Genomic Selection in Cereals

Just Jensen
Center for Quantitative Genetics and Genomics

Genomic selection in cereals

(Without formulas, using examples from wheat)

1. Genomic selection vs marker assisted selection
2. Cereal breeding programs
3. Experiment conducted
4. Genetic parameters and selection accuracy
5. Simplified GS
6. How to implement in breeding program
7. Summary

Introduction

Cereal breeding has recently moved from phenotypic selection to also rely on genotype-based selection.

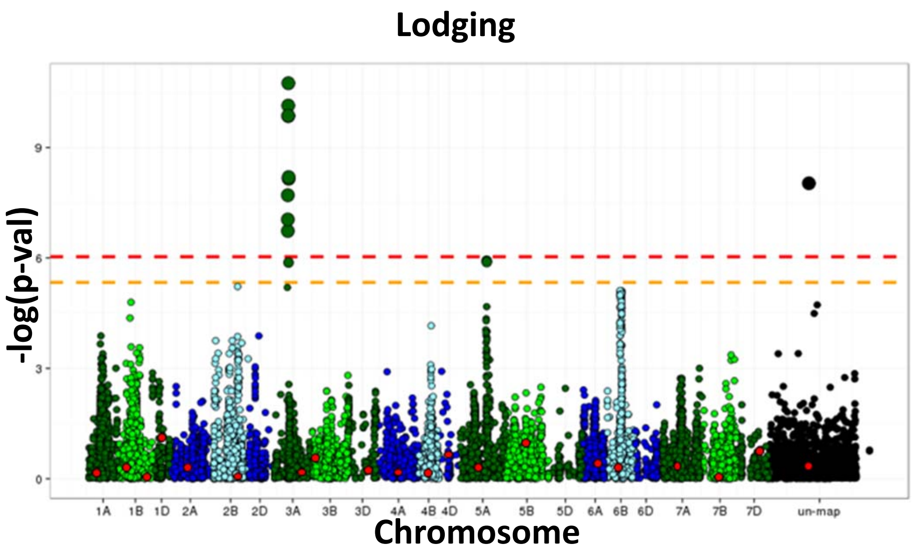
Marker assisted selection (MAS):
- Identify strong (single) marker/QTL associaiton
- Use the identified marker to screen the breeding material
- It works best (only) for traits with a simple genetic architecture

Genomic Prediction (GP):
- A dense marker set are used as predictors of performance
- More accurate predictions also for complex traits

Genomic Selection (GS):
- Use of genomic information in breeding program
- (prediction of breeding values, selection decisions

Marker Assisted Selection (MAS)

Use significant results in subsequent selection



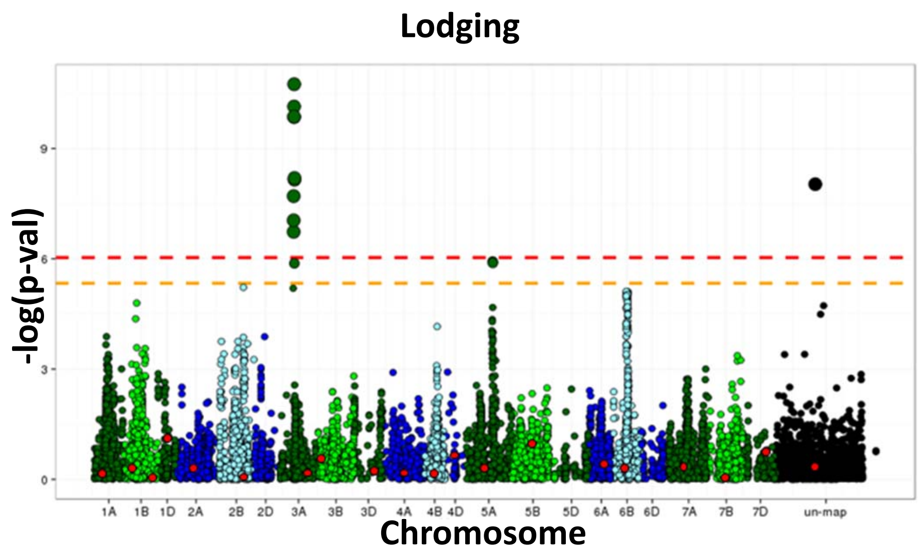
Genomic Selection

- Dataset with phenotypes and genotypes
- Train a model using ALL markers
  - (n << p) -> Complex regularization needed
- Predict future performance using all markers
- Summing over all <u>regularized</u> marker effects

