Genomic prediction using information from multiple populations

Yvonne Wientjes

INRA

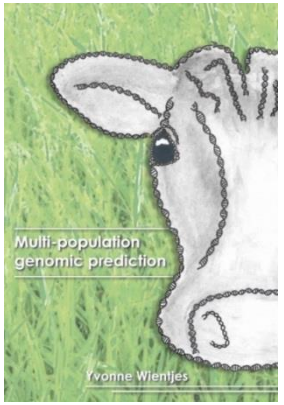
28 October 2019

WAGENINGEN
UNIVERSITY & RESEARCH

Who am I?

PhD - Wageningen University & Research

- Animal Breeding and Genomics
- 'Multi-population genomic prediction'



Postdoc - Wageningen University & Research

- Genomic prediction for crossbred performance
- Long-term effects of Genomic Selection

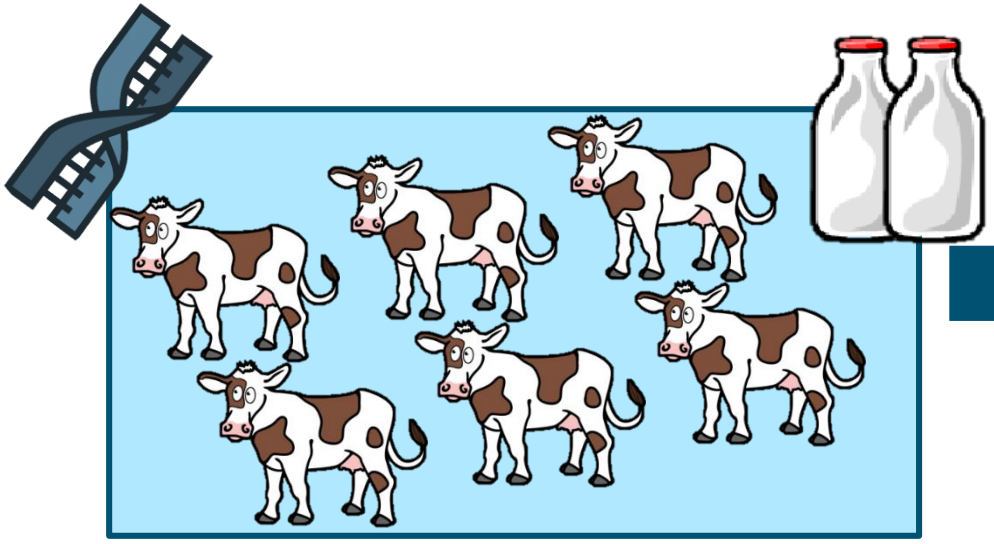
WAGENINGEN
UNIVERSITY & RESEARCH

NWO

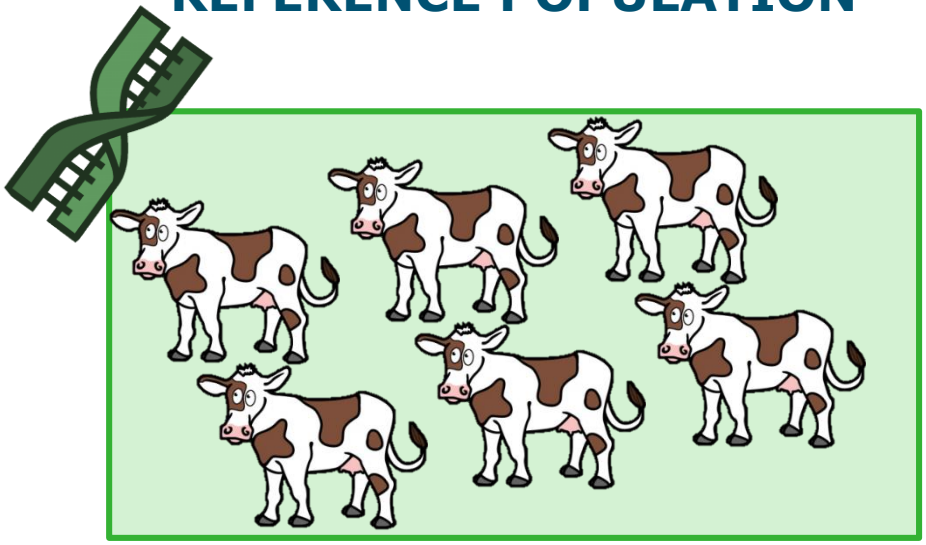
BREED4FOOD

CRV

Genomic selection



REFERENCE POPULATION

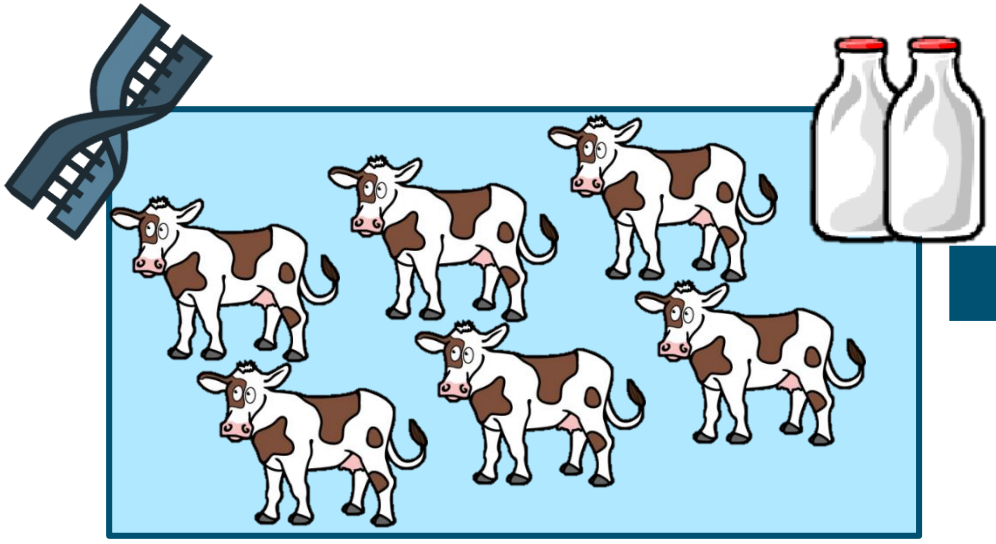


SELECTION CANDIDATES

Prediction equation with marker effects

Genomic breeding values

Reference population

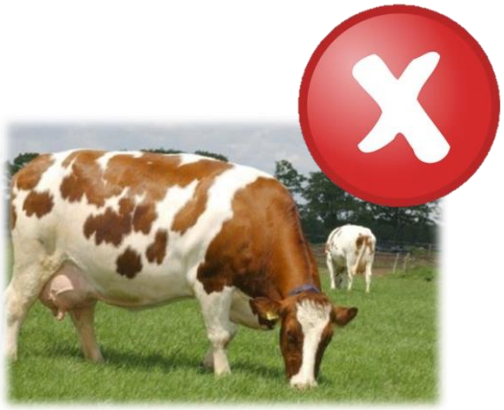


REFERENCE POPULATION

Prediction equation with marker effects

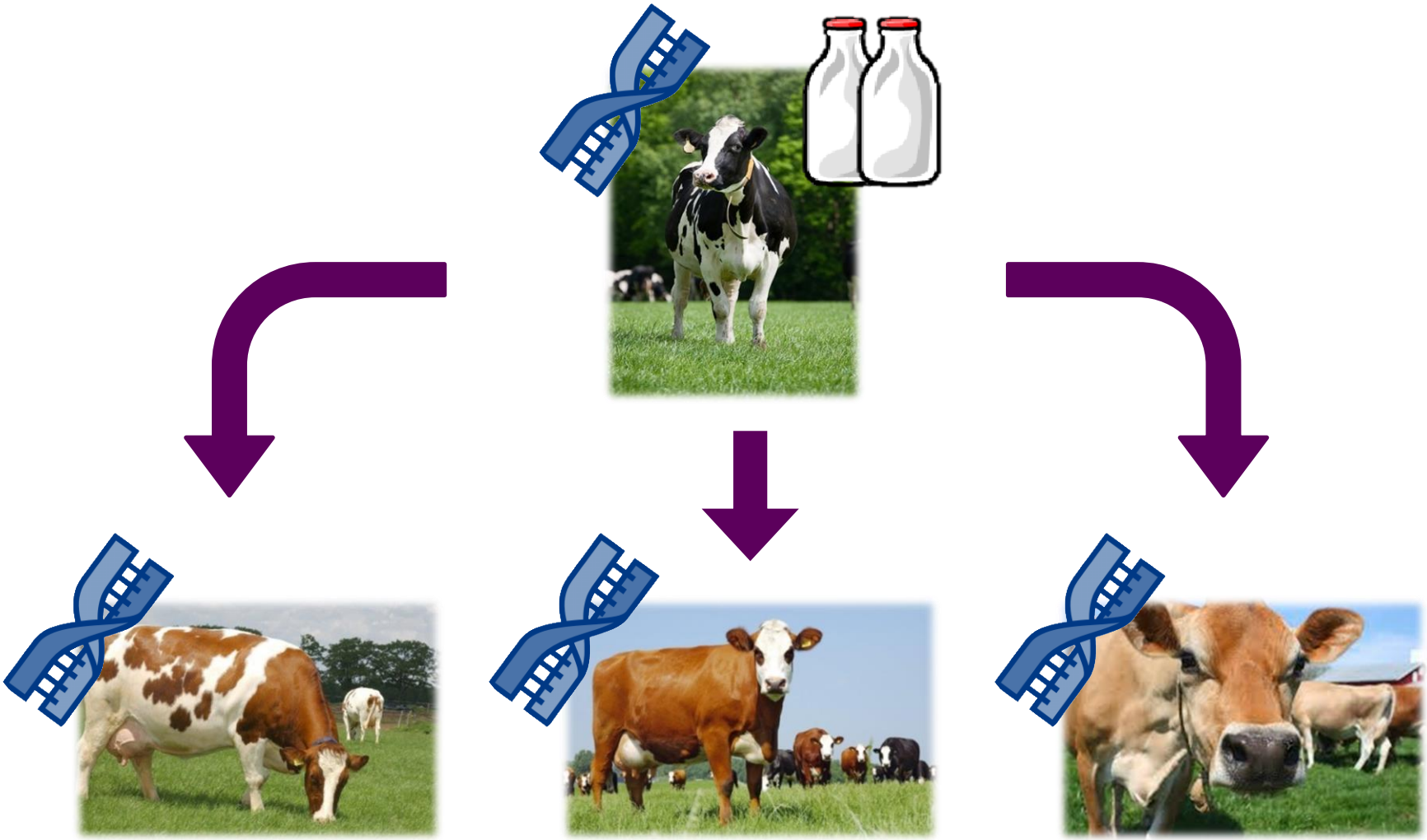
Reference population ↑ → ↑ **Accuracy of GEBVs**

Size of reference population



WAGENINGEN
UNIVERSITY & RESEARCH

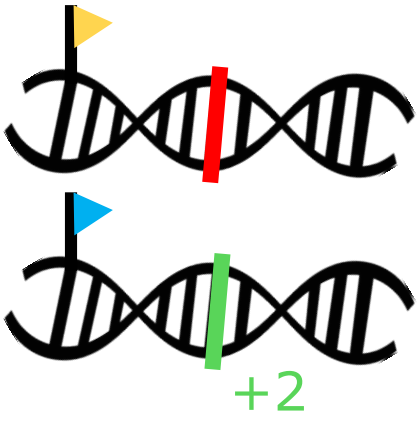
Use information across populations?

