Effect of linkage disequilibrium, haplotypes and family relations on reliability of genomic prediction

Yvonne Wientjes (yvonne.wientjes@wur.nl)
Roel Veerkamp
Mario Calus

30 August, 2012

EAAP, Bratislava, Slovakia

LIVESTOCK RESEARCH
WAGENINGEN UR

CRV

Animal Breeding &
Genomics Centre

Acknowledgements

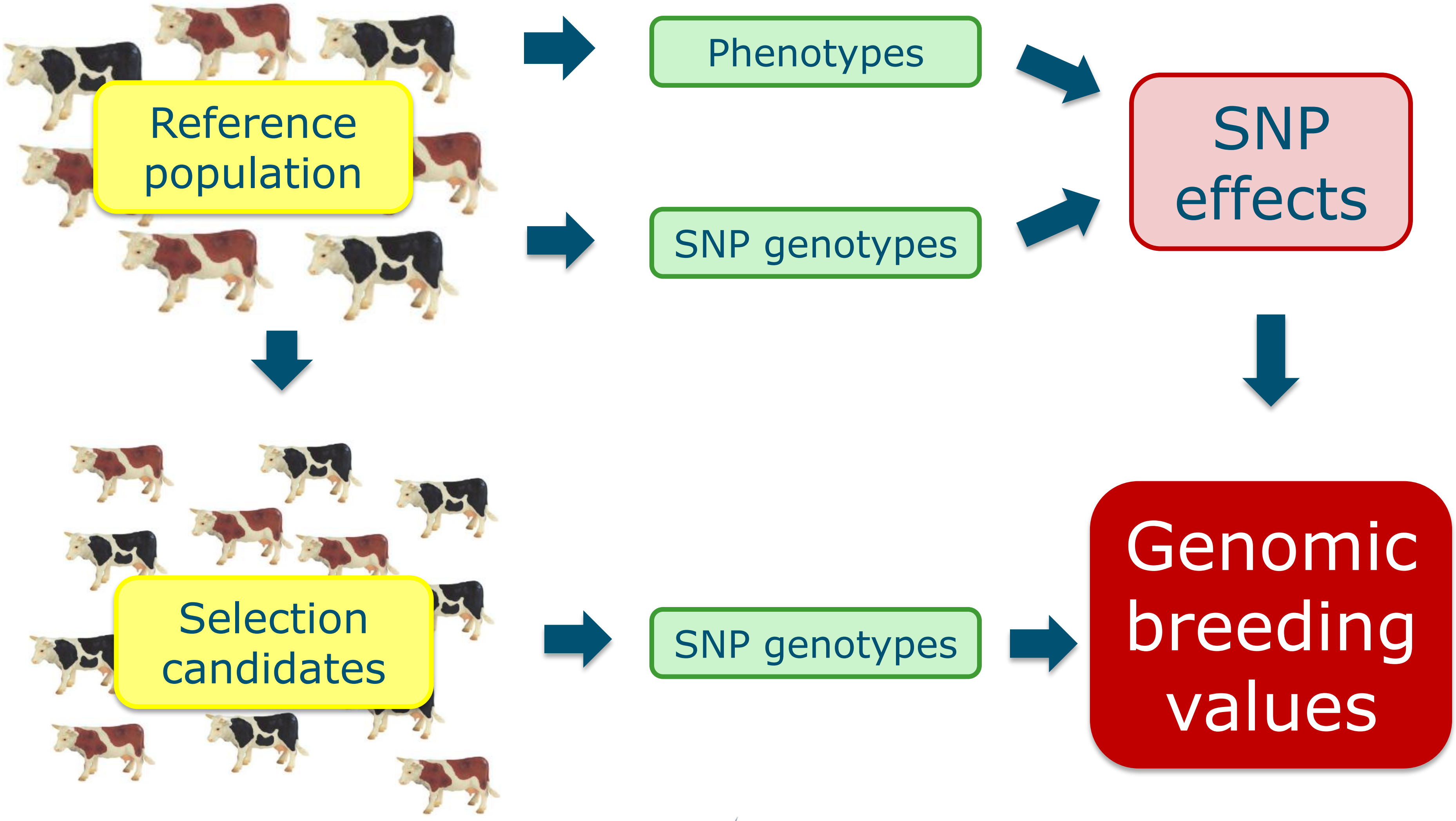
Discussions:

- Chris Schrooten (CRV B.V.)
- Henk Bovenhuis (WUR)

Financial support of

Introduction

Genomic selection



Reference population

Phenotypes

SNP genotypes

SNP effects

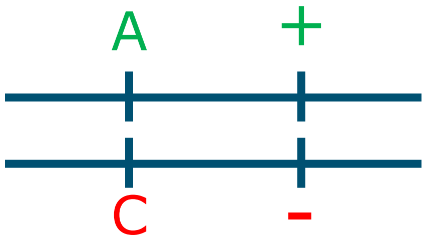
Selection candidates

SNP genotypes

Genomic breeding values

Reliability of genomic prediction

- Reliability of genomic breeding values due to:
  - Linkage disequilibrium between SNP and QTL (LD)
    - Meuwissen *et al.* (2001); Habier *et al.* (2007); Solberg *et al.* (2008); Habier *et al.* (2010)



  - Family relationships
    - Fernando (1998); Habier *et al.* (2007); Gianola *et al.* (2009); Habier *et al.* (2010)

Aim of this study

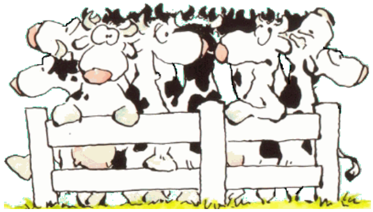
To investigate the **effects of LD** and **family relationships** on **reliability** of genomic prediction

Animal Breeding &
Genomics Centre

Materials & Methods

Reference population

- 529 cows
  - The Netherlands
- 35,002 SNPs
- Family relationships
  - 117 Parent-offspring relations
  - 48 Full-sib families, on average 2.27 animals
  - 69 paternal half-sib families, on average 7.23 animals
  - 65 maternal half-sib families, on average 2.65 animals



Selection candidates

- Simulate selection candidates
  - Four scenarios: FREQ, LD, HAP, FAM
  - 529 animals per scenario
- Information used from the reference population to simulate selection candidates:

	Allele frequencies	LD-pattern	Haplotypes	Family relationships
FREQ	X			
LD	X	X		
HAP	X	X	X	
FAM	X	X	X	X

Methods to predict reliability

VanRaden (2008)	**Daetwyler *et al.* (2008; 2010)**
G-BLUP	G-BLUP
Reliability **per individual**	**Mean** reliability of population
Selection index theory	Population parameters

Formula of VanRaden (2008)

$$r^2 = \mathbf{c}\left[\mathbf{G} + \mathbf{I}\left(\frac{\sigma_e^2}{\sigma_a^2}\right)\right]^{-1}\mathbf{c}'$$

G → Genomic relationship matrix of the **reference population**

c → Genomic relationship vector between **reference individuals** and one **selection candidate**

Formula of Daetwyler *et al.* (2008; 2010)

$$r^2 = \frac{N_p h^2}{N_p h^2 + M_e} \longrightarrow M_e = \frac{1}{Var(\boldsymbol{C} - \boldsymbol{A})}$$

Relationship matrix between **reference individuals** and **all selection candidates**:

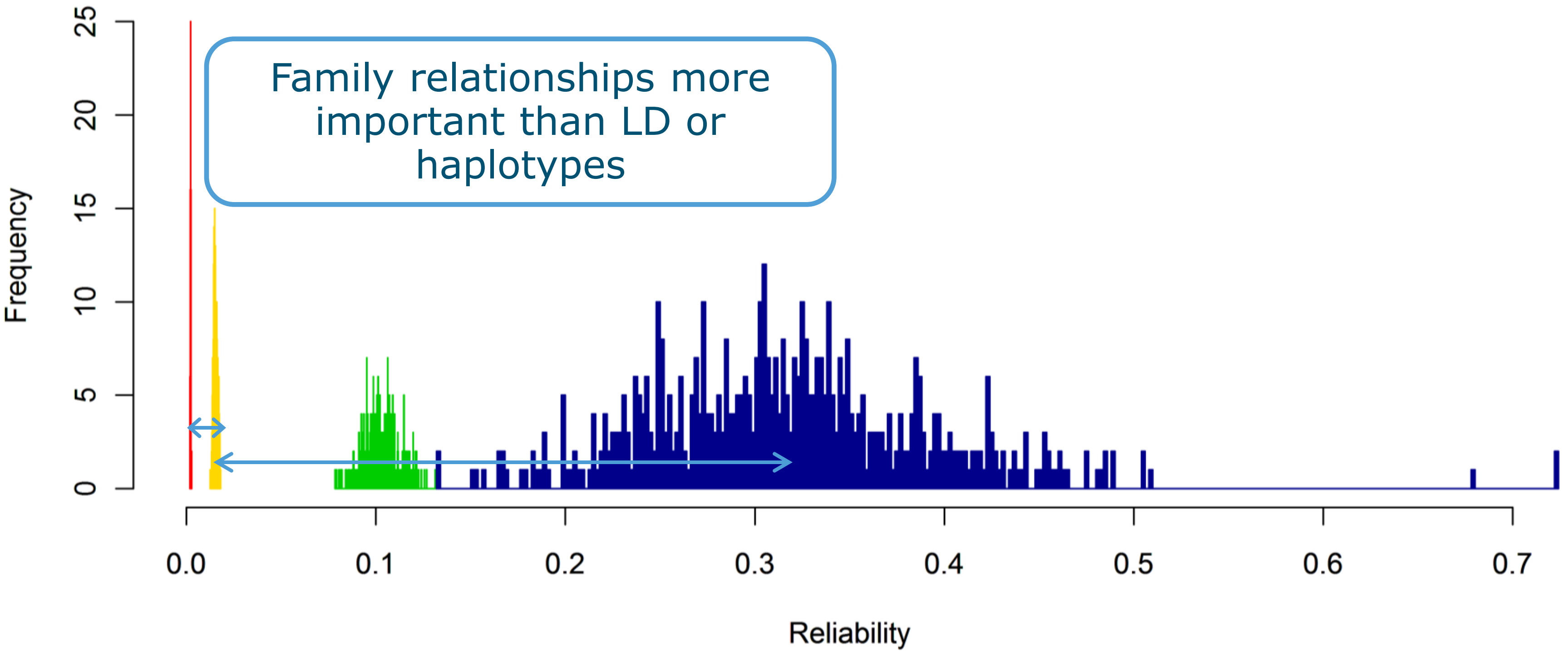
C → **Genomic**

A → **Additive genetic**

Results

Reliability per scenario – VanRaden (2008)

