Predicting the accuracy of multi-population genomic prediction

Yvonne Wientjes
Piter Bijma, Roel Veerkamp, Mario Calus

2nd September, 2015

LIVESTOCK RESEARCH
WAGENINGEN UR

Acknowledgements

BREED4FOOD

www.breed4food.com

CRV

www.crv4all.com

Designing breeding programs

Maximize response to selection

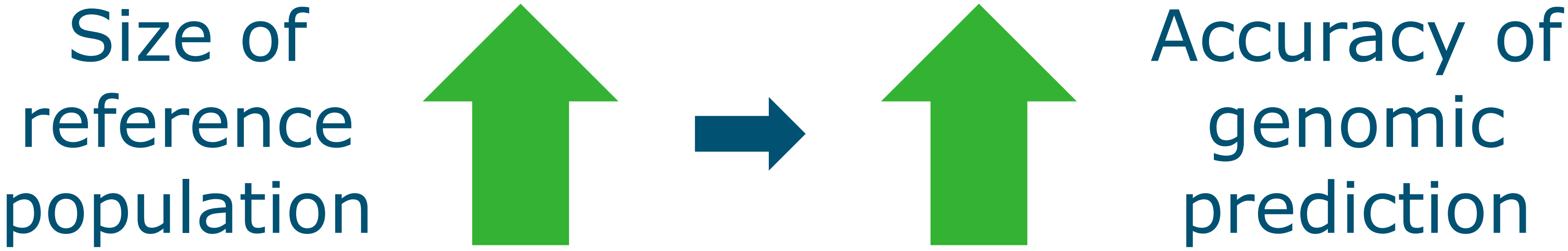
Designing breeding programs

Maximize response to selection

⬆

Accuracy for different designs

Accuracy of genomic selection



Size of reference population ↑ → ↑ Accuracy of genomic prediction

Accuracy of genomic selection

Size of reference population ↑ → ↑ Accuracy of genomic prediction

Combine populations:

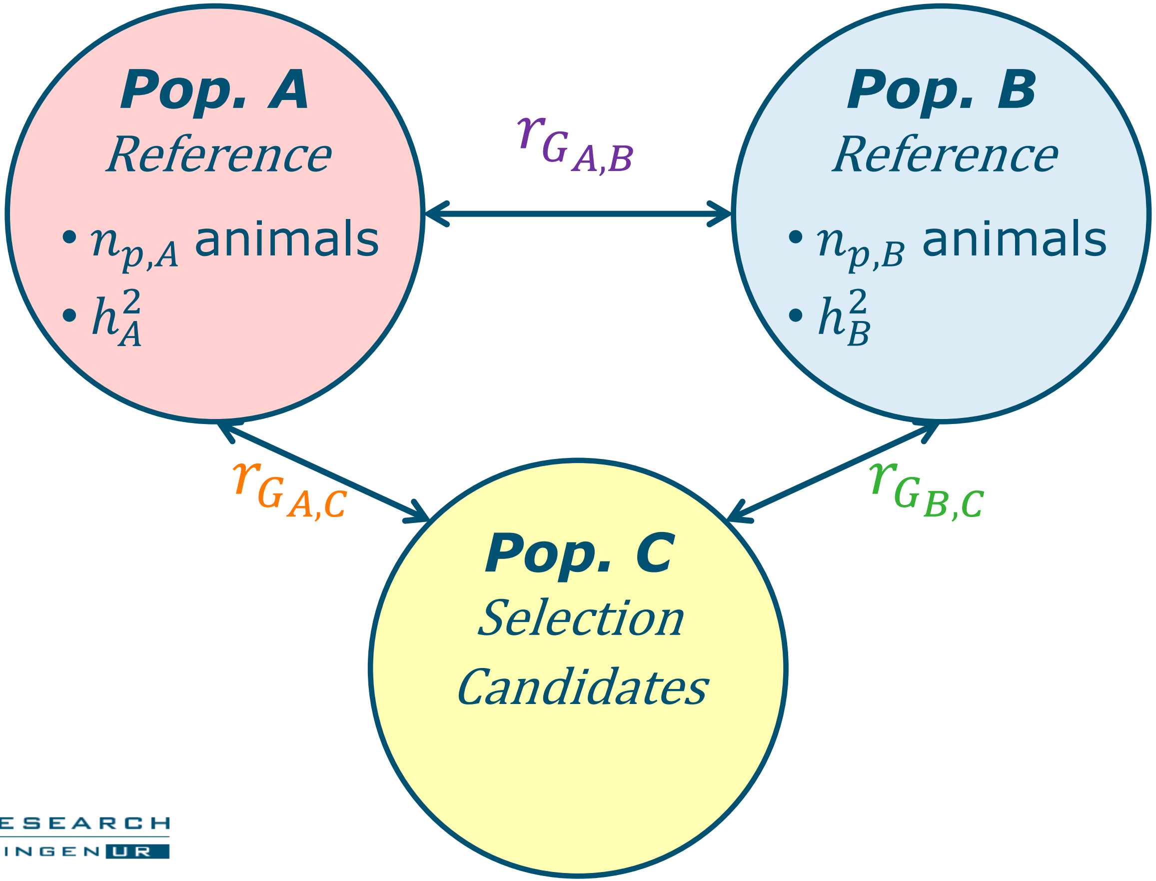
- Countries
- Lines
- Breeds
- Bulls/Cows