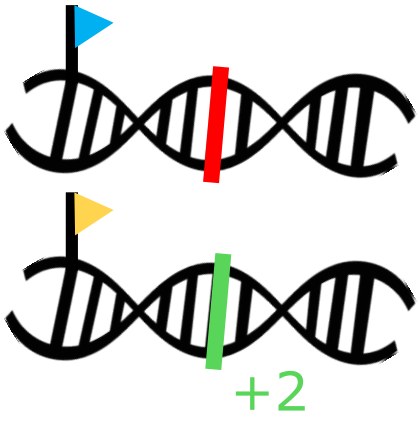


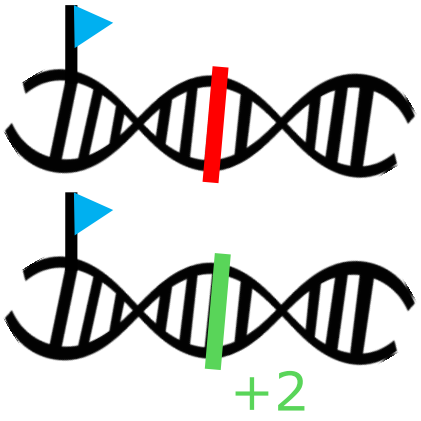
Differences between populations

- Linkage disequilibrium between markers and causal variants

Differences between populations in LD







WAGENINGEN
UNIVERSITY & RESEARCH

Differences between populations

- Linkage disequilibrium between markers and causal variants
- Allele frequencies of causal variants
- Effects of causal variants
  - Environment different
  - Non-additive effects

  → *Genetic correlation between populations*

Differences between populations

- Linkage disequilibrium between markers and causal variants
- Allele frequencies of causal variants
- Effects of causal variants
  - Environment different
  - Non-additive effects

  → *Genetic correlation between populations*
- Close family relationships are absent

Across-population genomic prediction

1033 Holstein (HF)

105 Groningen White Headed (GWH)

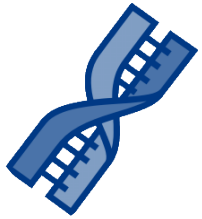
147 Meuse-Rhine-Yssel (MRY)

Across-population genomic prediction

Holstein (HF)

Groningen White Headed (GWH)

Meuse-Rhine-Yssel (MRY)



