

Multi-trait genomic selection - comparison of methods

Mario Calus (mario.calus@wur.nl)

Roel Veerkamp

Session 28 – Abstract nr. 4027

Animal Breeding & Genomics Centre



ANIMAL SCIENCES GROUP
WAGENINGENUR

Animal Breeding &
Genomics Centre

Introduction I

- Genomic selection is becoming common practice in animal breeding
- Key point is prediction of genomic breeding values (GEBV) using a reference population
- Only single trait implementations have been reported



Introduction II

- In 'classical' breeding value estimation, multi-trait (MT) application was breakthrough
- MT allows use of indicator traits to increase reliability of hard to measure or low heritability traits
- Can we implement MT genomic breeding value estimation?



Objectives

- Develop different methods to estimate genomic breeding values in a MT model
- Compare accuracy of GEBV obtained from different MT models



Four different MT models¹ were applied

Name	Model	Modelling of SNP variances
A	Polygenic using pedigree based rel. matrix	SNP not included
GRM	Polygenic using marker based rel. matrix	Equal for all SNP
BayesA	Effects are estimated for each SNP	Drawn from 1 distribution
BayesC	Effects are estimated for each SNP	Drawn from 2 distributions ²

¹ Variances are estimated in all models simultaneously with the effects

² One distribution for SNP that are (not) associated with a QTL

Implementation of MT BayesC

- Early implementation was unstable for 'unequal' design (some reference animals do not have phenotypes for all traits)
- Used implementation involves canonical transformation using an EM step to predict unknown phenotypes for reference animals



Simulation

- 5.655 SNPs / 5 M / 10 chrom. (11.3 SNPs / cM)
- r^2 between adjacent SNPs was 0.32
- 200 QTL equally spaced across the genome
- QTL effects drawn from multivariate normal distribution
- Two traits: $h^2(\text{tr. 1}) = 90\%$ & $h^2(\text{tr. 2}) = 60\%$
- 3 genetic correlations (r_g): 0.2, 0.5 & 0.8



Simulation ('unequal' design)

After 1000 generations ($N_e = 500$) to generate LD:

Generation	# animals	trait 1 ($h^2 = 0.9$)	trait 2 ($h^2 = 0.6$)
1	500	Phen.	Phen. / Unphen.
2	500	Phen.	Phen. / Unphen.
3	500	Unphen.	Unphen.
4	500	Unphen.	Unphen.

=> Reference population is 1000 (tr 1) & 500 (tr 2)



Results (average of 10 replicates)

Accuracy trait 1 ($h^2 = 0.9$):

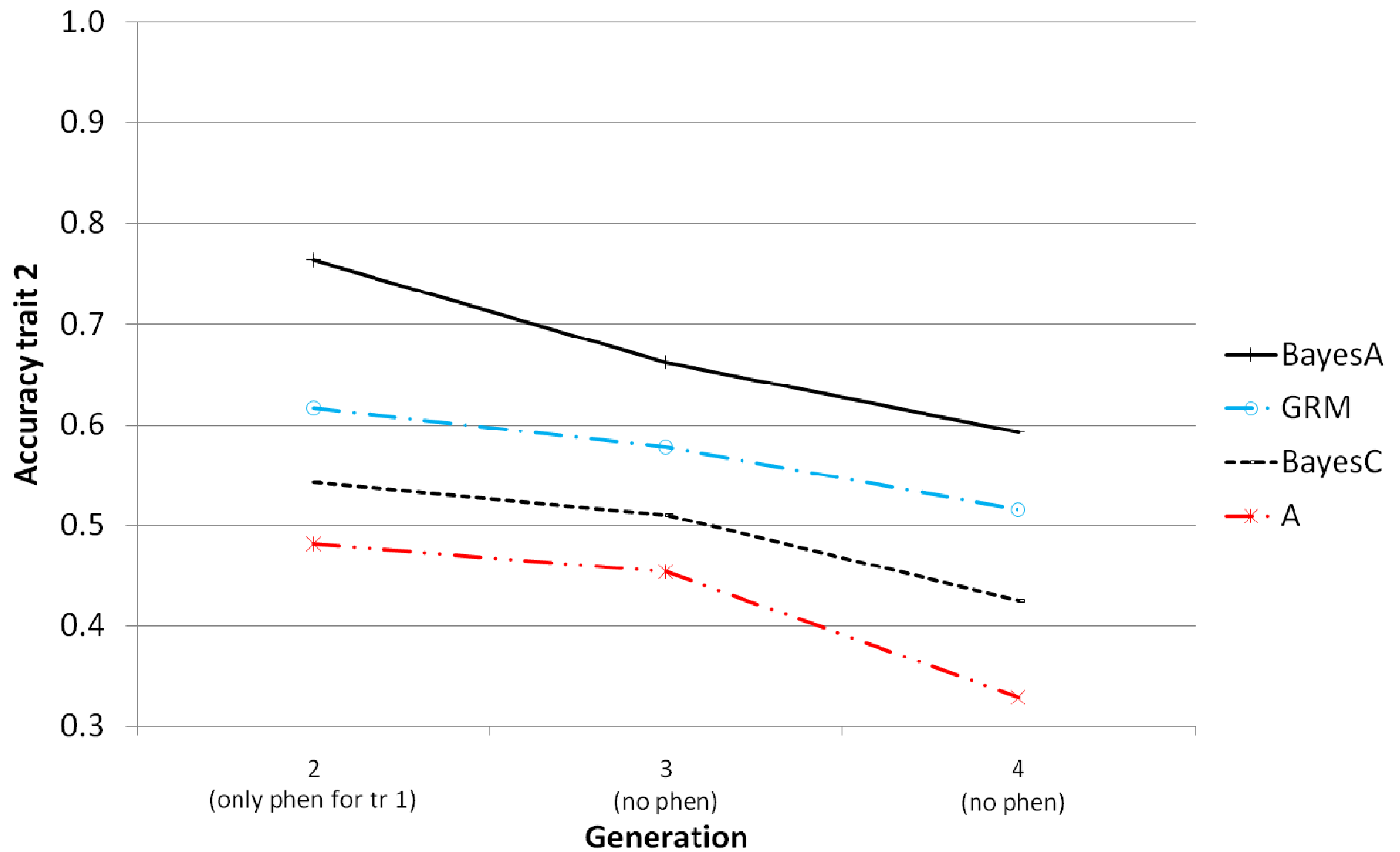
- Differences between models very small (not shown)

Accuracy trait 2 ($h^2 = 0.6$):

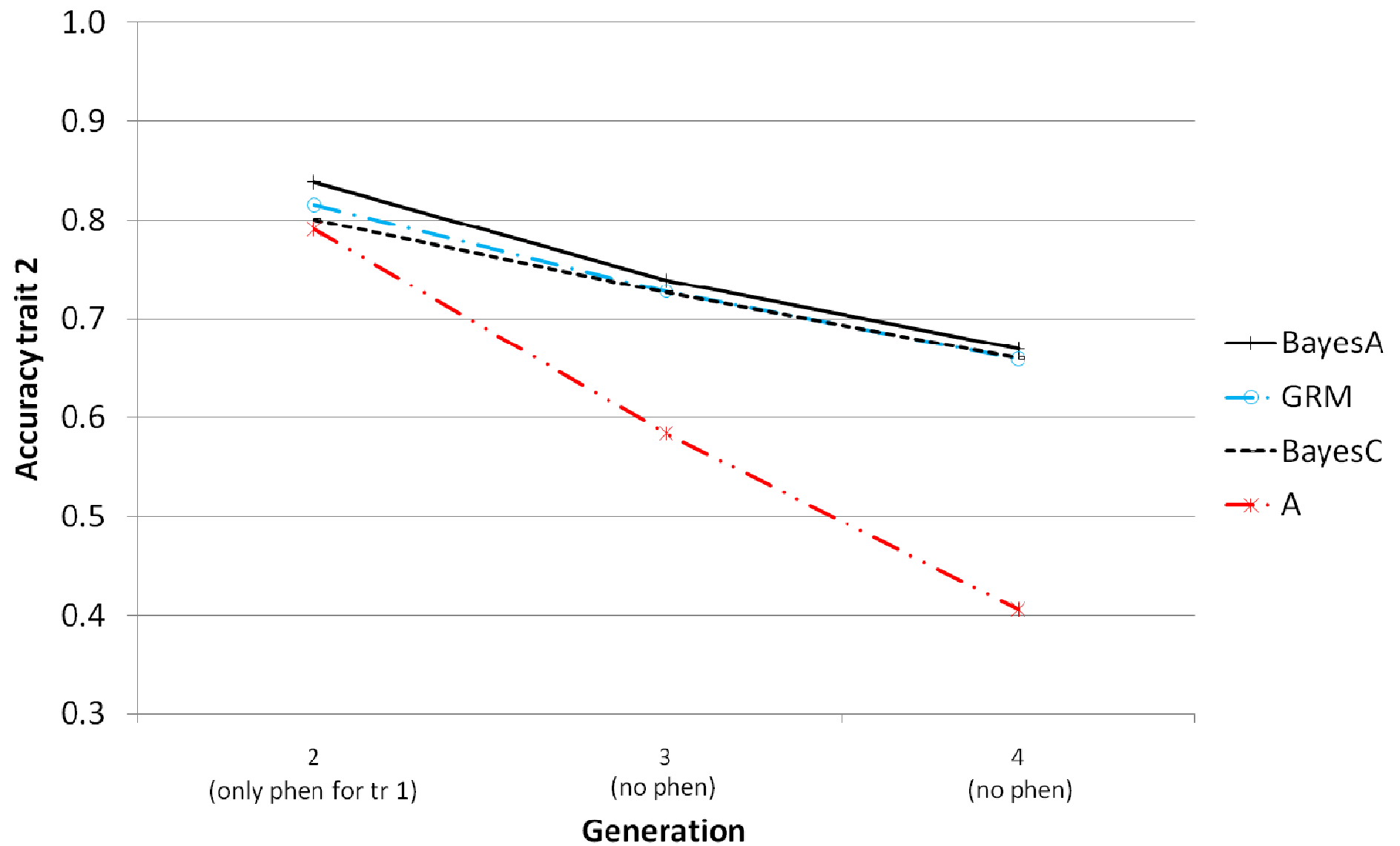
- See following slides for r_g of 0.2 and 0.8:
 - Generation 2: only phenotypes for trait 1
 - Generation 3 & 4: no phenotypes



$$r_g = 0.2$$



$$r_g = 0.8$$



Results summarized

- BayesA performs good across values of r_g
- At high r_g all models using SNPs perform similar
- BayesC has lowest accuracy at low r_g
- Low accuracy (BayesC) at low r_g possibly due to implementation of algorithm (canonical transformation & EM step)



Conclusions

- MT GEBV have substantial higher accuracy than 'classical' MT EBV
- BayesA performed best
- The presented implementation of BayesC is competitive for high r_g
- GRM performs good, despite the strong assumptions (equal variance per SNP)



Acknowledgements

- Involved companies:
Hendrix Genetics, CRV (HG)
- Netherlands Organisation for Scientific Research
(NWO – Casimir)
- RobustMilk

