Accuracy of MEBV

Effect of number of QTL and distribution of QTL variance

Albart Coster, John Bastiaansen, Mario Calus, Johan van Arendonk, Henk Bovenhuis

`albart.coster@wur.nl`
Animal Breeding and Genomics Centre,

Wageningen University, Wageningen, the Netherlands

Introduction

- Breeding values estimated with markers: MEBV.
  - Accurate estimated of BV without progeny testing.

This study:

1. Assess effect of QTL distribution on accuracy of MEBV.
2. Assess effect of method for calculating MEBV on accuracy of MEBV.

QTL distribution

Simulation studies assessing accuracy of MEBV:

- Simulated QTL according to Hayes and Goddard [2001]:
  - Few large QTL and many small QTL
- Genetics underlying traits differ:
  - Many genes involved in growth (*e.g.* human length, [Weedon et al., 2008]).
  - Single gene responsible for large proportion of genetic variation of milk fat composition, [Grisart et al., 2002].

- Need to identify effect of QTL distribution on accuracy of MEBV.

Methods for calculating MEBV

- Bayesian method of Meuwissen et al. [2001], BM;
- Least Angle Regression (LARS);
- Partial least square regression (PLSR);

- Do methods perform differently depending on distribution of QTL?

This study

- Effect of QTL distribution:
  - Number of QTL: low, intermediate, high.
  - QTL variance: all QTL equal; few large, many small.

Number of QTL	**QTL variance**
Low nQTL	unequal
Int. nQTL	unequal
High nQTL	unequal
Low nQTL	equal
Int. nQTL	equal
High nQTL	equal

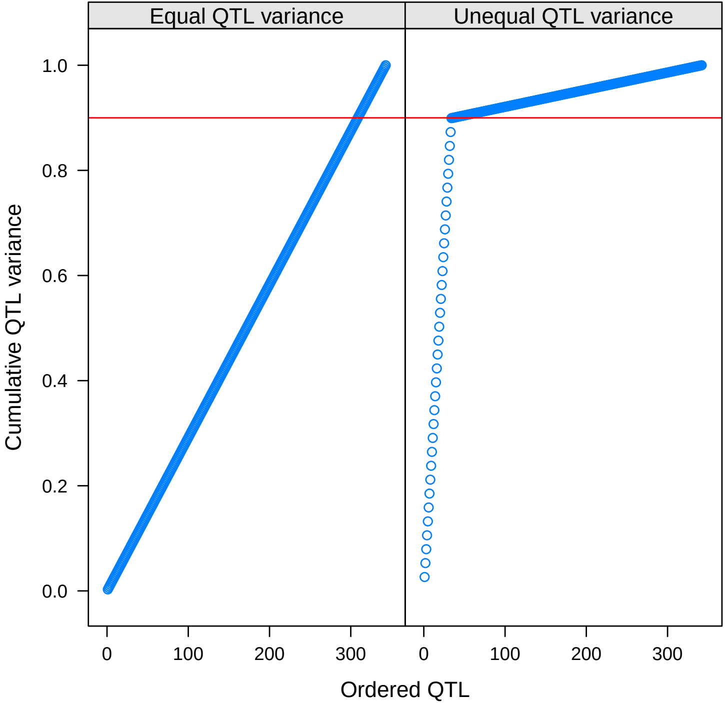
- Effect of method:
  - BM, LARS, PLSR.

Simulations

- Population of size 100 in mutation drift equilibrium.
  - Simulated for 5000 generations.
- Population of size 500 used for training.
- Offspring population of size 500 for evaluating MEBV.
- 5 chromosomes of 1 Morgan each.
- Average of 1431 SNP.
- 60 replicates.
- Heritability equal to 0.5.
- Simulations performed using package `HaploSim` in R.

Simulated data

variable	low nQTL	int. nQTL	high nQTL
nSNP	1431	1431	1431
LD SNP (r^2)	0.048	0.048	0.048
nQTL	35	172	343


Equal QTL variance
Unequal QTL variance
Cumulative QTL variance
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
0
100
200
300
Ordered QTL

Accuracies of MEBV

Method	**unequal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	0.77	0.67	0.60
LARS	0.75	0.67	0.65
PLSR	0.66	0.66	0.67

Method	**equal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	0.71	0.67	0.67
LARS	0.65	0.63	0.63
PLSR	0.68	0.67	0.66

Accuracies of MEBV

Method	**unequal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	0.77	0.67	0.60
LARS	0.75	0.67	0.65
PLSR	0.66	0.66	0.67

Method	**equal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	0.71	0.67	0.67
LARS	0.65	0.63	0.63
PLSR	0.68	0.67	0.66

Accuracies of MEBV

Method	**unequal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	0.77	0.67	0.60
LARS	0.75	0.67	0.65
PLSR	0.66	0.66	0.67

Method	**equal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	0.71	0.67	0.67
LARS	0.65	0.63	0.63
PLSR	0.68	0.67	0.66

Genetic variances

Method	**unequal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	55%	35%	14%
LARS	56%	55%	49%
PLSR	77%	73%	76%

Method	**equal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	51%	43%	37%
LARS	48%	45%	46%
PLSR	81%	81%	75%

Table: Variance of MEBV estimated by each method as percentage of simulated genetic variance.

Genetic variances

Method	**unequal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	55%	35%	14%
LARS	56%	55%	49%
PLSR	77%	73%	76%

Method	**equal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	51%	43%	37%
LARS	48%	45%	46%
PLSR	81%	81%	75%

Table: Variance of MEBV estimated by each method as percentage of simulated genetic variance.

Genetic variances

Method	**unequal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	55%	35%	14%
LARS	56%	55%	49%
PLSR	77%	73%	76%

Method	**equal QTL variance**		
	low nQTL	int. nQTL	high nQTL
BM	51%	43%	37%
LARS	48%	45%	46%
PLSR	81%	81%	75%

Table: Variance of MEBV estimated by each method as percentage of simulated genetic variance.

Conclusions

- Number and distribution of genes is an important factor for accuracy of MEBV.
- In general: accuracies decreased when number of genes increased.
- Accuracies obtained with PLSR almost not affected by QTL scenario.

Acknowledgements

Stichting Tecnologie en Wetenschap (Dutch Technology Foundation): stw

References

B. Grisart, W. Coppieters, F. Farnir, L. Karim, C. Ford, P. Berzi, N. Cambisano, M. Mni, S. Reid, P. Simon, R. Spelman, M. Georges, and R. Snell. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Research*, 12(2):222–231, 2002.

B. Hayes and M. E. Goddard. The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*, 33(3):209–229, 2001.

T. H. E. Meuwissen, B. J. Hayes, and M. E. Goddard. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157(4):1819–1829, 2001.

M. N. Weedon, H. Lango, C. M. Lindgren, C. Wallace, D. M. Evans, M. Mangino, R. M. Freathy, J. R. B. Perry, S. Stevens, A. S. Hall, N. J. Samani, B. Shields, I. Prokopenko, M. Farrall, A. Dominiczak, Diabetes Genetics Initiative, The Wellcome Trust Case Control Consortium, Johnson T., S. Bergmann, P. Beckmann J. S. Vollenweider, D. M. Waterworth, V. Mooser, C. N. A. Palmer, A. D. Morris, W. H. Ouwehand, GEM Consortium, M. Caulfield, P. B. Munroe, M. I. Hattersley, A. T. McCarthy, and M. Frayling. Genome-wide association analysis identifies 20 loci that influence adult height. *Nature Genetics*, 40(5):575–583, 2008.