Aim of this study

Derive and **validate**:

Prediction equation for accuracy of GEBVs using multiple populations in the reference population

Schematic overview



Pop. A

Reference

- $n_{p,A}$ animals
- h_A^2

$r_{G_{A,B}}$

Pop. B

Reference

- $n_{p,B}$ animals
- h_B^2

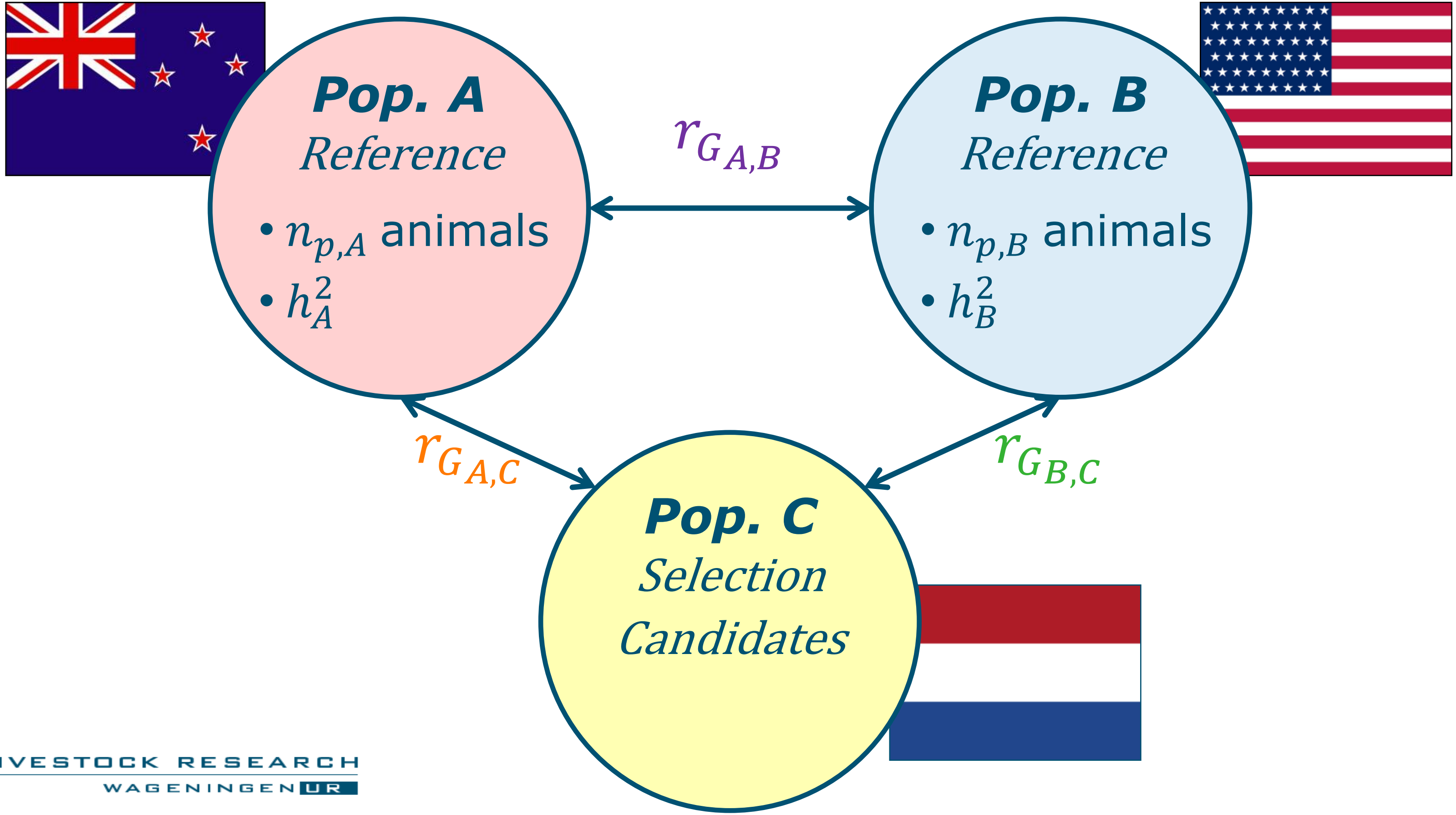
$r_{G_{A,C}}$

$r_{G_{B,C}}$

Pop. C

Selection Candidates

Schematic overview – an example



Pop. A

Reference

- $n_{p,A}$ animals
- h_A^2

$r_{G_{A,B}}$

Pop. B

Reference

- $n_{p,B}$ animals
- h_B^2

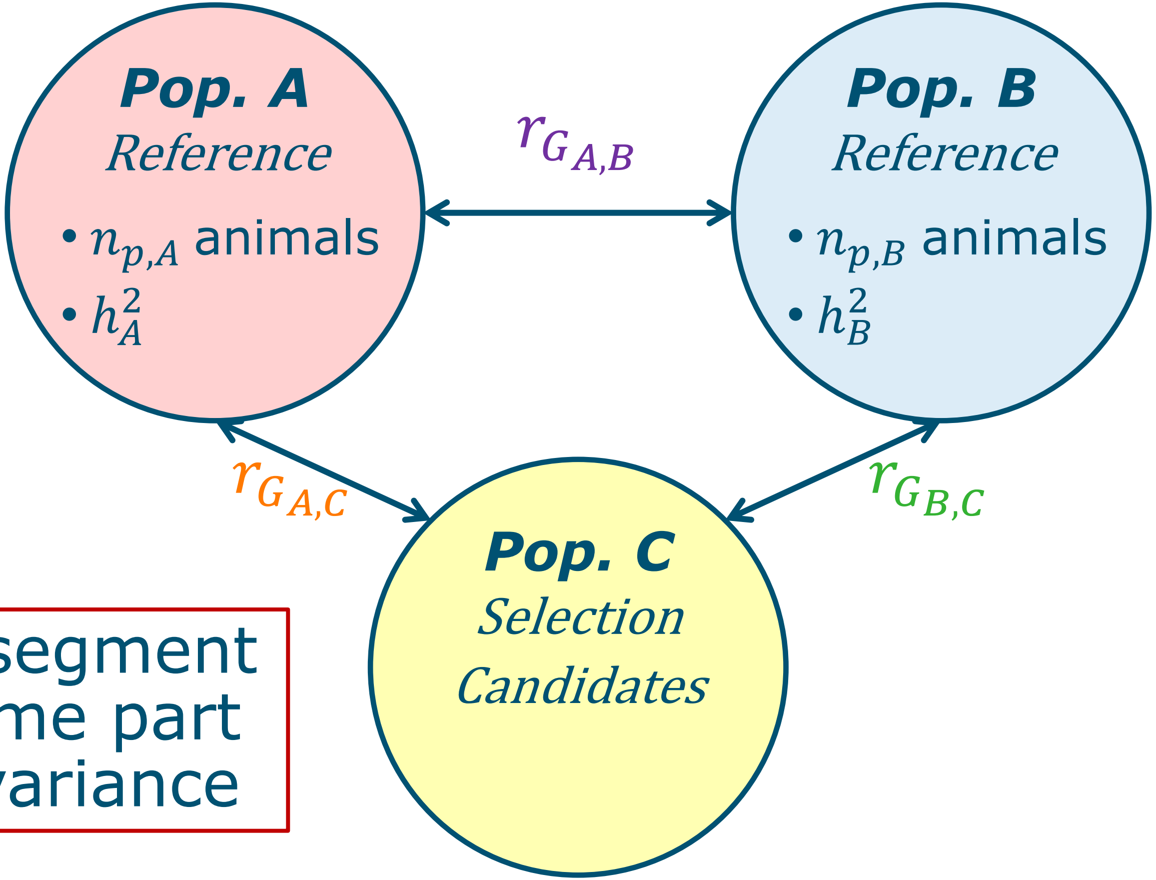
$r_{G_{A,C}}$

$r_{G_{B,C}}$

Pop. C

Selection Candidates

Schematic overview



Pop. A

Reference

- $n_{p,A}$ animals
- h_A^2

Pop. B

