Inclusion of cows in the reference population
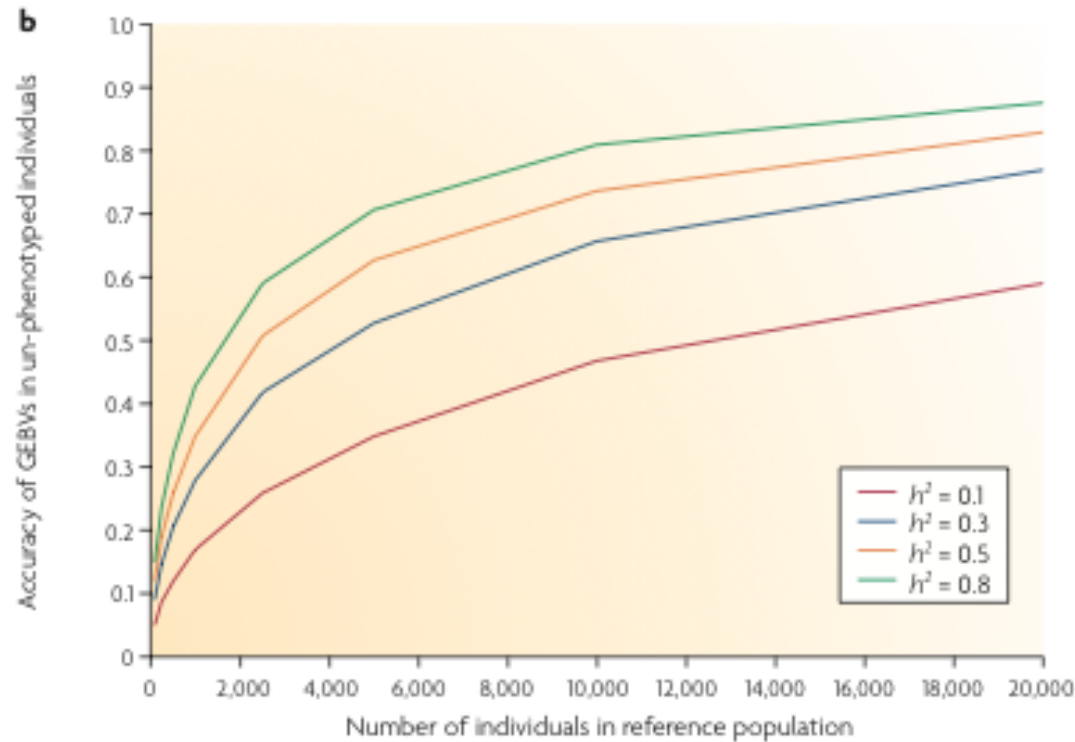
Beat Bapst, Christine Baes and Birgit Gredler
Qualitas AG
Swiss Animal Breeding Technology Platform (SABRE-TP)
12th of June, 2014

Why genotype cows? (I)

QUALITAS

(Goddard and Hayes, 2009)

- Current reference population is large; "a few" additional bulls do not affect reliabilities
- Searching for alternatives …

Why genotype cows? (II)

QUALITAS

- Population level
  - Enlarge reference population to increase accuracy of genomic selection
    → especially important for small breeds
  - Reveal population structure (e.g. for imputation)
  - Detect genetic defects (e.g. BH2 in Brown Swiss)

- Farm level
  - Selection at an early stage of life
  - Information for mating
  - Parentage verification
  - Detect genetic defects
  - Reveal genetic setup (e.g. polled animal: homozygous? heterozygous?)