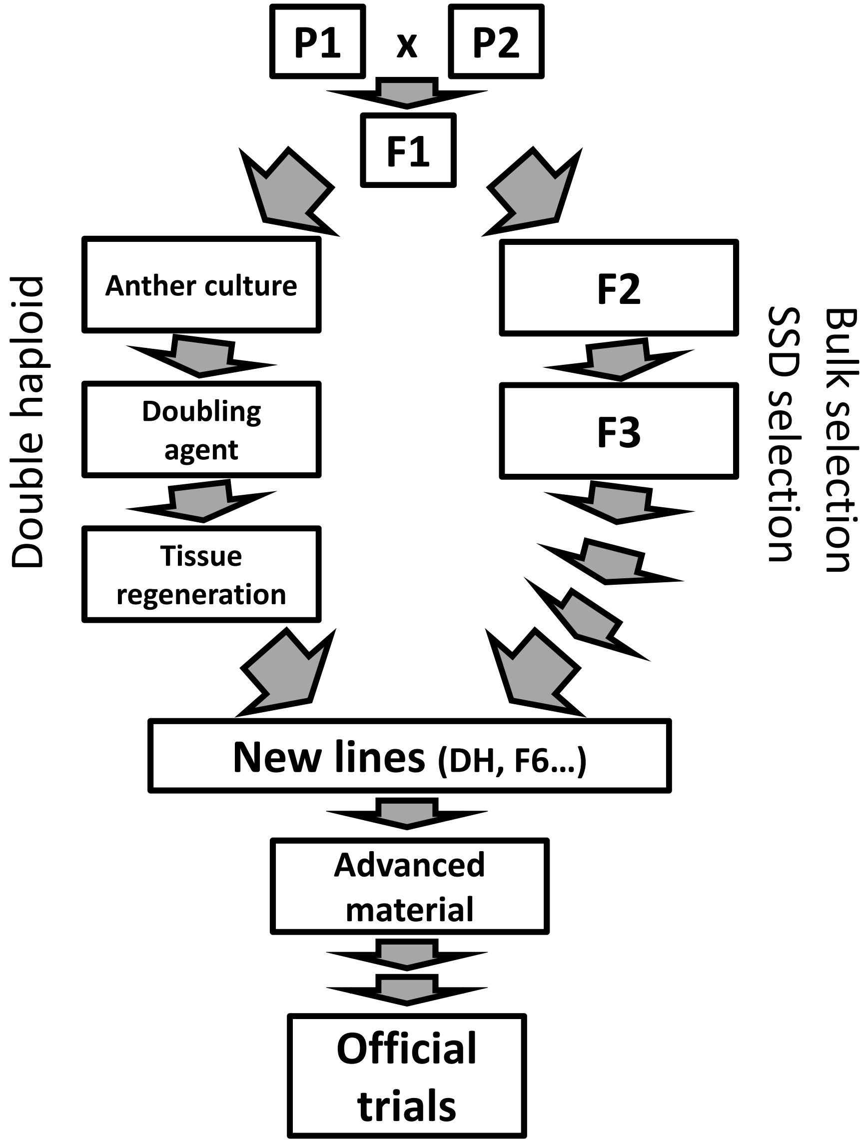
$$\mathbf{E}(\sum_{i=1}^{m} m_i \hat{\alpha}_i) = g$$

Genomic Prediction (GP)

