Initiating genomic selection in tetraploid potato

Elsa Sverrisdóttir

Department of Chemistry and Bioscience

GenSAP meeting 15. November 2017

Potatoes

- Third most important food crop worldwide

- 380 million tons produced yearly

- Nutritious: rich in carbohydrates, antioxidants, vitamins, iron

- **Space-efficient**: produce twice the amount of calories per hectare compared to cereals

- Autotetraploid, highly diverse, heterozygous

- Diseases and pests – late blight (kartoffelskimmel)

- Fertilisers and pesticides

- Lack of breeding gain

- Slow breeding cycle: 10-15 years to develop a new variety
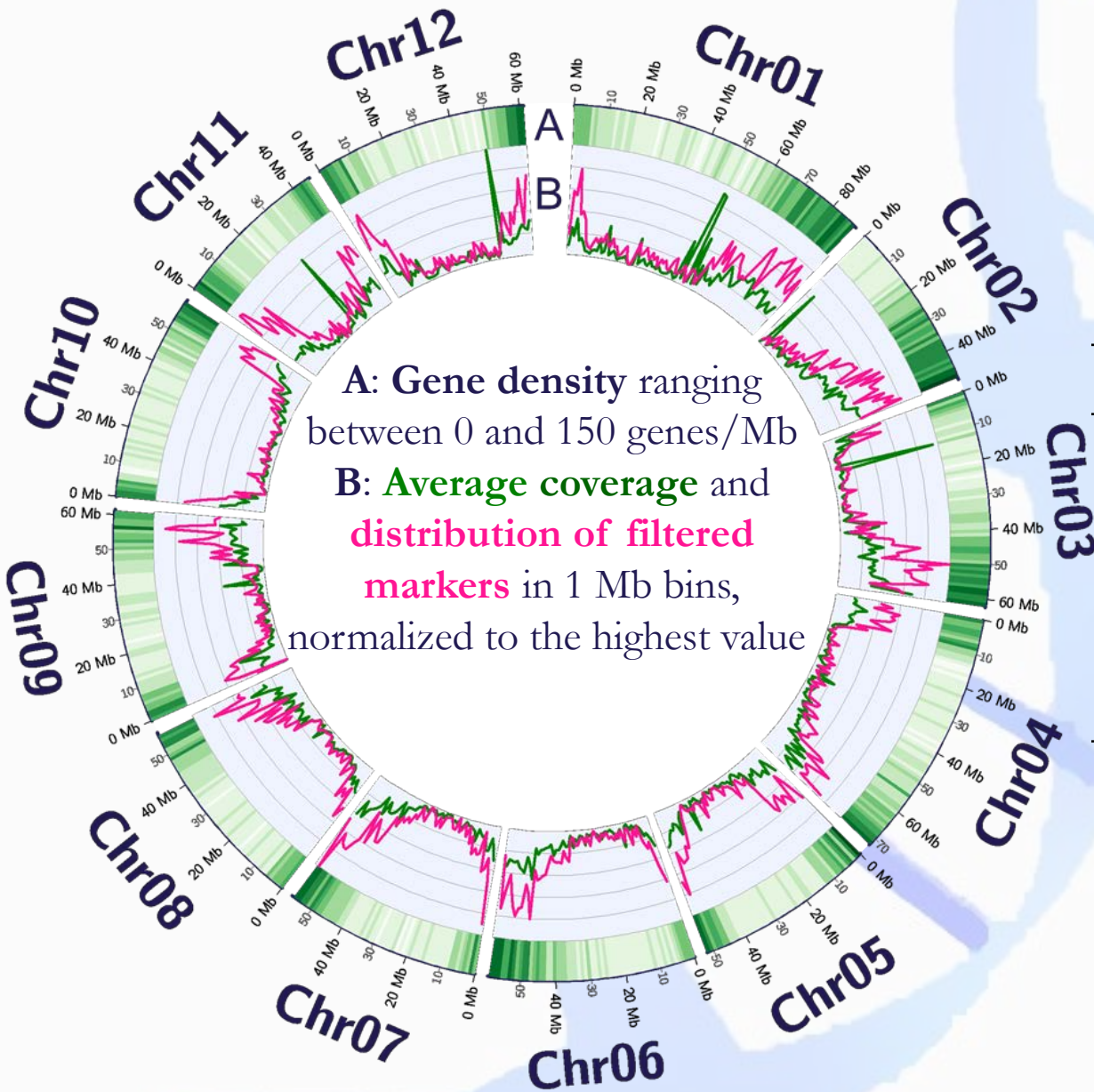
AALBORG UNIVERSITY
DENMARK