Heritability of 0.6

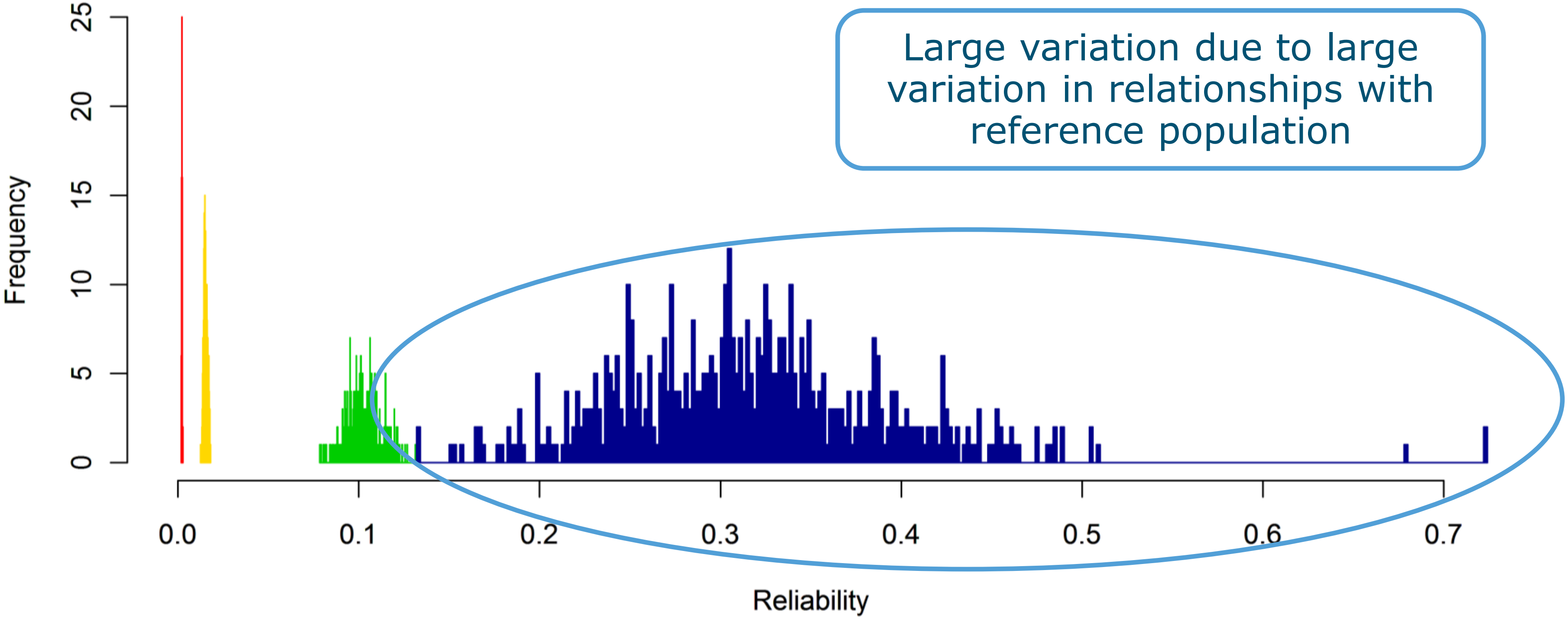


Family relationships more important than LD or haplotypes

🟥 = FREQ; 🟨 = LD; 🟩 = HAP; ■ = FAM

Reliability per scenario – VanRaden (2008)

Heritability of 0.6



Large variation due to large variation in relationships with reference population

🟥 = FREQ; 🟨 = LD; 🟩 = HAP; 🟦 = FAM

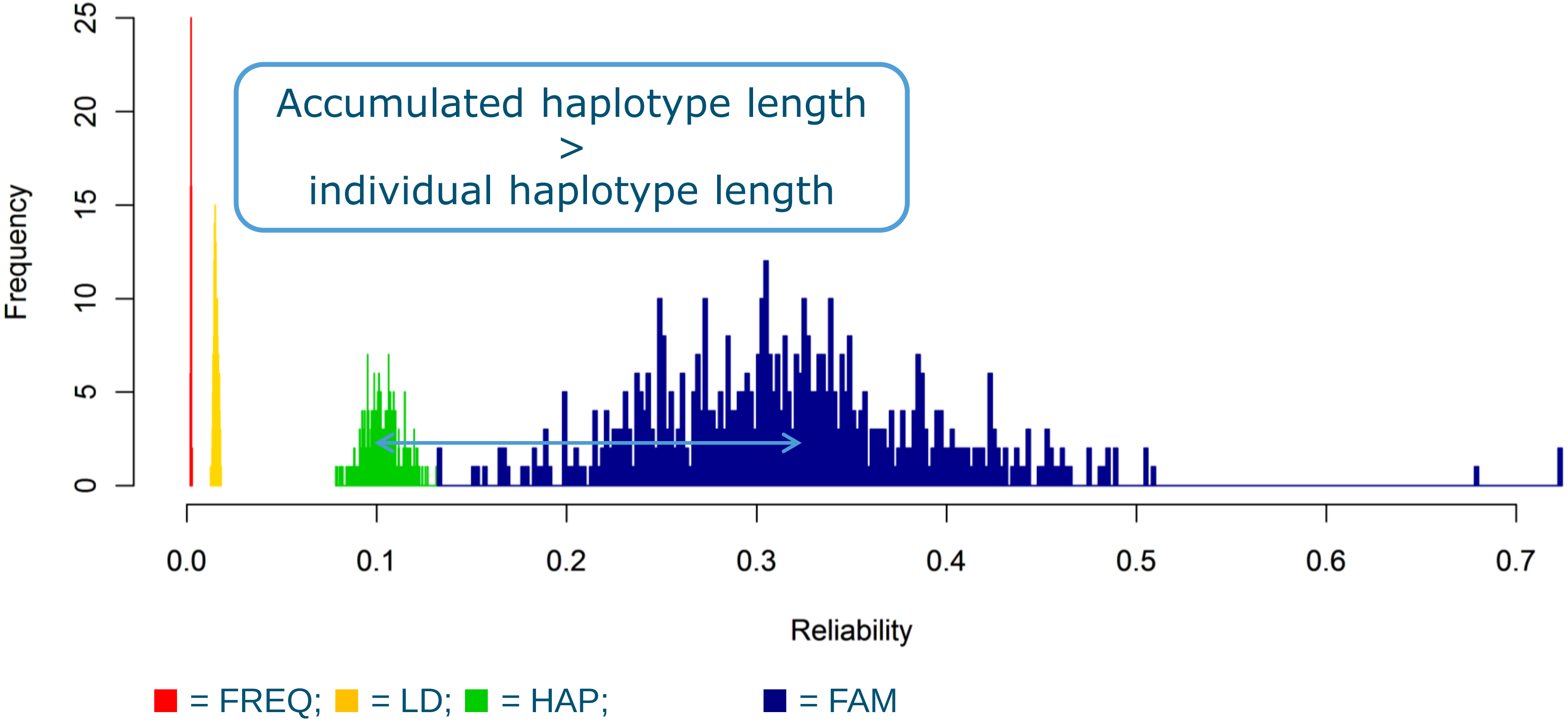
LIVESTOCK RESEARCH
WAGENINGEN UR

CRV

Animal Breeding &
Genomics Centre

Reliability per scenario – VanRaden (2008)