Reference

- $n_{p,B}$ animals
- h_B^2

Pop. C

Selection Candidates

$r_{G_{A,B}}$

$r_{G_{A,C}}$

$r_{G_{B,C}}$

Each locus/segment explains same part of genetic variance

Input parameters

Reference population

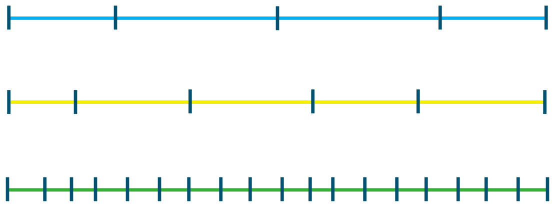
- Number of animals (n_P)
- Heritability (h^2)
- Genetic correlation (r_G)

Selection candidates – Reference population

- Genetic correlations (r_G)
- Number of independent chromosome segments shared (M_e)

M_e across populations

‘Effective number of estimated effects’



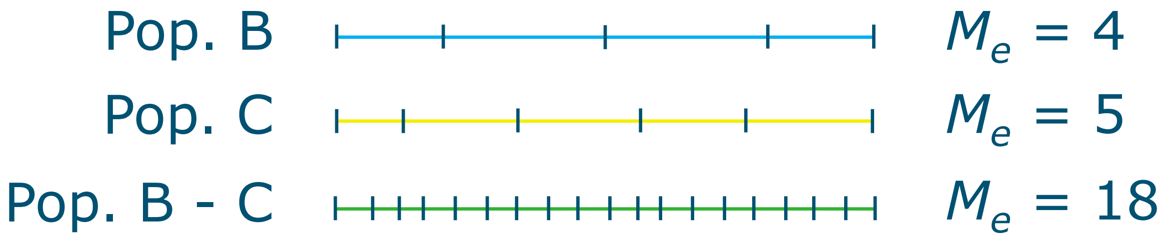
Pop. B $M_e = 4$

Pop. C $M_e = 5$

Pop. B - C $M_e = 18$

M_e across populations

'Effective number of estimated effects'