Overall aim

- Construct genomic prediction statistical models for important agricultural traits

- Investigate prediction performance across different populations of tetraploid potato

- Traits:
  - **Starch content/dry matter content**
  - **Chipping quality (colour of chip after frying)**
  - Yield
  - Late blight resistance

- Populations:
  - Main: MASPOT population: 762 offspring from 18 parents
  - Test panel DK: 74 elite cultivars and breeding clones (+ 18 parents)
  - Test panel UK: 292 elite cultivars and breeding clones

AALBORG UNIVERSITY
DENMARK

Methods

- Genotyping-by-sequencing with *Ape*KI

- Illumina sequencing: HiSeq 2500

- Statistical models in R: **GBLUP**, BayesA, BayesC

- Correction of phenotypic data:

$$y_{ijk} = \mu + genotype_i + year_j + location_k + e_{ijk}$$

AALBORG UNIVERSITY
DENMARK

A: **Gene density** ranging between 0 and 150 genes/Mb
B: **Average coverage** and **distribution of filtered markers** in 1 Mb bins, normalized to the highest value

Filtering step	Markers
Raw	3.4 million
MAF >1%, missing data <50%	505,321
Coverage >5	186,757
Coverage <60	171,859

Dry matter predictions

Prediction set / Training set	MASPOT	Test panel DK	Test panel UK	Combined
MASPOT [755]	**0.74 [1.04]**	0.67 [1.41]	0.62 [1.55]	0.75 [0.98]
Test panel DK [80]	0.71 [1.91]	**0.82 [1.49]**	0.63 [2.85]	0.83 [1.07]
Test panel UK [290]	0.57 [1.64]	0.37 [2.20]	**0.72 [1.58]**	0.76 [1.32]

AALBORG UNIVERSITY
DENMARK
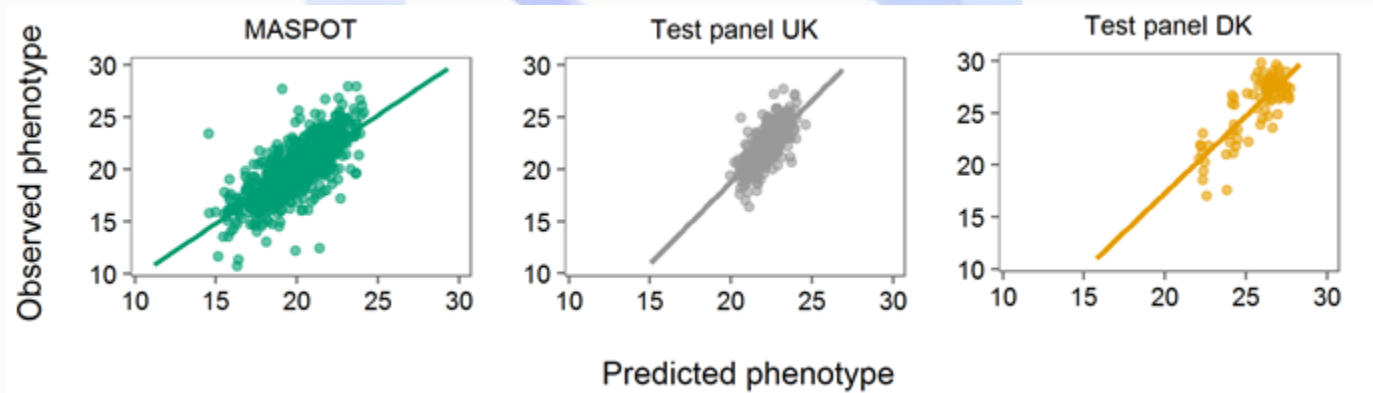
Chipping quality predictions

