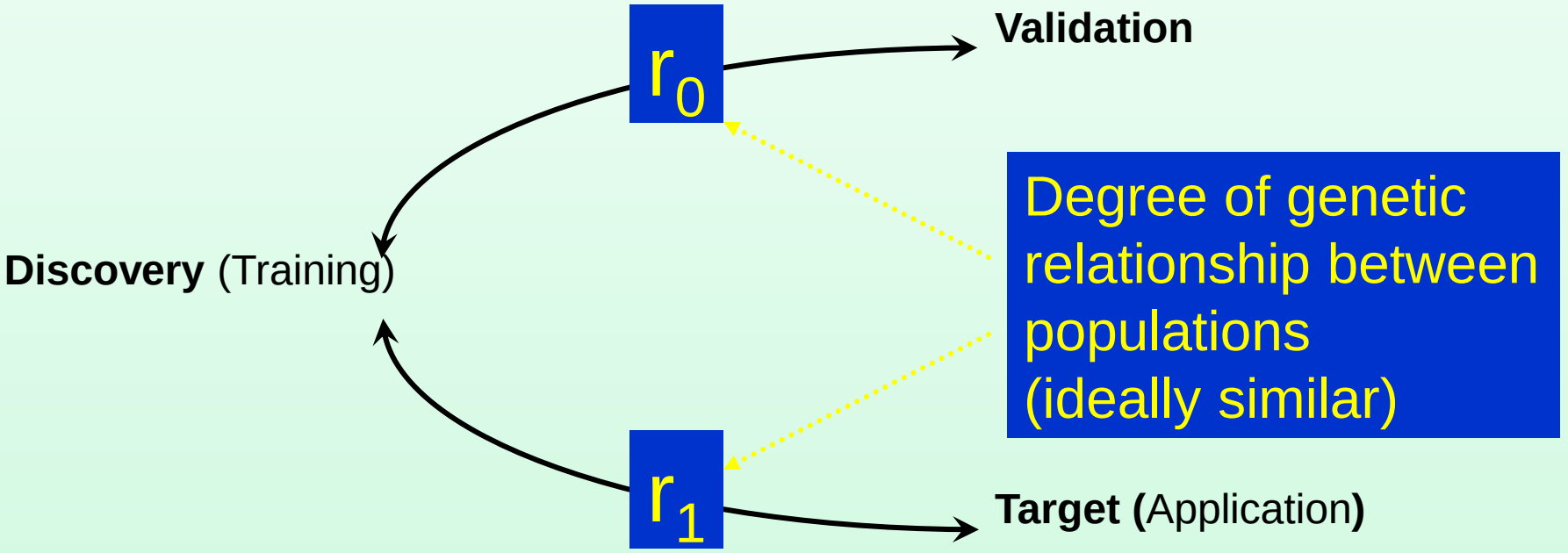


Figure: For low-heritability traits, a very large population size of training population will be required in the to achieve high accuracies of GEBV in target breeding population.

Goddard 2007

Factors affecting accuracy of GS: Relationship of training and validation population