Summing over all **regularized** markers



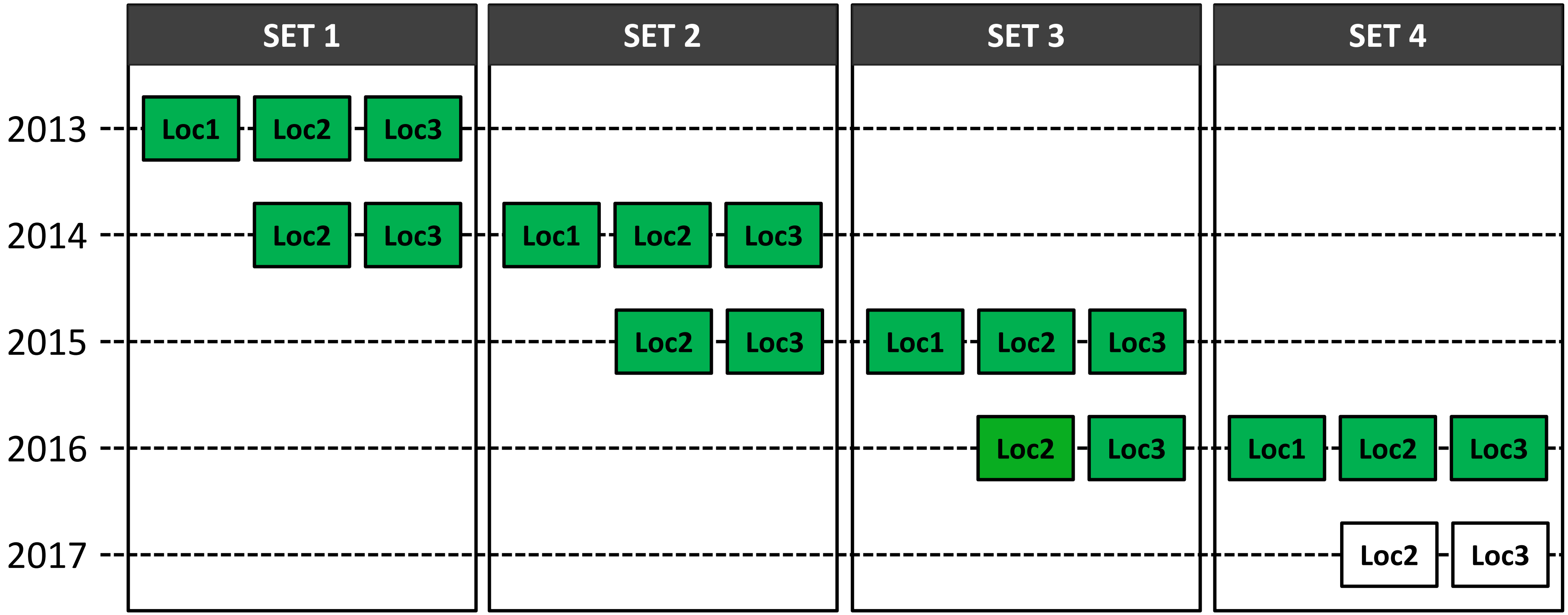
Introduction - cereal breeding scheme



General ideas

1. Variation is created by crossing parental lines
2. Homozygosity is restored
3. Large number of new lines are tested (few replicates) – first selection
4. Few hundreds of advanced lines are phenotyped in several replicates (different locations and years)

Experimental design



- Four sets of ~330 advanced lines from four different breeding cycles
- The lines were tested on three different Danish locations for four consecutive years – 2-3 reps per each environment
- Yield (Y), Lodging (L) and Starch Content (SC) were considered for this work

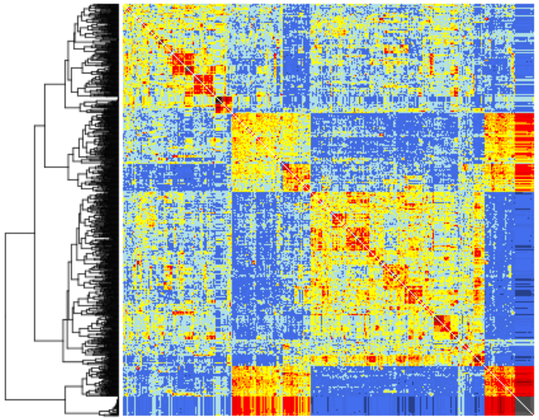
Experimental design: Data

- Four sets of ~330 advanced lines from four different breeding cycles (~ 1300 lines)
- The lines were tested on three different Danish locations for two consecutive years – 2-3 reps per each environment (~ 17000 plots)
- Traits
  - Yield (Y)
  - Lodging (L)
  - Starch Content (SC)

Experimental design – Genomic information

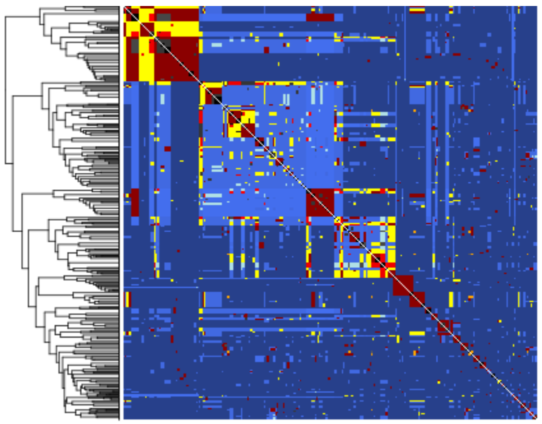
SNP markers

- 15K SNPs (Illumina® wheat SNPs array)
- ~11k SNPs after editing
- G-matrix (**G**): trace the realized genomic relationship between lines