800k chip

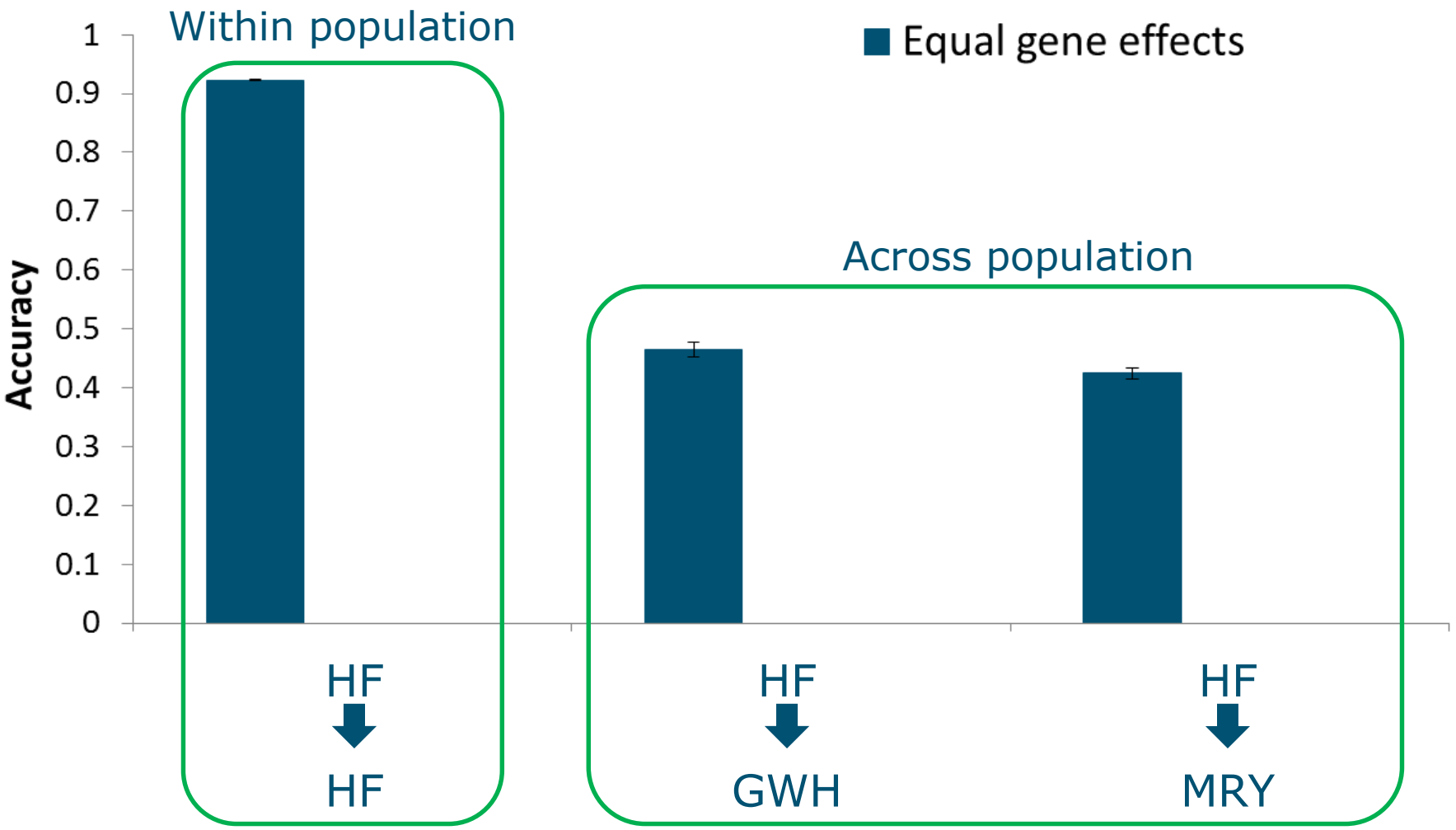
300 causal variants

~ 25 000 markers

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

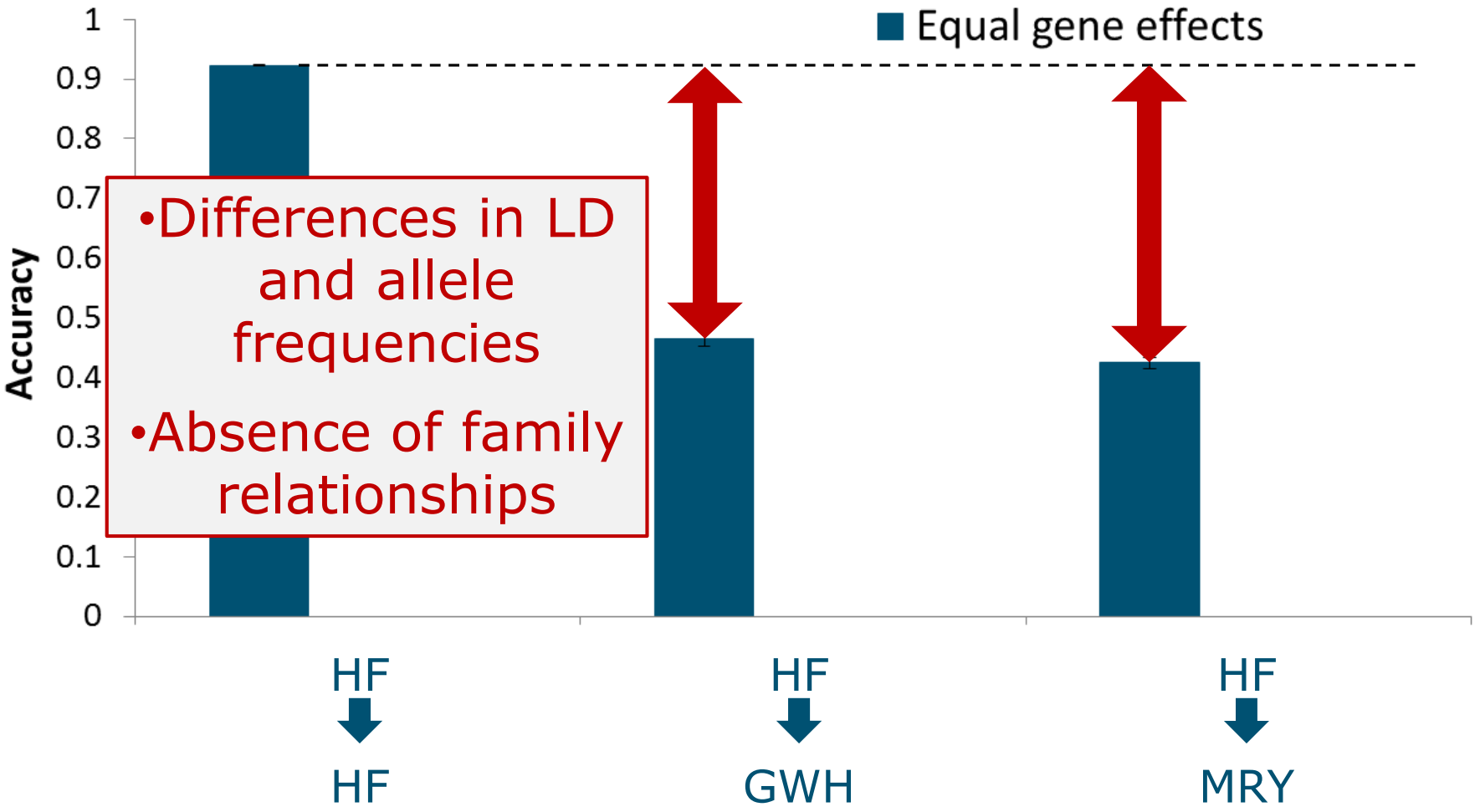
Across-population genomic prediction



WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

Across-population genomic prediction



WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

Across-population genomic prediction

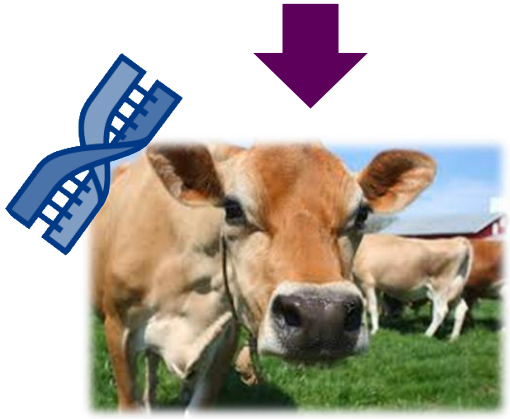
- Accuracy of across-population genomic prediction is much lower than within-population genomic prediction
- Differences in effects of causal variants reduce accuracy

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

Effect of properties causal variants

Holstein



Jersey

Causal variants

Moderately low allele frequency (~1 in 8)

Very low allele frequency (~1 in 13)

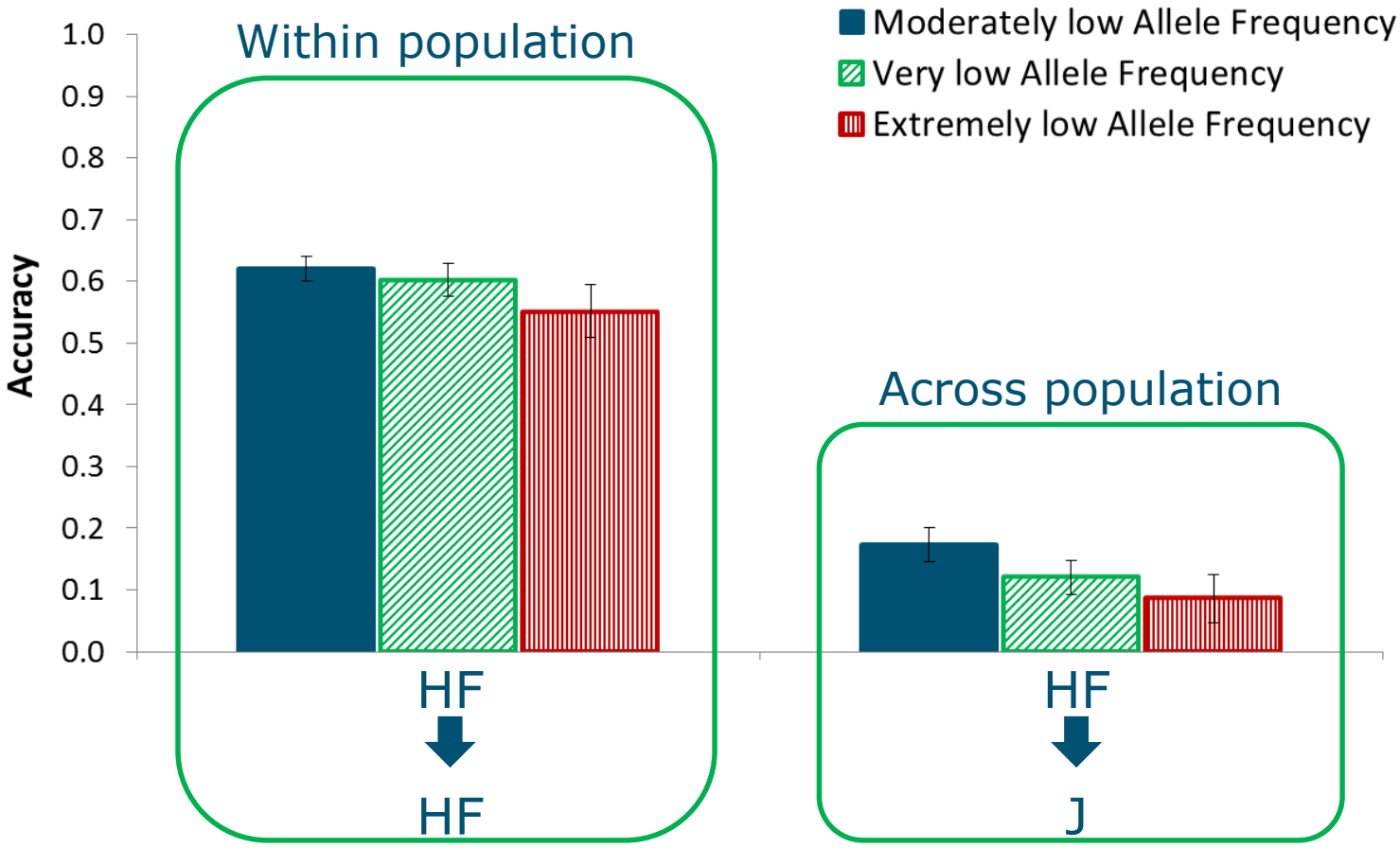
Extremely low allele frequency (~1 in 63)

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

Accuracies of predicting breeding values

Gene effects randomly sampled

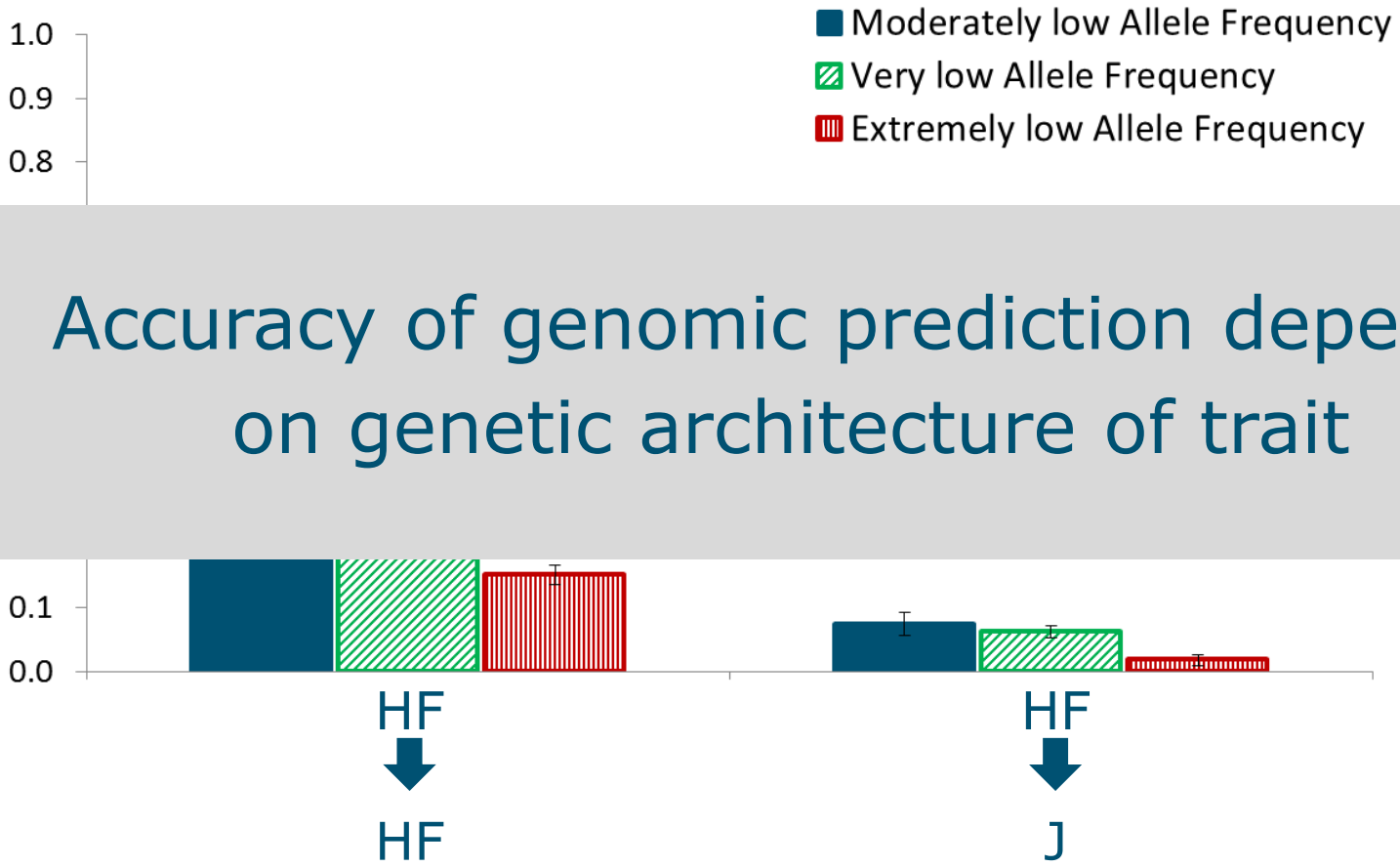


WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

Accuracies of predicting breeding values

Larger effects for genes with lower frequency



Accuracy of genomic prediction depends on genetic architecture of trait

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

Multi-population genomic prediction

Accuracy of across-population genomic prediction is low…

, but what if we combine populations in one reference population?