Factors affecting accuracy of GS: GS model

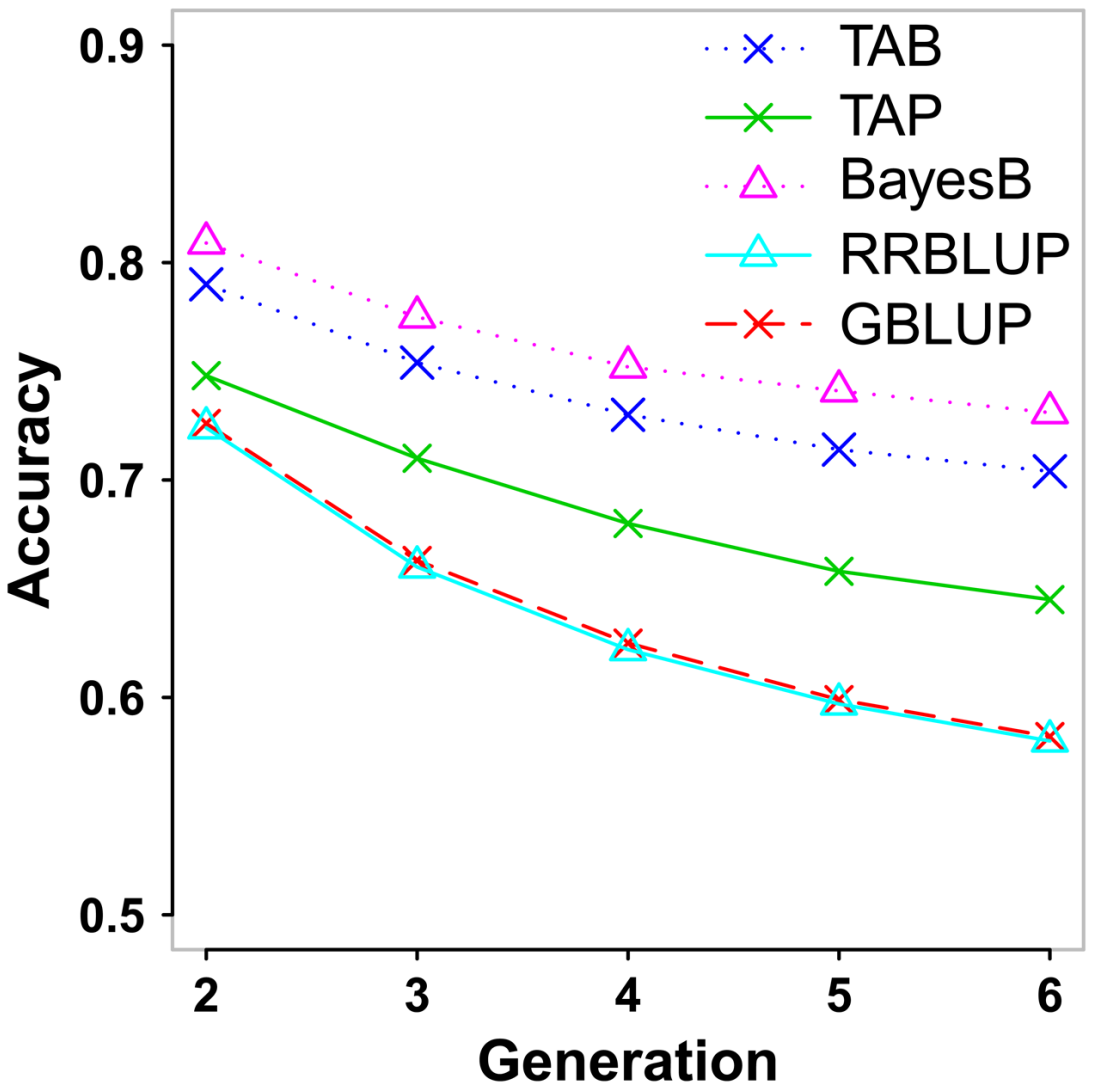


Figure: The graph shows the correlation between **estimated (GEBVs) and true breeding values in generations 2–6**

- **GS models have different accuracy of prediction**
- **Over generations prediction accuracy of GS models decrease**

Critical considerations for success of genomic selection

Good understanding of trait and accurate phenotyping

Phenotyping with special attention to Genotype x environment interaction: appearance and performance in general, in particular, response to environment

Data recording and management

Is breeding obsolete?

- Usually **GMO techniques** are used to **manipulate single gene** that could also be easily manipulated using marker assisted (conventional) breeding

- Marker-assisted breeding can manipulate **multiple traits simultaneously**

- Marker-assisted breeding can manipulate genetically **complex "quantitative traits" with small effects**--- traits that are influenced by the environment

- Marker-assisted breeding can bring about directed changes (provided **genetic variation exist for the trait of interest**)

Thank you for your attention!

???? are welcome

Look forward to collaborate on dissecting genetic basis of complex adaptation and abiotic stress tolerance

awais.khan@cgiar.org

Next generation sequenced based genotyping for *Ipomea trifida (2x)*

X

CIP 460410 (DLP 4653)

CIP460377 (DLP4597)

M9 x M19

Mapping population

Next generation sequenced based genotyping for *Ipomea trifida*

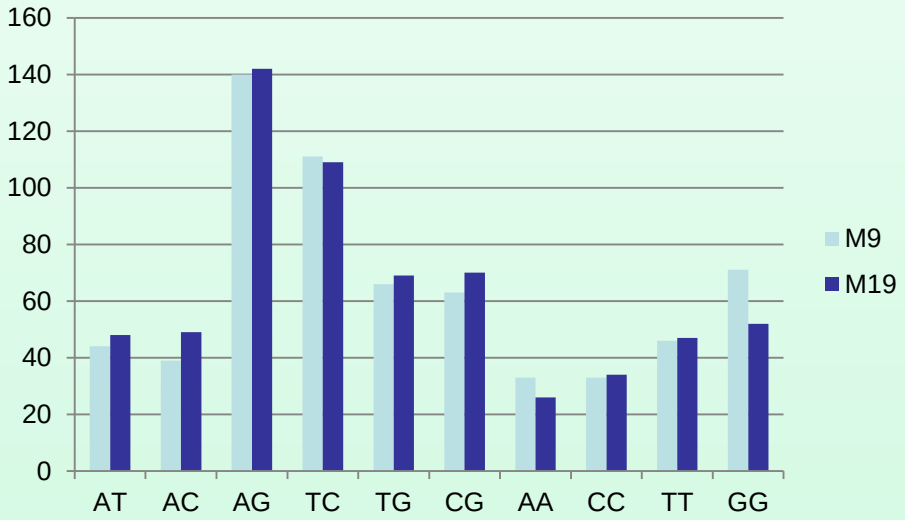
~ 3 Million sequence reads in total → and ~1.3 are good reads

SNPs without filtering → 5466

SNPs after eliminating NN→ 3643

SNPs after eliminating SNPs does not match between replicates→ 3210

SNPs that are polymorphic and segregating in the mapping population → 646



Maria David and Raul Eyzaguirre