Pedigree

- Pedigree records were available up to the grand-parent lines or further back.
- A-matrix (**A**): trace expected relationship between lines



Heritability, Wheat Traits

Trait	Classic heritability	Genomic heritability
Yield	0.72	0.72
Starch Content	0.71	0.57
Lodging	0.71	0.65

Genomic Prediction

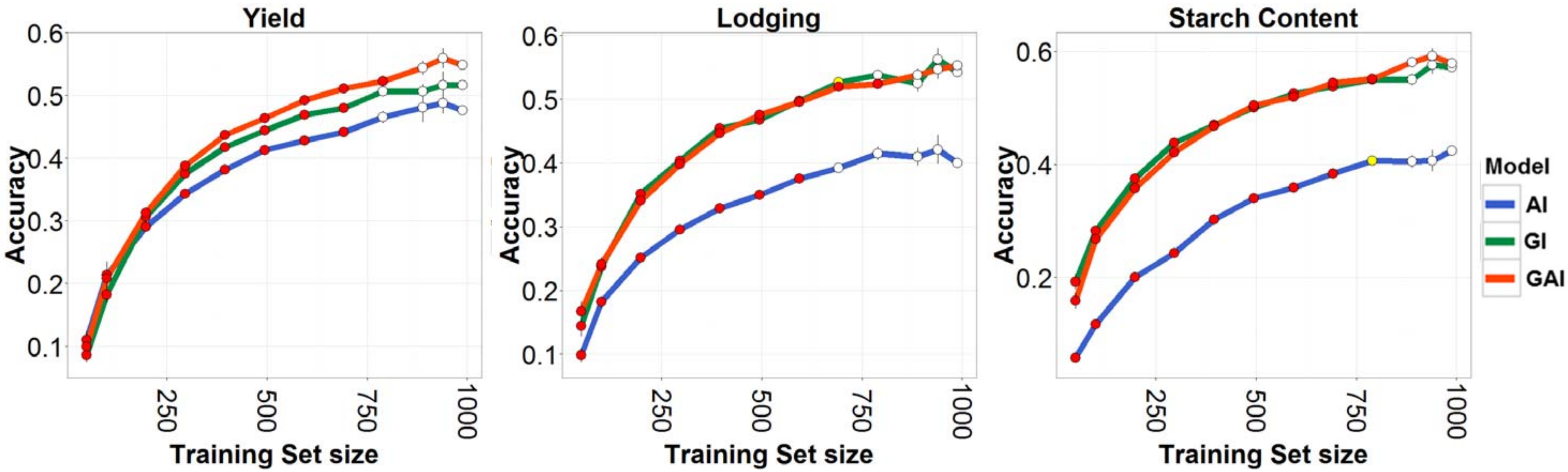
- Training population of lines with genotypes and phenotypes
- Predict new lines with genotypes only
- Model validation by cross validation
  - Leave one line out (LOO)
  - Train model based on all other lines
  - Predict left out line
  - Repeat for all lines
  - Compute accuracy of predicting left out phenotypes

LOO Accuracy of GP in wheat

Trait	Accuracy $r(\hat{g},\bar{p})$	h $r(\bar{p},g)$	Accuracy $r(\hat{g},g)$
Yield	0.52	0.85	0.61
Starch Content	0.56	0.75	0.74
Lodging	0.57	0.81	0.71

