AN EQUATION TO PREDICT THE ACCURACY OF MULTI-BREED GENOMIC PREDICTION WITH DIFFERENTIALLY WEIGHTED MARKERS

Biaty Raymond, Aniek C. Bouwman, Yvonne C.J. Wientjes, Chris Schrooten, Jeanine Houwing-Duistermaat and Roel F. Veerkamp

GENOMIC PREDICTION (GP)

- Accuracy depends on size of reference population (N)

- Numerically small breeds: Multi-breed GP?

Objectives:

- Test a new multi-breed genomic prediction model with differentially weighted markers

- Deterministically predict the accuracy of the new model

PROPOSED MODEL FOR MULTI-BREED GENOMIC PREDICTION

Ref. pop. (J)

$$rG(J,J) = 1$$

Valid. Pop. (J)

+

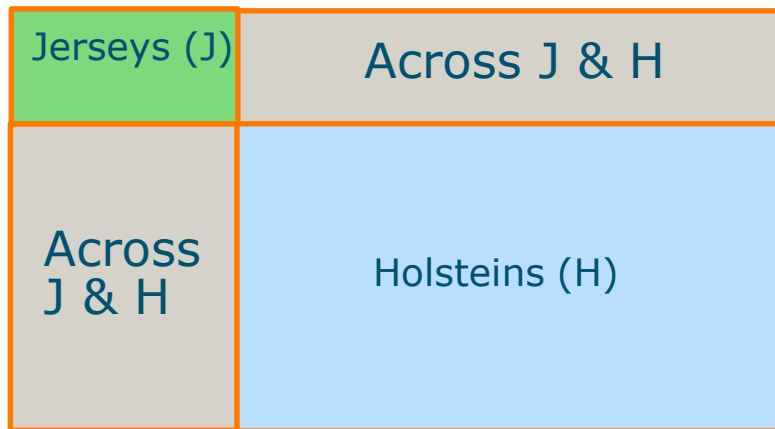Ref. pop. (H)

$$rG(J,H) < 1$$
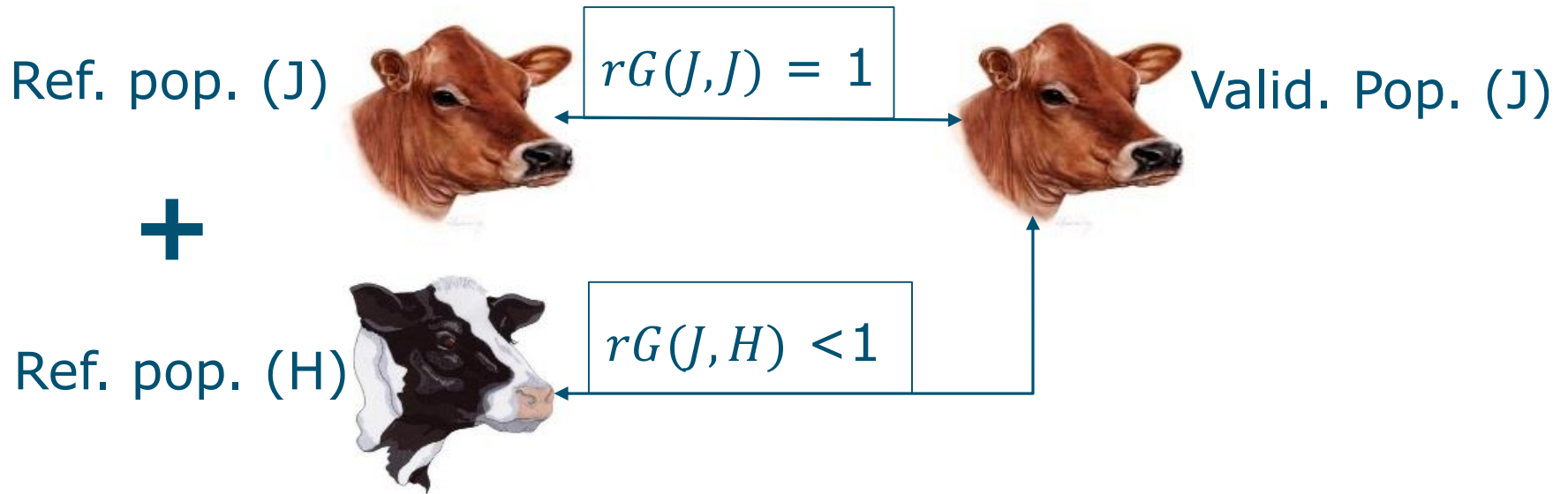
2 GRMs fitted simultaneously in the model:

- GRM1 (pre-selected markers from e.g. GWAS)

- GRM2 (remaining markers)

Multi-breed, multi-GRM (**MBMG**) model
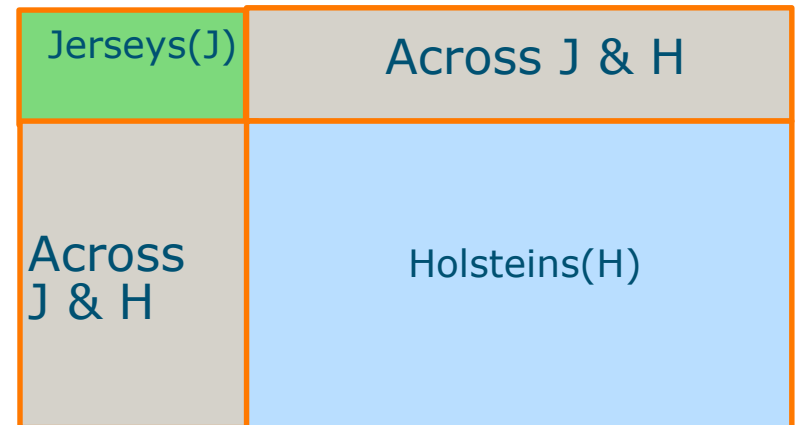
MULTI-BREED, MULTI-GRM MODEL

GRM1: Pre-selected markers

Jerseys (J)	Across J & H
Across J & H	Holsteins (H)

GRM2: Remaining markers

Jerseys(J)	Across J & H
Across J & H	Holsteins(H)

$$\begin{bmatrix} \mathbf{y}_J \\ \mathbf{y}_H \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_J \\ \mu_H \end{bmatrix} + \begin{bmatrix} \mathbf{W}_{1J} & 0 \\ 0 & \mathbf{W}_{1H} \end{bmatrix} \begin{bmatrix} \mathbf{g}_{1J} \\ \mathbf{g}_{1H} \end{bmatrix} + \begin{bmatrix} \mathbf{W}_{2J} & 0 \\ 0 & \mathbf{W}_{2H} \end{bmatrix} \begin{bmatrix} \mathbf{g}_{2J} \\ \mathbf{g}_{2H} \end{bmatrix} + \begin{bmatrix} \mathbf{e}_J \\ \mathbf{e}_H \end{bmatrix}$$
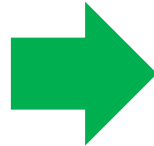
SELECTION INDEX THEORY

True breeding values

(Breed J)

MBMB model with:

2 Breeds (J & H)

2 GRMs (1 & 2)

4 Information sources:

$GEBV_{(J),1}$

$GEBV_{(H),1}$

$GEBV_{(J),2}$

$GEBV_{(H),2}$

$$I = b_{(J),1}\, GEBV_{(J),1} + b_{(H),1}\, GEBV_{(H),1} + b_{(J),2}\, GEBV_{(J),2} + b_{(H),2}\, GEBV_{(H),2}$$

ACCURACY OF SELECTION INDEX (1)

2 GRMs fitted simultaneously

- 0 cov between $\mathbf{g}$'s

$$\mathbf{P} = \begin{bmatrix} r_{J,1}^2 & r_{J,1}^2 r_{H,1}^2 & 0 & 0 \\ r_{J,1}^2 r_1^2 & r_{H,1}^2 & 0 & 0 \\ 0 & 0 & r_{J,2}^2 & r_{J,2}^2 r_{H,2}^2 \\ 0 & 0 & r_{J,2}^2 r_{H,2}^2 & r_{H,2}^2 \end{bmatrix}, \quad \mathbf{g} = \begin{bmatrix} r_{J,1}^2 \\ r_{H,1}^2 \\ r_{J,2}^2 \\ r_{H,2}^2 \end{bmatrix}$$

