Genomic selection using information from multiple populations
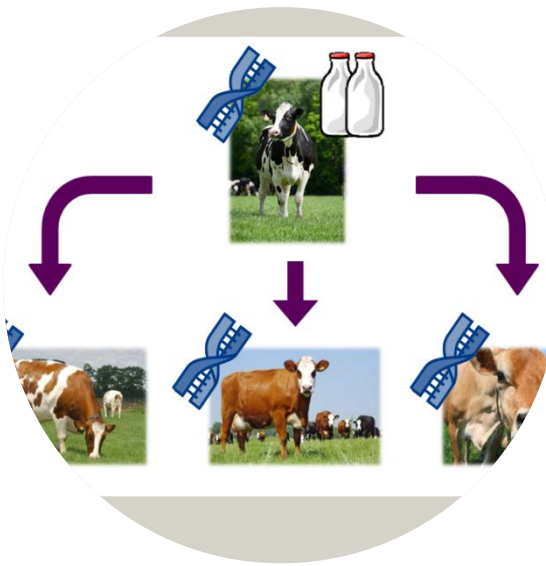
Yvonne Wientjes

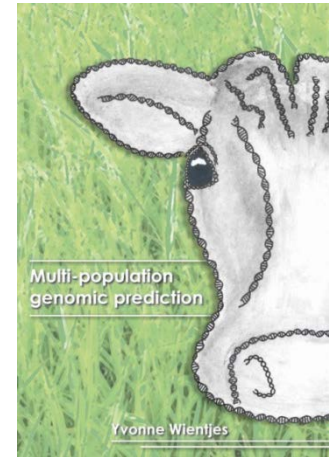*WIAS Lunch Lecture*   *27 June 2017*

WAGENINGEN
UNIVERSITY & RESEARCH

Who am I?

PhD GenomXL (ABG, WLR)

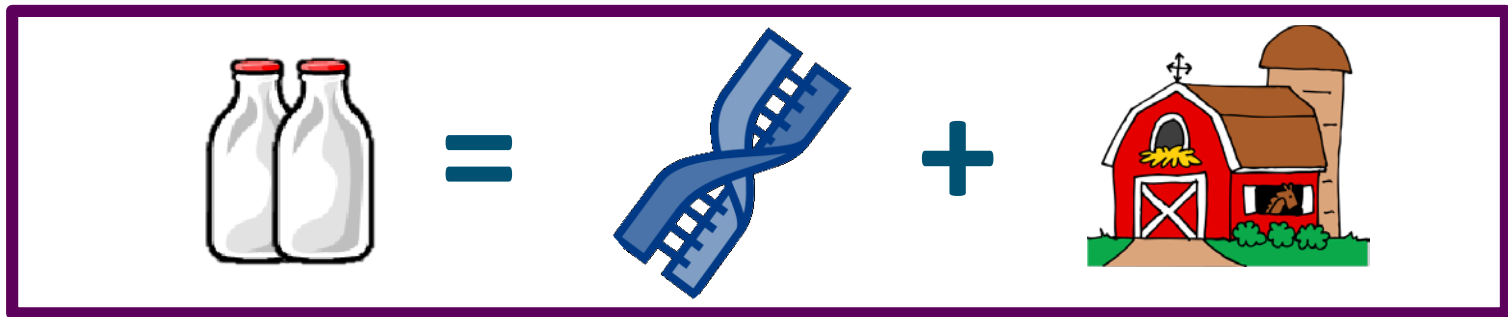- 'Multi-population genomic prediction'
- Mario Calus & Roel Veerkamp

Postdoc GenoMiX project (ABG, WU)

- Genomic prediction for crossbred performance
- Mario Calus, Piter Bijma & Pascal Duenk

Aim of animal breeding

Identify the GENETICALLY BEST animals to produce the next generation

Genomic selection

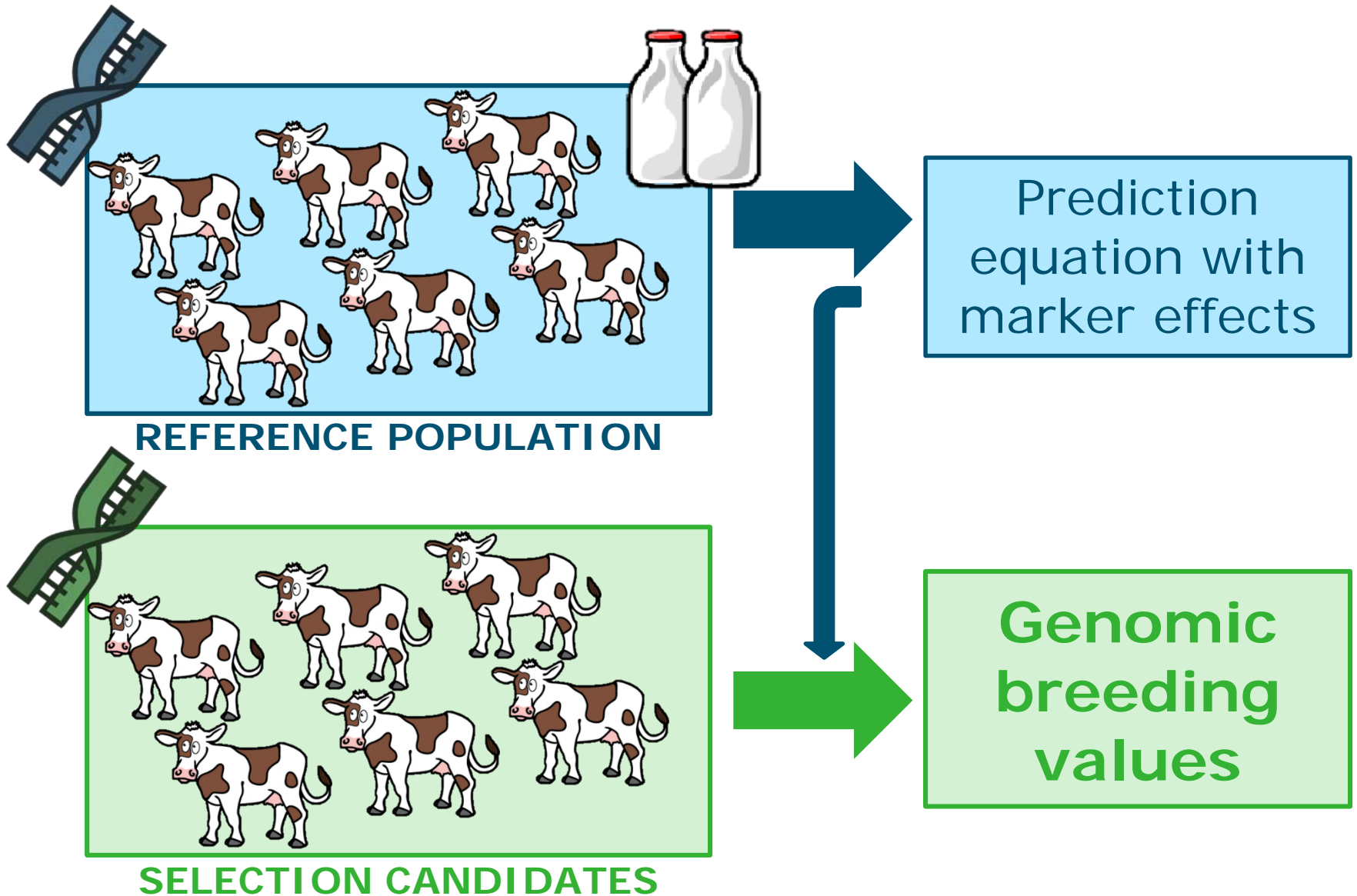## Select individuals based on DNA markers

- 'Genomic breeding values'

