

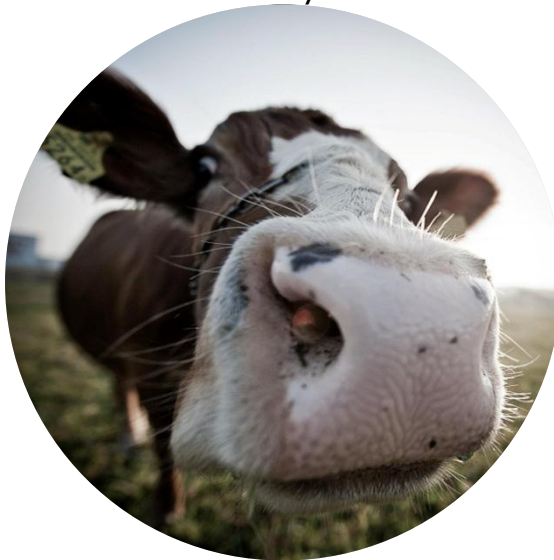
Reference population designs affects reliability of selection for (un)genotyped animals

Marcin Pszczola^{1,2,3}, Tomasz Strabel³, Johan van Arendonk², and Mario Calus¹

¹Wageningen UR Livestock Research, Animal Breeding and Genomics Centre, The Netherlands

²Wageningen University, Animal Breeding and Genomics Centre, The Netherlands

³Poznan University of Life Sciences, Department of Genetics and Animal Breeding, Poland



LIVESTOCK RESEARCH
WAGENINGEN UR



Poznań University of Life Sciences

In this talk:



Design of reference population & **relationship to** reference population **affects reliability** of breeding values



Average squared relationship to reference population – good **reliability proxy**



Genotyping selection candidates is more important

Acknowledgments

Organizers - EAAP SCHOLARSHIP

Faculty of Animal Breeding and Biology (PULS, Poland)

Koepon Stichting (Leusden, the Netherlands)



The GreenHouseMilk is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-238562. This publication represents the views of the authors, not the European Commission, and the Commission is not liable for any use that may be made of the information.

National Science Centre (Krakow, Poland); DEC-2011/01/N/NZ9/00614

Genomic selection - accuracy

Number of animals in the
reference population

Genomic selection



Traits difficult and/or expensive to measure



Methane emission



Feed intake



Longevity

Data

Simulation:



Genome: 3x1M; 1 SNP/0.001cM



Historic cattle population



Reference populations:



Highly, moderately, lowly and randomly related



Small (n=2,000)



Cows only



Selection candidates (n=1,000)

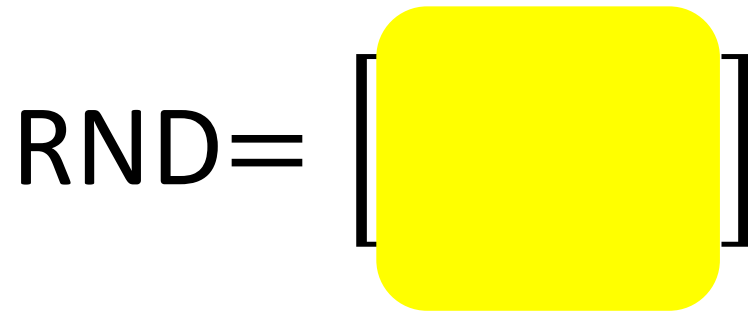
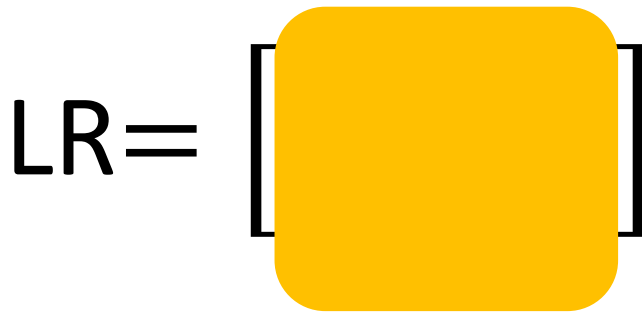
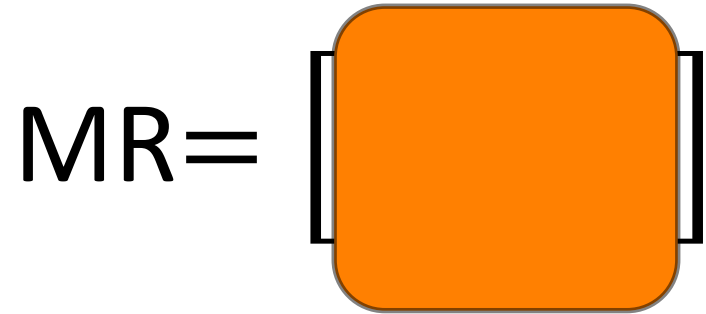
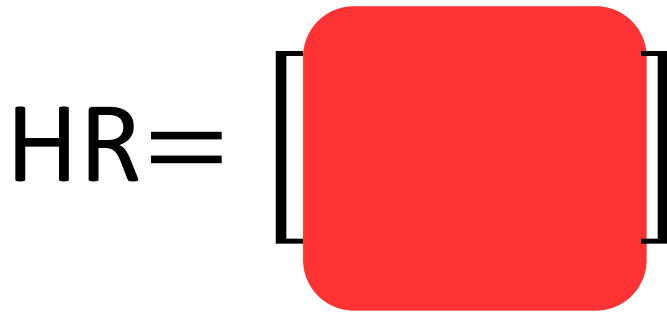


$h^2=0.3$


Reference populations' structure


Relationship within reference population	Number of	
	♂	♀
High (HR)	5	2000
Moderate (MR)	20	2000
Low (LR)	40	2000
Random (RND)	50	2000