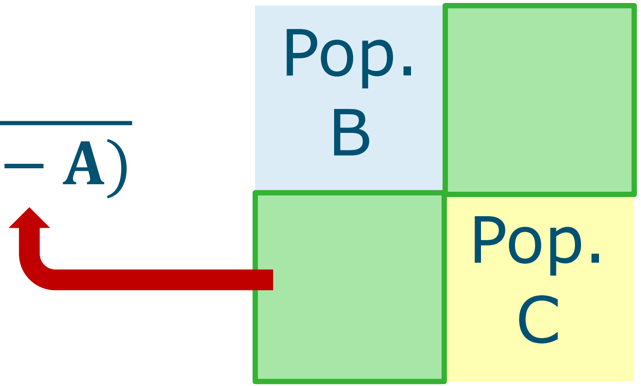
Pop. B $M_e = 4$

Pop. C $M_e = 5$

Pop. B - C $M_e = 18$

Relationship matrix

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



Input parameters

Reference population

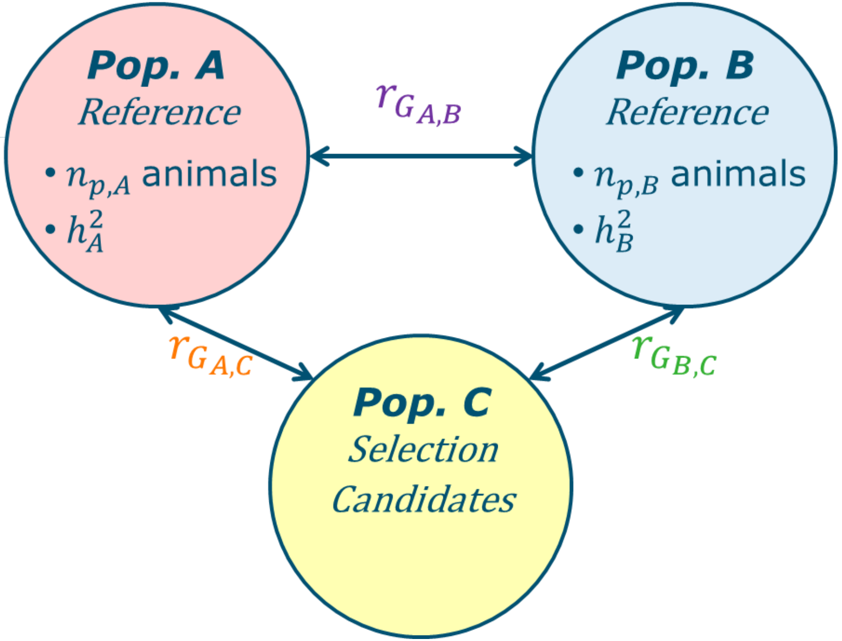
- Number of animals (n_P)
- Heritability (h^2)
- Genetic correlation (r_G)

Selection candidates – Reference population

- Genetic correlations (r_G)
- Number of independent chromosome segments shared (M_e)

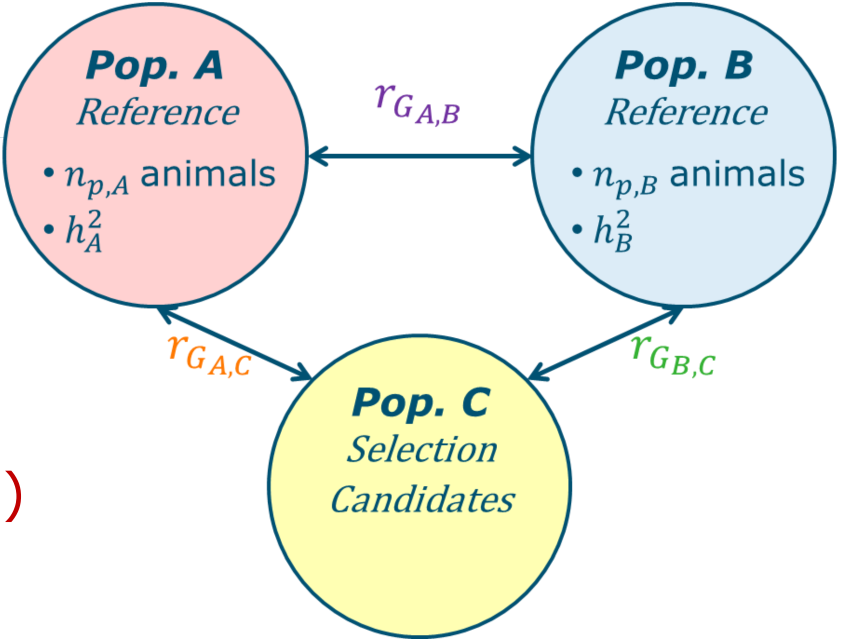
$$r_{GEBV_{A+B,C}} = \sqrt{\begin{bmatrix} r_{G_{A,C}}\sqrt{\frac{h_A^2}{M_{e_{A,C}}}} & r_{G_{B,C}}\sqrt{\frac{h_B^2}{M_{e_{B,C}}}} \end{bmatrix} \begin{bmatrix} \frac{h_A^2}{M_{e_{A,C}}} + \frac{1}{n_{p,A}} & r_{G_{A,B}}\frac{\sqrt{h_A^2 h_B^2}}{\sqrt{M_{e_{A,C}} M_{e_{B,C}}}} \\ r_{G_{A,B}}\frac{\sqrt{h_A^2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