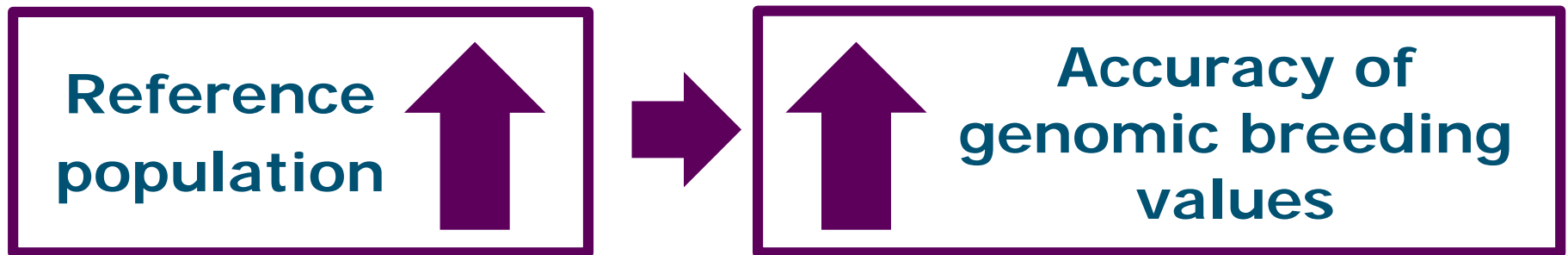**Can double genetic improvement per year in dairy cattle**
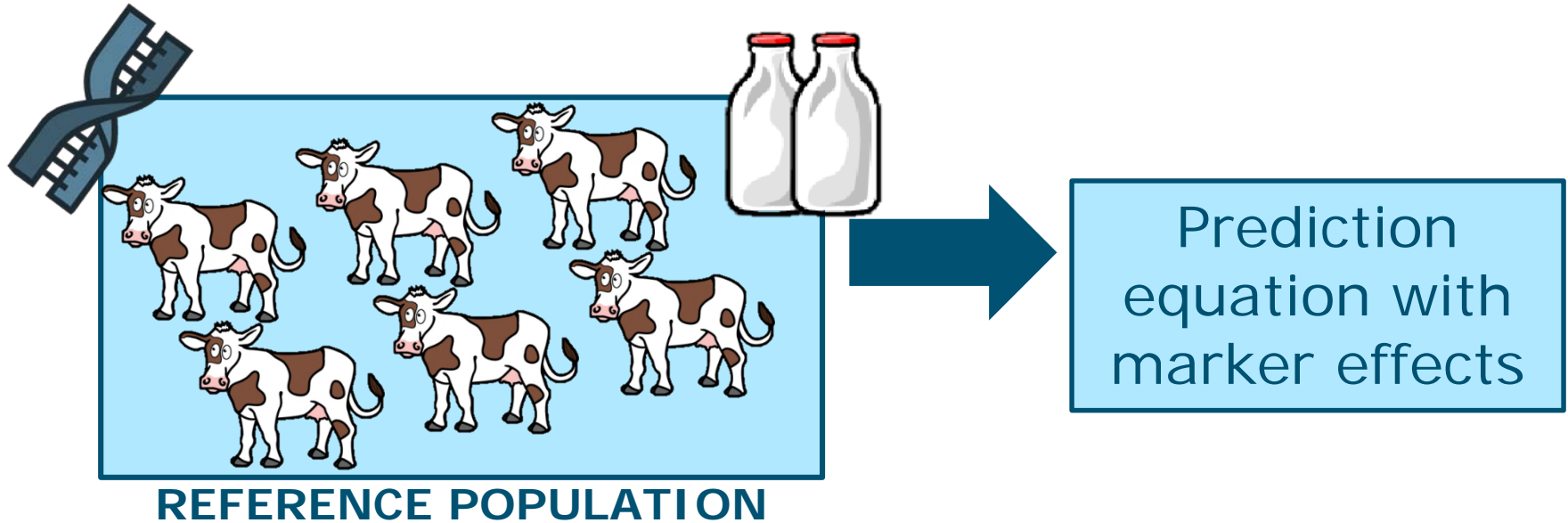
Genomic selection

> 50,000 markers

+2

WAGENINGEN
UNIVERSITY & RESEARCH

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 genomic breeding values**
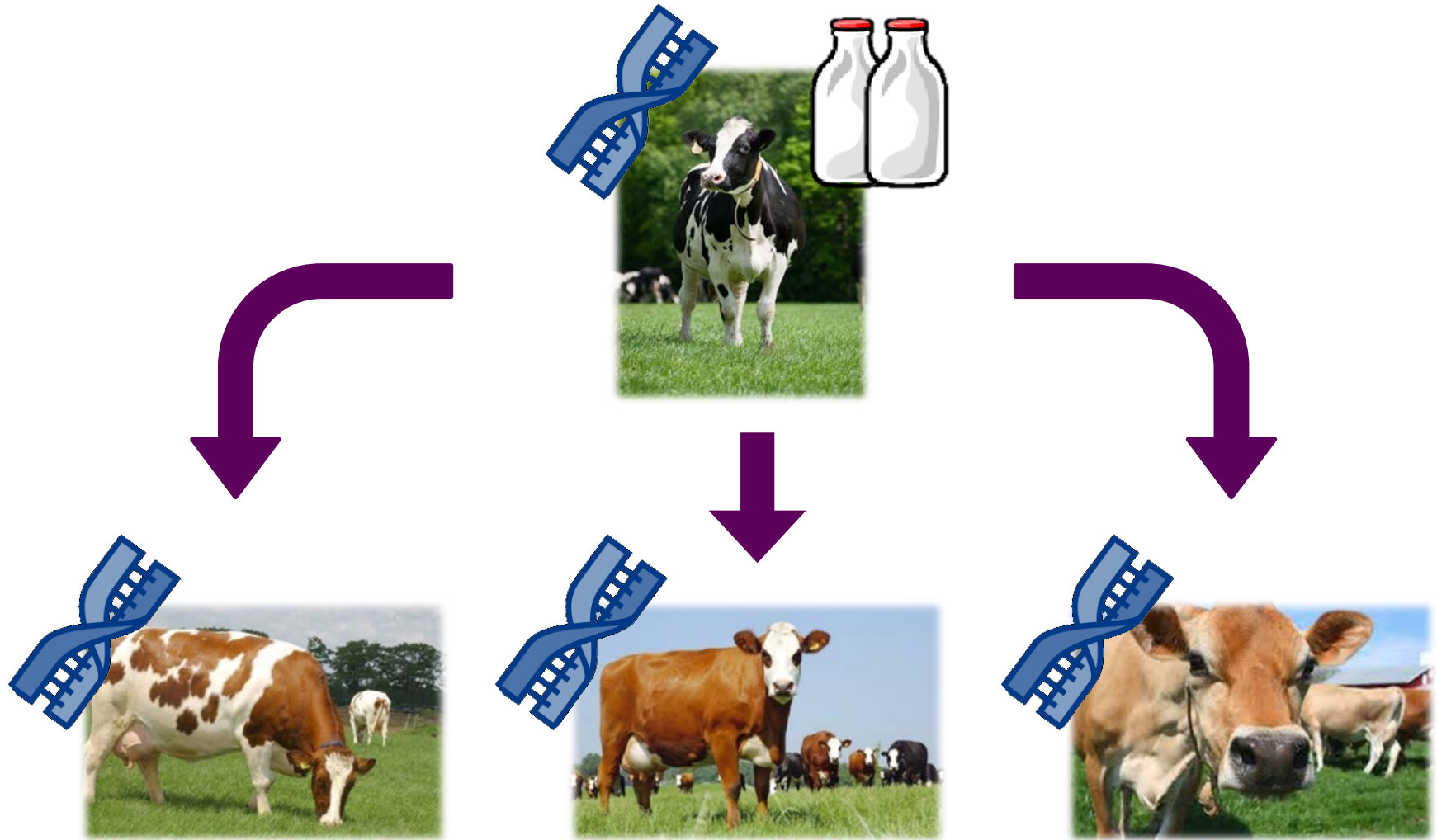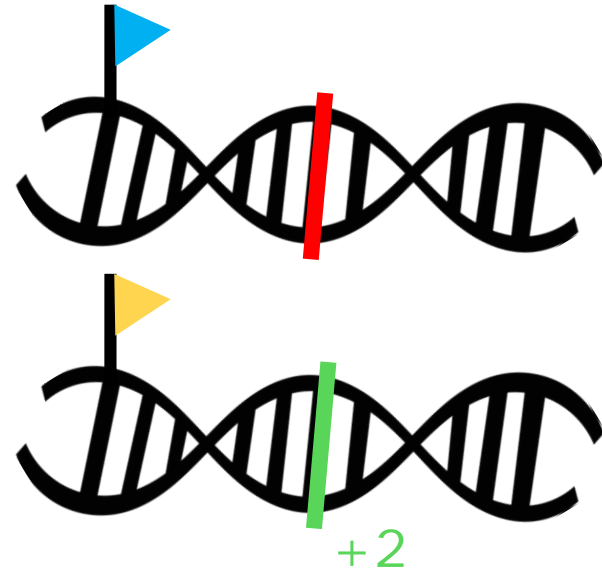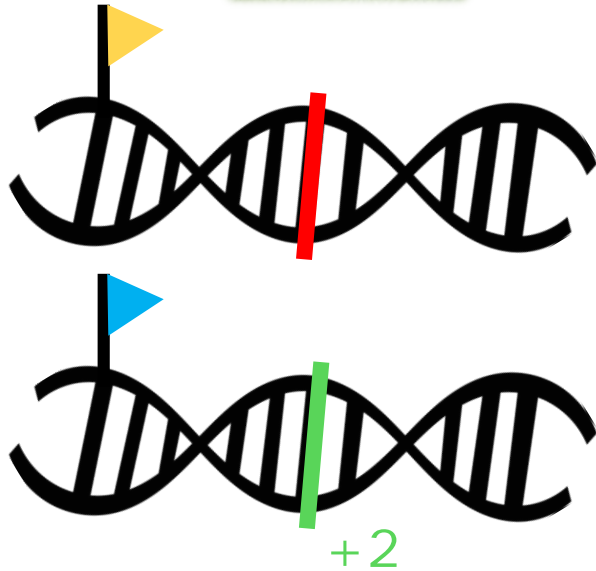
Size of reference population