Prediction set / Training set	MASPOT	Test panel DK	Test panel UK	Combined
MASPOT [524]	**0.56 [1.09]**	0.35 [1.32]	0.30 [0.31]	0.55 [0.80]
Test panel DK [40]	0.48 [1.76]	**0.17 [1.21]**	0.42 [0.63]	0.39 [0.77]
Test panel UK [290]	0.43 [2.04]	0.28 [3.79]	**0.79 [0.93]**	0.79 [0.86]
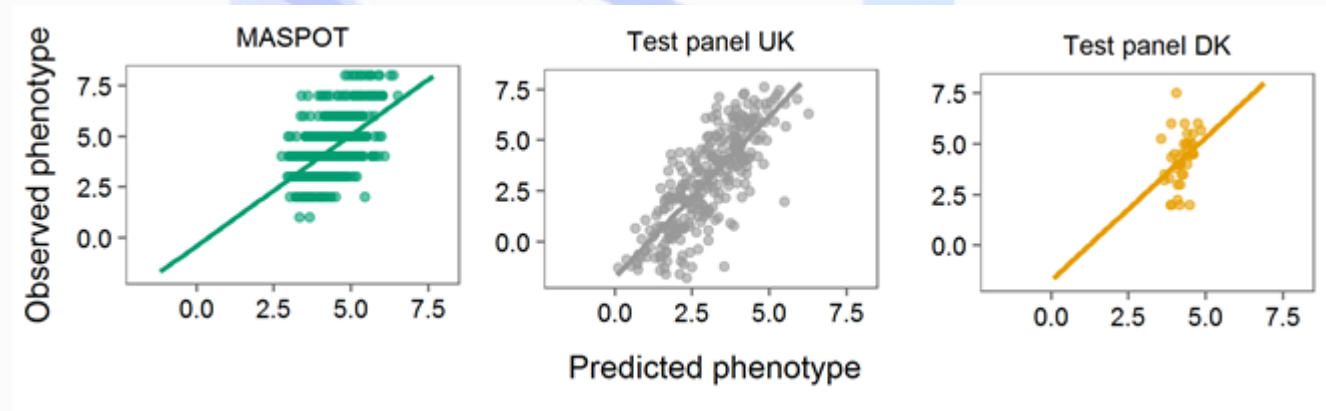
Heritability (pedigree) for MASPOT population
➢Chipping quality: 78%
➢Dry matter content: 90%

AALBORG UNIVERSITY
DENMARK

Predictions within the MASPOT population

Trait / Test set	Cross-validation	GBLUP Correlation	Bias
Chipping quality			
MASPOT [524]	8-fold	0.56	1.10
MASPOT [524]	Leave-sibs-out	0.47	1.41
Starch content			
MASPOT [755]	8-fold	0.73	1.04
MASPOT [755]	Leave-sibs-out	0.68	1.50