Derived equation



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

Derived equation



Pop. A
Reference
- $n_{p,A}$ animals
- h_A^2

$r_{G_{A,B}}$

Pop. B
Reference
- $n_{p,B}$ animals
- h_B^2

$r_{G_{A,C}}$

$r_{G_{B,C}}$

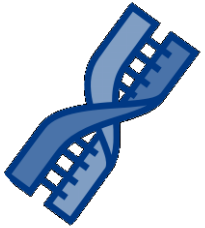
Pop. C
Selection Candidates

1 population in reference

→ Equal to Daetwyler *et al.* (2008)

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

Simulations for validation



Genotypes of

1033 HF animals

422 405 SNPs

50 000 candidate QTL

(Random or Low MAF)

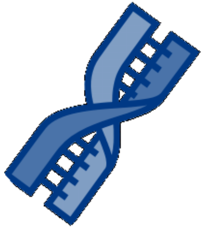
Phenotypes

Simulated

(4000 QTL)

TBV + E

Simulations for validation



Genotypes of 1033 HF animals

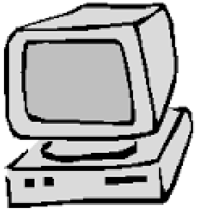
422 405 SNPs

50 000 candidate QTL (Random or Low MAF)

Phenotypes

Simulated (4000 QTL)