$$r_{I,H} = \sqrt{\frac{\mathbf{g'}\mathbf{P}^{-1}\mathbf{g}}{var(H)}} \implies \sqrt{\left(\frac{r_{J,1}^2 + r_{H,1}^2 - 2r_{J,1}^2 r_{H,1}^2}{\left(1 - r_{J,1}^2 r_{H,1}^2\right)}\right) + \left(\frac{r_{J,2}^2 + r_{H,2}^2 - 2r_{J,2}^2 r_{H,2}^2}{\left(1 - r_{J,2}^2 r_{H,2}^2\right)}\right)}$$
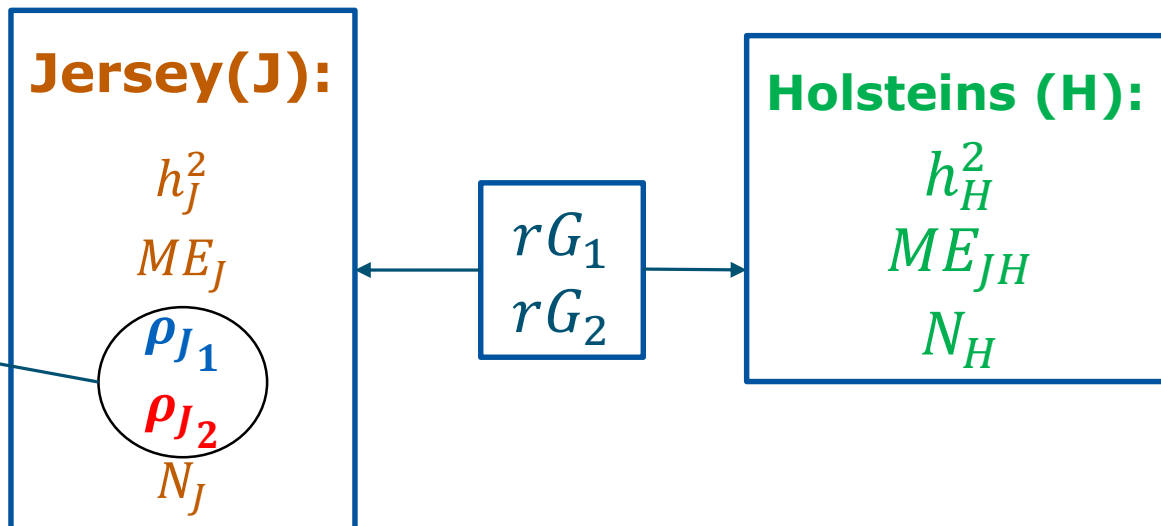
Deterministic $\implies$ Equation to predict accuracy of GEBVs