AALBORG UNIVERSITY
DENMARK
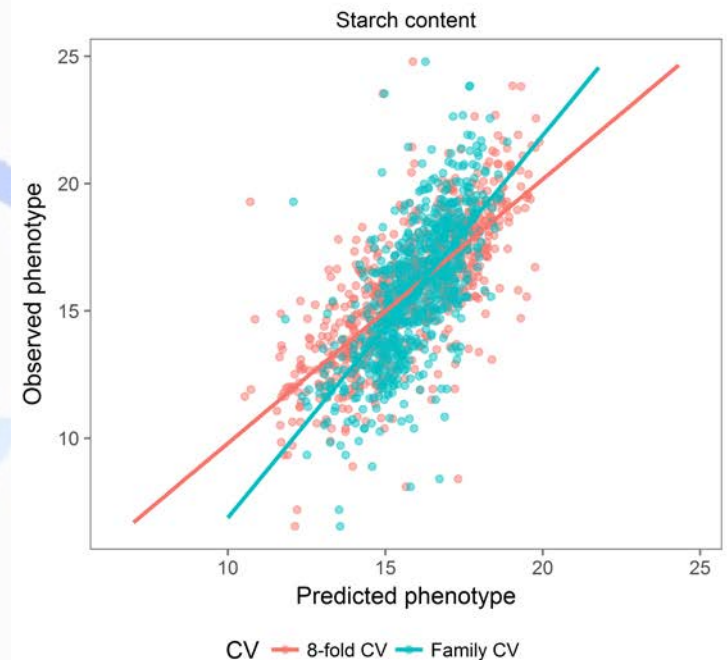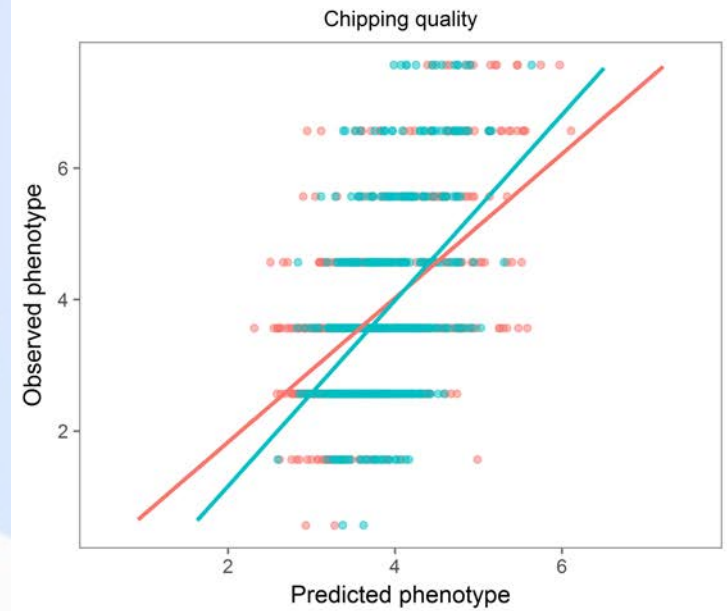
GWAS: Genome-wide association study