Use information across breeds?

Differences between breeds

- Link between markers and causal variants (genes)
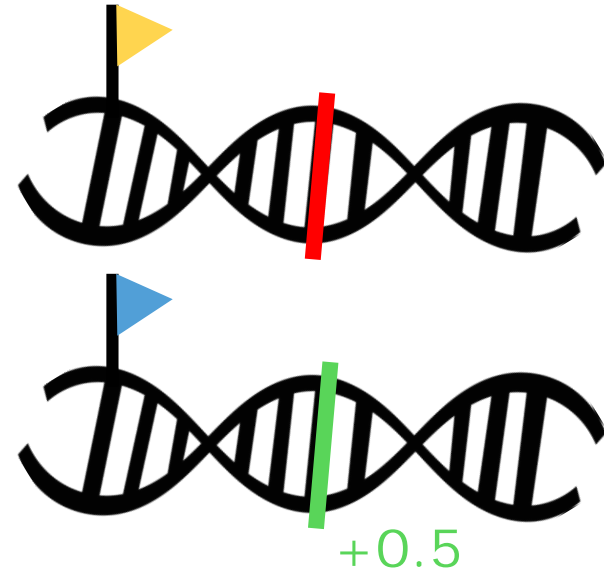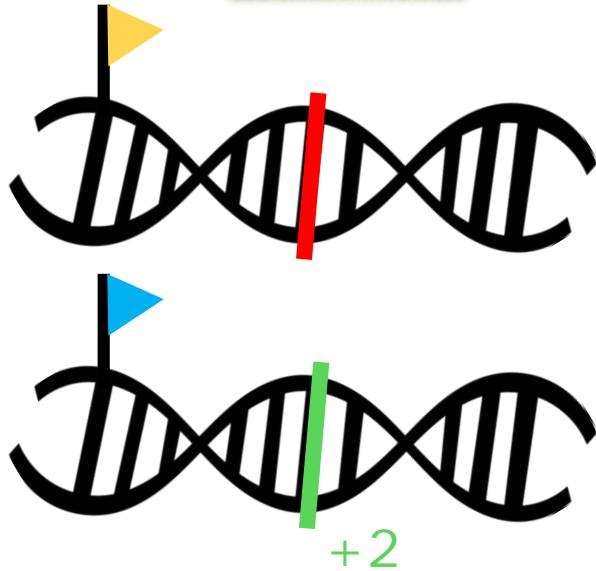 - *Linkage disequilibrium*

Differences between breeds

Differences between breeds

- Link between markers and causal variants (genes)
  - *Linkage disequilibrium*

- Allele frequencies of causal variants

- Effects of causal variants
  - Environment different
  - *Genetic correlation between breeds*

Differences between breeds

+2

+0.5

Differences between breeds

- Link between markers and causal variants (genes)
  - *Linkage disequilibrium (LD)*