- Daetwyler et al. (2008) ⎤
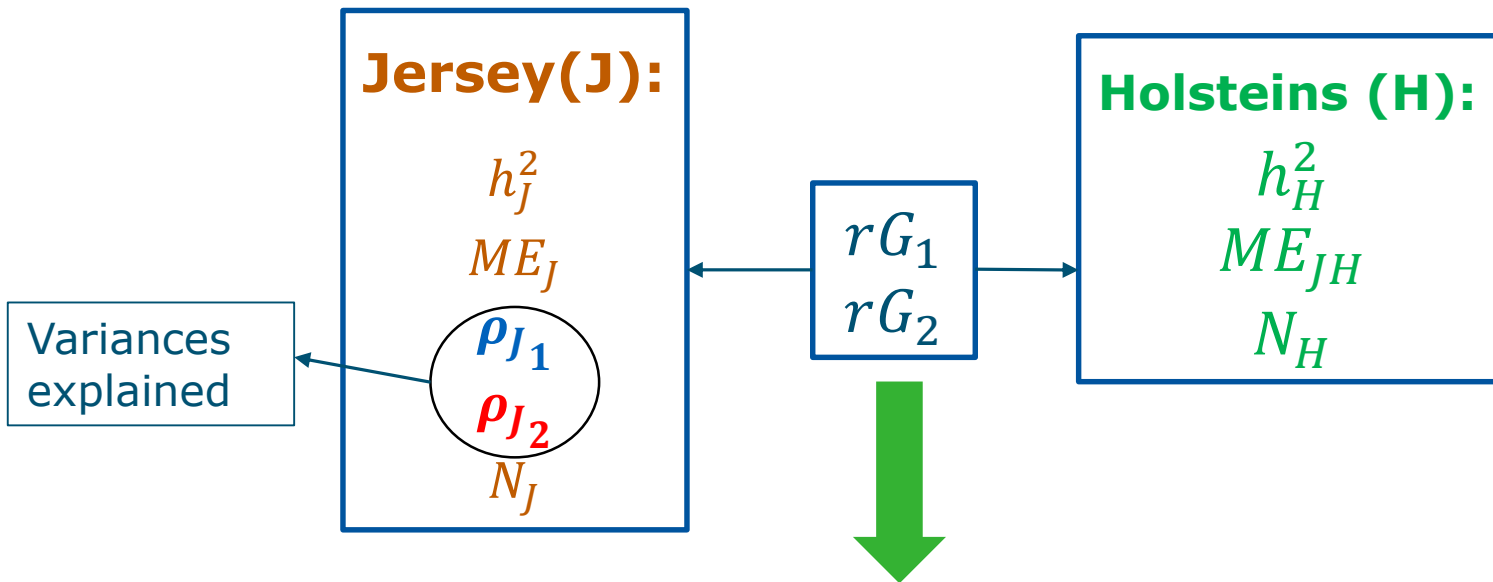- Wientjes et al. (2015) ⎦ Pop. parameters

ACCURACY OF SELECTION INDEX (2)

$$r_{I,H} = \sqrt{ \begin{bmatrix} \rho_{J_1}\sqrt{\dfrac{h_J^2}{ME_{J_1}}} \\ \rho_{J_1}rG_1\sqrt{\dfrac{h_H^2}{ME_{JH_1}}} \\ \rho_{J_2}\sqrt{\dfrac{h_J^2}{ME_{J_2}}} \\ \rho_{J_2}rG_2\sqrt{\dfrac{h_H^2}{ME_{JH_2}}} \end{bmatrix}^T \cdot \begin{bmatrix} \dfrac{h^2J}{ME_{J_1}}+\dfrac{1}{N_J} & \rho^2 J_1 rG_1\dfrac{\sqrt{h_J^2 h_B^2}}{\sqrt{ME_{J_1}.ME_{JH_1}}} & 0 & 0 \\ \rho^2 J_1 rG_1\dfrac{\sqrt{h_J^2 h_H^2}}{\sqrt{ME_{J_1}.ME_{JH_1}}} & \dfrac{h^2 