Pedigree-based average relationship within the reference populations




Pedigree-based average relationship within the reference populations

HR = 

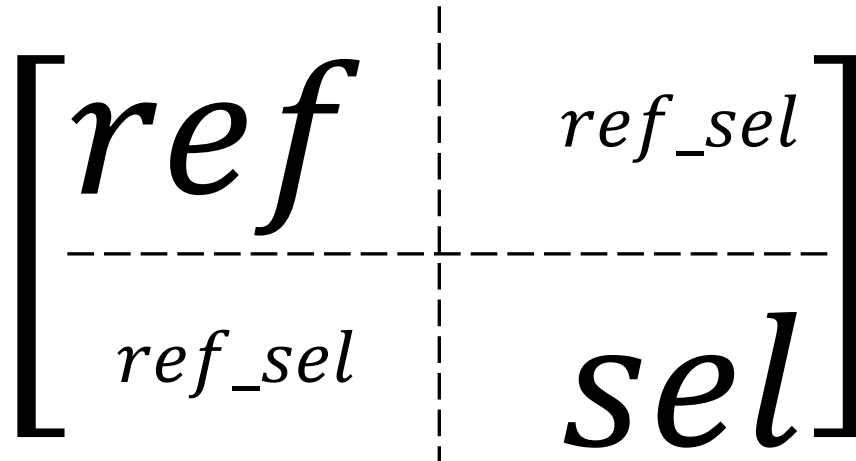
MR = 

LR = 

RND = 

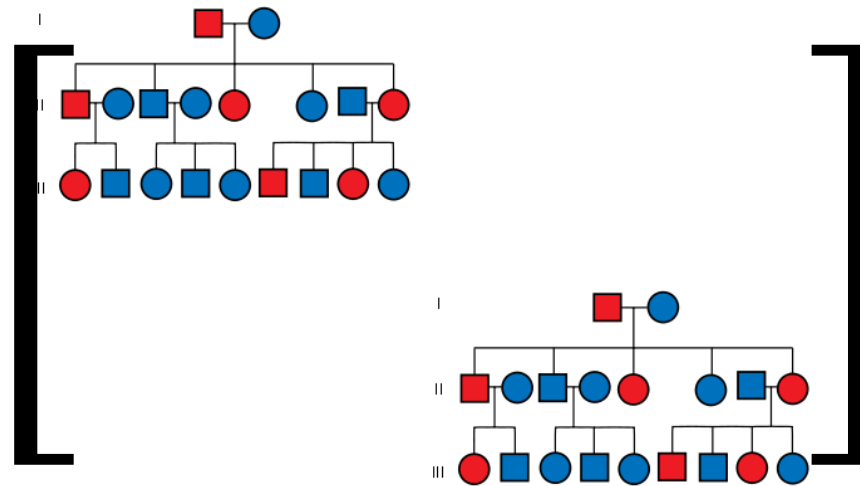
Scenarios

Different groups genotyped



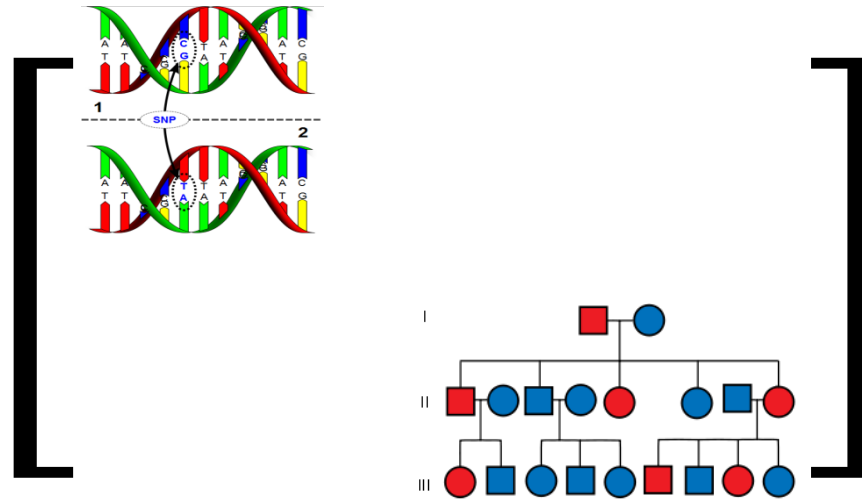
Scenarios

AA =



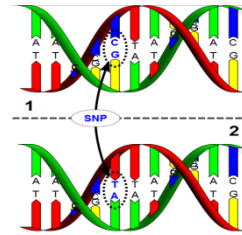
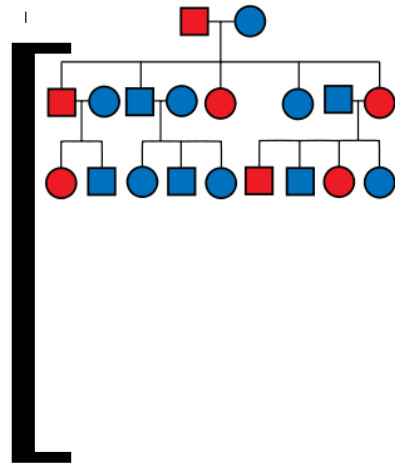
Scenarios