TBV + E



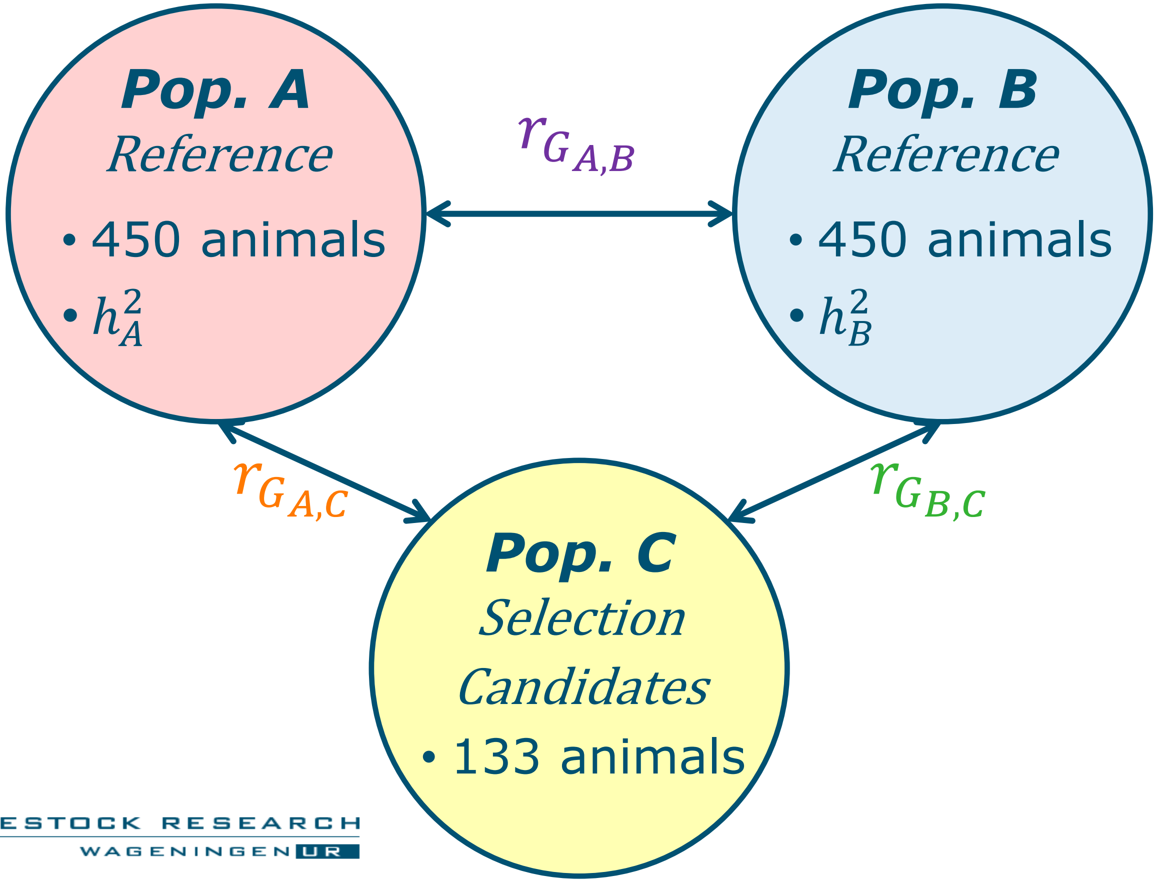
Breeding value (GEBV) estimation

Multi-trait GBLUP

Empirical Accuracy

Correlation GEBV with TBV

Populations for validation



Pop. A

Reference

- 450 animals
- h_A^2

$r_{G_{A,B}}$

Pop. B

Reference

- 450 animals
- h_B^2

$r_{G_{A,C}}$

$r_{G_{B,C}}$

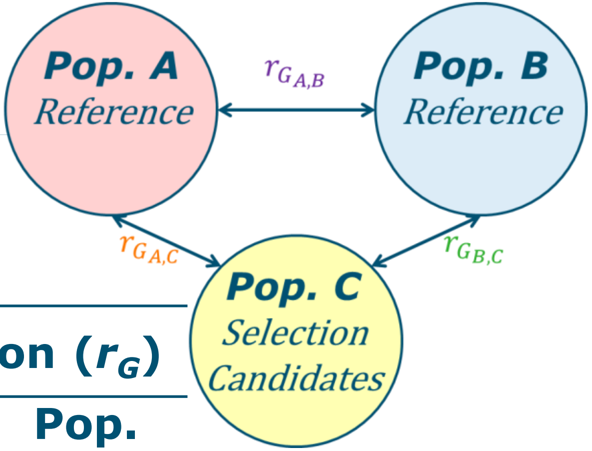
Pop. C

Selection Candidates

- 133 animals

Division:
Half-sib families kept in same population