H}{ME_{JH_1}}+\dfrac{1}{N_H} & 0 & 0 \\ 0 & 0 & \dfrac{h^2 J}{ME_{J_2}}+\dfrac{1}{N_J} & \rho^2 J_2 rG_2\dfrac{\sqrt{h_A^2 h_H^2}}{\sqrt{ME_{J_2}.ME_{JH_2}}} \\ 0 & 0 & \rho^2 J_2 rG_2\dfrac{\sqrt{h_J^2 h_H^2}}{\sqrt{ME_{J_2}.ME_{JH_2}}} & \dfrac{h^2 H}{ME_{JH_2}}+\dfrac{1}{N_J} \end{bmatrix}^{-1} \cdot \begin{bmatrix} \rho_{J_1}\sqrt{\dfrac{h_J^2}{ME_{J_1}}} \\ \rho_{J_1}rG_1\sqrt{\dfrac{h_J^2}{ME_{JH_1}}} \\ \rho_{J_2}\sqrt{\dfrac{h_J^2}{ME_{J_2}}} \\ \rho_{J_2}rG_2\sqrt{\dfrac{h_J^2}{ME_{JH_2}}} \end{bmatrix} }$$
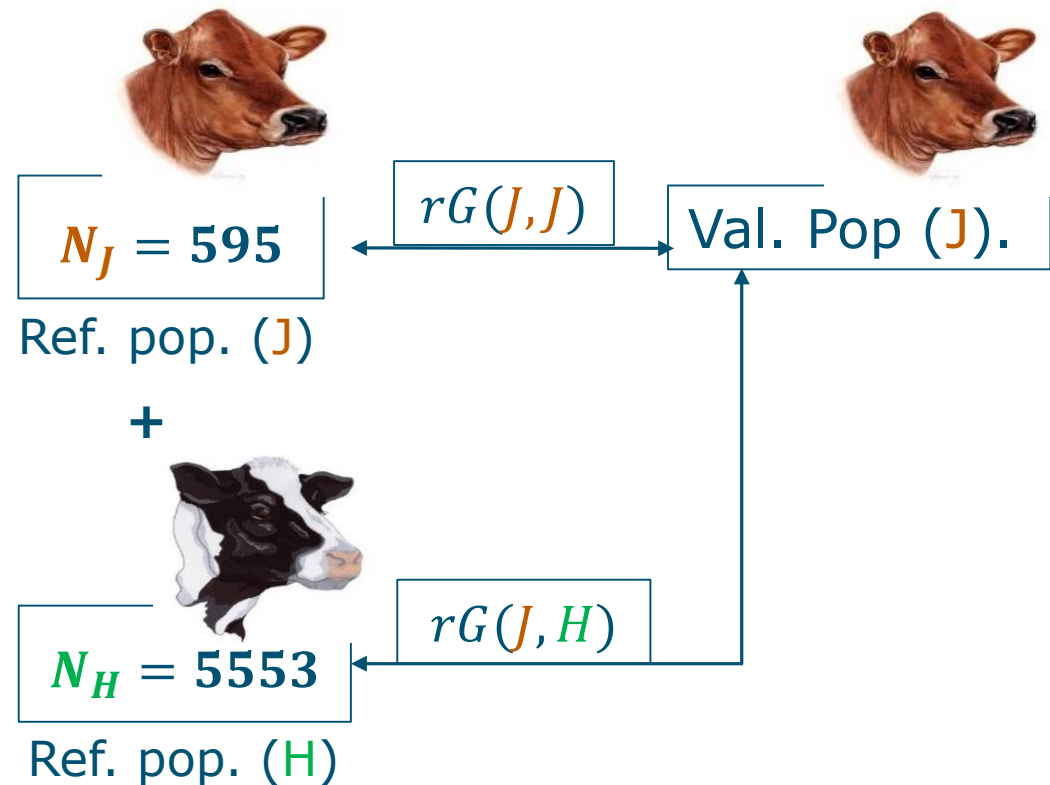
Input/population parameters