GA =



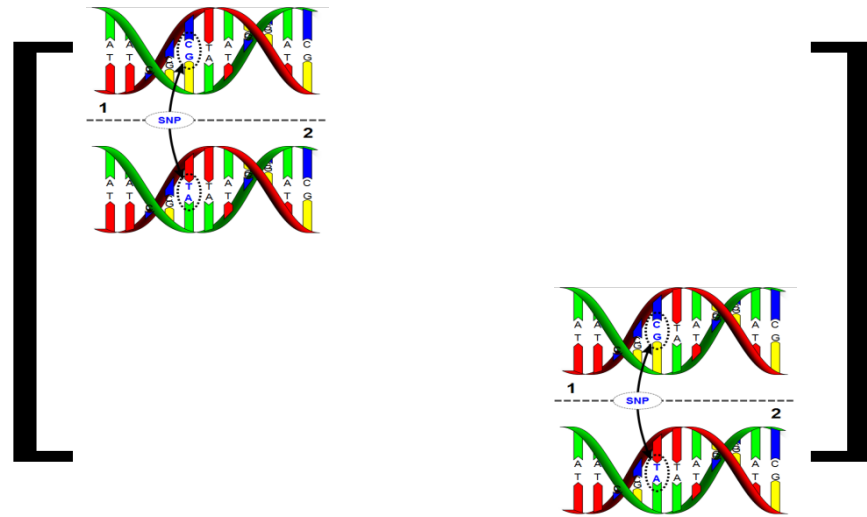
Scenarios

AG =



Scenarios

GG =



Scenarios

$$AA = \left[\begin{array}{c} \text{Pedigree} \\ \mathbf{A} \\ \text{Pedigree} \end{array} \right]$$

$$GA = \left[\begin{array}{c} \text{DNA} \\ \text{Pedigree} \end{array} \right]$$

$$AG = \left[\begin{array}{c} \text{Pedigree} \\ \text{DNA} \end{array} \right]$$

$$GG = \left[\begin{array}{c} \text{DNA} \\ \text{DNA} \end{array} \right]$$

Scenarios

$$AA = \left[\begin{array}{c} \text{Pedigree} \\ \text{A} \\ \text{Pedigree} \end{array} \right]$$

$$GA = \left[\begin{array}{c} \text{DNA} \\ \text{Pedigree} \end{array} \right]$$

$$AG = \left[\begin{array}{c} \text{Pedigree} \\ \text{DNA} \end{array} \right]$$

$$GG = \left[\begin{array}{c} \text{DNA} \\ \text{G} \\ \text{DNA} \end{array} \right]$$

Scenarios

$$AA = \left[\begin{array}{c} \text{[Pedigree Chart]} \\ \text{A} \\ \text{[Pedigree Chart]} \end{array} \right]$$

$$GA = \left[\begin{array}{c} \text{[DNA Helix]} \\ \text{H} \\ \text{[Pedigree Chart]} \end{array} \right]$$

$$AG = \left[\begin{array}{c} \text{[Pedigree Chart]} \\ \text{H} \\ \text{[DNA Helix]} \end{array} \right]$$

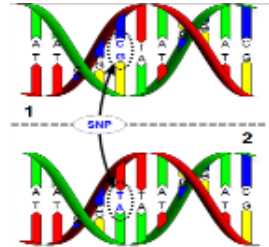
$$GG = \left[\begin{array}{c} \text{[DNA Helix]} \\ \text{G} \\ \text{[DNA Helix]} \end{array} \right]$$

G matrix:

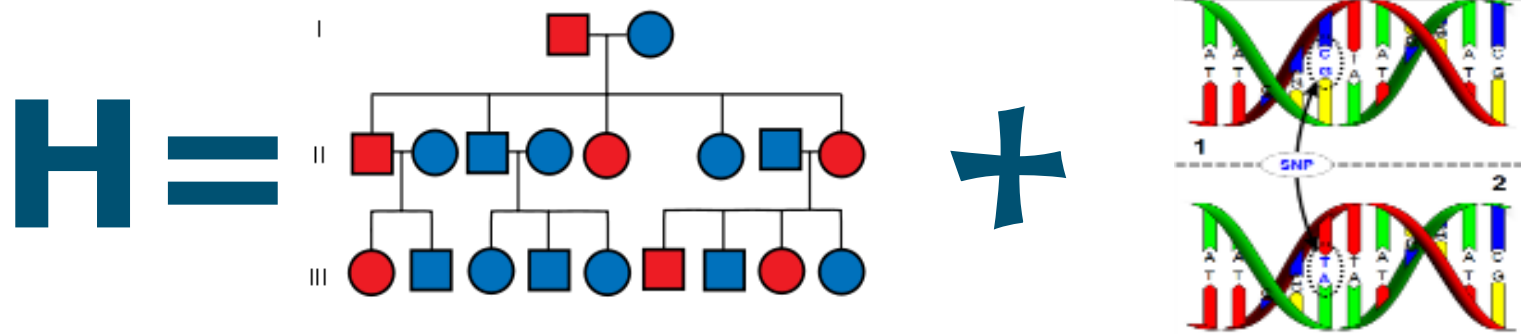
- VanRaden (2008)

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum p_i(1-p_i)}$$

- Current allele frequencies (p_i)



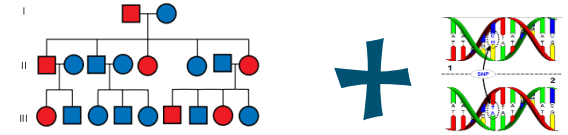
H matrix:



Legarra et al. (2009)

H matrix:

A & G compatibility:



- A & G on the same base pop.
 - Wright's F-statistics (Powell et al.)
- G regressed to A
 - Bins of relationships (0-0.10, >0.10-0.25, >0.25, and >0.50)
 - Parent-offspring & diagonals not regressed

Reliability

$$r^2 = \mathbf{C}_{ref_sel} \left[\mathbf{C}_{ref} + I \left(\frac{\sigma_e^2}{\sigma_a^2} \right) \right]^{-1} \mathbf{C}'_{ref_sel}$$