- Allele frequencies of causal variants

- Effects of causal variants
  - Environment different
  - *Genetic correlation between breeds*
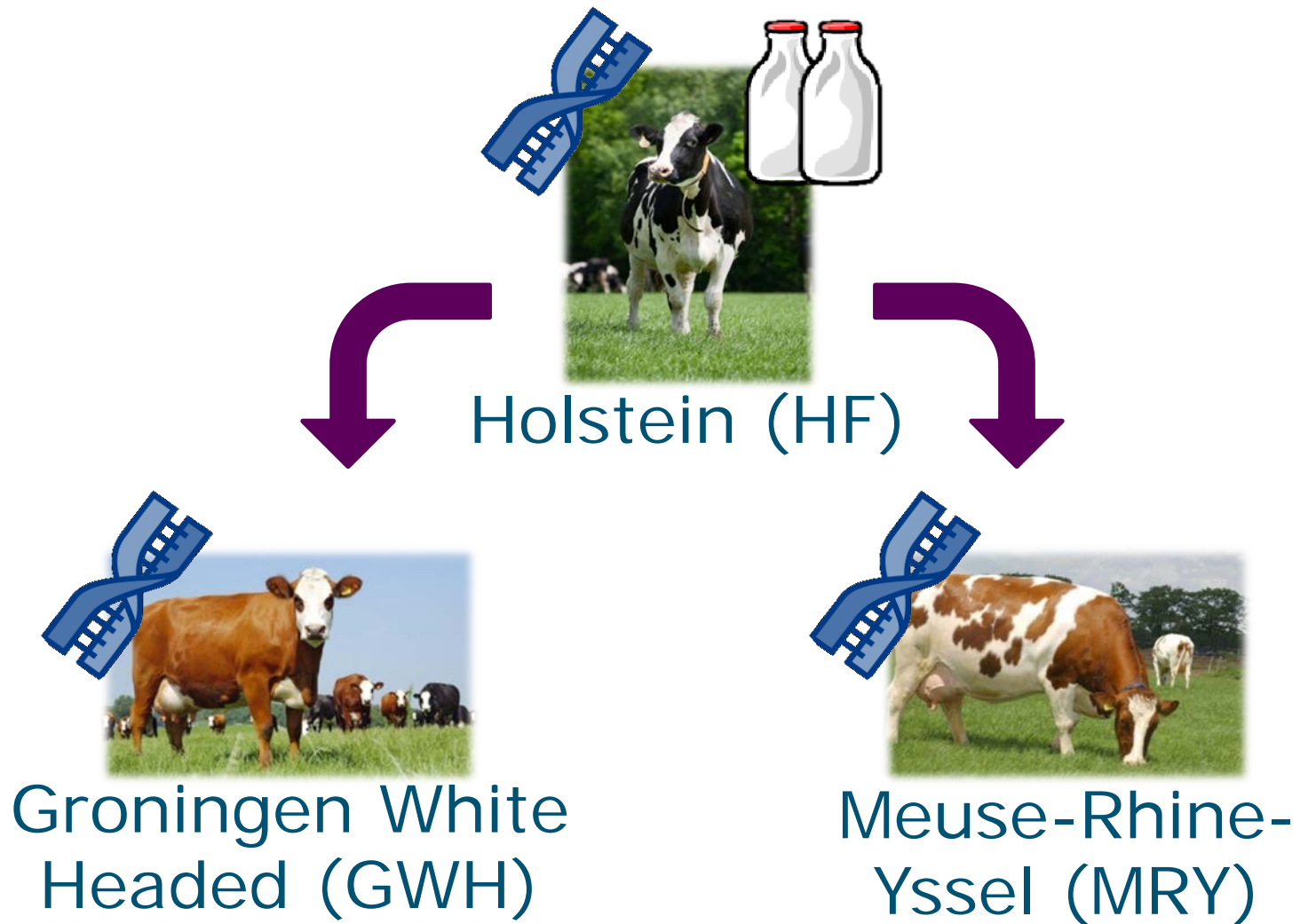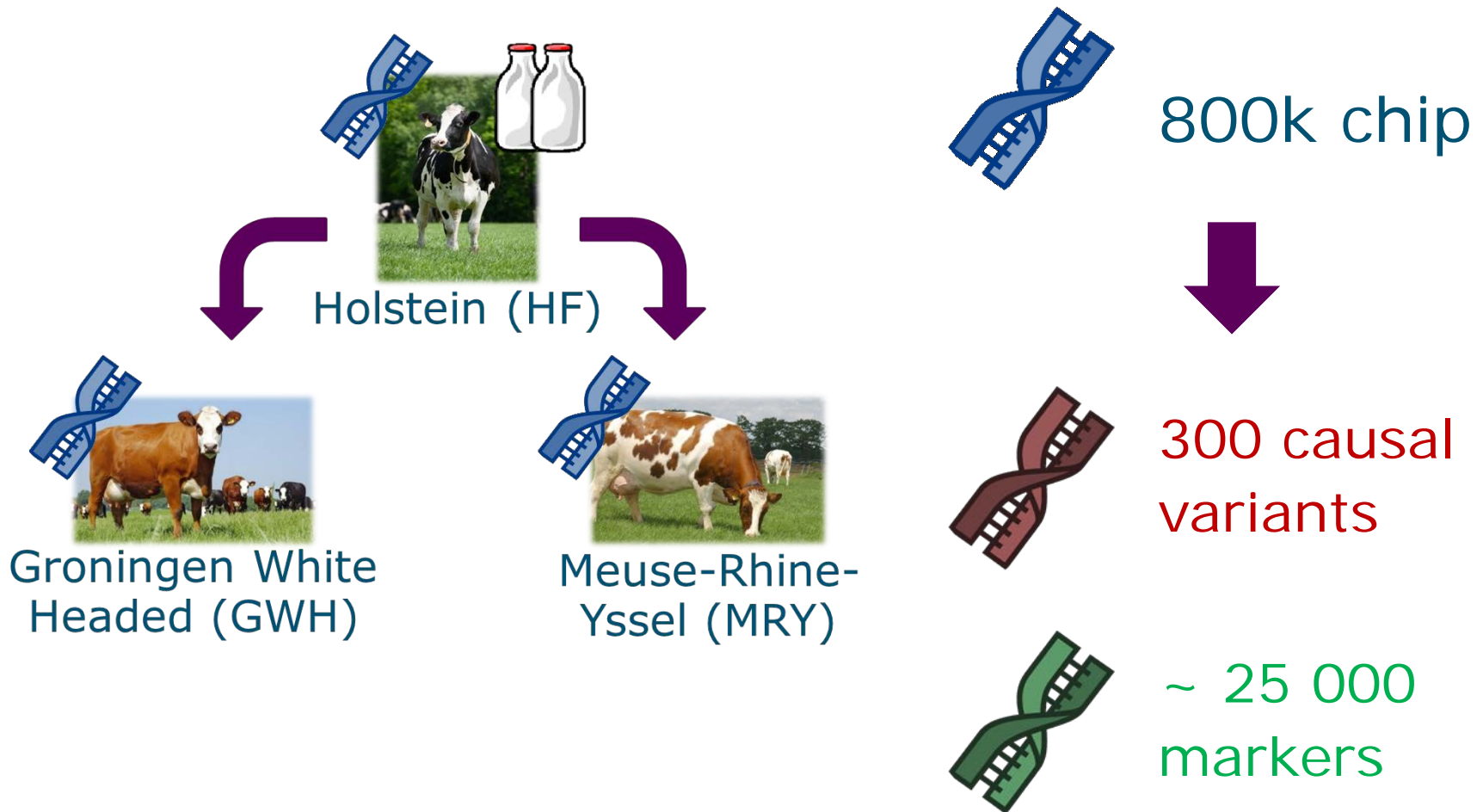
- Close family relationships are absent

WAGENINGEN
UNIVERSITY & RESEARCH

Accuracy of across-breed genomic prediction

Holstein (HF)

Groningen White Headed (GWH)

Meuse-Rhine-Yssel (MRY)

WAGENINGEN
UNIVERSITY & RESEARCH

Accuracy of across-breed genomic prediction

Holstein (HF)

Groningen White Headed (GWH)

Meuse-Rhine-Yssel (MRY)