Jersey(J):

h_J^2

ME_J

ρ_{J_1}

ρ_{J_2}

N_J

Variances explained

rG_1

rG_2

Holsteins (H):

h_H^2

ME_{JH}

N_H

INPUT/POPULATION PARAMETERS

Jersey(J):

h_J^2

ME_J

$\boldsymbol{\rho_{J_1}}$

$\boldsymbol{\rho_{J_2}}$

N_J

Variances explained

rG_1

rG_2

Holsteins (H):

h_H^2

ME_{JH}

N_H

Marker set	rG (Hol. vs Jer.)
Pre-selected (n=357)	**0.44**
Remaining (~48k)	0.22
All	0.29
Pre-selected & Remaining	**0.88** & 0.21

DATA ANALYSED

- Real phenotype
  - Stature DRPs

- Simulated phenotype
  - 500 SNP QTLs
  - $h^2{}_J = 0.4$
  - $h^2{}_H = 0.6$
  - $rG(J,H) = 0.8$
  - $\mathbf{y} = \mathbf{TBV} + \mathbf{e}$
  - 100 replicates

$N_J = 595$

Ref. pop. (J)

$rG(J,J)$

Val. Pop (J).

+

$N_H = 5553$

Ref. pop. (H)

$rG(J,H)$

MBMG MODELS EVALUATED

Real phenotypes: stature

Model	GRM1	GRM2
1	Pre-selected marker (n=357)	Remaining markers (n~48k)

LETTERS
https://doi.org/10.1038/s41588-018-0056-5

nature genetics

Meta-analysis of genome-wide association studies for cattle stature identifies common genes that regulate body size in mammals
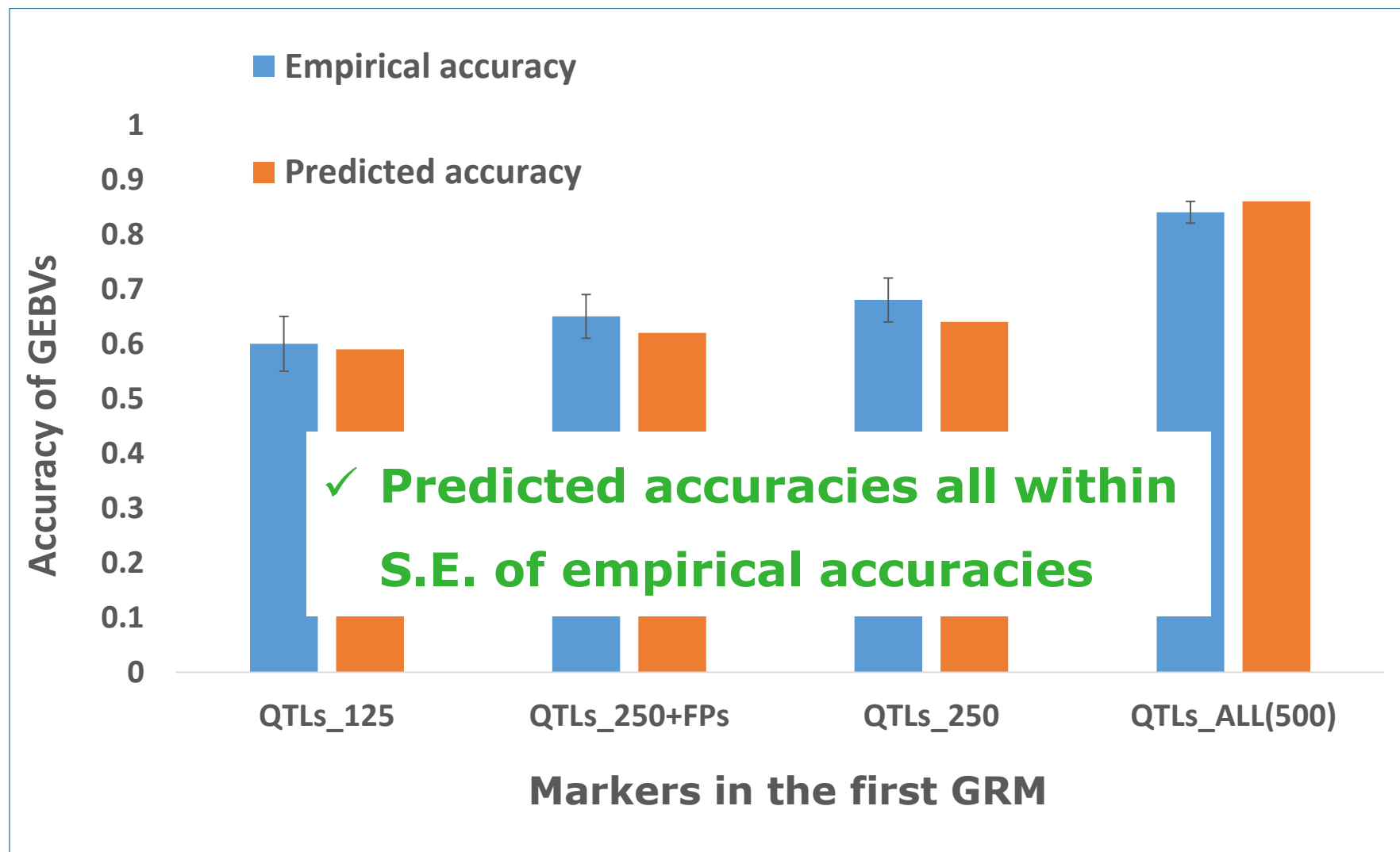
Simulated phenotypes

Model	GRM1	GRM2
1	QTLs_125	NON_QTLs
2	QTLS_250 + 250 NON_QTL (FPs)	NON_QTLs
3	QTLs_250	NON_QTLs
4	QTLs_500	NON_QTLs

RESULTS (REAL PHENOTYPES)