Scenarios for validation



Pop. A
Reference

Pop. B
Reference

Pop. C
Selection Candidates

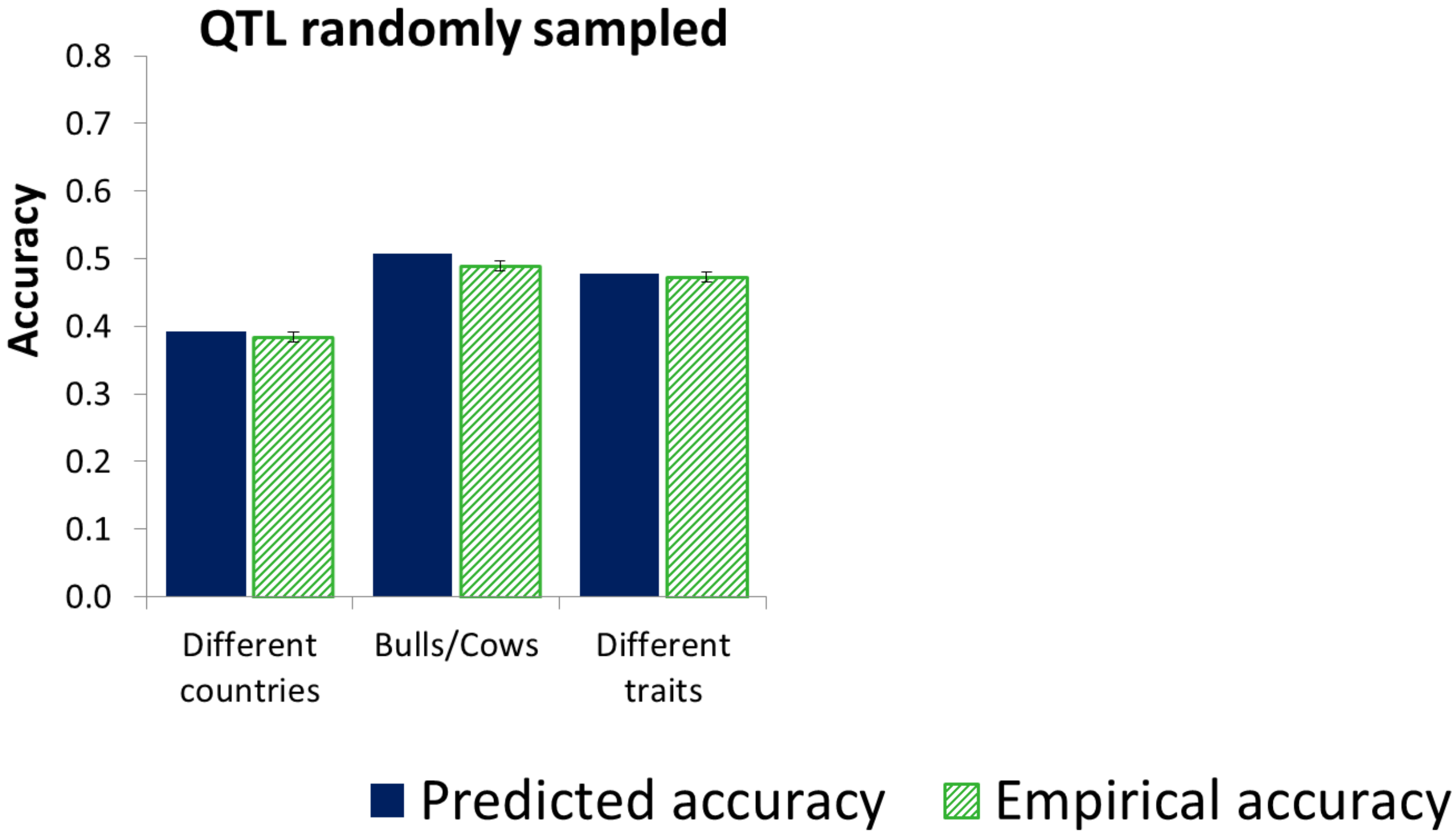
$r_{G_{A,B}}$

$r_{G_{A,C}}$

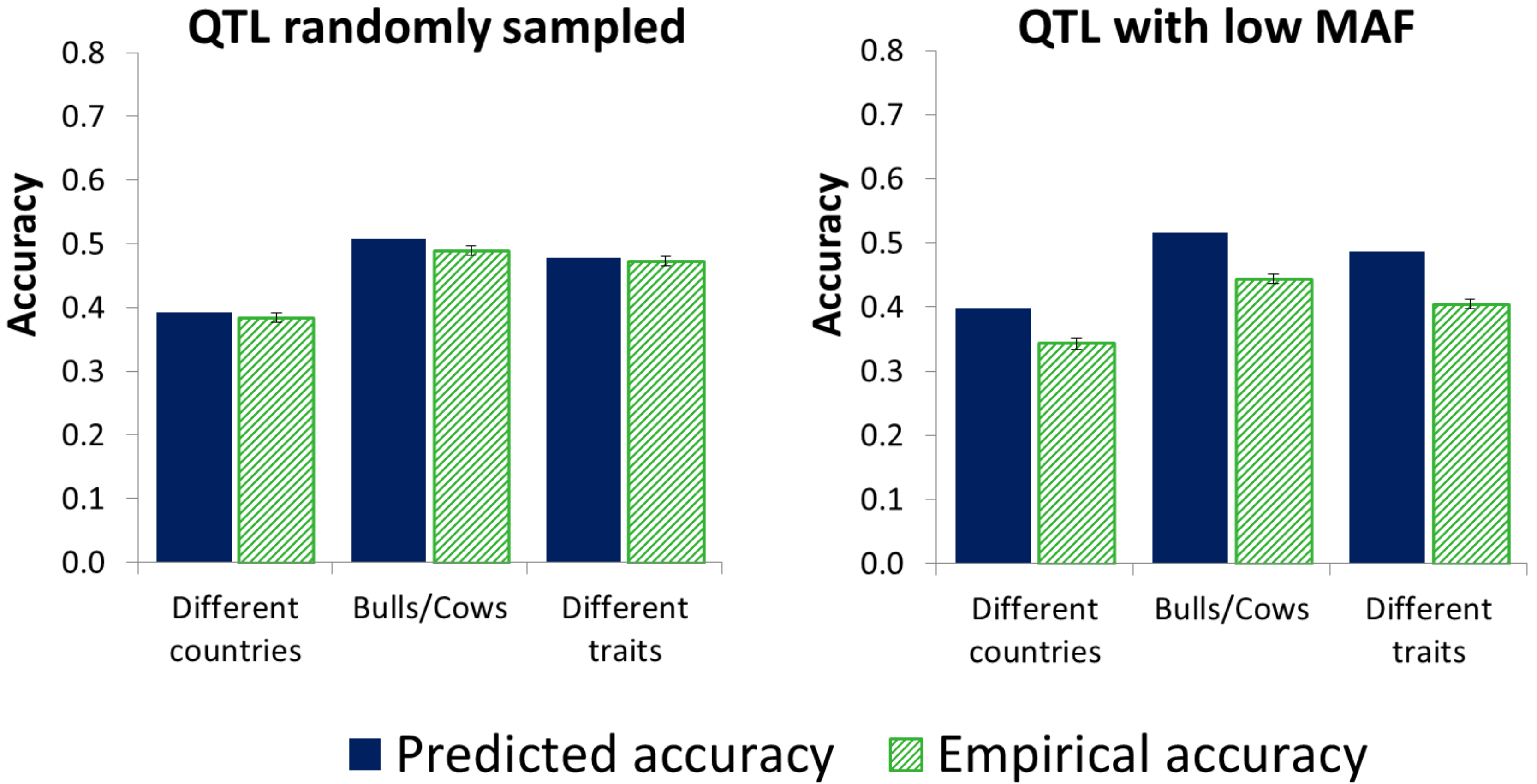
$r_{G_{B,C}}$

	Heritability (h^2)		Genetic correlation (r_G)		
Scenarios	**Pop. A**	**Pop. B**	**Pop. A - B**	**Pop. A – C**	**Pop. B - C**
Different countries	0.95	0.95	0.6	0.8	0.4
Bulls/cows	0.95	0.30	1.0	1.0	1.0
Different traits	0.95	0.30	0.6	1.0	0.6