Optimize Training population size

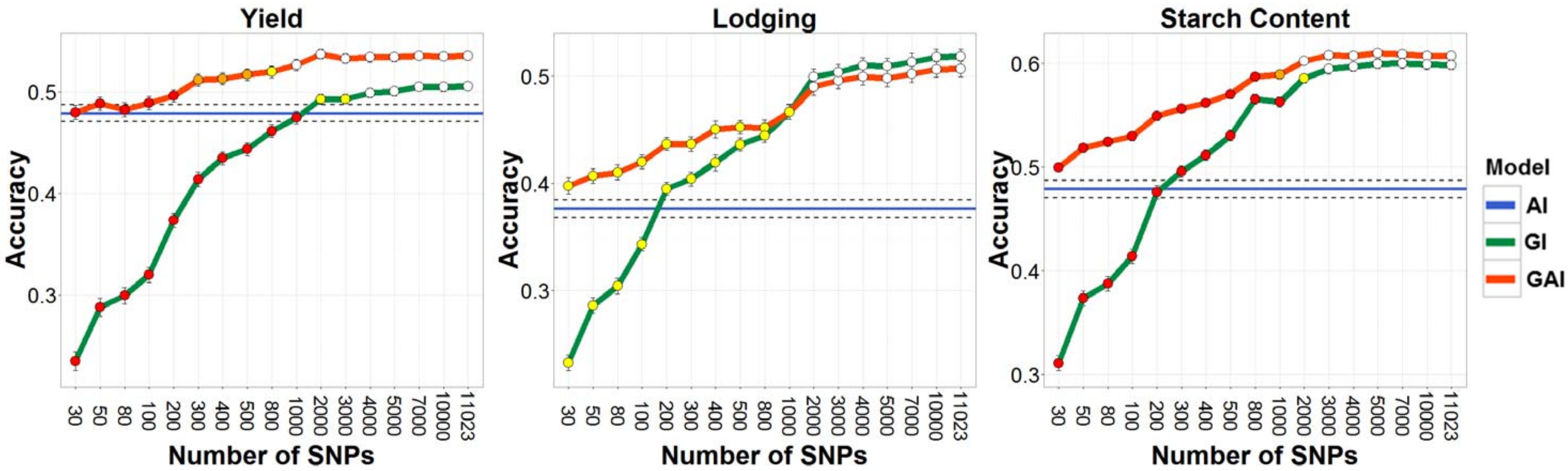
Randomly reduced training population size



1. Models including **G** always outperform the one including only **A**
2. Around 700 lines are enough to maximize the prediction accuracy (~220 per breeding cycles)
3. Phenotyping few hundred of new lines per breeding cycles will allow to predict the others trough GP

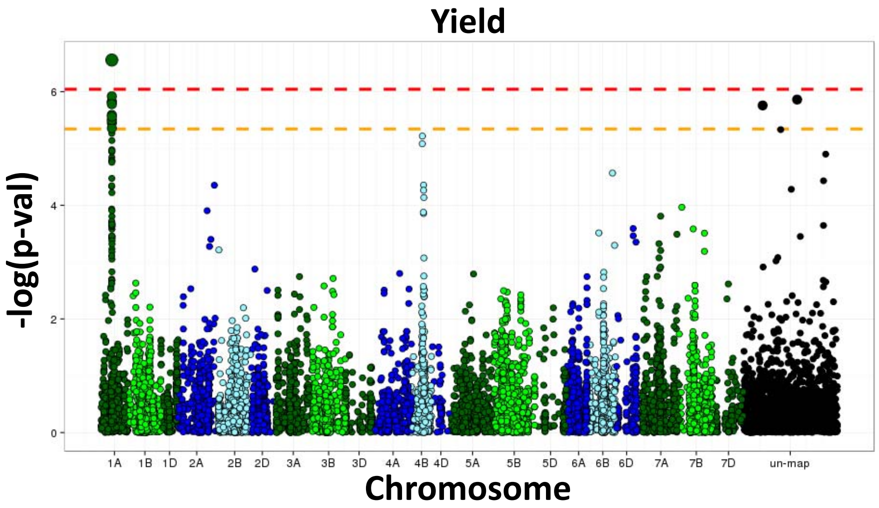
Marker set size: random sampling

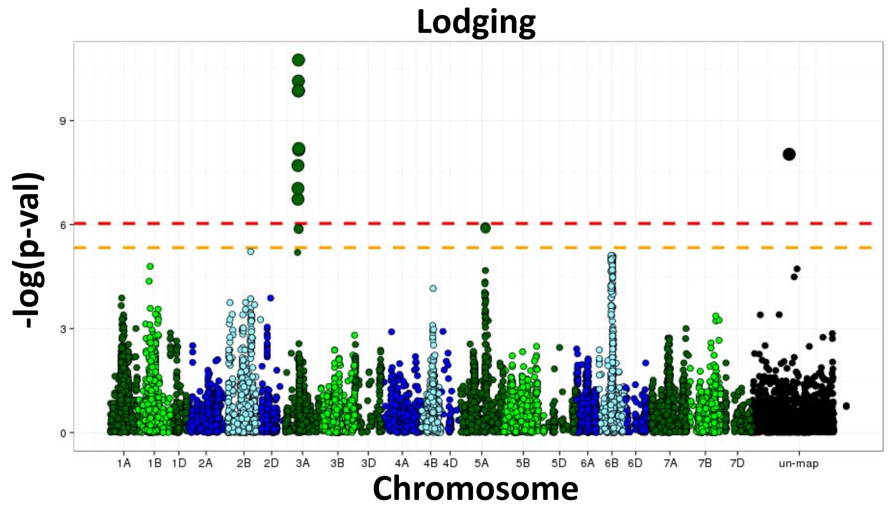
Sets with increasing number of randomly selected markers were used to make predictions