$$y = 1\mu + x\beta + g + e$$

x: marker genotype vector

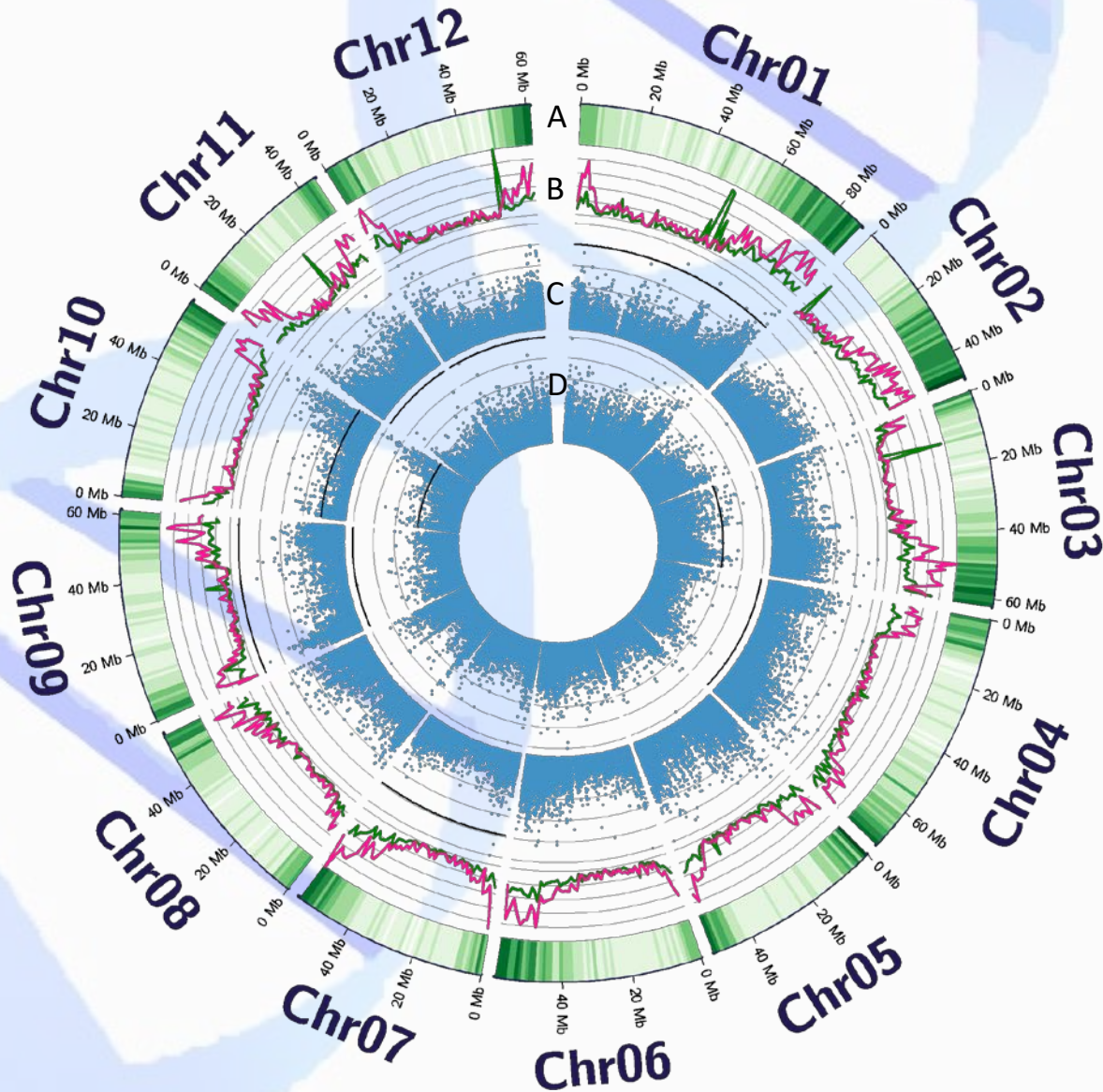β: marker effect

→p-values for each marker position

→FDR: significance threshold

→372 SNPs for chipping quality

→612 SNPs for starch content

-log(p) normalised between 0 and 1
C: Starch content
D: Chipping quality

AALBORG UNIVERSITY
DENMARK

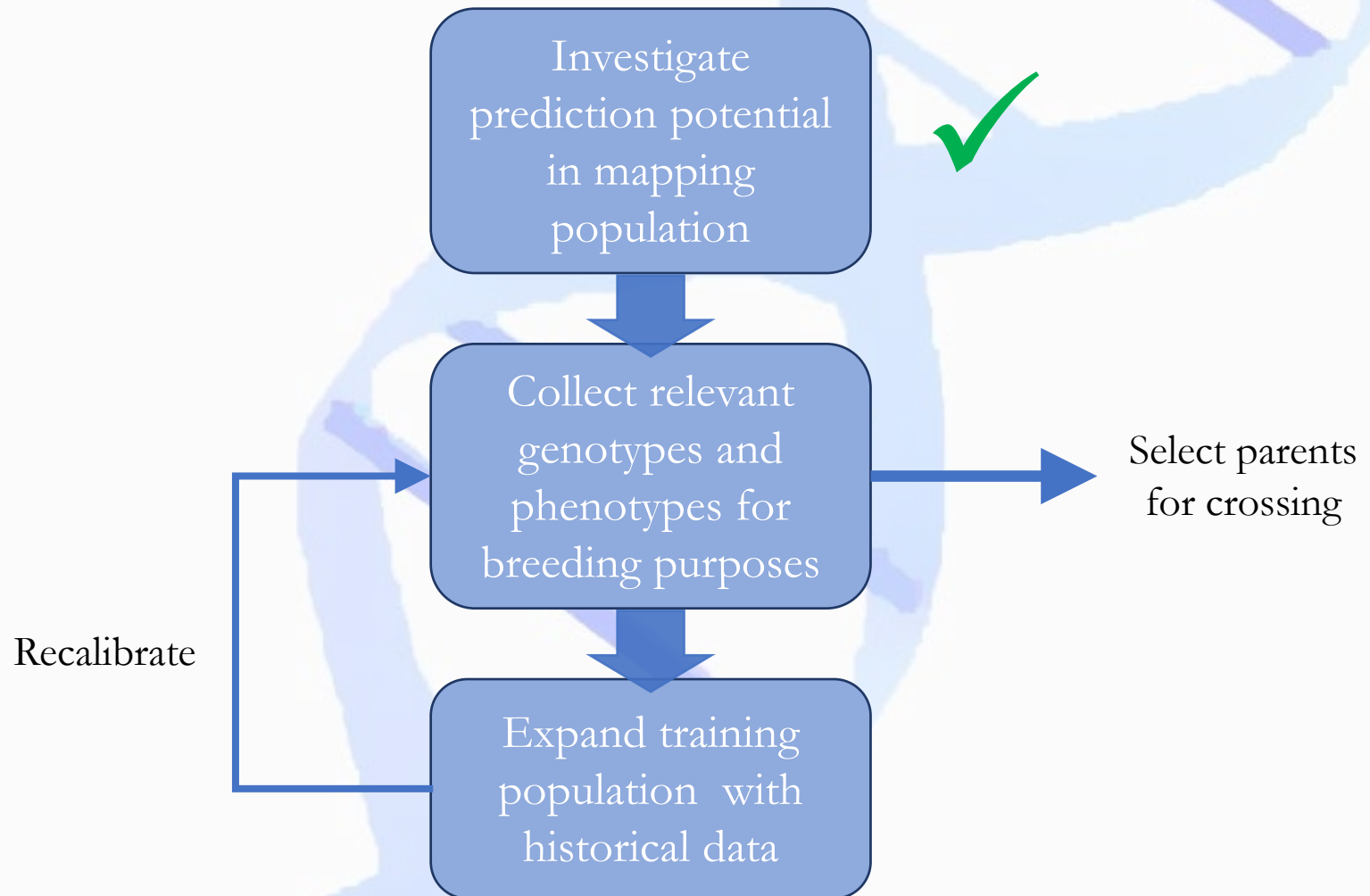| Trait / Test set	Training set	Markers	GBLUP Correlation	Bias
Chipping quality				
Test panel [30]	MASPOT	All [171,859 SNPs]	0.30	1.47
Test panel [30]	MASPOT	GWAS [372 SNPs]	0.17	0.29
Test panel [30]	Combined*	All [171,859 SNPs]	0.37	1.32
Test panel [30]	Combined*	GWAS [372 SNPs]	0.30	0.63
Starch content				
Test panel [63]	MASPOT	All [171,859 SNPs]	0.42	1.26
Test panel [63]	MASPOT	GWAS [612 SNPs]	0.11	0.19
Test panel [63]	Combined*	All [171,859 SNPs]	0.65	1.04
Test panel [63]	Combined*	GWAS [612 SNPs]	0.34	0.48

AALBORG UNIVERSITY
DENMARK
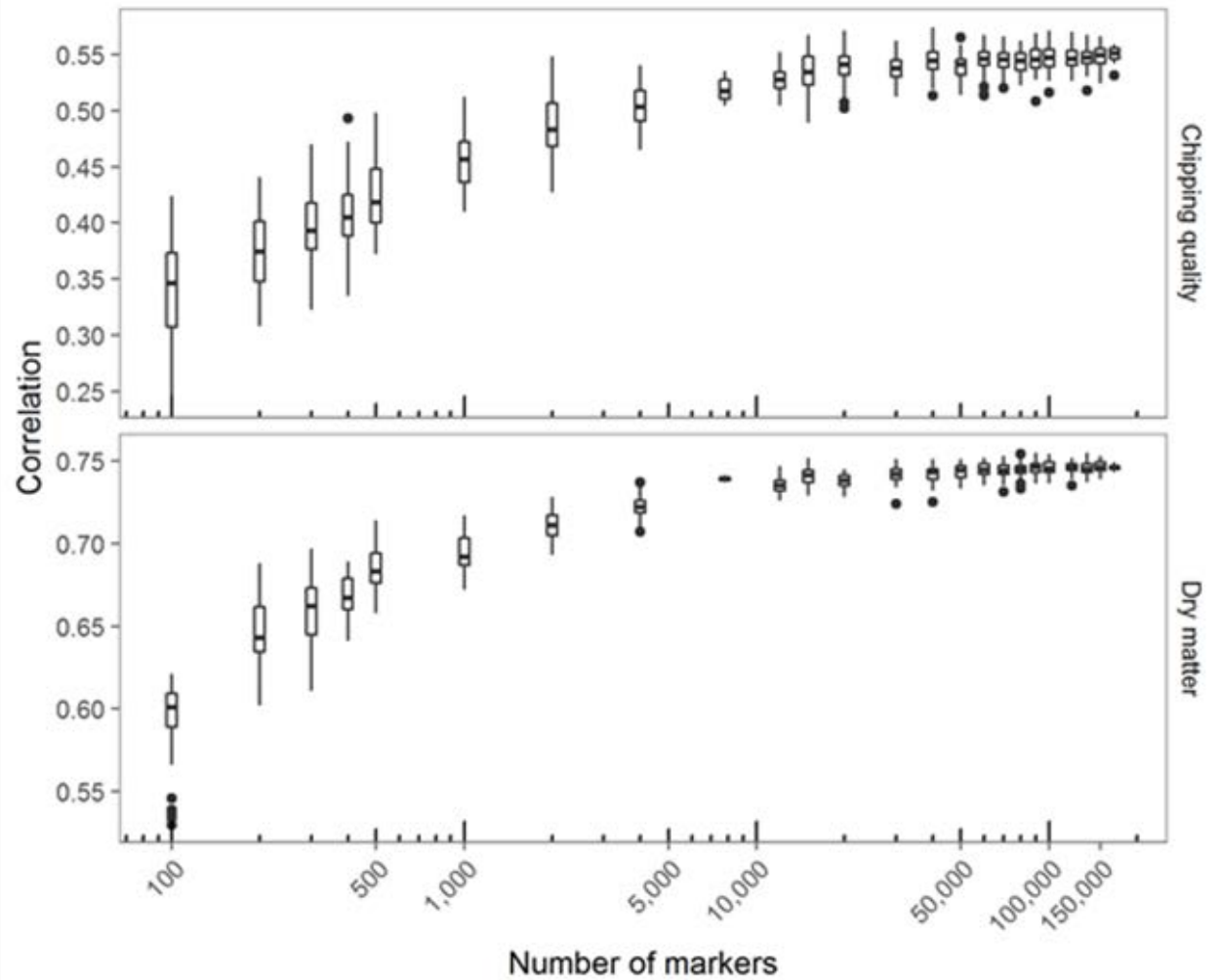
Marker number

Marker number
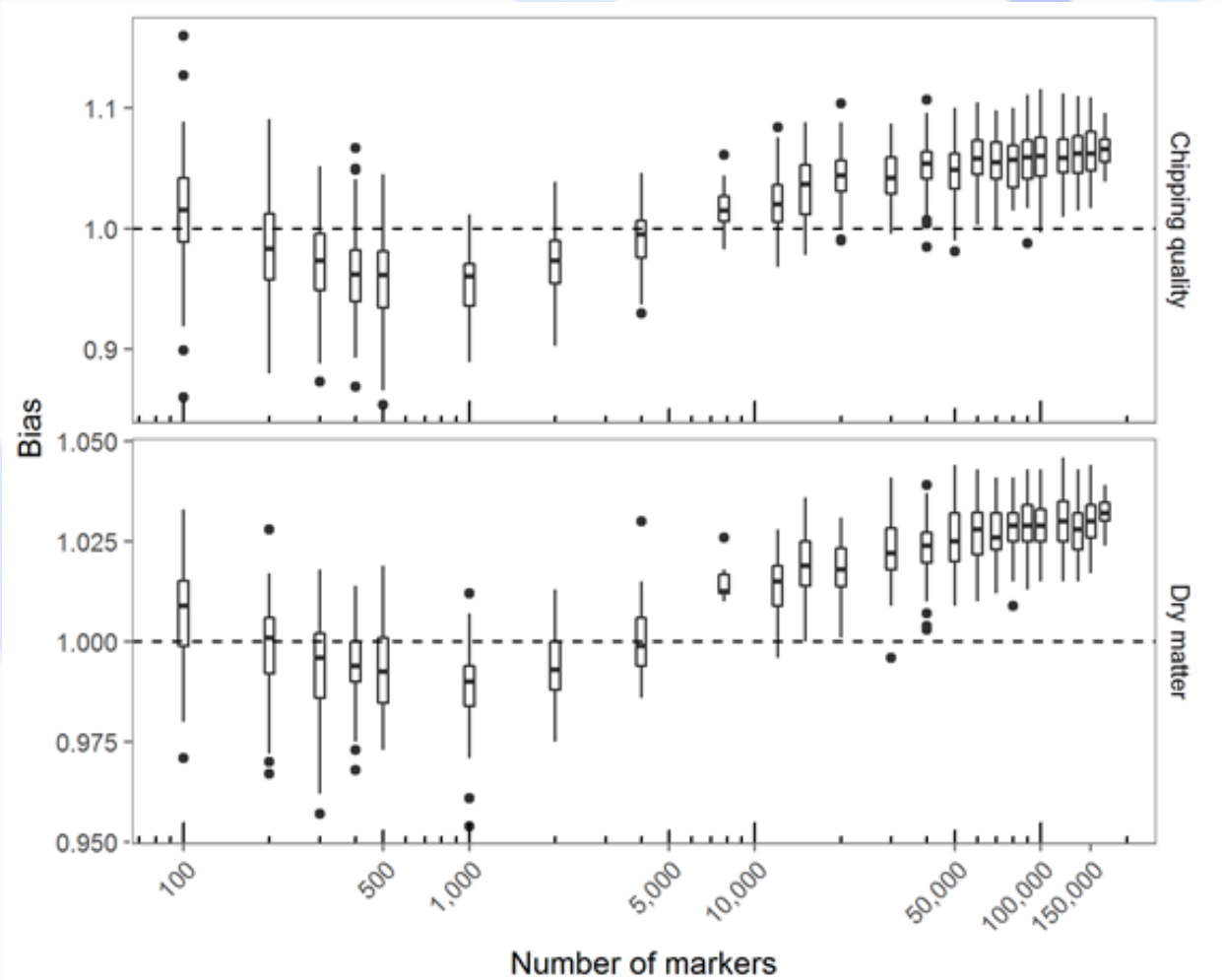
Conclusions

- Chipping quality and starch content/dry matter content can be predicted with moderate to high accuracy within the same population

- Low to moderate prediction accuracies are obtained across populations with large biases

- Maximal prediction accuracy could be obtained for all populations simultaneously if relevant genotypes are included in the model

- Predictions could not be improved by only using significant SNPs selected with GWAS

- SNPs selected with GWAS in one population are not (necessarily) significant for other populations

AALBORG UNIVERSITY
DENMARK

Implementing genomic selection in tetraploid potato

Investigate prediction potential in mapping population ✓

Collect relevant genotypes and phenotypes for breeding purposes

Select parents for crossing

Recalibrate

Expand training population with historical data

AALBORG UNIVERSITY
DENMARK

Acknowledgements

Kåre Lehmann Nielsen, Aalborg University
Functional Genomics group

Torben Asp, Aarhus University
Stephen Byrne, Aarhus University (Present: Teagasc, Ireland)
Luc Janss, Aarhus University

Glenn Bryan, The James Hutton Institute

Hanne Grethe Kirk, LKF Vandel (Present: Danespo)
Jens Kristian, LKF Vandel (Present: Danespo)

THANK YOU FOR YOUR ATTENTION

AALBORG UNIVERSITY
DENMARK