Heritability of 0.6



Accumulated haplotype length
\>
individual haplotype length

🟥 = FREQ; 🟨 = LD; 🟩 = HAP; 🟦 = FAM

Comparison of reliability from two formulas

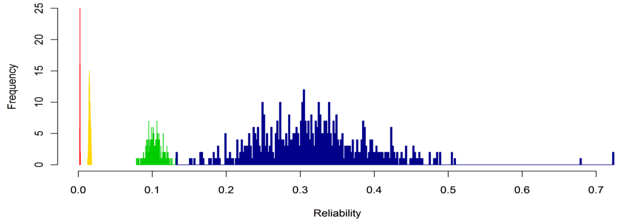
h^2	Scenario	M_e	VanRaden (2008)	Daetwyler *et al.* (2008)
0.6	**FREQ**	122,116	0.002	0.003
0.6	**LD**	16,292	0.015	0.019
0.6	**HAP**	2,064	0.102	0.133
0.6	**FAM**	837	0.318	0.275
0.1	**FREQ**	122,116	0.0004	0.0004
0.1	**LD**	16,292	0.003	0.003
0.1	**HAP**	2,064	0.021	0.025
0.1	**FAM**	837	0.104	0.059

≈

Conclusion

Conclusion

- Reliability of genomic prediction:



h^2	Scenario	VanRaden (2008)	Daetwyler *et al.* (2008)
0.6	**FREQ**	0.002	0.003
0.6	**LD**	0.015	0.019
0.6	**HAP**	0.102	0.133
0.6	**FAM**	0.318	0.275
0.1	**FREQ**	0.0004	0.0004
0.1	**LD**	0.003	0.003
0.1	**HAP**	0.021	0.025
0.1	**FAM**	0.104	0.059

≈

- **Family relationships** contribute more than **linkage disequilibrium** per se
- Effect of **accumulated length** of shared haplotypes **>** length of **individual haplotypes** shared
- Can be **predicted accurately** with prediction formula using M_e based on **genomic** and **additive genetic relationship matrices** for populations with **complex family structures**

Questions

Thanks!

Questions?

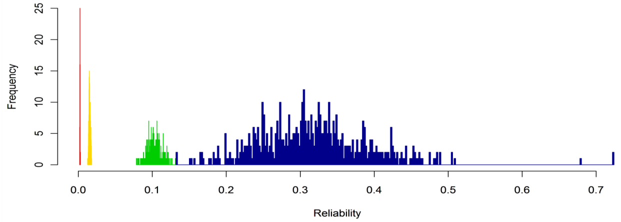
LIVESTOCK RESEARCH
WAGENINGEN UR

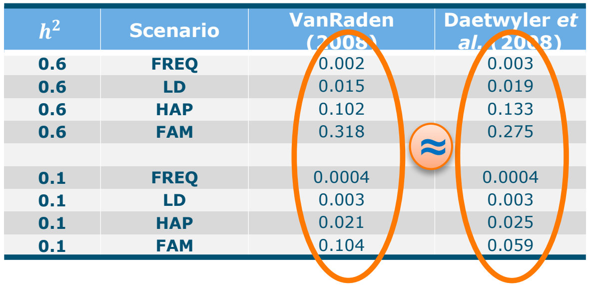
CRV

Animal Breeding &
Genomics Centre

Conclusion

- Reliability of genomic prediction:





h^2	Scenario	VanRaden (2008)	Daetwyler *et al.* (2008)
0.6	**FREQ**	0.002	0.003
0.6	**LD**	0.015	0.019
0.6	**HAP**	0.102	0.133
0.6	**FAM**	0.318	0.275
0.1	**FREQ**	0.0004	0.0004
0.1	**LD**	0.003	0.003
0.1	**HAP**	0.021	0.025
0.1	**FAM**	0.104	0.059

- **Family relationships** contribute more than **linkage disequilibrium** per se
- Effect of **accumulated length** of shared haplotypes **>** length of **individual haplotypes** shared
- Can be **predicted accurately** with prediction formula using M_e based on **genomic** and **additive genetic relationship matrices** for populations with **complex family structures**