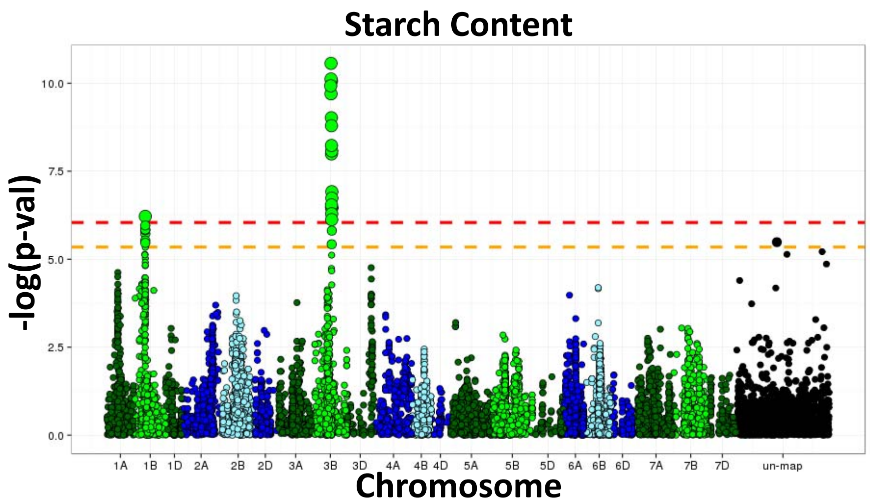


1. Around 1K markers are enough to maximize the prediction accuracy
2. Adding **A** strongly mitigate the loss in accuracy at low markers set size
3. Where is the optimum?

Can we do better that a random sampling?





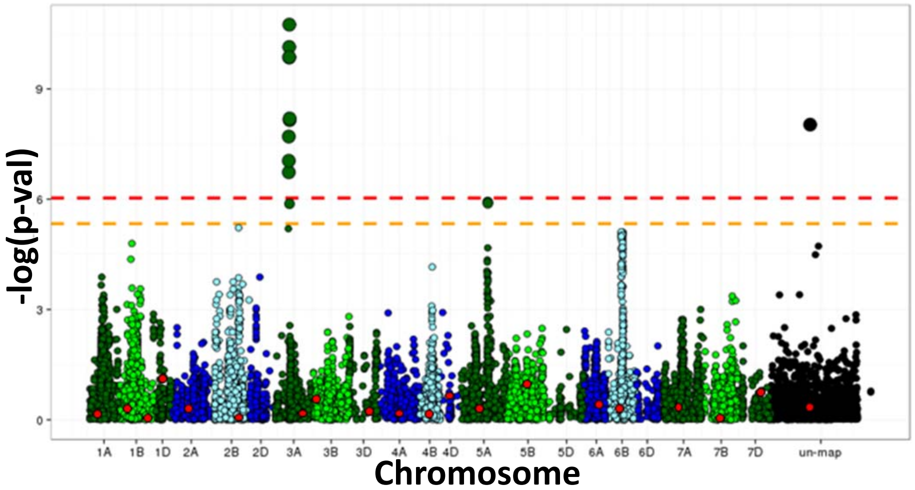


1. It was possible identify some SNP/QTL association
2. Overall the significant markers were explaining 7.4%, 9.5% and 10.1% of the total genetic variance for respectively Y, L and SC