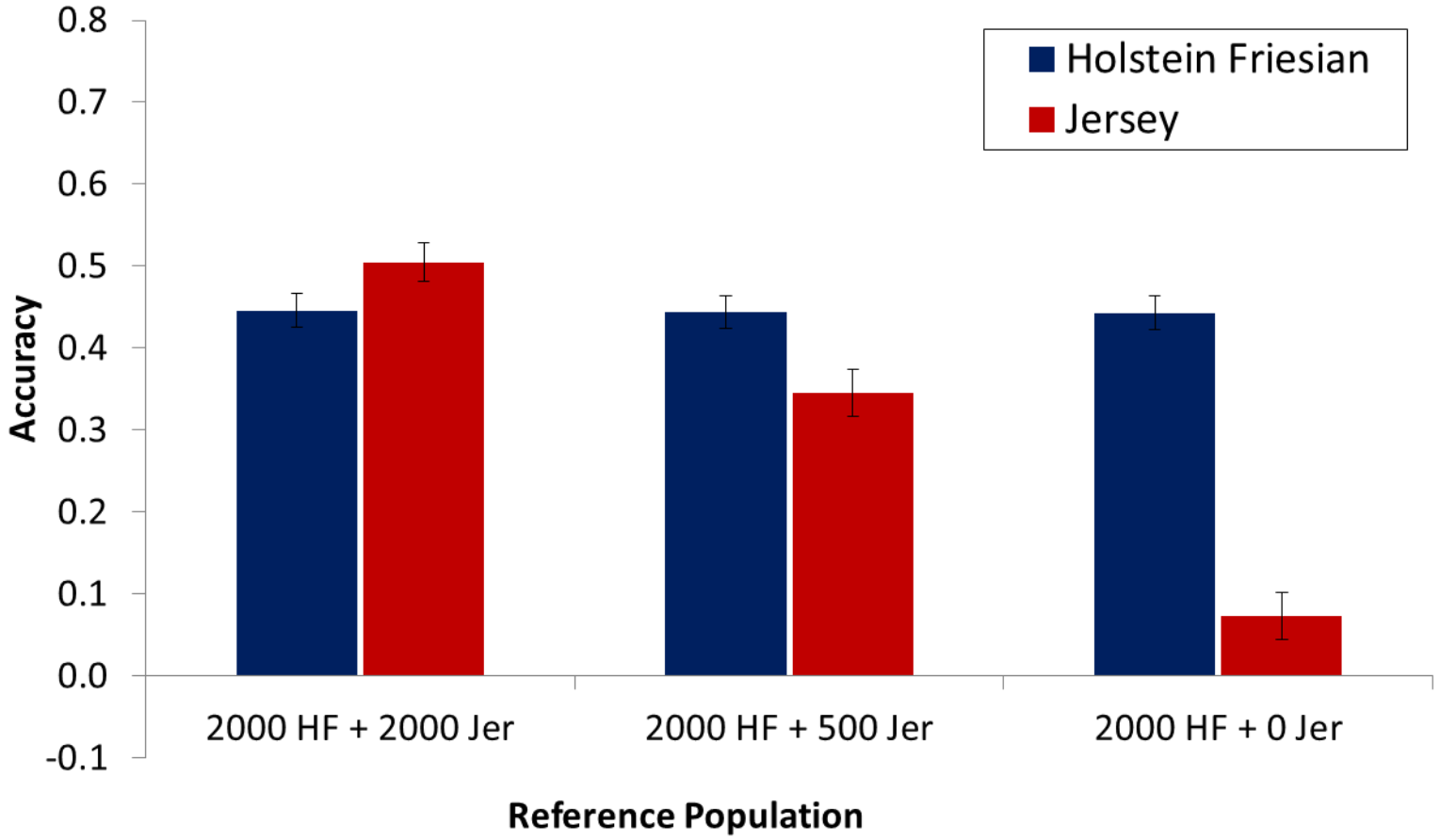
Multi-population genomic prediction

Prediction equation with marker effects

Genomic breeding values

Wientjes *et al.*, GSE, 2015

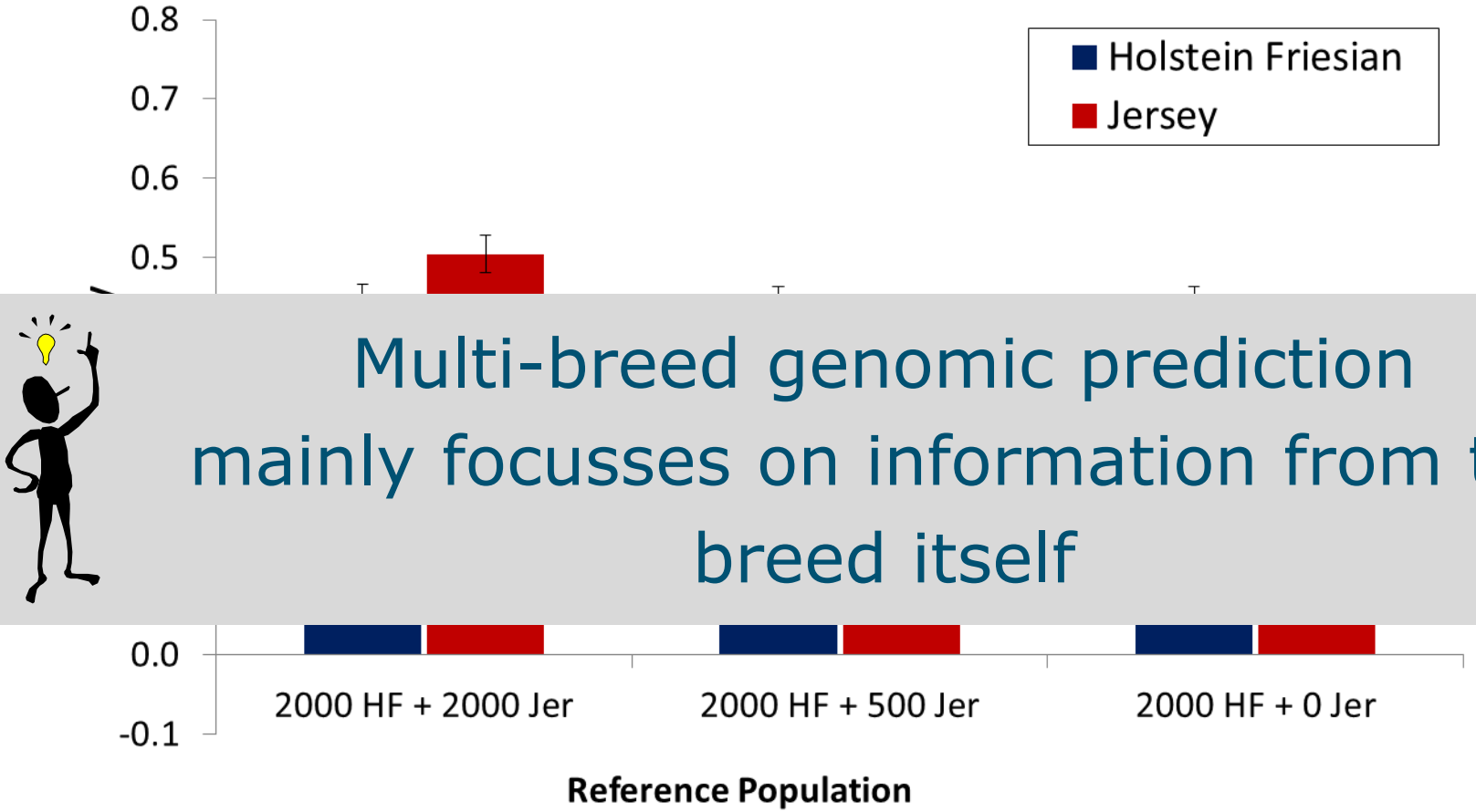
Multi-population genomic prediction



WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

Multi-population genomic prediction



Multi-breed genomic prediction mainly focusses on information from the breed itself

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2015

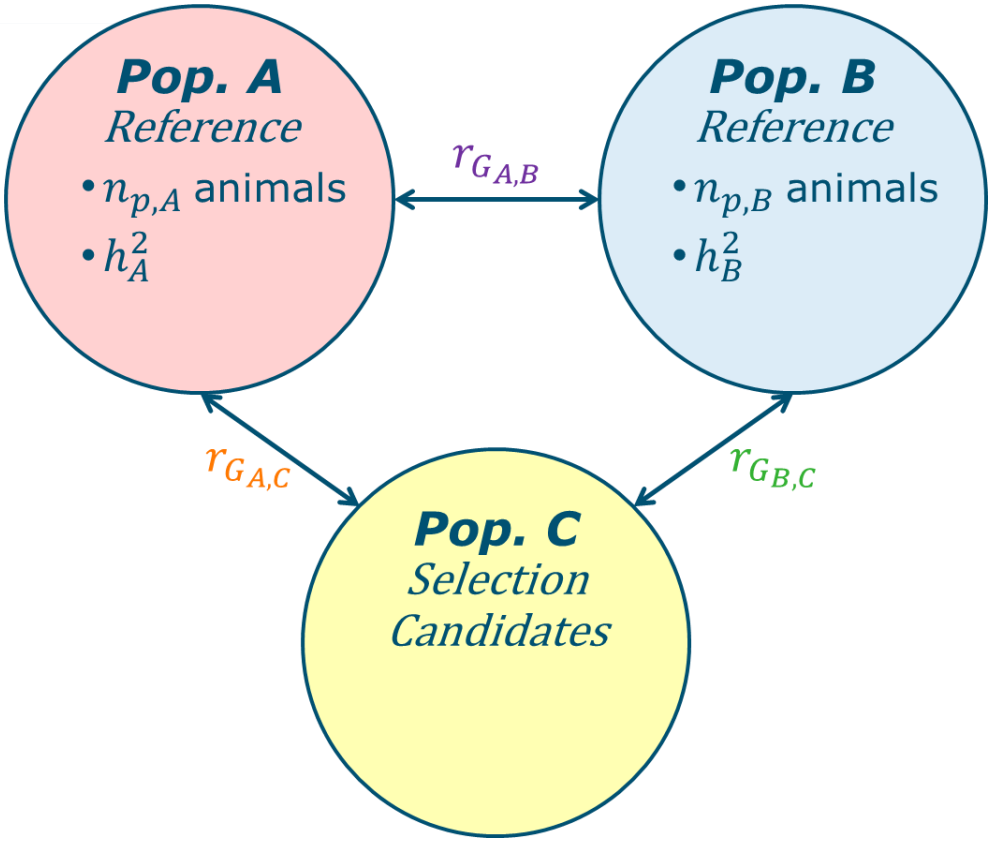
Can we predict accuracy?