Reliability

$$r^2 = \mathbf{C}_{ref_sel} \left[\mathbf{C}_{ref} + I \left(\frac{\sigma_e^2}{\sigma_a^2} \right) \right]^{-1} \mathbf{C}'_{ref_sel}$$



h^2

Reliability

$$r^2 = \mathbf{C}_{ref_sel} \left[\mathbf{C}_{ref} + I \left(\frac{\sigma_e^2}{\sigma_a^2} \right) \right]^{-1} \mathbf{C}'_{ref_sel}$$


$$\mathbf{C} = \begin{bmatrix} \mathit{ref} & \mathit{ref_sel} \\ \mathit{ref_sel} & \mathit{sel} \end{bmatrix}$$

(A or G or H)

Reliability

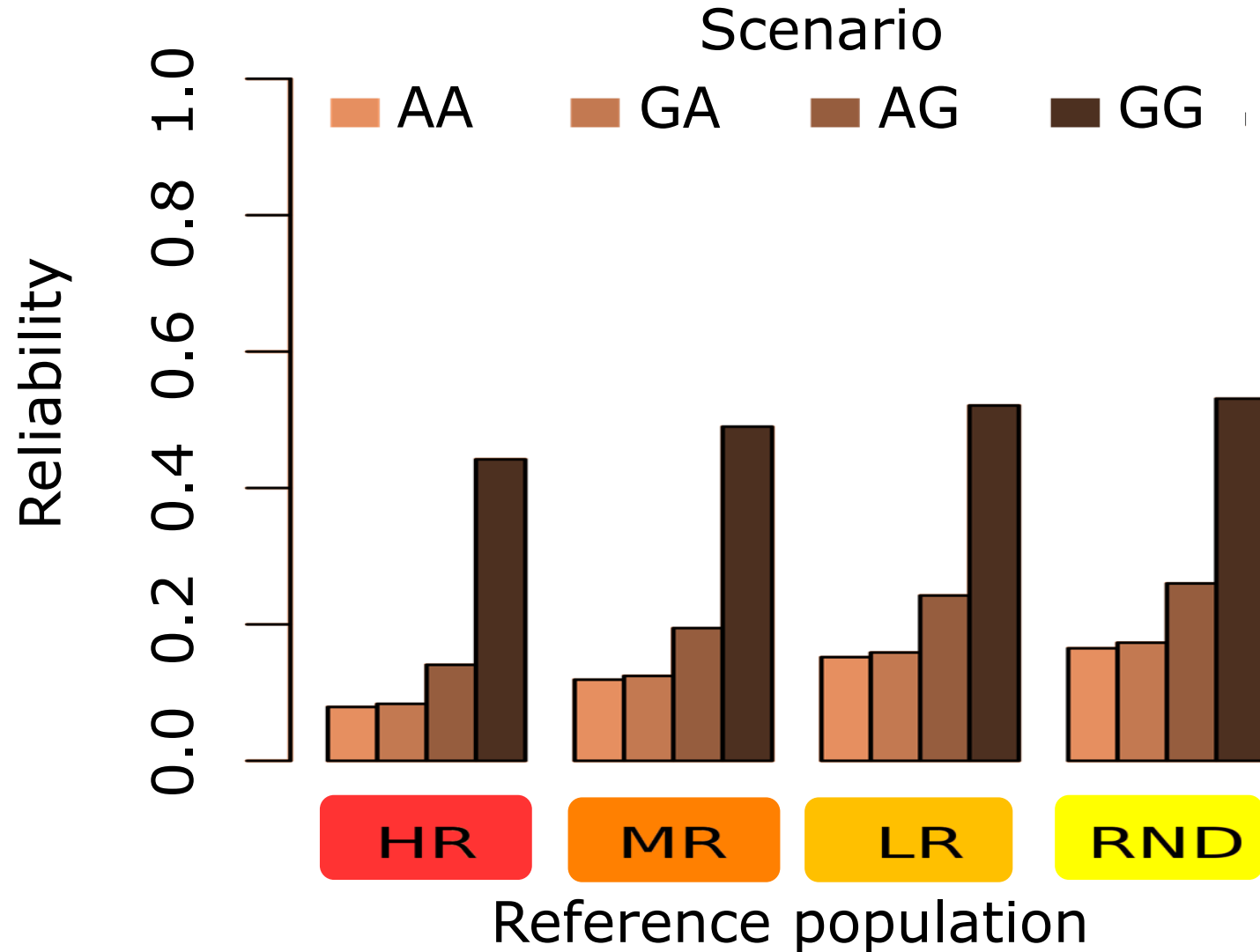
$$r^2 = \mathbf{C}_{ref_sel} \left[\mathbf{C}_{ref} + I \left(\frac{\sigma_e^2}{\sigma_a^2} \right) \right]^{-1} \mathbf{C}'_{ref_sel}$$

$$\mathbf{C} = \begin{bmatrix} \mathit{ref} & \mathit{ref_sel} \\ \mathit{ref_sel} & \mathit{sel} \end{bmatrix}$$