800k chip

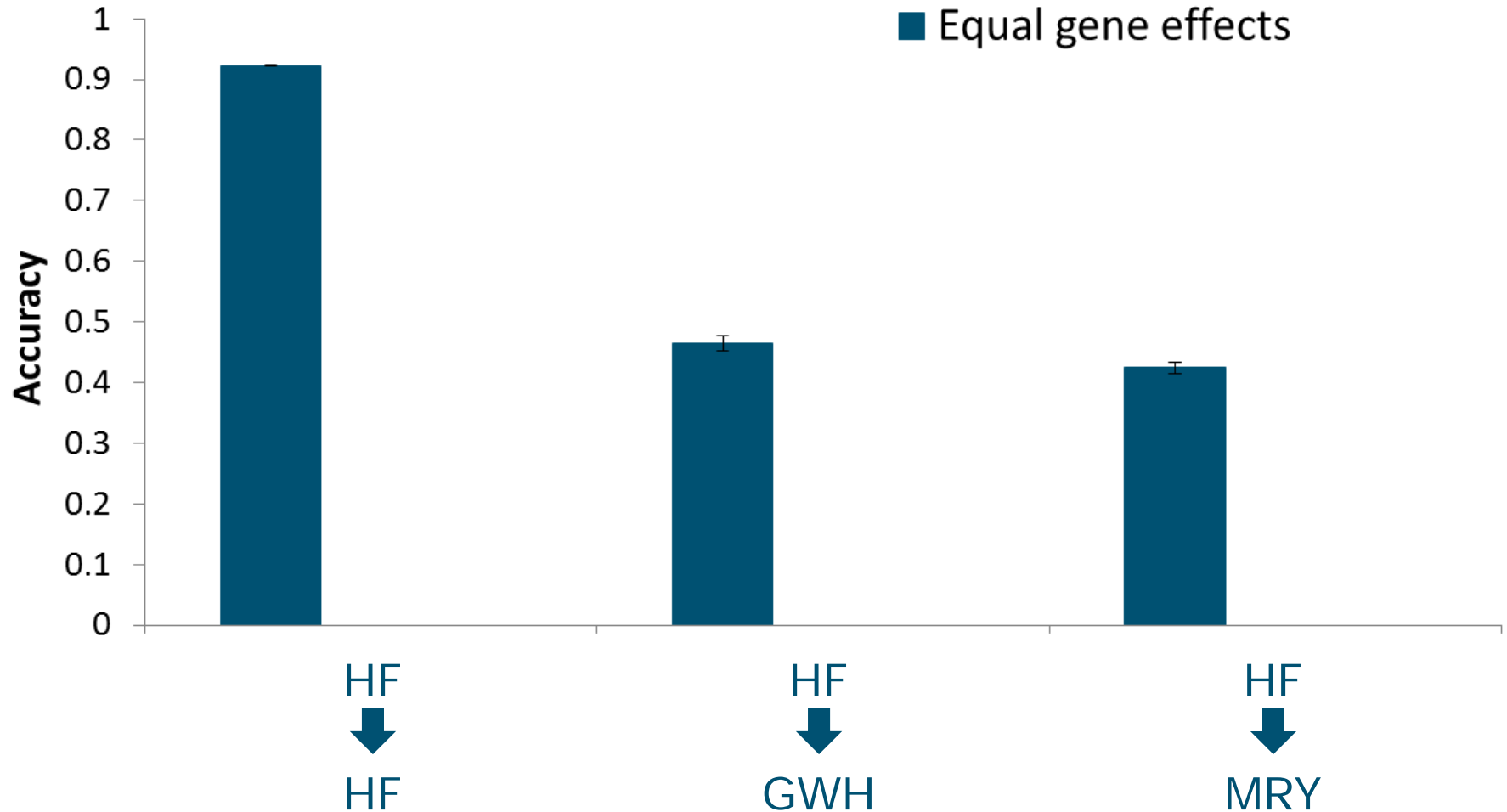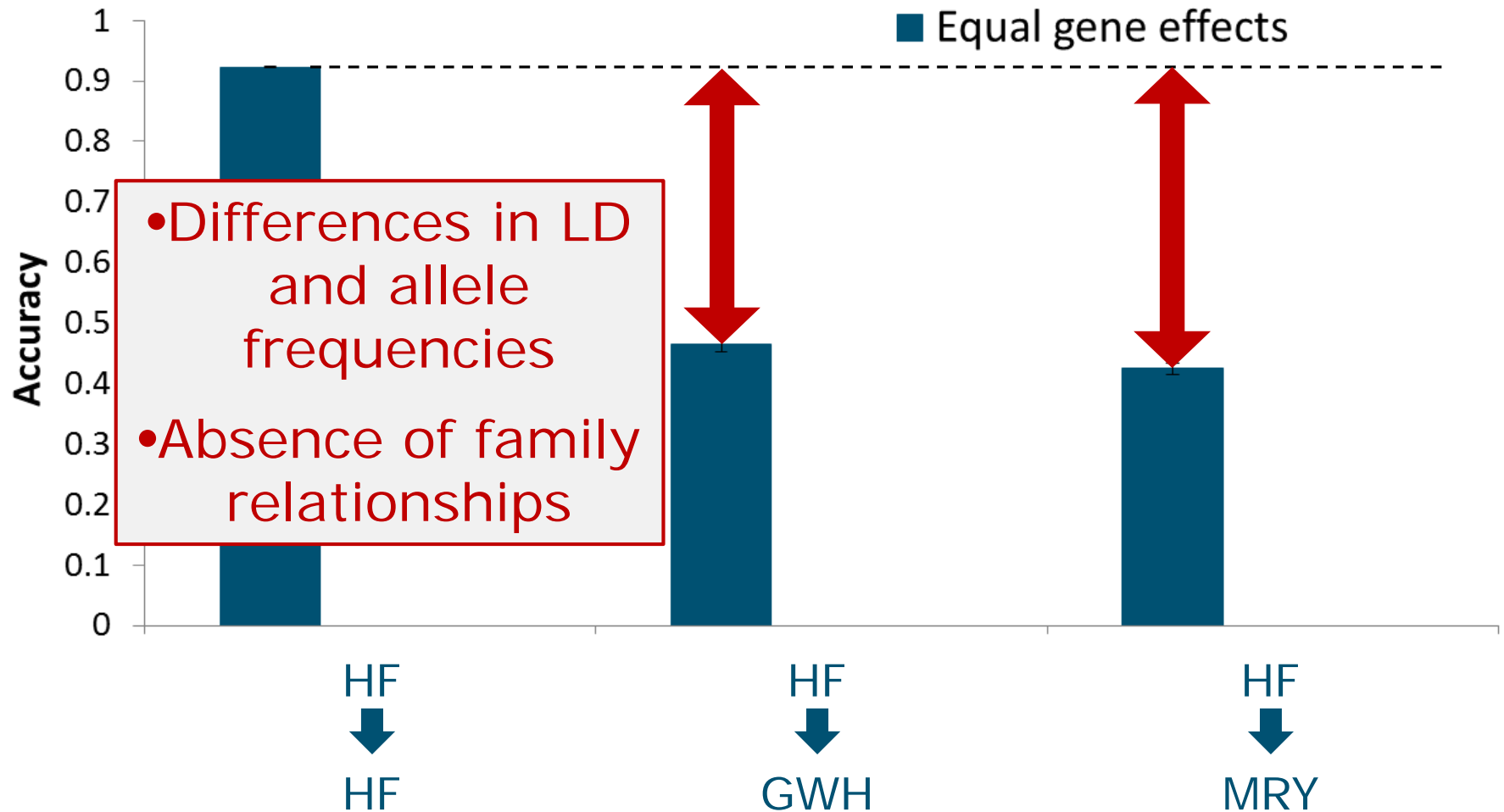300 causal variants

~ 25 000 markers

Wientjes *et al.*, GSE, 2015

Accuracy of across-breed genomic prediction

Accuracy of across-breed genomic prediction

Accuracy of across-breed genomic prediction

Legend:
- ■ Equal gene effects
- ▨ Correlation gene effects 0.8
- ▥ Correlation gene effects 0.4

80% **80%**

❖ Accuracy of across-breed genomic prediction is much lower than within-breed genomic prediction

❖ Differences in effects of causal variants reduce accuracy

WAGENINGEN
UNIVERSITY & RESEARCH

Multi-breed genomic prediction

Accuracy of across-breed genomic prediction is low…

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

Multi-breed genomic prediction

Prediction equation with marker effects
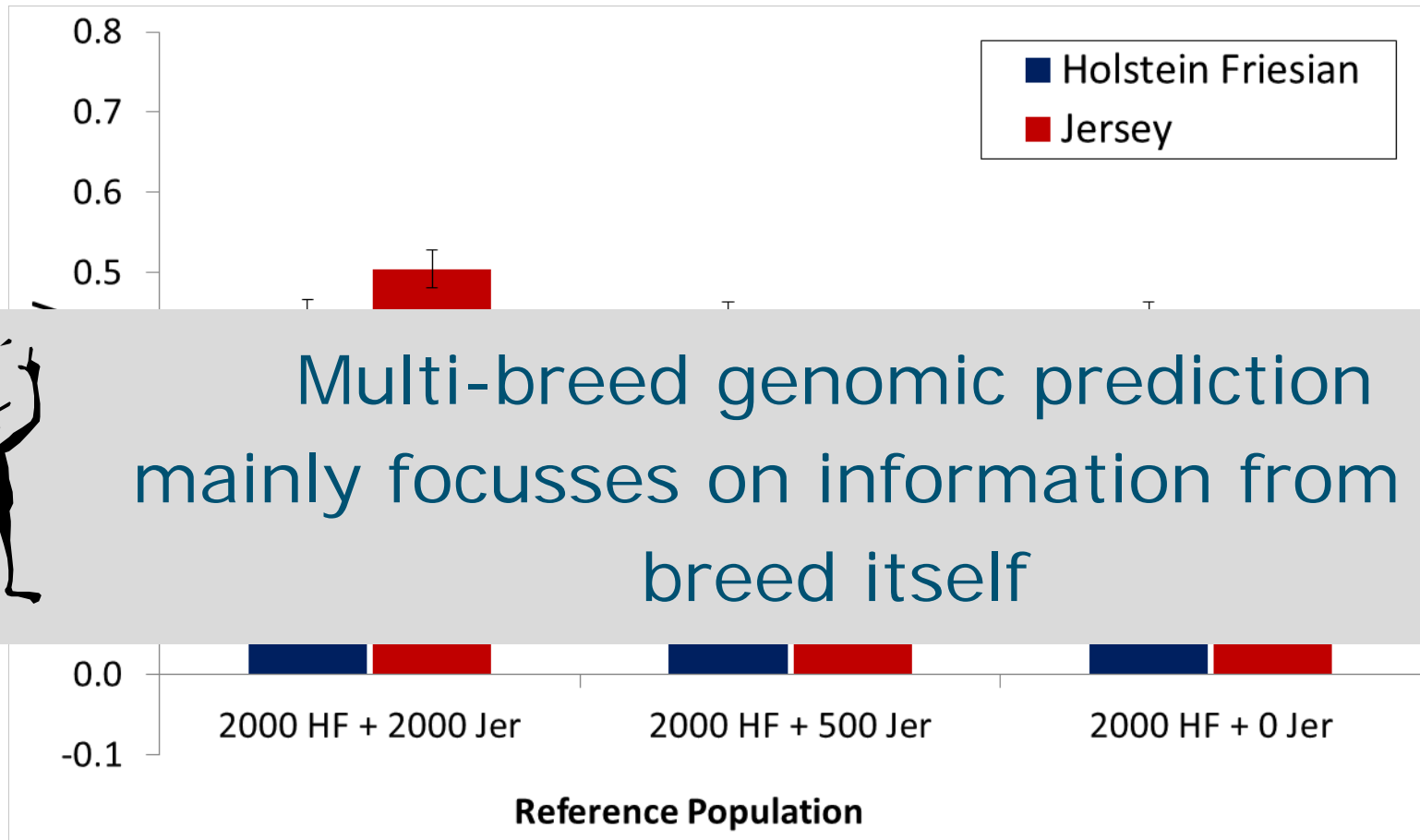
Genomic breeding values

Wientjes *et al.*, GSE, 2015

Accuracy of multi-breed genomic prediction

Accuracy of multi-breed genomic prediction

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

Can we predict accuracy?

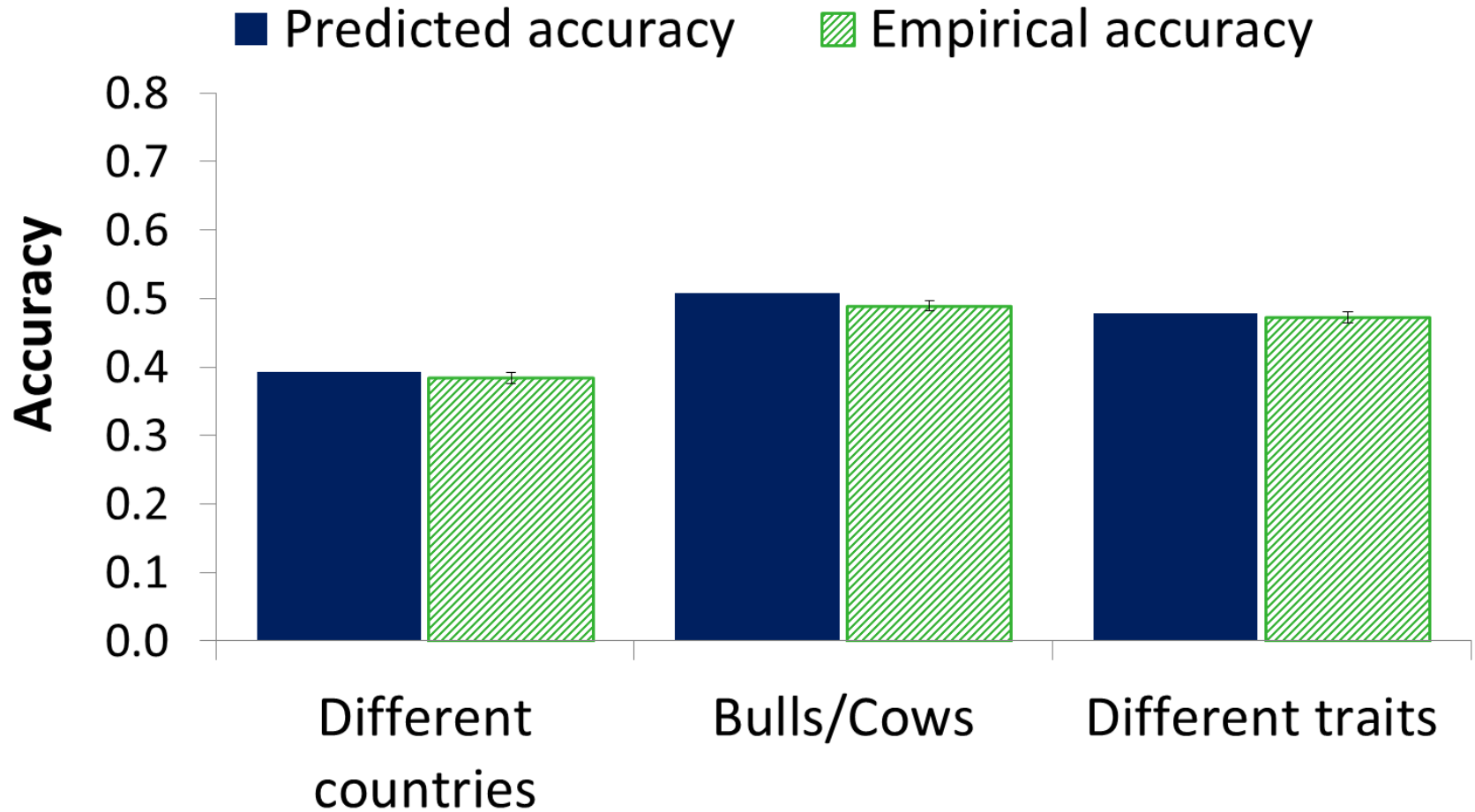Important when designing breeding programs
- Who to genotype?