Important when designing breeding programs

- Who to genotype?

Wientjes *et al.*, Genetics, 2016

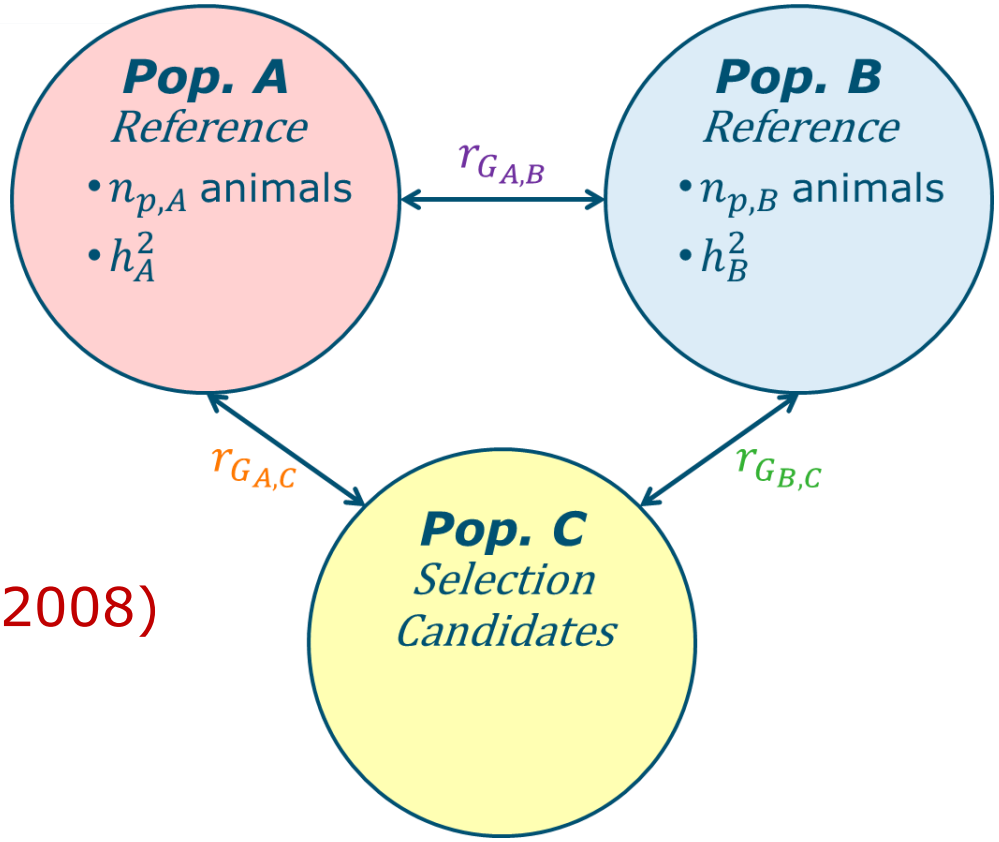
Derived equation



$$r_{GEBV_{A+B,C}} = \sqrt{\begin{bmatrix} r_{G_{A,C}}\sqrt{\frac{h_A^2}{M_{e_{A,C}}}} & r_{G_{B,C}}\sqrt{\frac{h_B^2}{M_{e_{B,C}}}} \end{bmatrix} \begin{bmatrix} \frac{h_A^2}{M_{e_{A,C}}} + \frac{1}{n_{p,A}} & r_{G_{A,B}}\frac{\sqrt{h_A^2 h_B^2}}{\sqrt{M_{e_{A,C}} M_{e_{B,C}}}} \\ r_{G_{A,B}}\frac{\sqrt{h_A^2 h_B^2}}{\sqrt{M_{e_{A,C}} M_{e_{B,C}}}} & \frac{h_B^2}{M_{e_{B,C}}} + \frac{1}{n_{p,B}} \end{bmatrix}^{-1} \begin{bmatrix} r_{G_{A,C}}\sqrt{\frac{h_A^2}{M_{e_{A,C}}}} \\ r_{G_{B,C}}\sqrt{\frac{h_B^2}{M_{e_{B,C}}}} \end{bmatrix}}$$

Wientjes *et al.*, Genetics, 2016

Derived equation



1 population in reference

→ Equal to Daetwyler *et al.* (2008)

$$r_{GEBV_{A+B,C}} = \sqrt{\begin{bmatrix} r_{G_{A,C}}\sqrt{\frac{h_A^2}{M_{e_{A,C}}}} & r_{G_{B,C}}\sqrt{\frac{h_B^2}{M_{e_{B,C}}}} \end{bmatrix} \begin{bmatrix} \frac{h_A^2}{M_{e_{A,C}}} + \frac{1}{n_{p,A}} & r_{G_{A,B}}\frac{\sqrt{h_A^2 h_B^2}}{\sqrt{M_{e_{A,C}} M_{e_{B,C}}}} \\ r_{G_{A,B}}\frac{\sqrt{h_A^2 h_B^2}}{\sqrt{M_{e_{A,C}} M_{e_{B,C}}}} & \frac{h_B^2}{M_{e_{B,C}}} + \frac{1}{n_{p,B}} \end{bmatrix}^{-1} \begin{bmatrix} r_{G_{A,C}}\sqrt{\frac{h_A^2}{M_{e_{A,C}}}} \\ r_{G_{B,C}}\sqrt{\frac{h_B^2}{M_{e_{B,C}}}} \end{bmatrix}}$$

Wientjes *et al.*, Genetics, 2016

M_e across populations

'Effective number of estimated effects'

Pop. B  $M_e = 4$

Pop. C  $M_e = 5$

Pop. B - C  $M_e = 18$

M_e across populations

'Effective number of estimated effects'