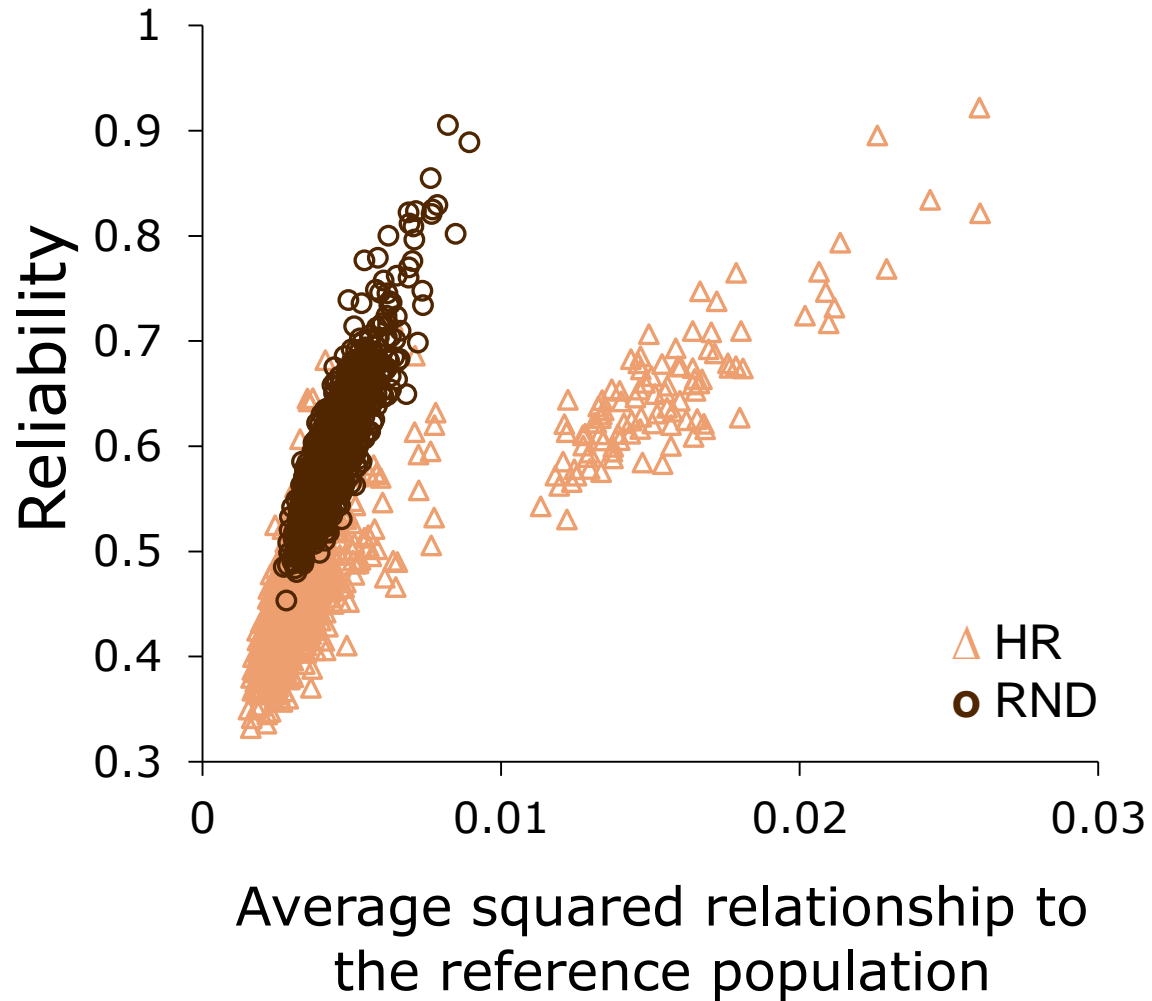
(A or G or H)

Average reliabilities of selection candidates

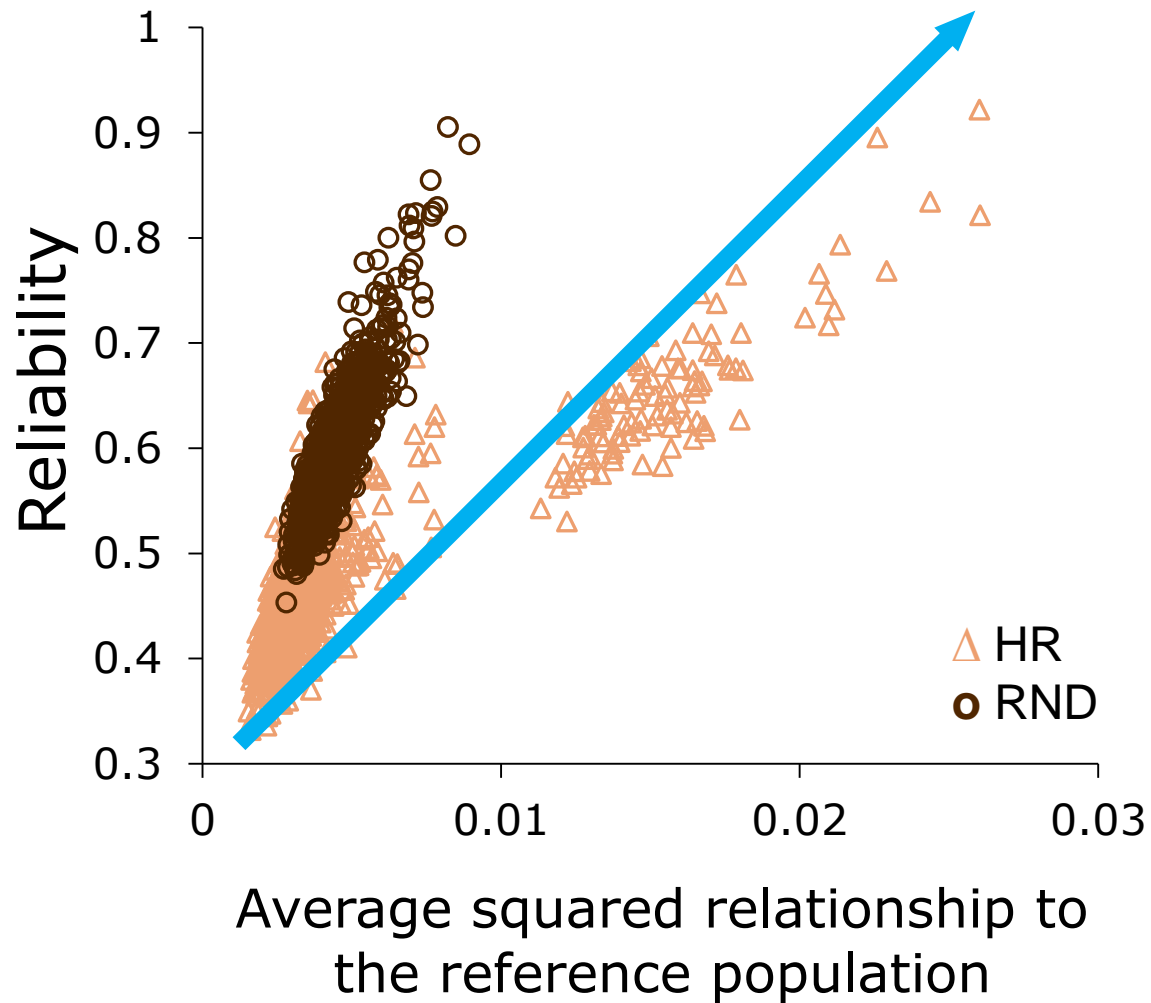
(n=1000) across differently designed reference populations for different (un)genotyped groups