Prediction equation:

$$r_{GEBV_{A+B,C}} = \sqrt{\left[r_{G_{A,C}}\sqrt{\dfrac{h_A^2}{M_{e_{A,C}}}} \quad r_{G_{B,C}}\sqrt{\dfrac{h_B^2}{M_{e_{B,C}}}} \right] \left[\begin{array}{cc} \dfrac{h_A^2}{M_{e_{A,C}}}+\dfrac{1}{n_{p,A}} & r_{G_{A,B}}\dfrac{\sqrt{h_A^2 h_B^2}}{\sqrt{M_{e_{A,C}}M_{e_{B,C}}}} \\ r_{G_{A,B}}\dfrac{\sqrt{h_A^2 h_B^2}}{\sqrt{M_{e_{A,C}}M_{e_{B,C}}}} & \dfrac{h_B^2}{M_{e_{B,C}}}+\dfrac{1}{n_{p,B}} \end{array} \right]^{-1} \left[\begin{array}{c} r_{G_{A,C}}\sqrt{\dfrac{h_A^2}{M_{e_{A,C}}}} \\ r_{G_{B,C}}\sqrt{\dfrac{h_B^2}{M_{e_{B,C}}}} \end{array} \right]}$$

Input parameter:
- Heritability $\qquad (h^2)$
- Relatedness between populations $\qquad (M_e)$
- Genetic correlation $\qquad (r_G)$
- Number of animals $\qquad (n_P)$

Wientjes *et al.*, Genetics, 2016

Predicted *vs* Empirical accuracy

Predicted *vs* Empirical accuracy

Equation can accurately predict accuracy of multi-population genomic prediction

Combining populations?

Beneficial, when:

- Closely related populations

- Population itself is small

- A large number of individuals is added

What about chickens and pigs?
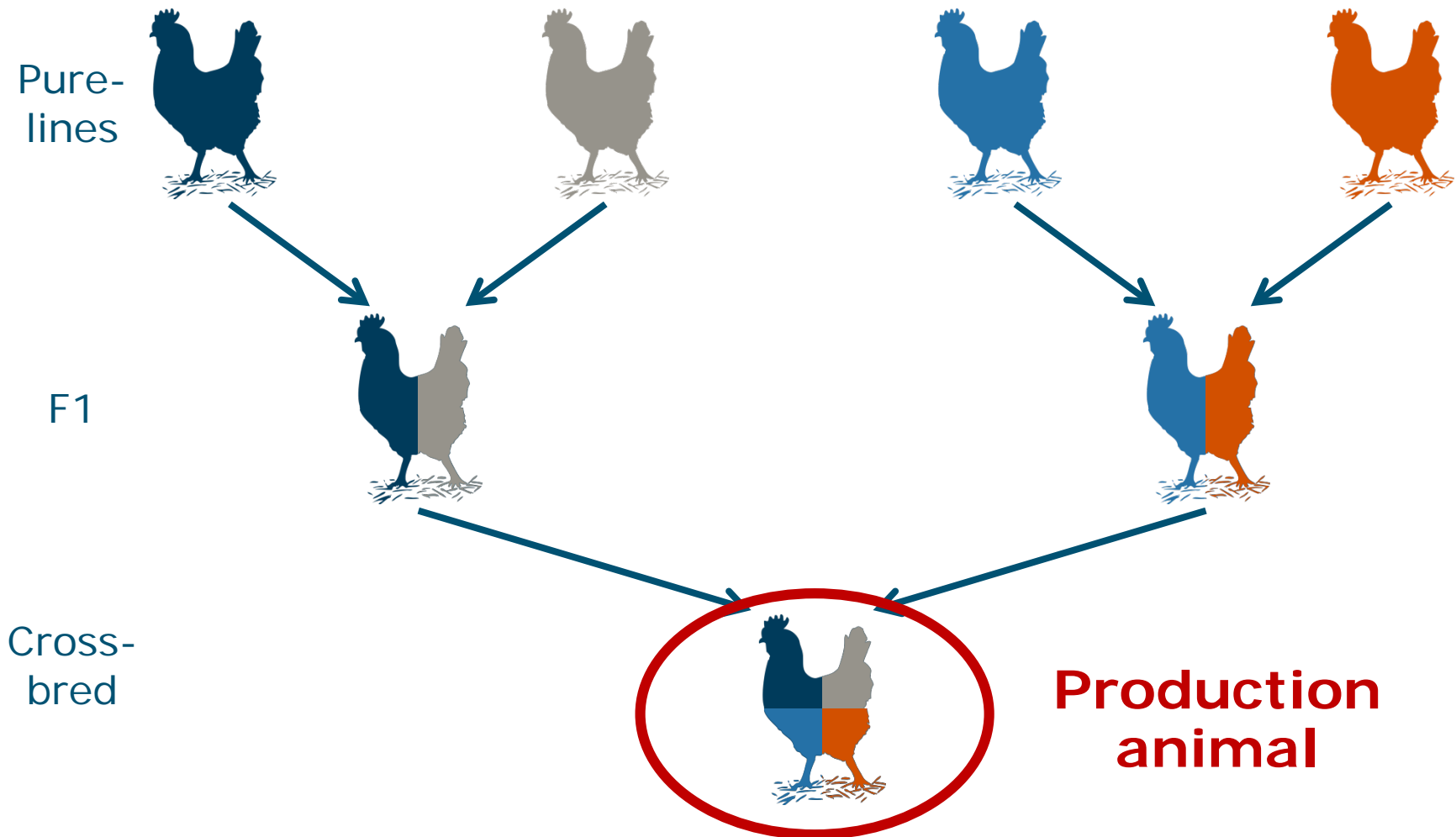
Potential of genomic selection

Dairy cattle

- Benefit mostly due to reduction generation interval
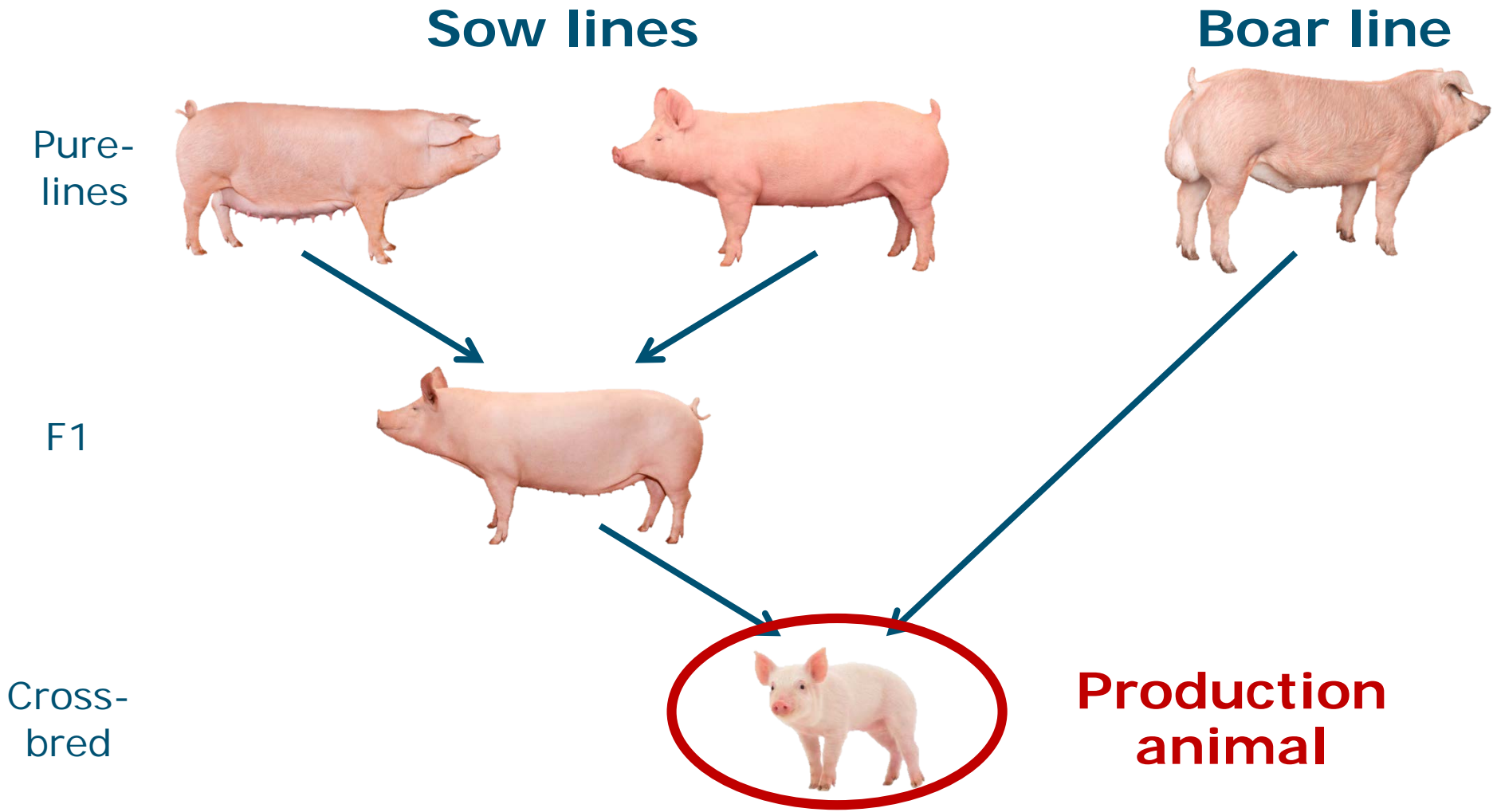- Dominated by one breed: Holstein Friesian
- Mostly purebreds