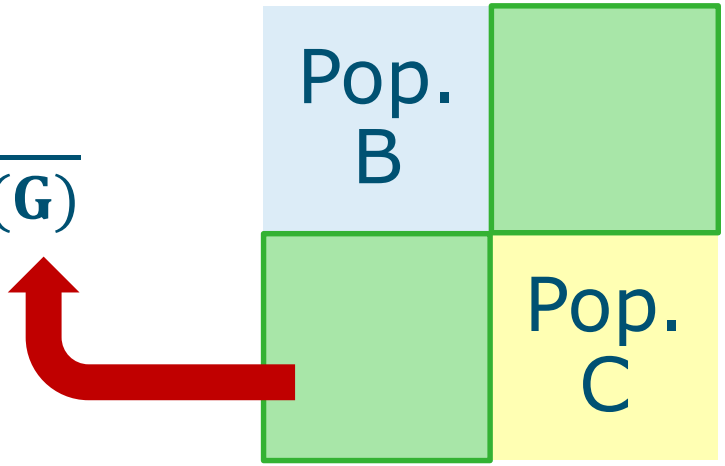
Pop. B — $M_e = 4$

Pop. C — $M_e = 5$

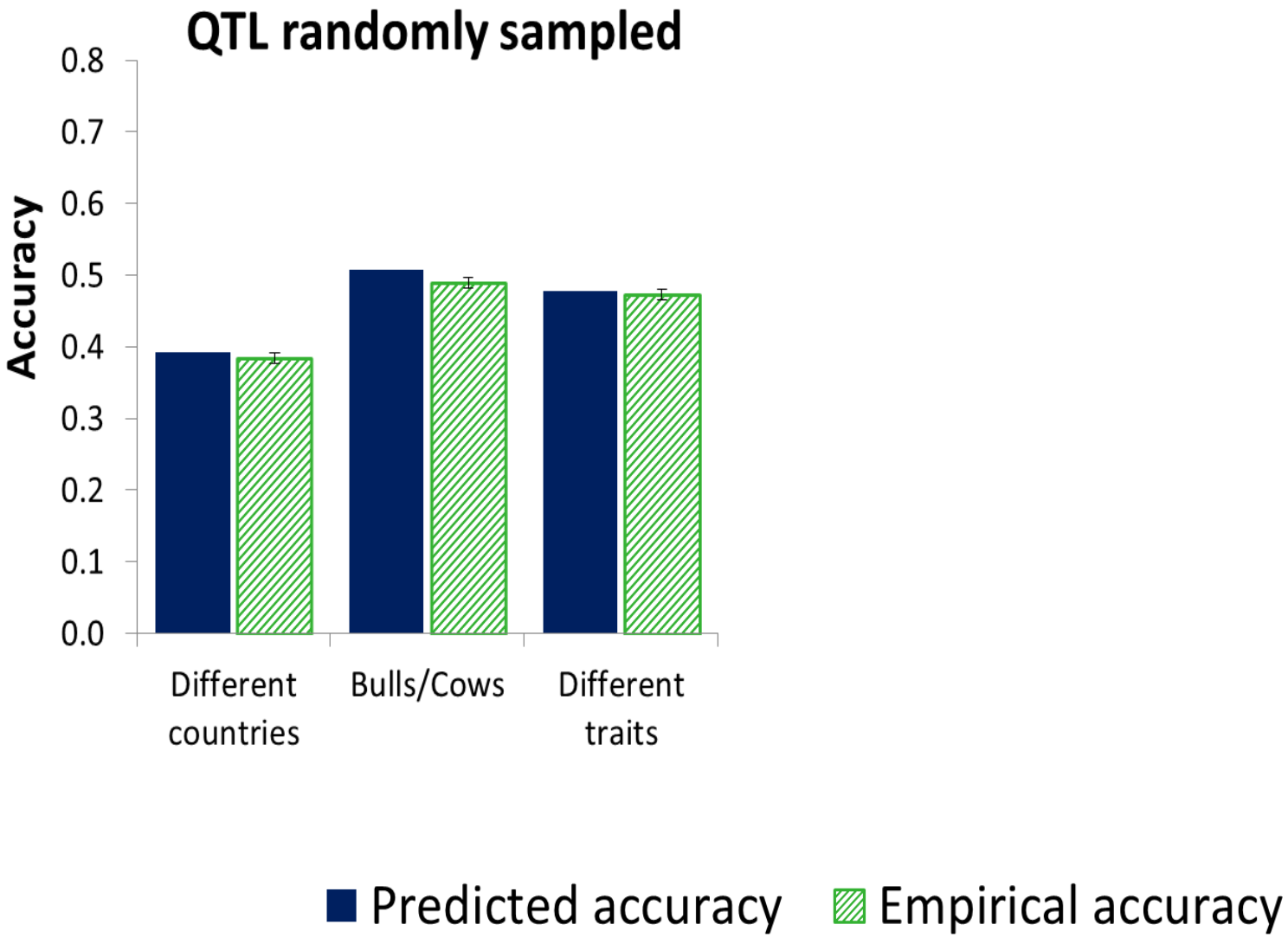
Pop. B - C — $M_e = 18$

Relationship matrix

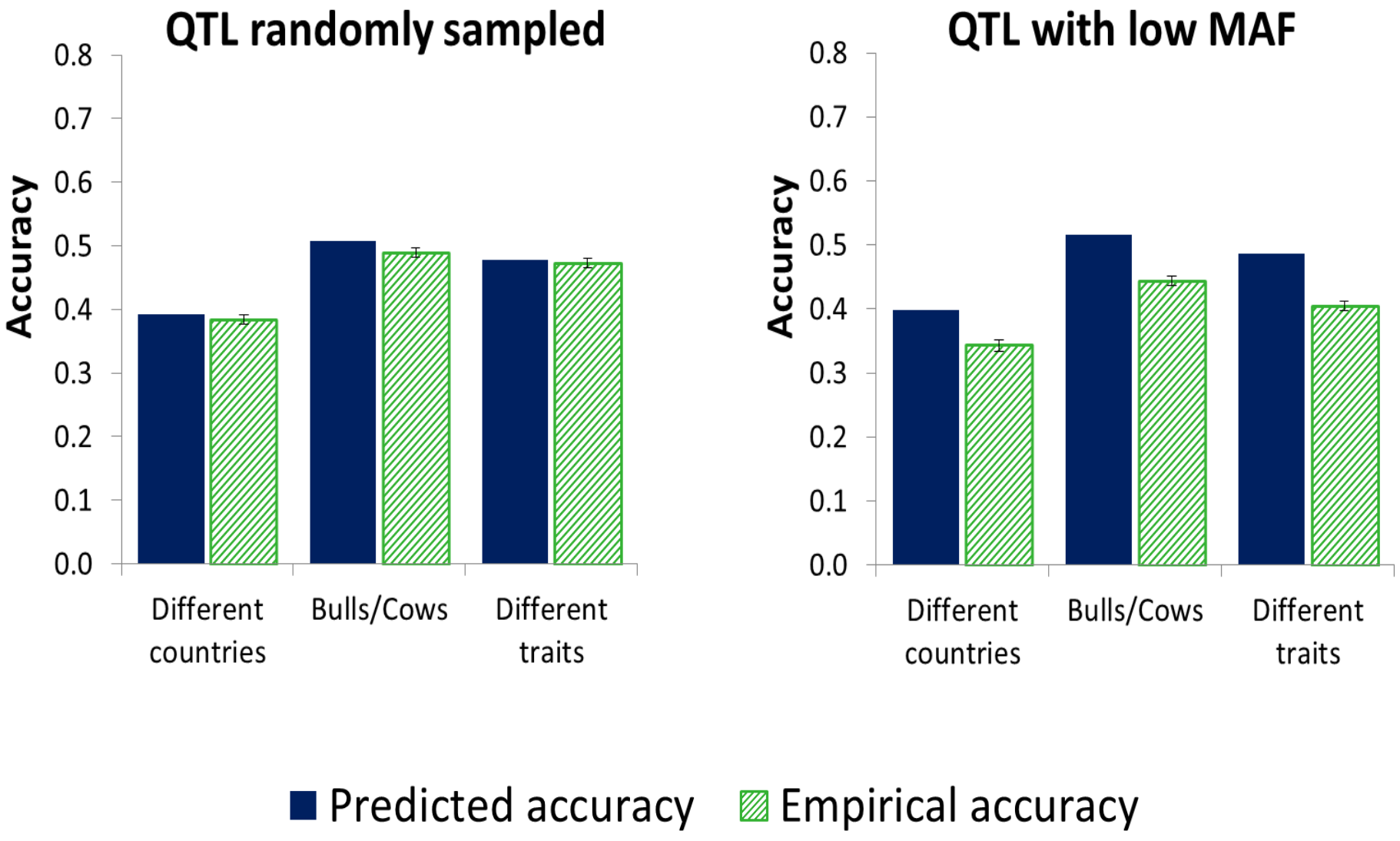
$$M_{e_{B,C}} = \frac{1}{Var(\mathbf{G})}$$