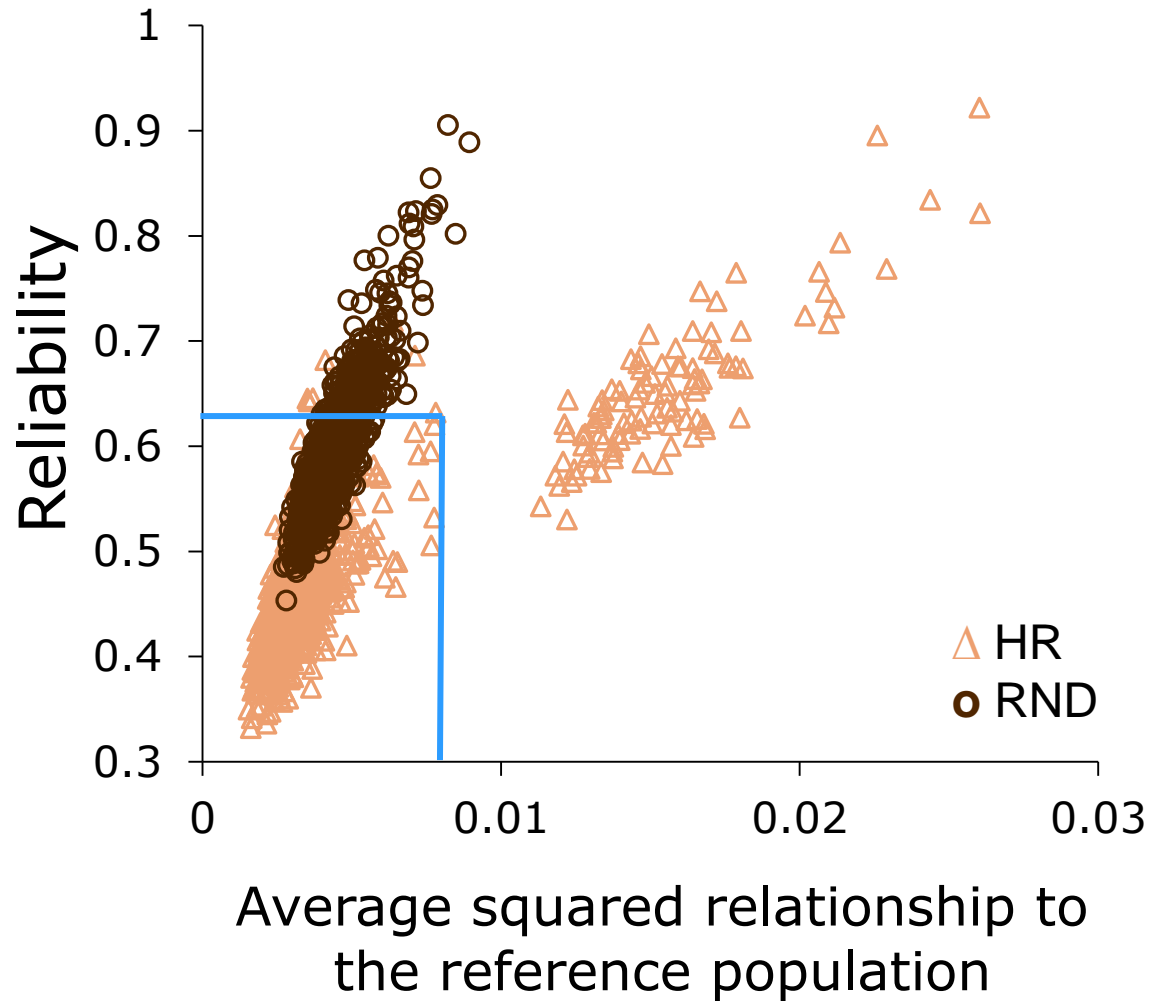
Individual reliabilities of selection candidates