595 Jerseys & 5503 Holsteins in ref. pop.
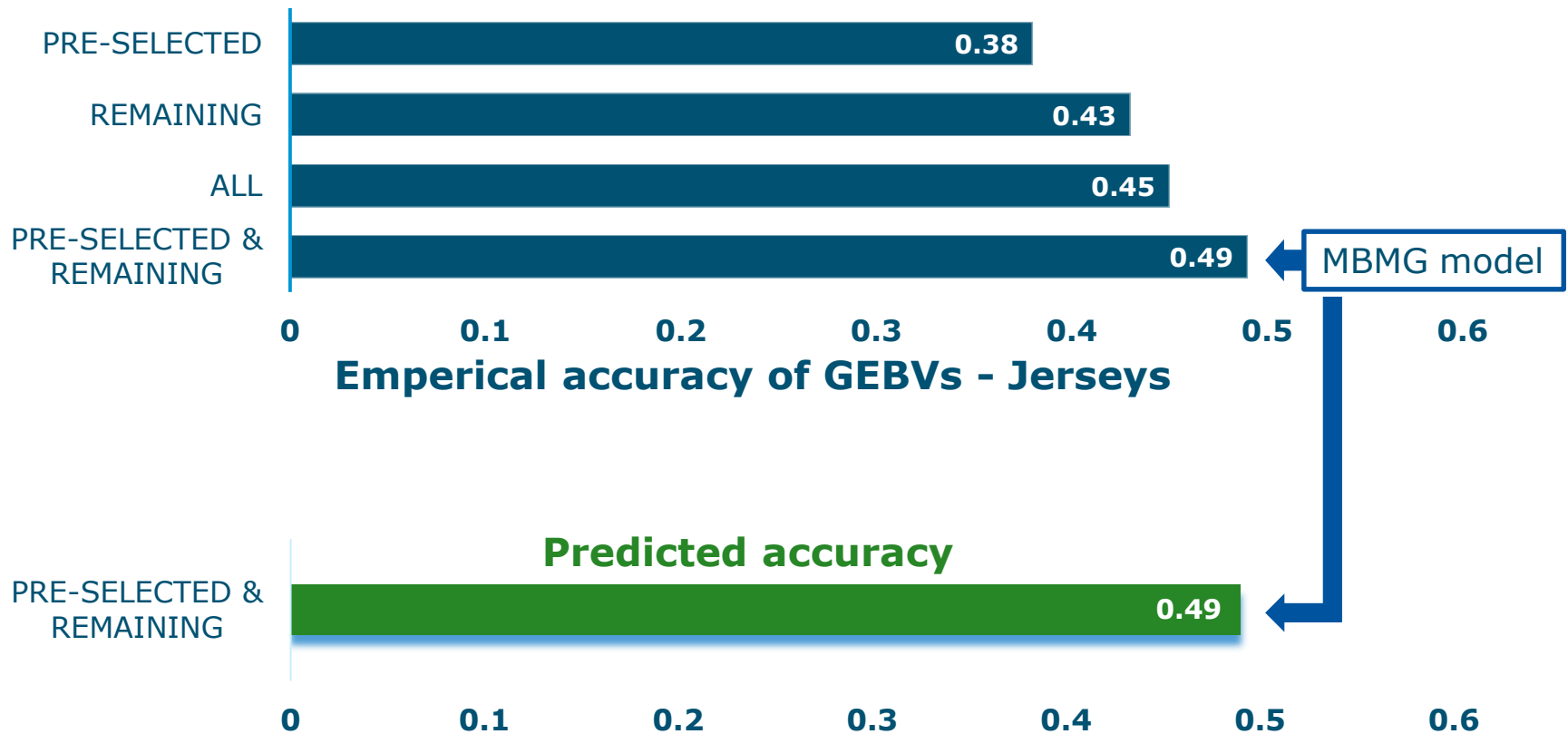
RESULTS (SIMULATION)

CONCLUSION

GENOMIC PREDICTION IN NUMERICALLY SMALL BREEDS

- Higher accuracy with MBMG model, when QTL are known

- Selection index theory:

  - Confirm the higher accuracy of MBMG

  - Developed equation to predict accuracy using population parameters

    → Tool for optimising genomic breeding programs