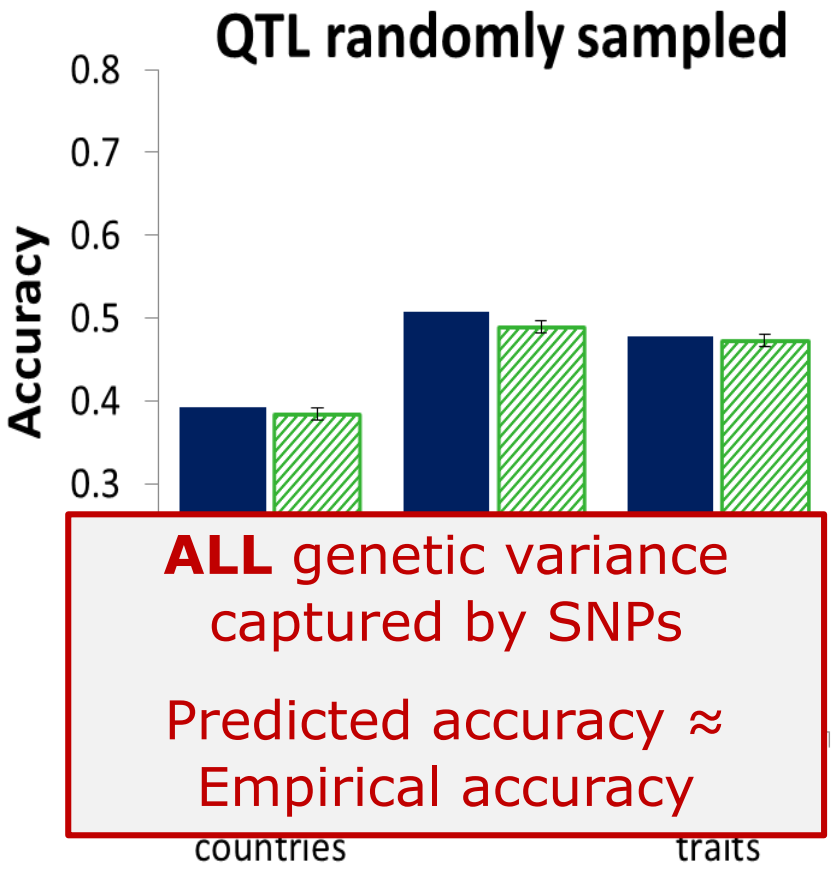
Results of validation

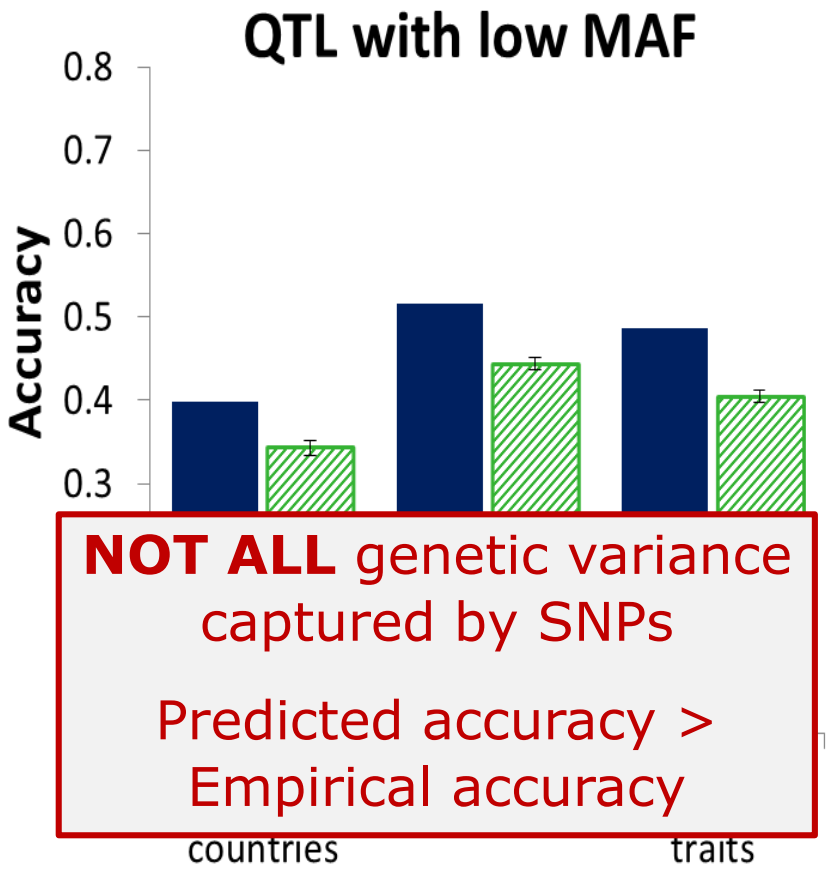


Results of validation



Results of validation





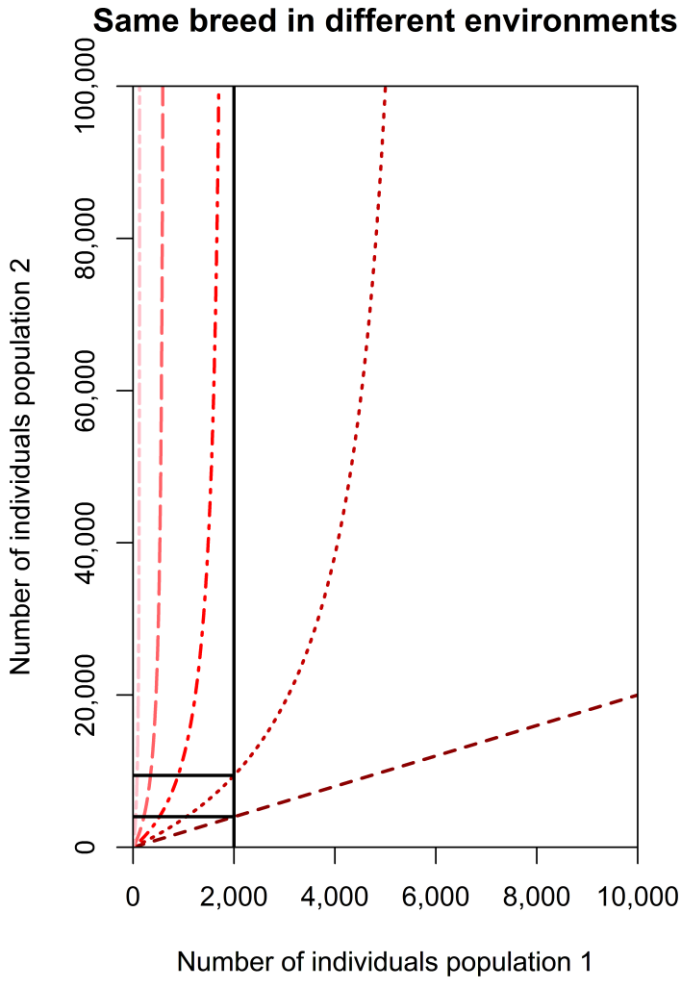
Predicted accuracy Empirical accuracy

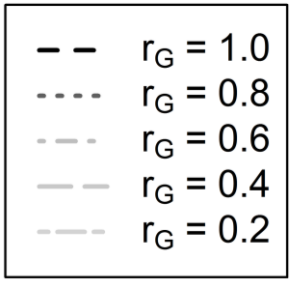
Combining populations

Populations of:

- Different countries
- Different breeds
- Bulls/Cows
- Measured for different traits
- Different generations
- …

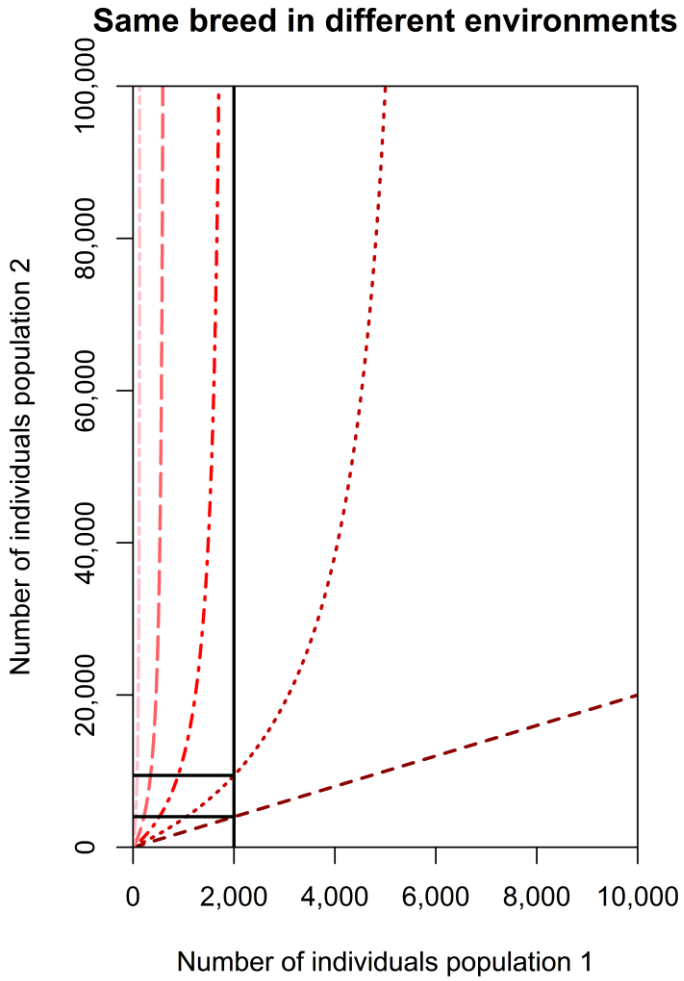
Combining populations?

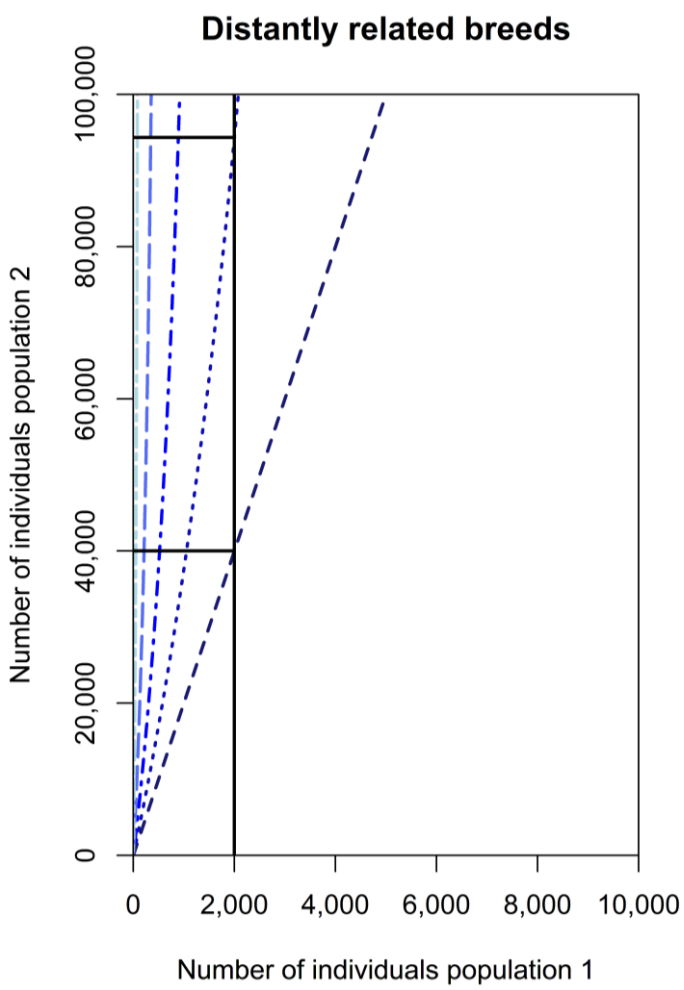


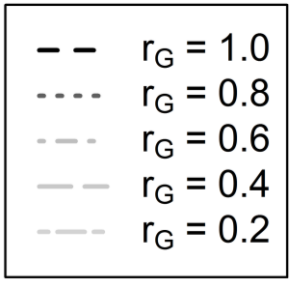


WAGENINGEN
UNIVERSITY & RESEARCH

Combining populations?







WAGENINGEN
UNIVERSITY & RESEARCH

Combining populations?

Beneficial, when:

- Closely related populations
- Population itself is small
- A large number of individuals is added

→ **Depends on genetic correlation between populations!!**

Genetic correlation between populations (r_g)

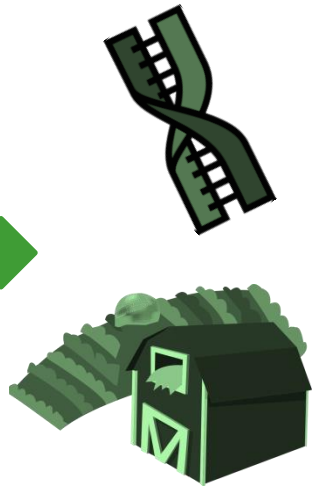
'Correlation between breeding values of two individuals with the same genotype in the two populations'

Pop. 1

+2
-1
-2
+1
-2

Pop. 2

+1
+2
-1
+1
-1



WAGENINGEN
UNIVERSITY & RESEARCH

Estimation of rg

- Multi-trait model
  - Each population different trait
- REML

Requires relationships between populations

Relationships between populations

Theoretically defined relationships:

$$\mathbf{G}\_\text{New} = \begin{bmatrix} \dfrac{\mathbf{Z}_1\mathbf{Z}_1'}{\sum 2p_{1i}(1-p_{1i})} & \dfrac{\mathbf{Z}_1\mathbf{Z}_2'}{\sqrt{\sum 2p_{1i}(1-p_{1i})}\sqrt{\sum 2p_{2i}(1-p_{2i})}} \\ \dfrac{\mathbf{Z}_2\mathbf{Z}_1'}{\sqrt{\sum 2p_{1i}(1-p_{1i})}\sqrt{\sum 2p_{2i}(1-p_{2i})}} & \dfrac{\mathbf{Z}_2\mathbf{Z}_2'}{\sum 2p_{2i}(1-p_{2i})} \end{bmatrix}$$

Wientjes *et al.*, Genetics, 2017

Relationships between populations

Theoretically defined relationships:

$$\mathbf{G}\_\text{New} = \begin{bmatrix} \dfrac{\mathbf{Z}_1\mathbf{Z}_1'}{\sum 2p_{1i}(1-p_{1i})} & \dfrac{\mathbf{Z}_1\mathbf{Z}_2'}{\sqrt{\sum 2p_{1i}(1-p_{1i})}\sqrt{\sum 2p_{2i}(1-p_{2i})}} \\ \dfrac{\mathbf{Z}_2\mathbf{Z}_1'}{\sqrt{\sum 2p_{1i}(1-p_{1i})}\sqrt{\sum 2p_{2i}(1-p_{2i})}} & \dfrac{\mathbf{Z}_2\mathbf{Z}_2'}{\sum 2p_{2i}(1-p_{2i})} \end{bmatrix}$$

VanRaden – method 1

Wientjes *et al.*, Genetics, 2017

Relationships between populations

Theoretically defined relationships:

$$\mathbf{G}\_\text{New} = \begin{bmatrix} \dfrac{\mathbf{Z}_1\mathbf{Z}_1'}{\sum 2p_{1i}(1-p_{1i})} & \dfrac{\mathbf{Z}_1\mathbf{Z}_2'}{\sqrt{\sum 2p_{1i}(1-p_{1i})}\sqrt{\sum 2p_{2i}(1-p_{2i})}} \\ \dfrac{\mathbf{Z}_2\mathbf{Z}_1'}{\sqrt{\sum 2p_{1i}(1-p_{1i})}\sqrt{\sum 2p_{2i}(1-p_{2i})}} & \dfrac{\mathbf{Z}_2\mathbf{Z}_2'}{\sum 2p_{2i}(1-p_{2i})} \end{bmatrix}$$

G_Chen: Scaling factor across populations:

$$\sum 2\sqrt{p_{1i}(1-p_{1i})p_{2i}(1-p_{2i})}$$

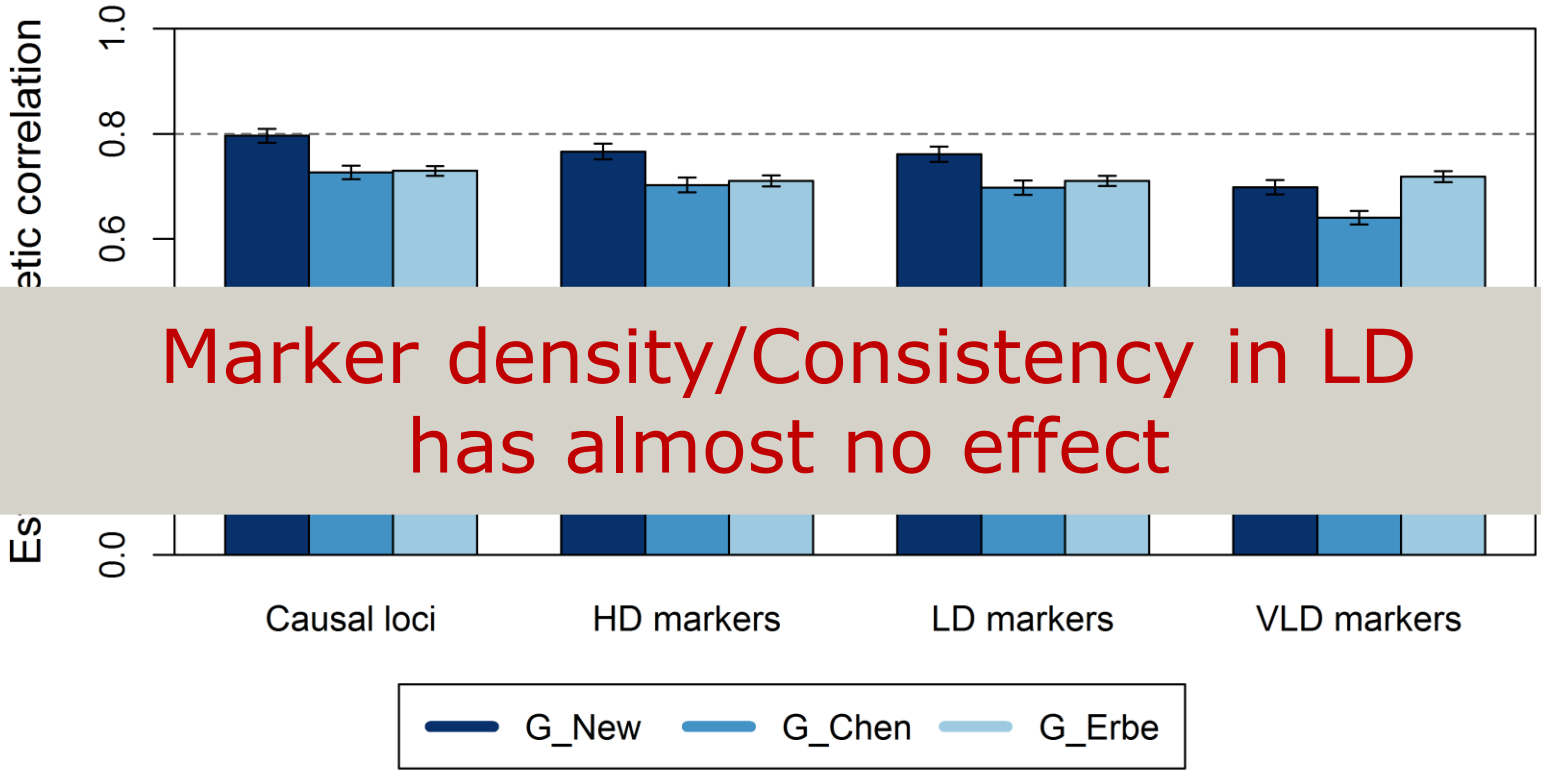
G_Erbe: Base when populations separated

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2018

Estimated genetic correlation

Allele frequency differences between populations of causal loci REPRESENTED by markers



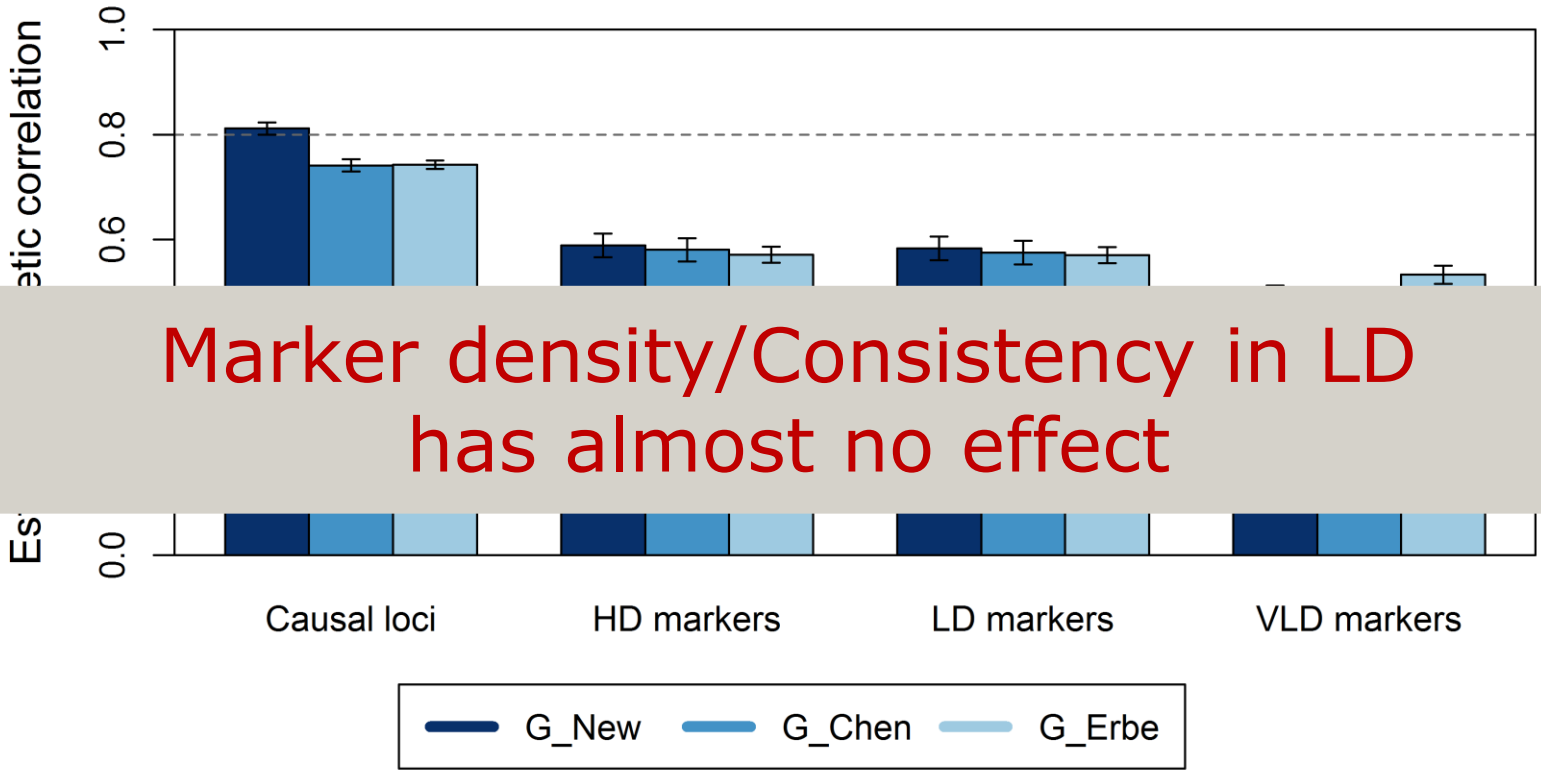
Marker density/Consistency in LD has almost no effect

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2018

Estimated genetic correlation

Allele frequency differences between populations of causal loci NOT REPRESENTED by markers



Marker density/Consistency in LD has almost no effect

WAGENINGEN
UNIVERSITY & RESEARCH

Wientjes *et al.*, GSE, 2018

Conclusion

Combining populations beneficial when:

- Closely related populations
- Population itself is small
- Many individuals are added

Genetic correlation between populations:

- Important parameter
- Unbiasedly estimated with **G**_New
  - markers represent properties causal variants

Acknowledgements

Mario Calus
Roel Veerkamp
Piter Bijma
Henk Bovenhuis
Pascal Duenk
Jeremie Vandenplas

Ben Hayes
Mike Goddard

Chris Schrooten

Financial support