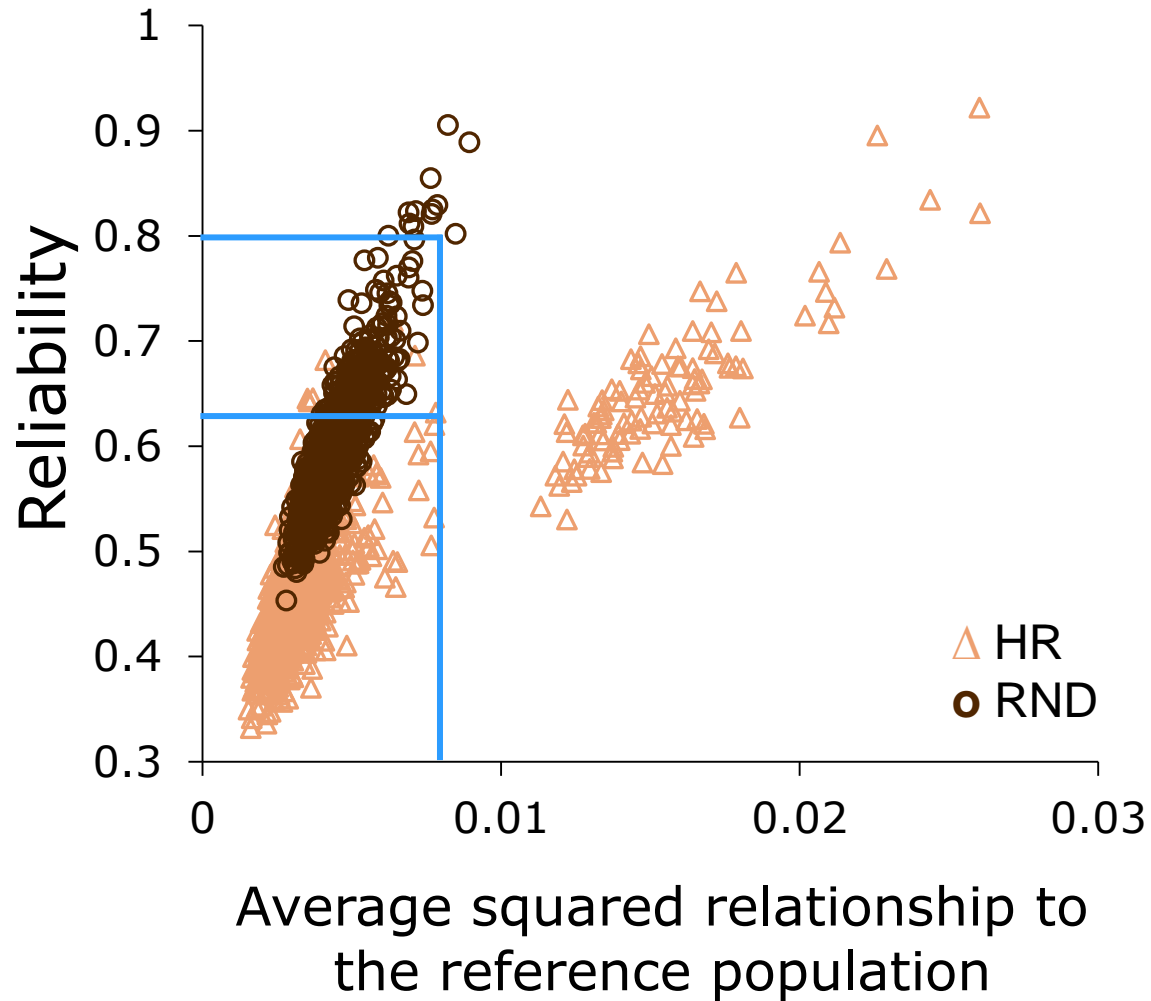
Individual reliabilities of selection candidates



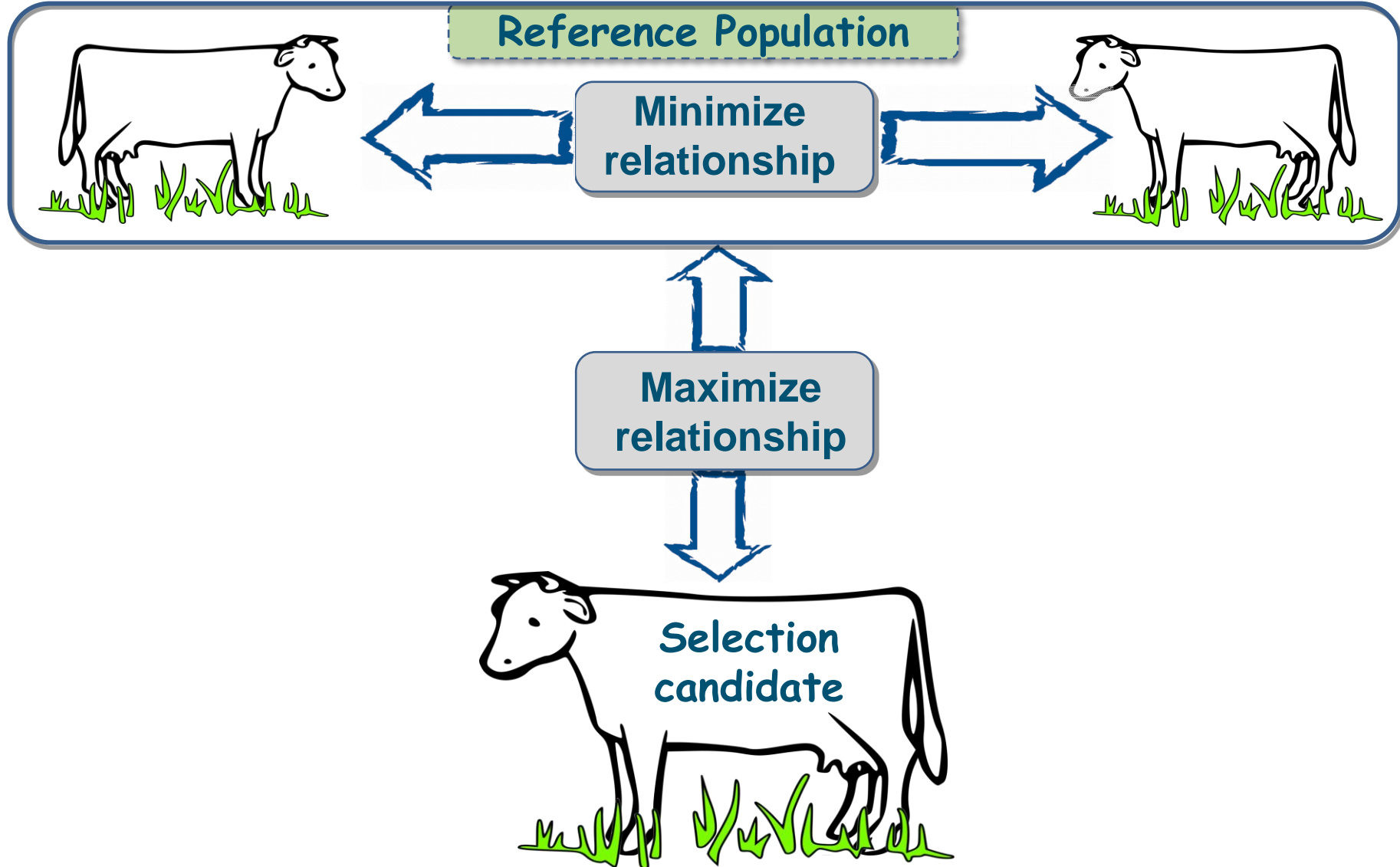
Individual reliabilities of selection candidates