Chicken/pig

- Generation interval already small!!
- Lots of different lines/breeds
- Crossbreeding

Chicken breeding design (4-way crossbred)

Pure-lines

F1

Cross-bred

Production animal

Pig breeding design (3-way crossbred)

Sow lines   **Boar line**

Pure-lines

F1

Cross-bred

Production animal

Challenge of crossbreeding

AIM: Select purebred animals to optimize crossbred performance

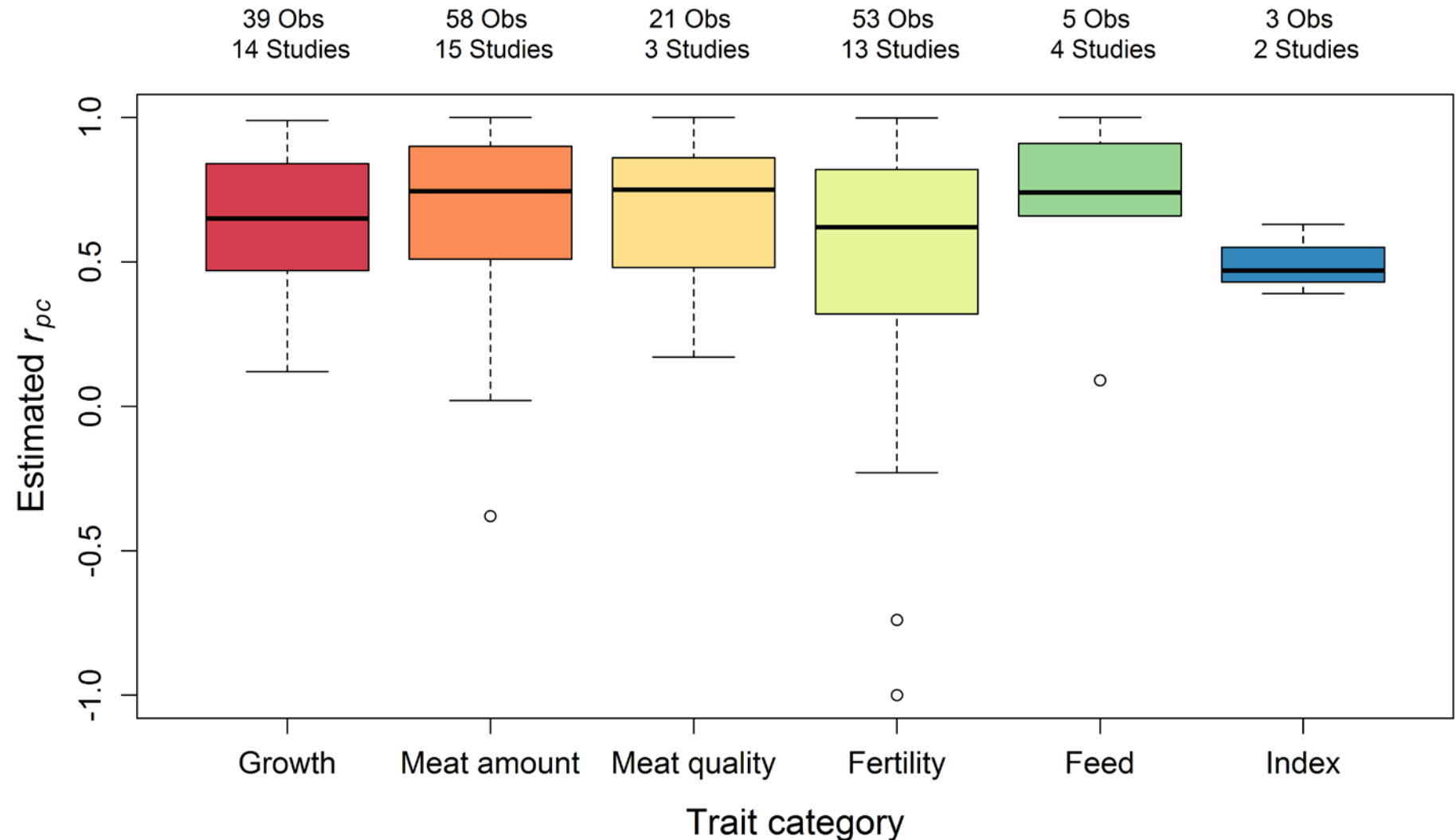**Differences purebred versus crossbred**

- Environment
- Genetic background

Purebred-crossbred correlation $(r_{pc}) < 1$

Best purebred animals don't necessarily produce the best crossbred offspring

r_{pc} for different trait groups in pigs

Wientjes and Calus, JAS, 2017

Current research

Optimizing breeding program designs and breeding value estimation to improve crossbred performance
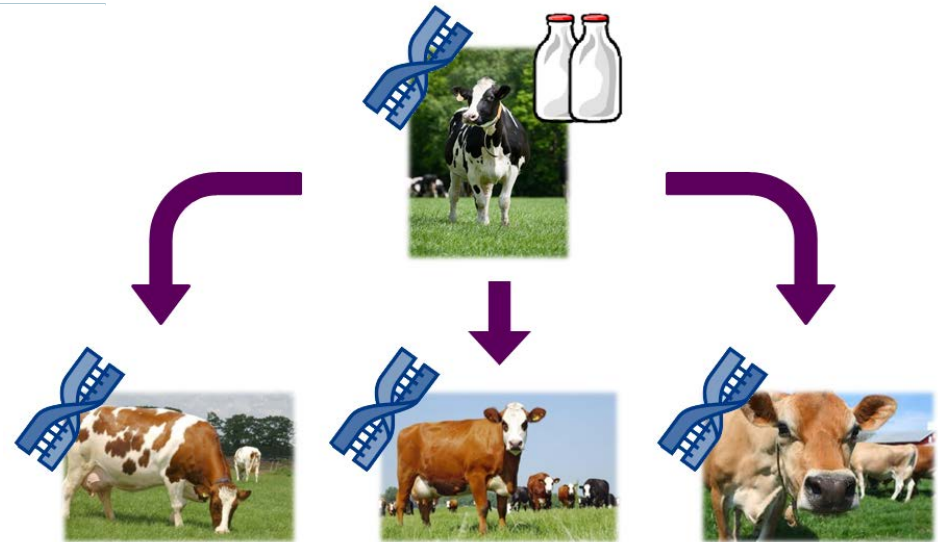
Conclusion

Populations differ

Combining information from populations, beneficial when:

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

Current challenge:

- Use genomic prediction to improve crossbred performance

Thank you!