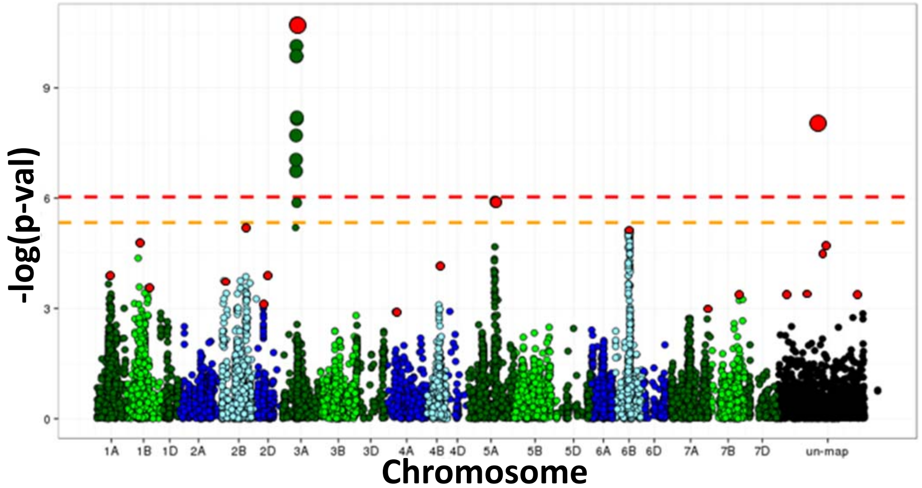
Can we do better that a random sampling?

