

HOS 6236 Molecular Marker Assisted Plant Breeding

Fall 2017

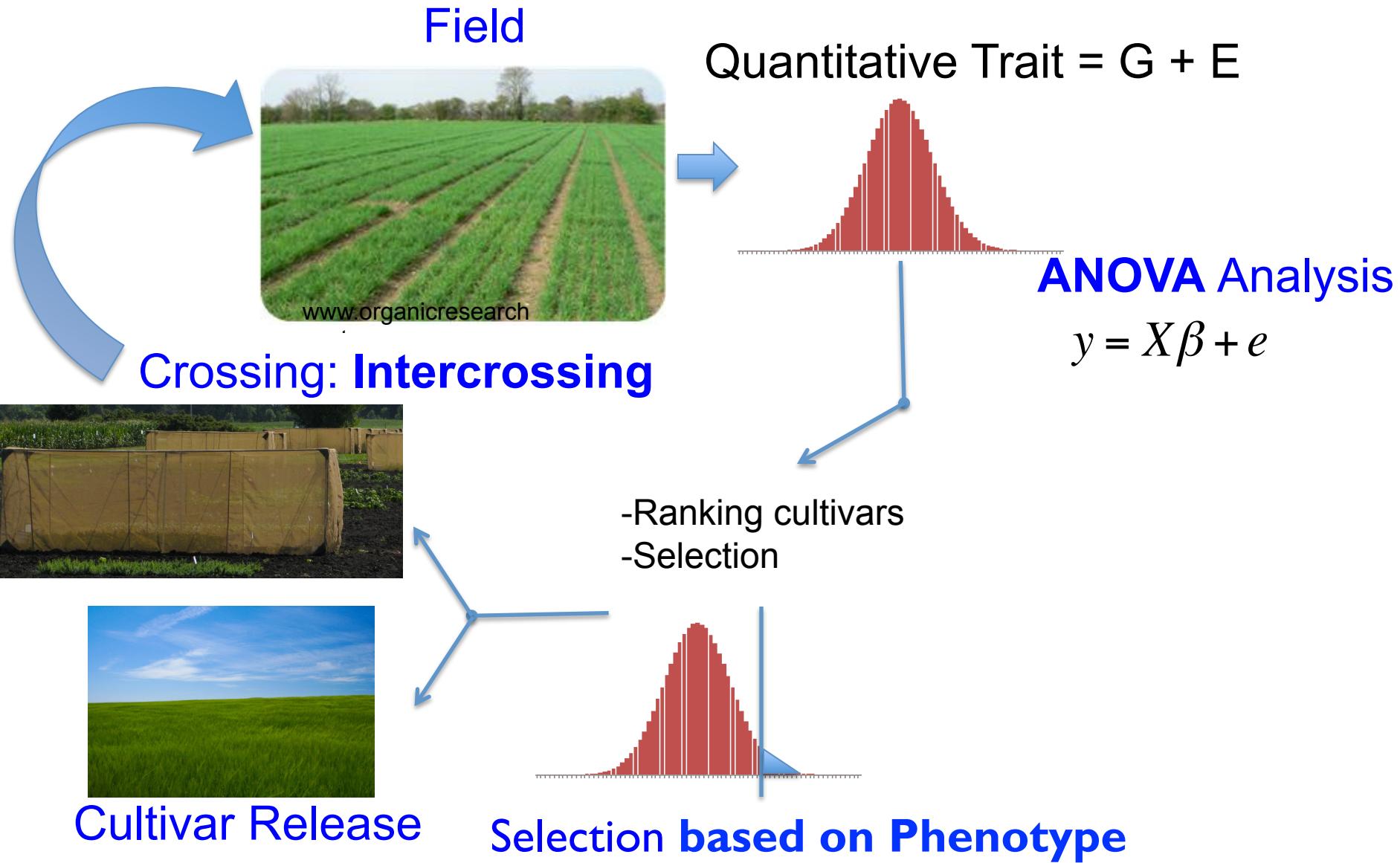
Last Class:

GWAS and DNA imputation

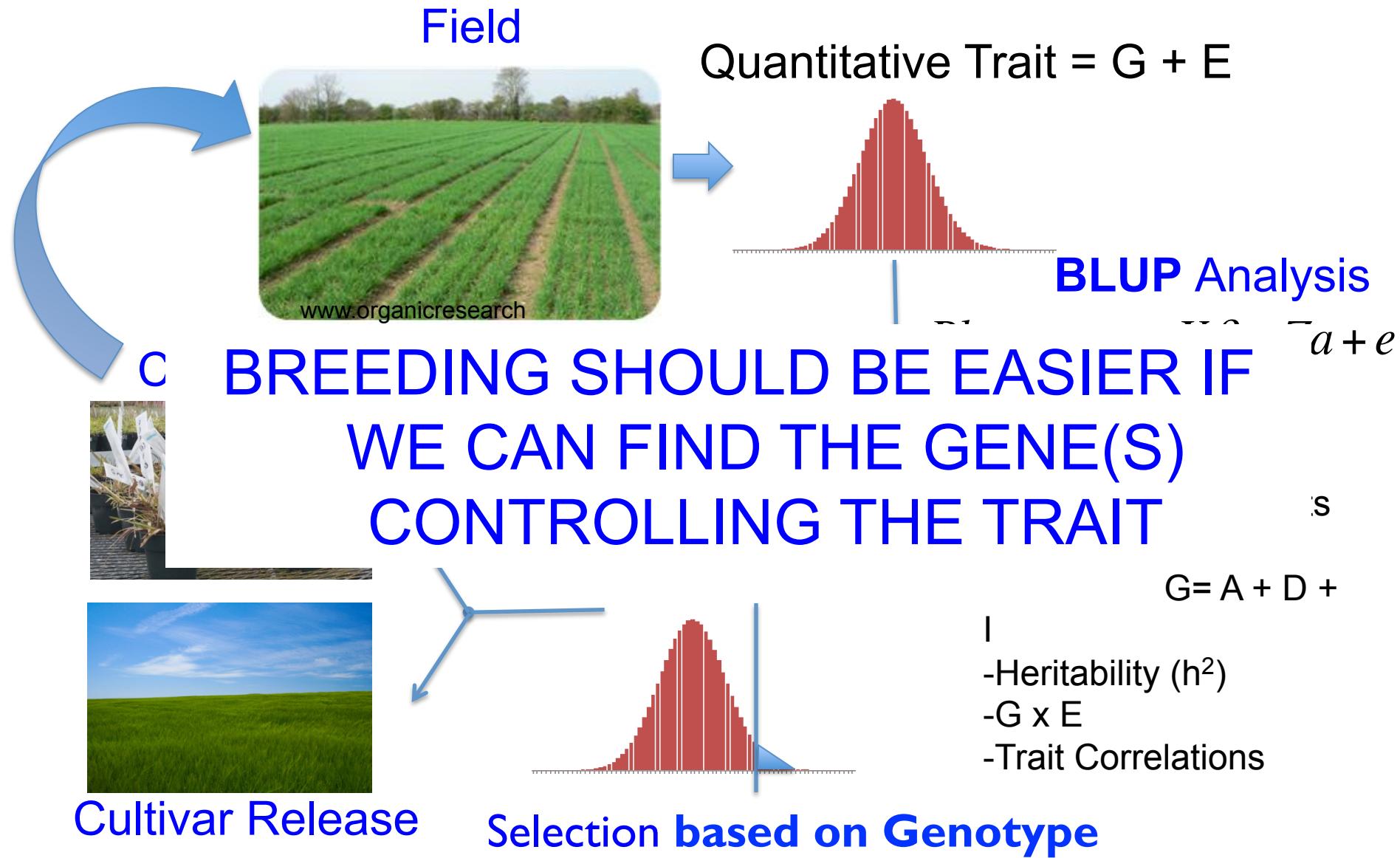
Today's Class

What is Genomic Selection?

Traditional Breeding



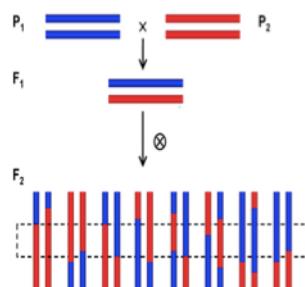
Recurrent Genotypic Selection



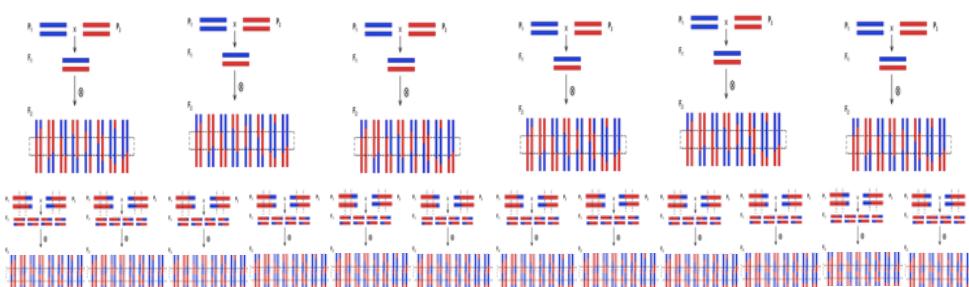
Genome-wide Selection

All based on linkage disequilibrium:

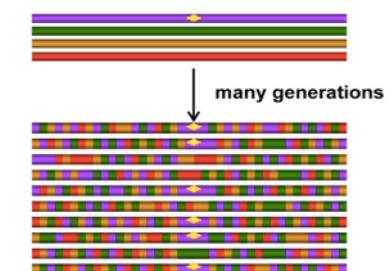
I - QTL analysis



III – Genome-wide selection



II – Genetic association



Resolution

Low

Medium

High

Linkage Blocks

Large

Medium

Small

(Meuwissen et al., 2001)

Genome-wide Selection

Copyright © 2001 by the Genetics Society of America

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes[†] and M. E. Goddard^{†,‡}

*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, [†]Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and [‡]Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Manuscript received August 17, 2000
Accepted for publication January 17, 2001

“Recent advances in molecular genetic techniques will make dense marker maps available and genotyping many individuals for these markers feasible”

However when this happens number of parameters (p) to be estimated (markers) will much larger than the number of individuals (n) to estimate those parameters

Under this scenario ($p \gg n$) the normal estimators are not efficient, are bias and have problems of overfitting

How to deal with this problem?

Subset selection: Based on the idea that high dimensional data is dangerous and many co-variates are non-informative or redundant..

Ideally you fit all possible models and select one based on model selection criteria (i.e. r^2 , AIC, BIC)... not possible when $p>>n$, so use:

- Single variable analysis
- Principal Component Analysis (PCA)
- Partial Least Square regression (PLS)
- others...

How to deal with this problem?

Penalized (shrinkage) estimation: uses all available co-variates simultaneously by shrinking the estimates:

- Best Linear Unbiased Predictor (BLUP)
- Least Absolute Shrinkage and Selection Operator (LASSO)
- Bayesian methods
- Others...

Advantages: can fit all co-variates simultaneously, accounting for collinearity and produce better prediction models

Genome-wide Selection

Copyright © 2001 by the Genetics Society of America

Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps

T. H. E. Meuwissen,* B. J. Hayes[†] and M. E. Goddard^{†,‡}

*Research Institute of Animal Science and Health, 8200 AB Lelystad, The Netherlands, [†]Victorian Institute of Animal Science, Attwood 3049, Victoria, Australia and [‡]Institute of Land and Food Resources, University of Melbourne, Parkville 3052, Victoria, Australia

Manuscript received August 17, 2000
Accepted for publication January 17, 2001

They simulate a genome with 50k markers and a few thousand individuals

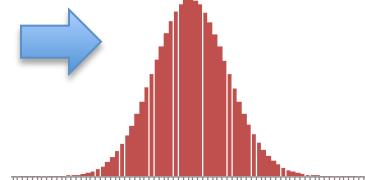
They compare PLS against RR-BLUP and Bayes A/B

“It was concluded that selection on genetic values predicted from markers could substantially increase the rate of genetic gain in animals and plants...”

What is Genomic Selection?

Construct prediction models using phenotype and molecular markers capturing most of the quantitative variation of interest in one breeding cycle

Quantitative phenotypic information

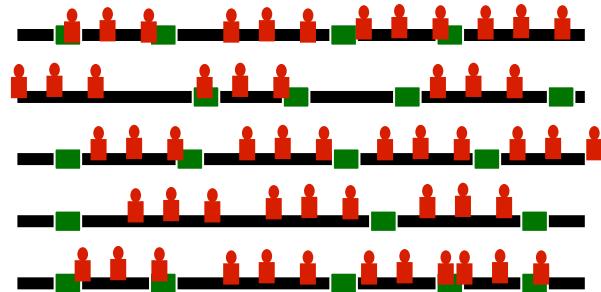


Breeding Value (BV) + Molecular Markers

Prediction model construction

$$BV = \mu + \sum_{j=1}^p X_j g_j + e$$

Genotypic information



(Meuwissen et al., 2001)

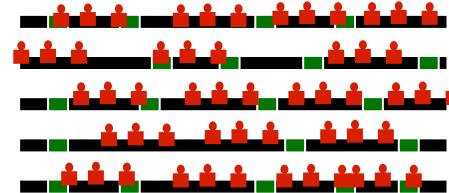
What is Genomic Selection?

Genotypes of new progeny are used to predict their performance and to select superior genotypes in future cycles

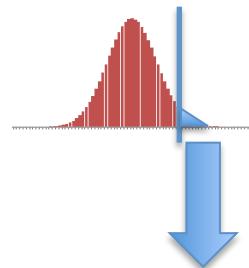
Genotypes Generation i



Molecular Markers



Selection Generation $i+1$



Deployment



Prediction

$$y_F = \sum_{j=1}^p \mathbf{X}_j \hat{g}_j$$