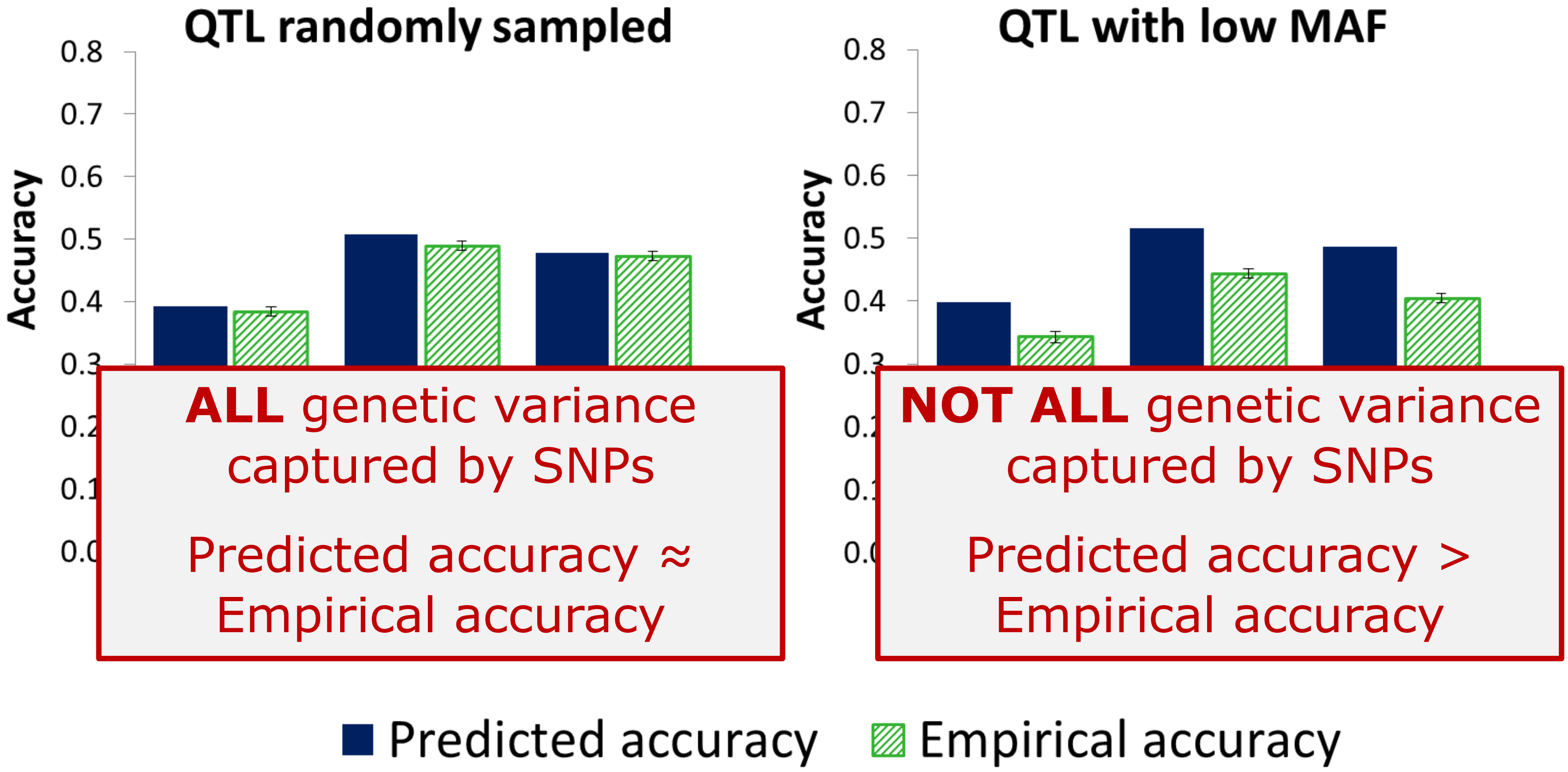
Results of validation



Results of validation



Results of validation



ALL genetic variance captured by SNPs

Predicted accuracy ≈ Empirical accuracy

NOT ALL genetic variance captured by SNPs

Predicted accuracy > Empirical accuracy

Conclusion

Equation accurately predicts GEBV accuracy

- Multiple populations:
    - Countries
    - Bulls/Cows
    - Traits
    - Breeds?
- All genetic variance captured

Conclusion

Equation accurately predicts GEBV accuracy

- Multiple populations:
    - Countries
    - Bulls/Cows
    - Traits
    - Breeds?

yvonne.wientjes@wur.nl

- All genetic variance captured

LIVESTOCK RESEARCH
WAGENINGEN UR

Thank you!