GWAS results can be used to guide the marker selection

Lodging – Random sampling

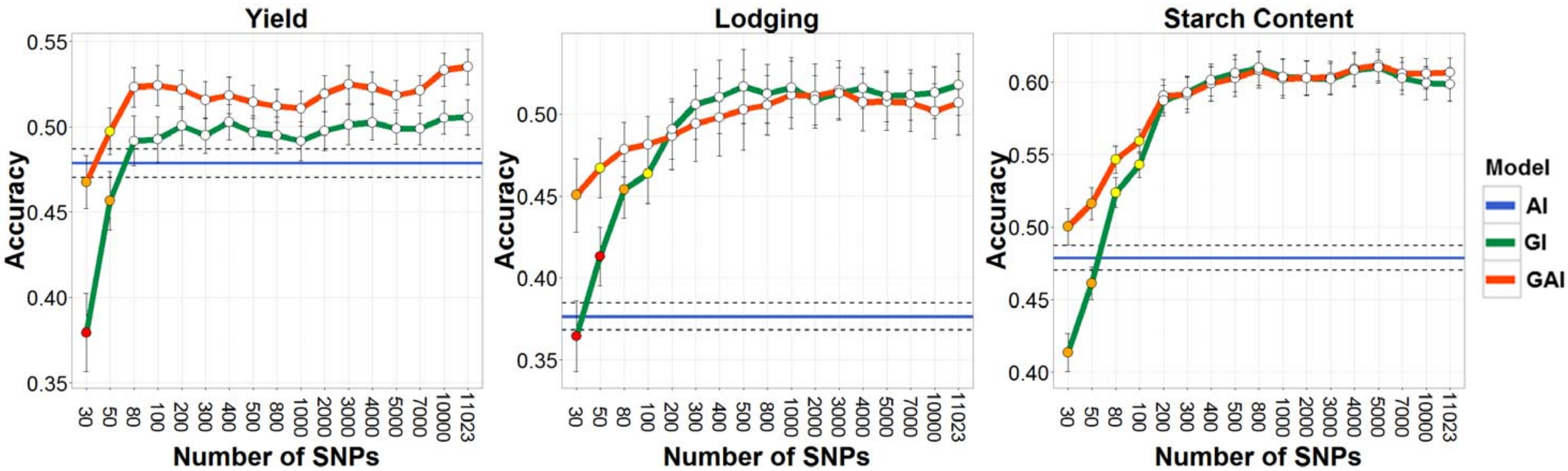


Lodging – GWAS-based sampling



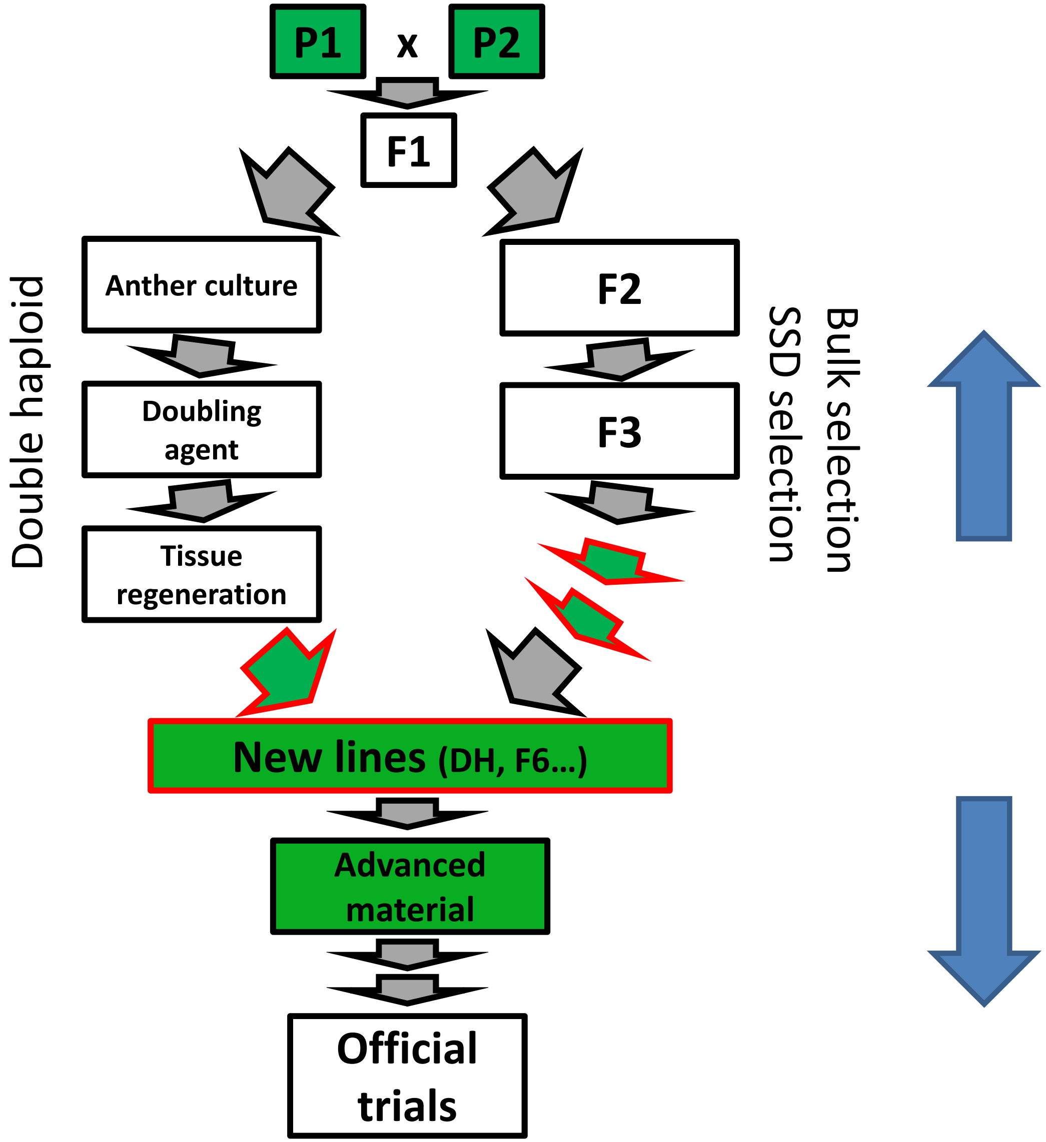
Marker set size: GWAS-based selection

Sets with increasing number of GWAS-based selected markers were used to make predictions



1. 100-200 markers are enough to maximize the prediction accuracy
2. Less markers are needed to maximize GP when selection is based on GWAS results
3. Dense markers still needed to find the significant gene regions in the training pop.

A Genomic Breeding Scheme in Cereals



- Selection of parents
- Speed breeding

- Selection of single plants after DH or SSD
- Selection of lines to test
- Selection of advanced material

Summary

- Genomic selection in cereals is possible
- Models need regular updating
- Genomic selection and MAS can be combined, but dense markers still needed
- Breeding programs need to be modified to take advantage of genomic selection

Acknowledgments

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS

Fabio Cericola, Luc Janss , Andrea Belucci, Biructawit Tessema, Jason Tsai, Just Jensen

Nordic Seed

Ahmed Jahoor, Jihad Orabi, Vahid Edriss, Nanna Hellum Nielsen, Peter Skov Kristensen, Jens Due Jensen, Jeppe Reitan Andersen

Miljø- og Fødevareministeriet gudp

AARHUS UNIVERSITY
DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS