Improvement of genomic prediction in advanced wheat breeding lines by including additive-by-additive epistasis

PhD student (MSc) Miguel Angel Raffo
Supervisor: Prof. Just Jensen
Center for Quantitative Genetics and Genomics (QGG)
Aarhus University - Denmark

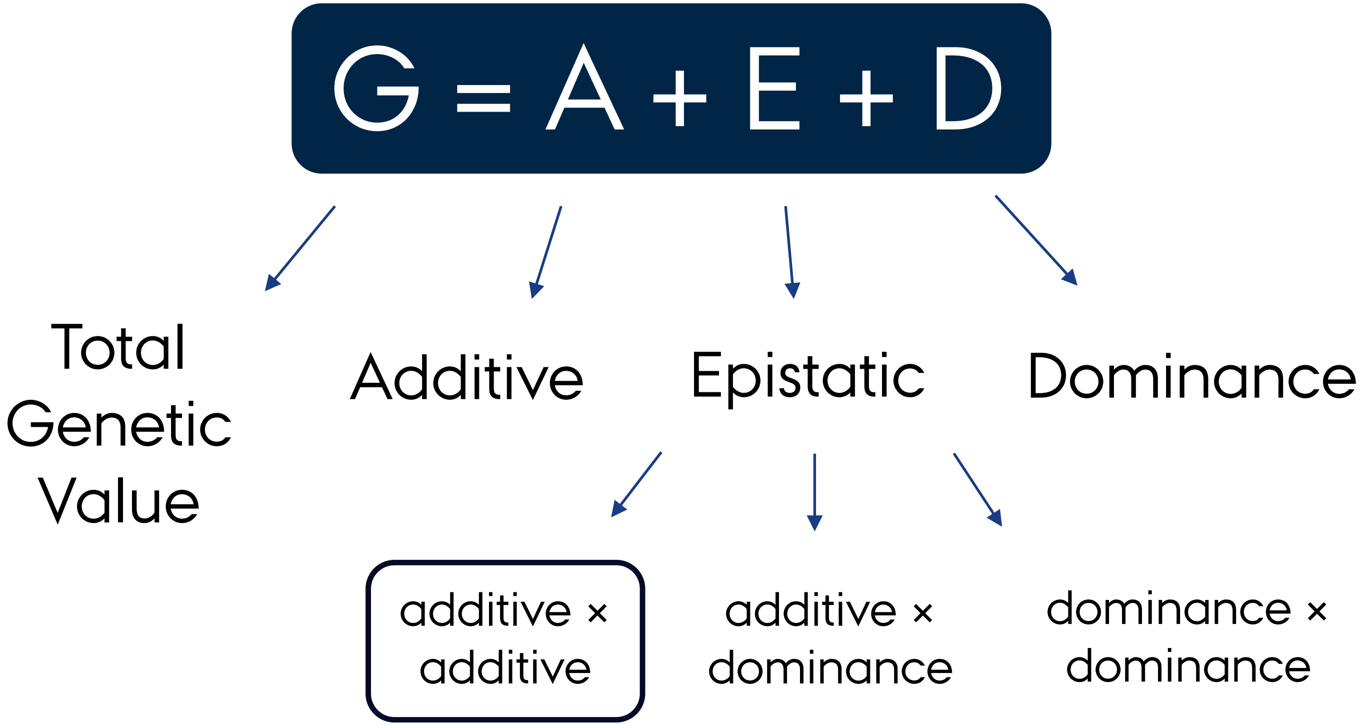
OUTLINE

- Background
- Statistical models
- Model validation
- Variance components
- Predictive performance
- Discussion
- Take home message
- Q&A

M&M

Results

BACKGROUND



$G = A + E + D$

Total Genetic Value

Additive

Epistatic

Dominance

additive × additive

additive × dominance

dominance × dominance

MOST RELEVANT NON-ADDITIVE EFFECT IN INBRED LINES

BACKGROUND

- Fundamental role in complex trait determination

(Huang and Mackay, 2016)

- Interactions are fixed in inbred lines

- Opportunity for a more efficient selection strategy

Parent for crosses based on additive effect

Cultivars based on total genetic merit

However ... *low implementation of models including epistasis in wheat breeding / literature reports variable results*

AIMS

1) To evaluate the performance of genetic models including additive and additive × additive epistatic effects for variance components (VC) estimation

2) To investigate whether including epistasis in genomic prediction enhances the predictive ability

PROPOSAL

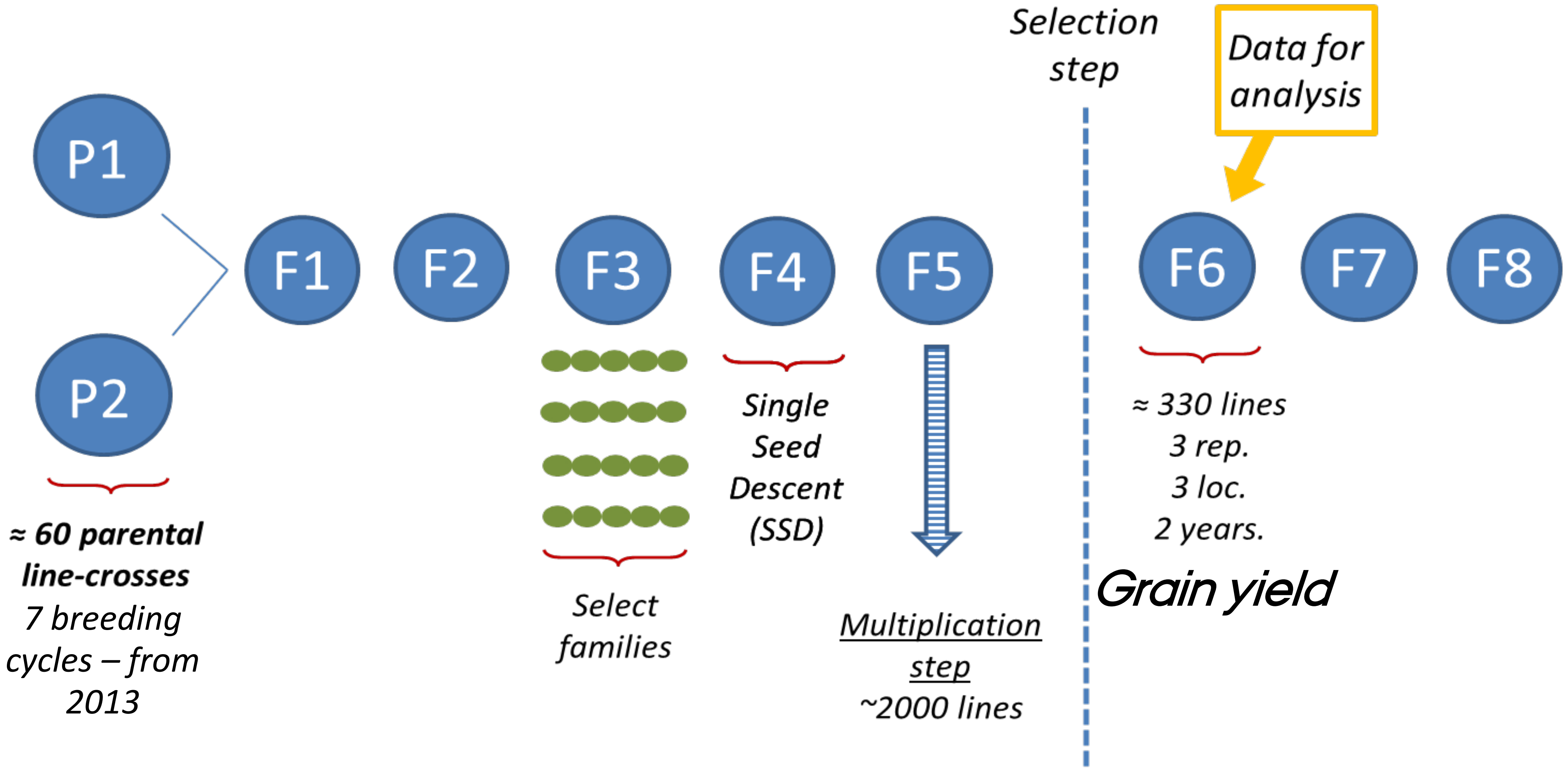
Natural and orthogonal interactions approach (NOIA)

(Alvarez-Castro and Carlborg 2007, Vitezica et al. 2017)

- Hardy Weinberg-equilibrium (HWE) ✓

- Linkage equilibrium (LE)

DATA – WHEAT BREEDING PROGRAM



Selection step

Data for analysis

≈ 60 parental line-crosses
7 breeding cycles – from 2013

Select families

Single Seed Descent (SSD)

Multiplication step
~2000 lines

≈ 330 lines
3 rep.
3 loc.
2 years.

Grain yield

STATISTICAL MODELS

1. *I-model (baseline) - without genomic information*

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{b} + \boldsymbol{Z}_1\boldsymbol{a} + \boldsymbol{Z}_2\boldsymbol{f} + \sum_{i=1}^{n=9}\boldsymbol{Z}_i\boldsymbol{s} + \boldsymbol{e}$$

Grain yield

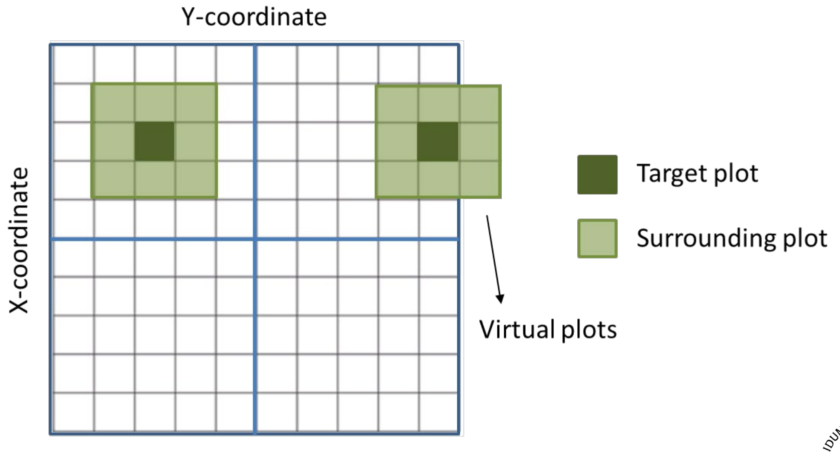
FE: Trial “block”

R: Line

R: Line x Env. “GxE”

R: Spatial effect

R: error



STATISTICAL MODELS

2. *G-model*

Baseline + Additive genomic effect based on NOIA parametrization (Alvarez-Castro and Carlborg 2007, Vitezica et al. 2017)

3. G_A+G_I-*model*

G-model + Epistatic genomic effects based on NOIA parametrization (Alvarez-Castro and Carlborg 2007, Vitezica et al. 2017)

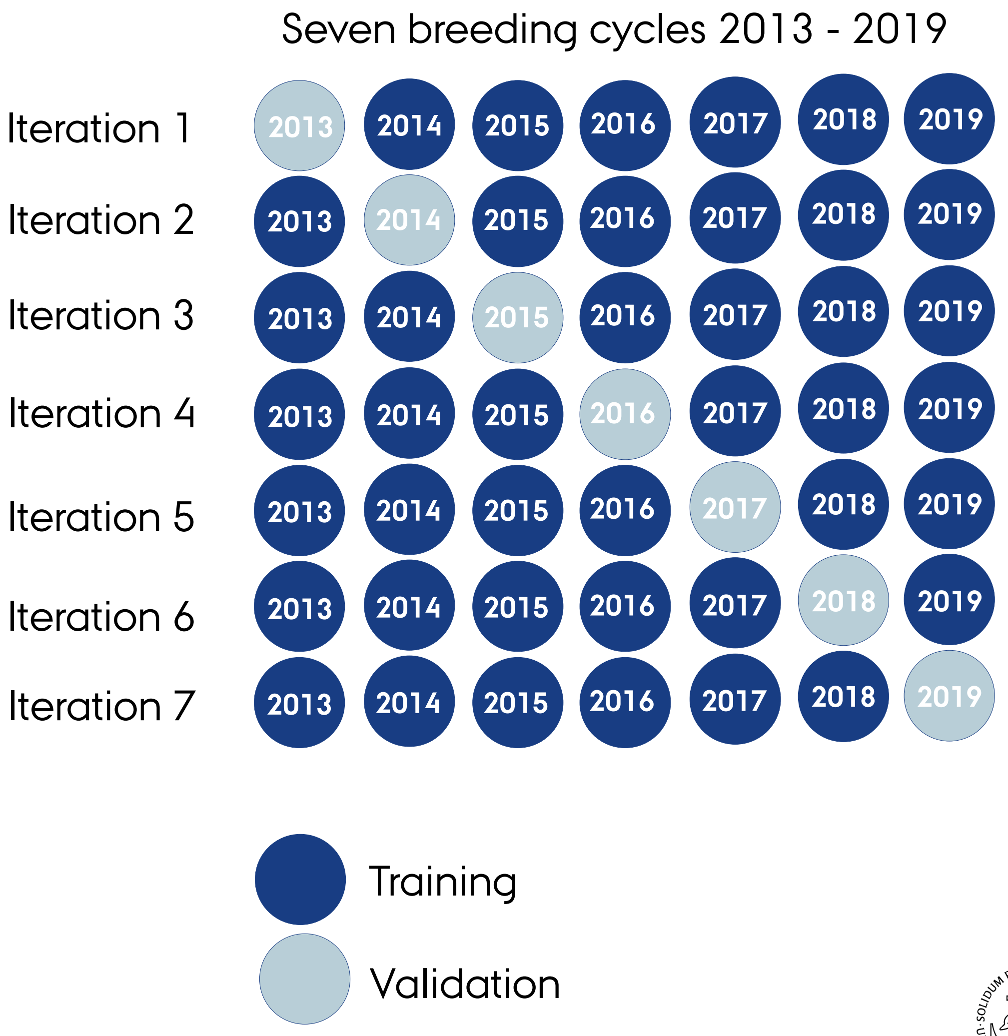
MODEL VALIDATION

Two cross-validation (CV) schemes:

1) Leave-one-line-out CV

- Highest possible relationships between training and validation sets
- Measure of the potential predictive performance of models

2) Leave-one-breeding-cycle-out CV



Seven breeding cycles 2013 - 2019

Iteration 1	○ 2013	● 2014	● 2015	● 2016	● 2017	● 2018	● 2019
Iteration 2	● 2013	○ 2014	● 2015	● 2016	● 2017	● 2018	● 2019
Iteration 3	● 2013	● 2014	○ 2015	● 2016	● 2017	● 2018	● 2019
Iteration 4	● 2013	● 2014	● 2015	○ 2016	● 2017	● 2018	● 2019
Iteration 5	● 2013	● 2014	● 2015	● 2016	○ 2017	● 2018	● 2019
Iteration 6	● 2013	● 2014	● 2015	● 2016	● 2017	○ 2018	● 2019
Iteration 7	● 2013	● 2014	● 2015	● 2016	● 2017	● 2018	○ 2019

● Training

○ Validation

RESULTS

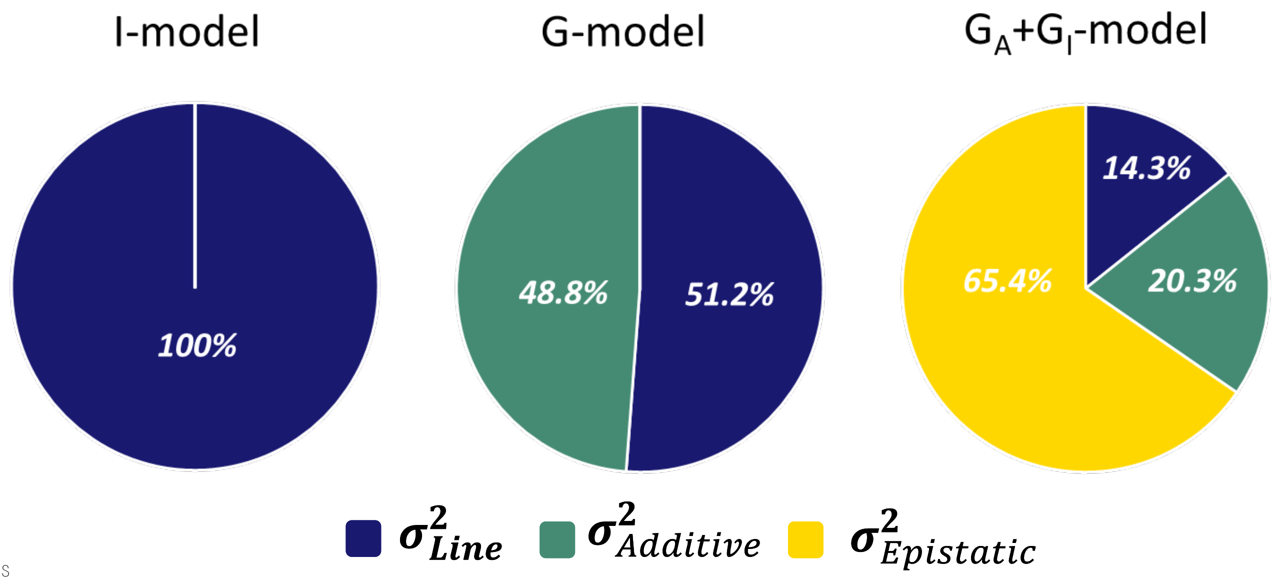
VARIANCE COMPONENTS ESTIMATION

Models	σ^2_{Line}	σ^2_{Add}	σ^2_{Epi}	$\sigma^2_{Spatial}$	σ^2_{GxE}	σ^2_{error}
1. I-model	0,089	-	-	0,043	0,131	0,057
2. G-model	0,053	0,051	-	0,044	0,131	0,057
3. G_A+G_I-model	0,014	0,020	0,064	0,044	0,131	0,057

RESULTS

VARIANCE COMPONENTS ESTIMATION

Models	σ^2_{Line}	σ^2_{Add}	σ^2_{Epi}	$\sigma^2_{Spatial}$	σ^2_{GxE}	σ^2_{error}
1. I-model	0,089	-	-	0,043	0,131	0,057
2. G-model	0,053	0,051	-	0,044	0,131	0,057
3. G_A+G_I-model	0,014	0,020	0,064	0,044	0,131	0,057



- *Lack of orthogonality*
- *Linkage Equilibrium is assumed in NOIA but not fulfilled*
- *Performance for genomic prediction ?*

RESULTS

GENOMIC PREDICTION

Predictive ability

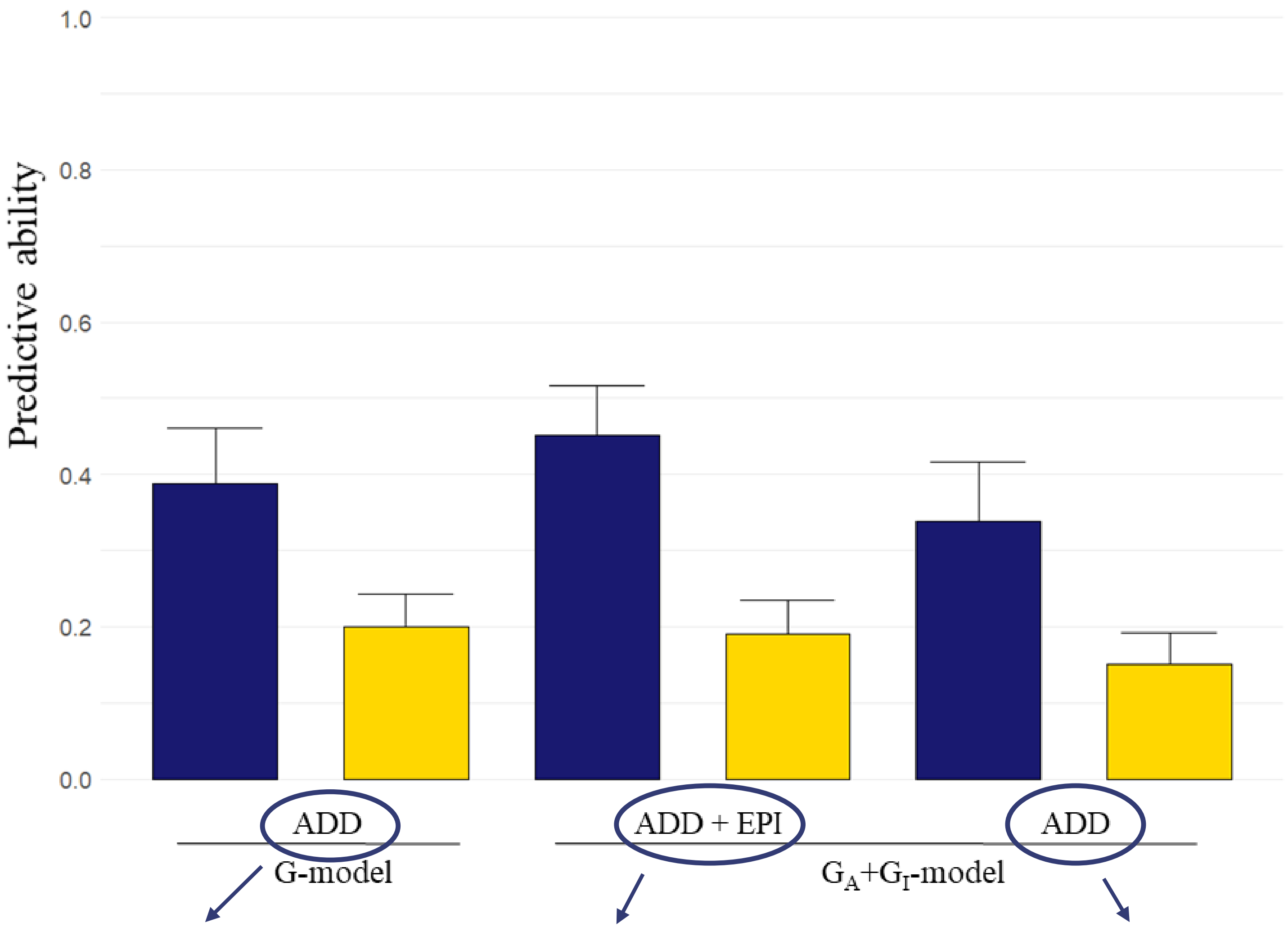
correlation (predictions, phenotypes corrected by fixed effects)

Cross-validations:

- 🟦 Leave-one-line-out
- 🟨 Leave-one-breeding-cycle-out

distribution of PAs bootstrap sampling (r=10,000)

Two-tailed paired t-test



GEBVs

GEBVs + epistatic values

GEBVs

RESULTS

GENOMIC PREDICTION

Predictive ability

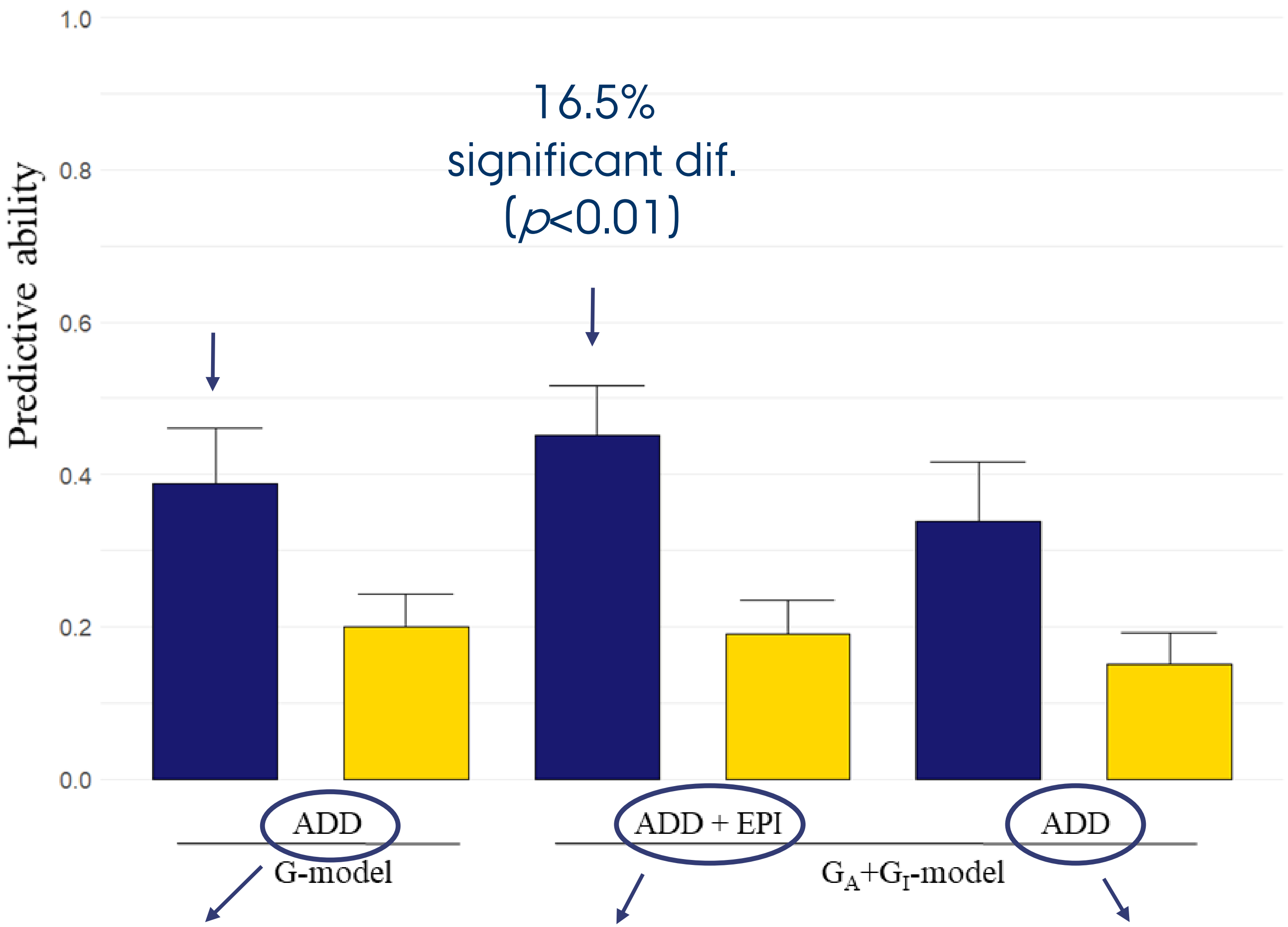
correlation (predictions, phenotypes corrected by fixed effects)

Cross-validations:

- 🟦 Leave-one-line-out
- 🟨 Leave-one-breeding-cycle-out

distribution of PAs bootstrap sampling (r=10,000)

Two-tailed paired t-test



16.5%
significant dif.
(p<0.01)

GEBVs

GEBVs + epistatic values

GEBVs

RESULTS

GENOMIC PREDICTION

Predictive ability

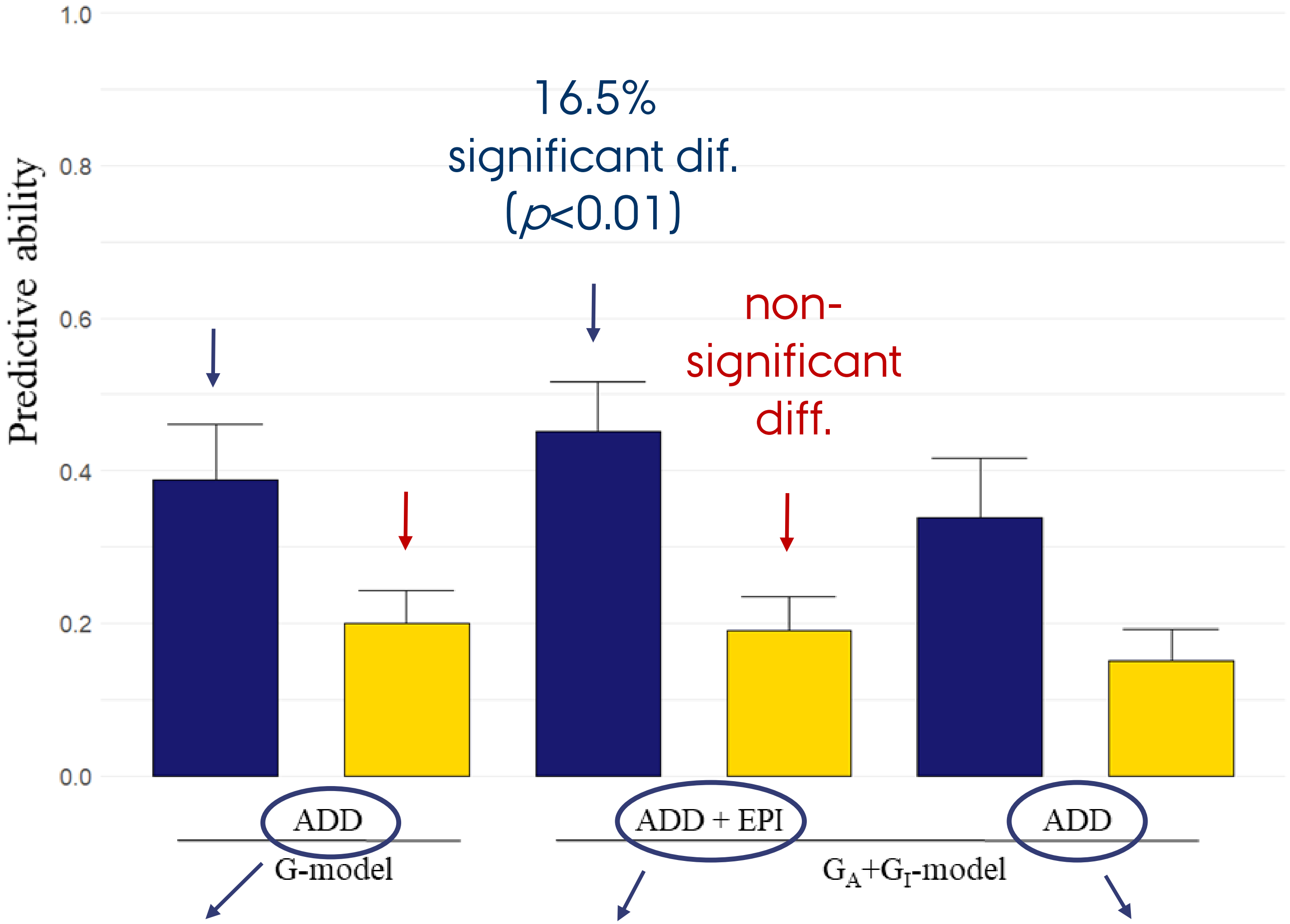
correlation (predictions, phenotypes corrected by fixed effects)

Cross-validations:

- 🟦 Leave-one-line-out
- 🟨 Leave-one-breeding-cycle-out

distribution of PAs bootstrap sampling (r=10,000)

Two-tailed paired t-test



16.5% significant dif. (p<0.01)

non-significant diff.

GEBVs

GEBVs + epistatic values

GEBVs

DISCUSSION

Lack of orthogonality of variance components estimation:

- *NOIA* model working under LE assumption, LD between loci/QTLs affect estimation

Effect of including epistasis in predictive ability:

leave-one-line-out (LOO) ≠ leave-one-breeding cycle out (LSO)

significant (16.5%) no significant

- additive-by-additive pairwise interaction

→ more likely pairs involved are present in close relatives

→ weaker LD depends on the genetic distance between each gene of the pair and their linked markers

TAKE HOME MESSAGE

“Orthogonal partitioning of genetic variations is not possible with current models, but additive-by-additive effects can be useful to enhance predictions of wheat breeding lines”

ACKNOWLEDGEMENTS

CENTER FOR QUANTITATIVE GENETICS AND GENOMICS QGG

Just Jensen

Xiangyu Guo

Huiming Liu

Nordic Seed

Pernille Sarup

Jeppe Reitan Andersen

Jihad Orabi

Ahmed Jahoor

TAG THEORETICAL AND APPLIED GENETICS

International Journal of Plant Breeding Research

https://rdcu.be/cEb51

ANII AGENCIA NACIONAL DE INVESTIGACIÓN E INNOVACIÓN

National Research and Innovation Agency of Uruguay

Questions ?

Email: mraffo@qgg.au.dk

LinkedIn: https://www.linkedin.com/in/miguel-angel-raffo-16b47a62/

AARHUS
UNIVERSITY