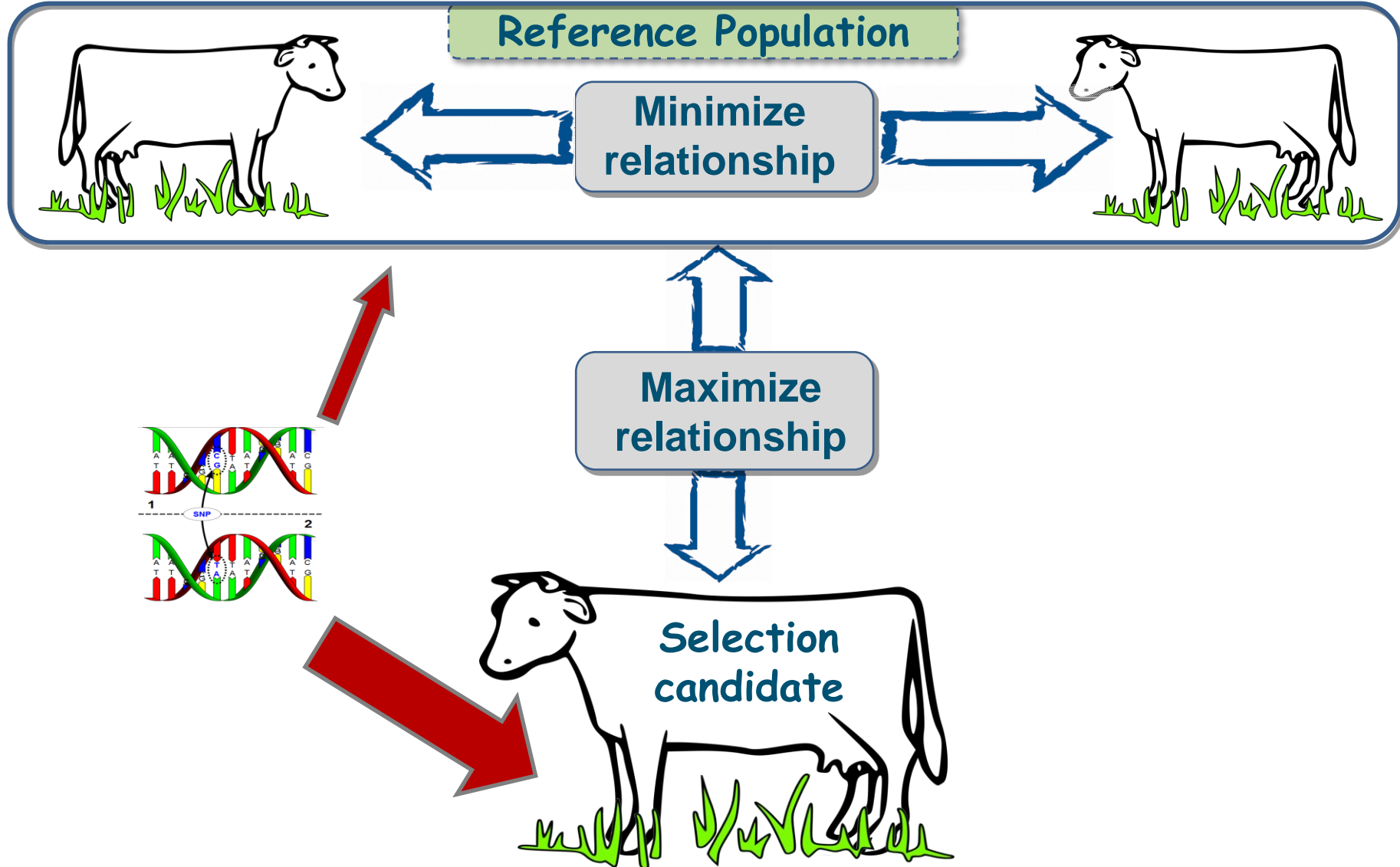
Individual reliabilities of selection candidates



Conclusions – Accuracy is increased by:



Conclusions – Accuracy is increased by:



Thank you for your attention!



Increasing GS
accuracy differently
than by increasing
the reference
population size is
(still) needed!



LIVESTOCK RESEARCH
WAGENINGEN UR

Marcin.Pszczola@wur.nl



Poznań University of Life Sciences