Including cows in the reference population: state of the art

QUALITAS

- Only a few countries have included genotyped cows in the reference population:
  - US: HOL, BSW, JER
  - AUS: HOL, JER
  - Planned:
    Increase in AUS; DK: JER

- Only a few investigations based on real data are published regarding the reliability gain:
  - Pryce et al. (2013)
  - ……

Motivation (A)

QUALITAS

- Available in Switzerland: 1429 ♀ and 6309 ♂ BSW genotypes

- How much does reliability increase by including cows in our Swiss Brown Swiss reference population?
  ⇒ step A1

- Which cows should be genotyped?
  ⇒ step A2

Motivation (B)
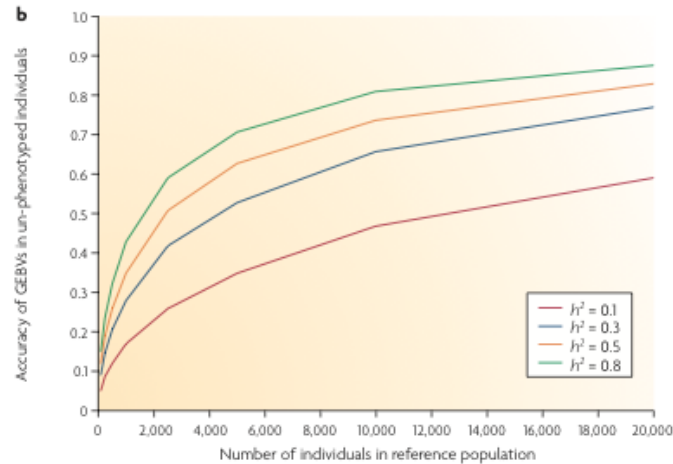
RESEARCH   **Open Access**

Imputation of non-genotyped individuals based on genotyped relatives: assessing the imputation accuracy of a real case scenario in dairy cattle

Aniek C Bouwman[1]*, John M Hickey[2], Mario PL Calus[1] and Roel F Veerkamp[1]

Abstract

Background: Imputation of genotypes for ungenotyped individuals could enable the use of valuable phenotypes created before the genomic era in analyses that require genotypes. The objective of this study was to investigate the accuracy of imputation of non-genotyped individuals using genotype information from relatives.

- Impute thousands of non-genotyped cows → include them in the reference population → Will the accuracy / reliability be affected?

⇒ **step B1**

Methods (A&B)

- Estimate marker effects and direct genomic values (dgv):
  - Predict breeding values from genomic data:
    Bayesian approach (BayesC), using the software GenSel
    (Fernando and Garrick, 2010)
  - Input ("phenotypic") data:
    Deregressed proofs (Garrick et al., 2009)

- Validation: accuracy: $r_{dgv,ebv}$

- Compute the genomic relationship matrix (gmatrix):
  VanRaden, 2008

- Compute the relationship coefficients:
  Pedig package (Boichard et al., 2010)

Methods (A&B): Validation

QUALITAS

- Forward prediction

- Random sample from bull subset (birth year 05 - 09 = youngest sires) → *validation set*
  old bulls + remaining part of the youngest sires → *reference population*
- Cows included in one run, one run without cows→ comparison of accuracies = 1 replication
- Replications:
  A: 120 for each version
  B: 30 for each version

Methods (B): Imputation

- Imputation of the full pedigree was done in 2 steps:

- FImpute (Sargolzaei et al. 2011)
  - Phasing of 9,826 50k genotyped bulls and cows
  - Haplotype library has been constructed based on these animals

- AlphaImpute (Hickey et al. 2012)
  - Imputation of the full pedigree
  - Haplotype library produced by FImpute was used for imputation
  - Results: Genotype probabilities for all animals in the pedigree

- Time-and computation intensive: Imputation of about 37,000 target cows has been split in 4 steps (time duration for one step: ~350 h)

Data (A): Number of geno- typed bulls and cows

QUALITAS

	Requirements	Origin	n
Bulls	Rel. production% $\geq$ 65	Braunvieh Schweiz	1650
Bulls	Rel. production% $\geq$ 65	bilateral exchange	2435
Bulls total			**4085**
Cows	Rel. production% $\geq$ 50	Braunvieh Schweiz elite cow program	264
Cows	Rel. production% $\geq$ 50	LIB project	972
Cows total			**1236**
Total			**5321**

Data (A): Different Strategies Composition of the reference populations

Strategy	Reference population	Milk kg			Somatic cell score		
		Bulls n	Cows n	Total n	Bulls n	Cows n	Total n
All	Bulls Rel % $\geq$ 65 Cows Rel % $\geq$ **50**	4085	1236	5321	3891	1162	5052
Cow55	Bulls Rel % $\geq$ 65 Cows Rel % $\geq$ **55**	4085	1189	5274	3891	1014	4905
Cow60	Bulls Rel % $\geq$ 65 Cows Rel % $\geq$ **60**	4085	1021	5106	3891	396	4287
Cow65	Bulls Rel % $\geq$ 65 Cows Rel % $\geq$ **65**	4085	243	4328	3891	45	3936

Data (B): Different Strategies Composition of the reference populations

QUALITAS

| Stra-tegy	Reference Population	Milk kg			Somatic cell score		
		Bulls n	Cows n	Total n	Bulls n	Cows n	Total n
C $\emptyset\,a_{xy}\,max$	Bulls Rel% ≥ 65 Cows Rel% ≥ 60	4914	5000	9914	4691	5000	9691
D $\emptyset\,a_{xy}\,min$	Bulls Rel% ≥ 65 Cows Rel% ≥ 60	4914	5000	9914	4691	5000	9691
E $max\,a_{xy}\,max$	Bulls Rel% ≥ 65 Cows Rel% ≥ 60	4914	5000	9914	4691	5000	9691
F $max\,a_{xy}\,min$	Bulls Rel% ≥ 65 Cows Rel% ≥ 60	4914	5000	9914	4691	5000	9691

a_{xy}= additive relationship coefficient

C: Average a_{xy} between cows und existing reference bulls is maximal (Pszczola et al., 2013)
D: Average a_{xy} between cows und existing reference bulls is minimal (Pszczola et al., 2013)
E: Maximum a_{xy} between cows und existing reference bulls is maximal (Habier et al., 2010)
F: Maximum a_{xy} between cows und existing reference bulls is minimal (Habier et al., 2010)

Results A1: Including geno-typed cows; Different reliability levels; Milk kg
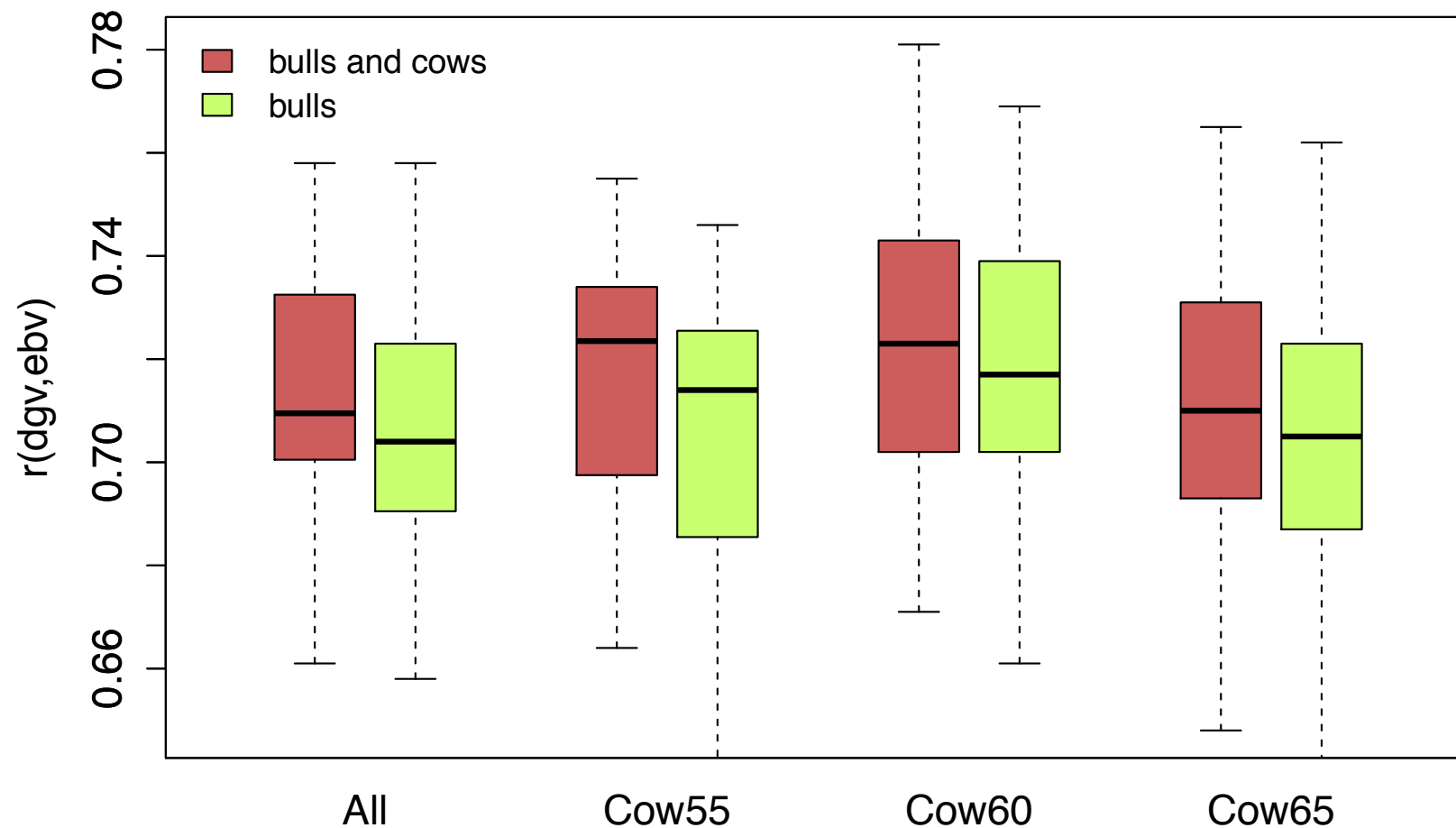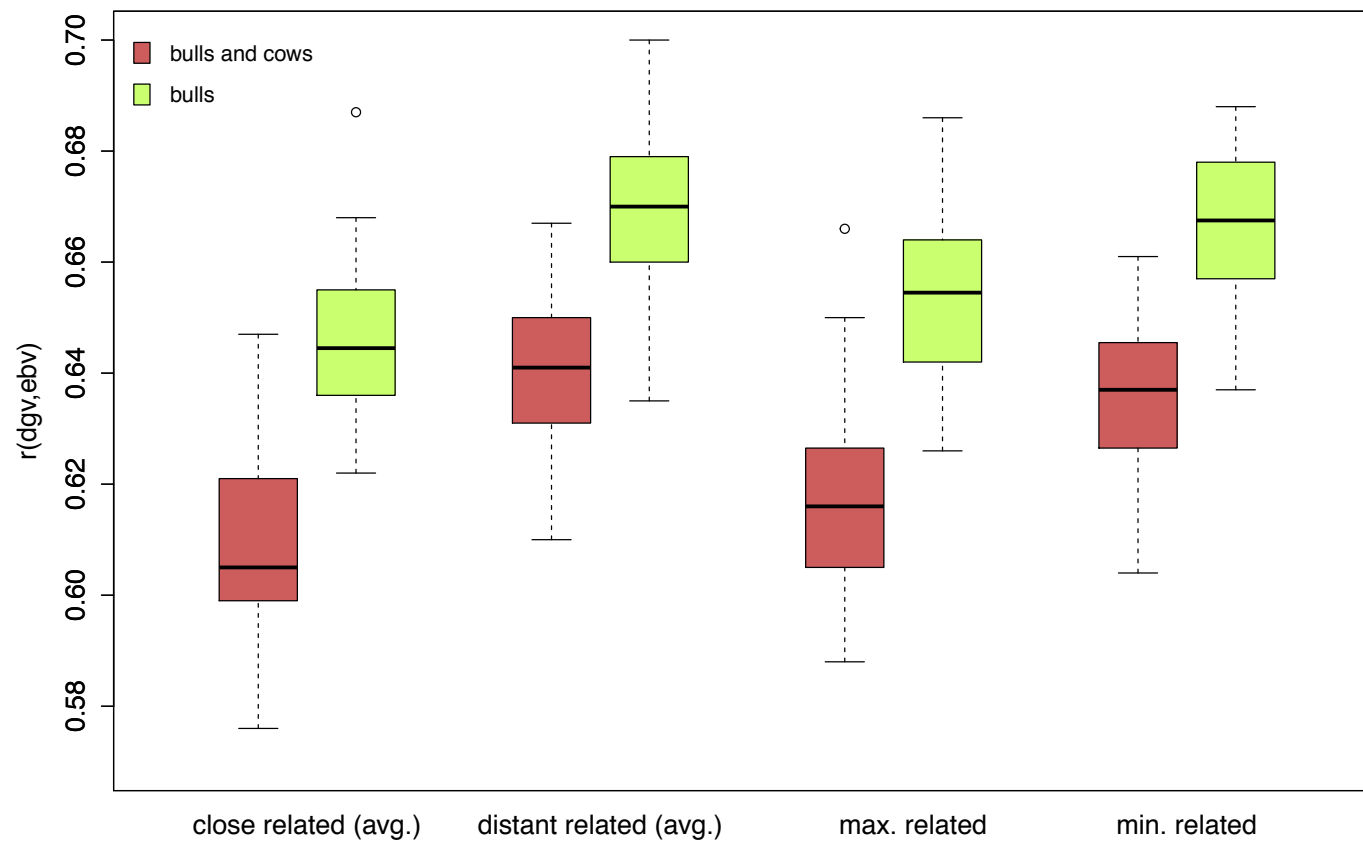
Results A1: Including geno-typed cows; Different reliability levels; SCS
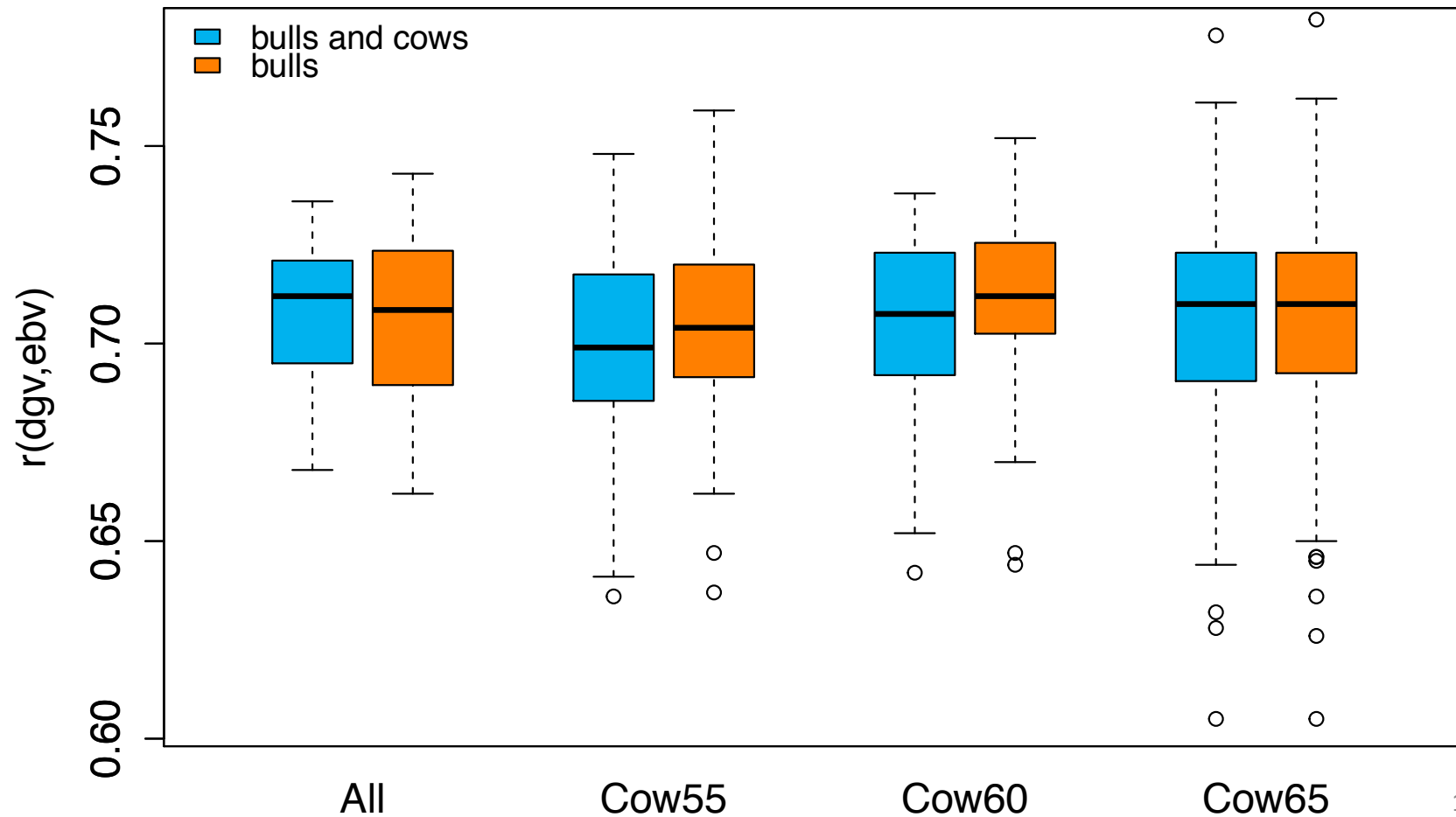
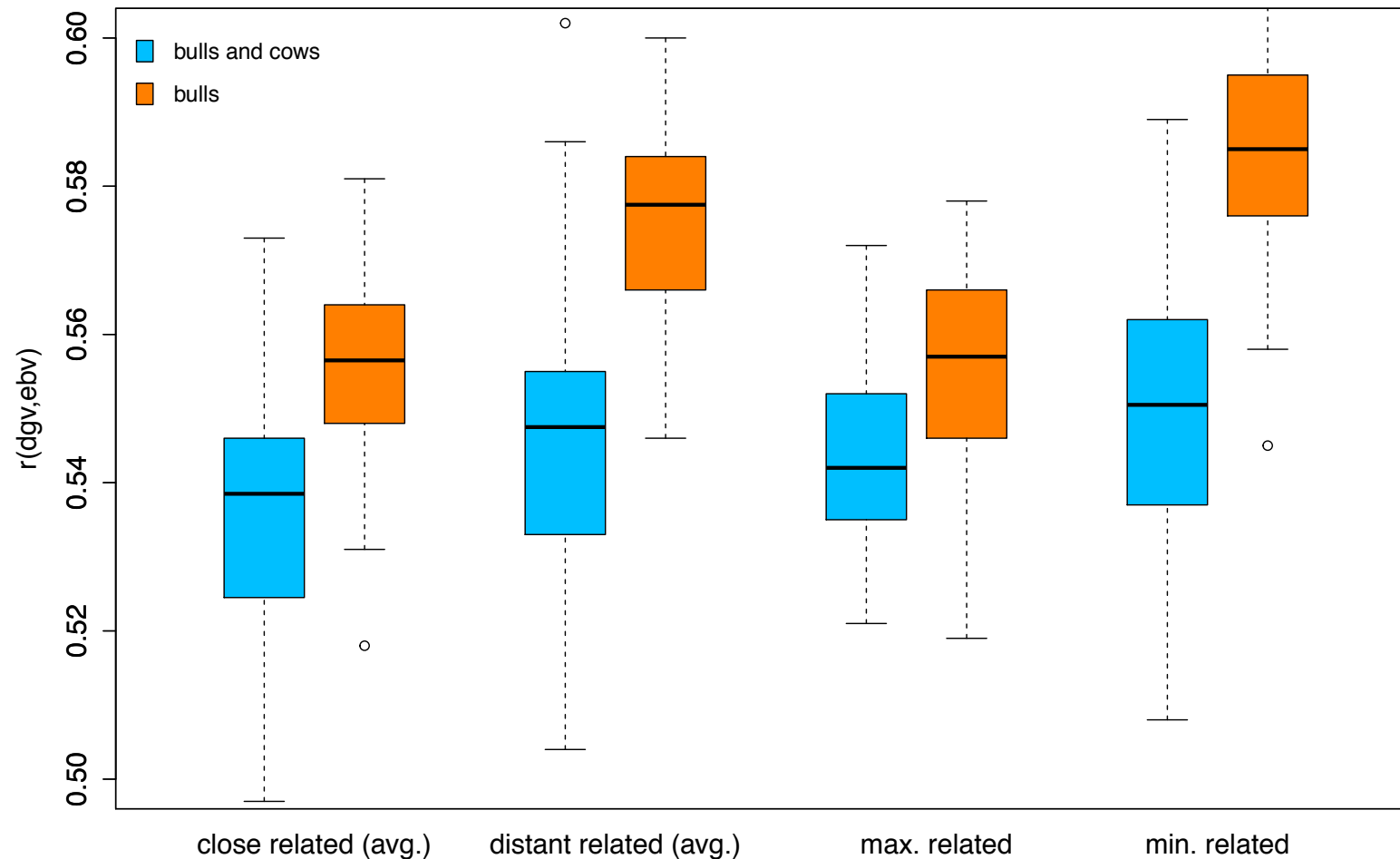# Results B1: : Including ungeno-typed imputed cows; Different relationship levels; Milk kg (first results)

Results B1: : Including ungeno-typed imputed cows; Different relationship levels; SCS (first results)

Discussion and outlook (I)

QUALITAS

- The effect of including 1000 cows in the existing CHE BSW reference population (about 4100 bulls) is small → not enough cows
  - Pryce et al. (2013): 3000 HOL bulls + 10000 HOL cows in AUS → +4-8% reliability gain
- Trait dependent
- Composition of the bull and cow set in the reference population seems to have an influence → important to genotype/include the appropriate cows
- Covering the whole genetic diversity of the population (Pszczola et al., 2012)

Discussion and outlook (II)

- Possibly the inclusion of non-genotyped imputed cows does not really work
  (first intermediate results)

- Reasons …
  - Reliability of cow ebvs too low?
  - Composition of the reference population?
    - Proportion bulls – cows
    - Relationships between bulls - cows
  - Imputation accuracy not high enough?

- Exploring the reasons

Discussion and outlook (III)

QUALITAS

- Genotyping cows must go on ☺ (cost/benefit)
    - The number of bulls with accurate ebvs is limited
    - Explore female specific traits
    - …see introduction

- Exchange cow genotypes (phenotypes: cow ebv exchange)

- Vision: Genotyping should be so cheap that herdbook registration requires a valid genotype

19

Thank you for your attention!

QUALITAS