A comparison of various methods for the computation of genomic breeding values of dairy cattle using software at genomicselection.net

R. Mrode, M.P. Coffey, I. Stradén, T.H.E. Meuwissen, J.B.C.H.M. van Kaam, J.F. Kearney and D.P. Berry

Introduction

SAC

- Software can usually be a major limitation to practical application of advances in methodologies
- Well resourced countries are usually first to implement and has implications for Interbull
- One of the possible hindrances to prediction of breed values using SNPs could be relevant and reliable software
- In an attempt to address this problem a number of interested scientists formed the so called club-ware
- Software developed and tested will be loaded on the website: www.genomicselection.net

Objectives

SAC

- Review the array of programmes at the website
- Summarise results from implementations using Irish dairy cattle data
- Computation time and accuracy of direct genomic breeding value from several methods are compared
- Several variables were computed at various allele frequencies to examine differences among evaluation methods

Data

- Data consisted of 1095 AI Holstein-Friesian Bulls genotyped with the Illumina 50K chip and have daughters in Ireland
- After all edits 42265 SNPs were available on these bulls
- Genotypic values of 0 and 2 were assigned to the two homozygotes and 1 to heterozygote for each locus
- DYDs for milk, fat and protein yields, calving interval and survival were analysed in this study

Genomic evaluation methods

- SNP effects and DGVs were estimated using linear and Bayesian methods

- Linear methods included
  - BLUP1 - iterating on data
  - BLUP2 – involved the use of the genomic relationship matrix

- Bayesian methods
  - BayesA
  - BayesA-P - Bayes A with polygenic effects included
  - BayesB - with about 34% of SNPs effects assumed to be zero
  - fastBayesB

Genomic evaluation methods

- The MCMC chains were 80000 cycles with the first 24,000 discarded as burn in period

- For BayesB 20 Metropolis-Hastings cycles were evaluated within each MCMC chain

- Analyses were carried out using a Sun workstation VG800 with 32GB of memory and eight 5GHZ processors

Genomic evaluation methods

- Is the relative information contributed by SNPs at various frequencies to DGV different among the evaluation methods?

- Several variables were computed at high(>0.80) , medium (0.4-0.79)and low (<0.40) allele frequencies to examine differences among evaluation methods

  – Mean SNP effects and variances
  – Mean SNP deviations (VanRaden, 2008) and associated weight (Wt).

  $wt_j = (z_j' r^{-1} z_j + \alpha)^{-1} z_j' r^{-1} z_j$

Accuracy of genomic evaluations

- The 1096 sires were divided to a training and validation dataset
- Training set
  - Only bulls with adjusted reliability > 40 were used in the training set
  - 755 sires for milk traits and 729 for CI and 642 for SUR
- Validation data set
  - Bulls have at least 40 daughters
  - 254 sires born post 1996, for milk traits
  - 187 sires born post 1995 with reliability ≥ 65% for CI
  - 116 sires born post 1994 with reliability ≥ 65% for SUR

Accuracy of genomic evaluations

- Statistics used to determine accuracy of evaluations based on the validation data set
- Correlation between DGVs and national EBVs
- Regression of national EBVs on DGVs
- Mean and std of differences between DGVs and national EBVs

Computing times

- Most were initially developed using the small example data. Thus this study was their first application to real data.

	BLUP1	BayesA	BayesA-P	FastBayesB	BayesB
Time:	5min	14.6hrs	22.9hrs	5-49min	65hrs

- Times will be acceptable for a national evaluation system apart from BayesB. Currently almost all countries uses BLUP at national level to compute DGV

- However data set is small and time is expected to increase with more animals included

Typical input file

- Inputfile for genotype information
- Inputfile for performance data
- Output file with SNP effects
- Outputfile for mean effect
- 2000 #no of iterations
- 10 # no of snps/markers
- 325 # no of animals
- 0.10 # proportion as burn in
- 10 # genetic variance
- 1 # type of prior 1=theo 2=xu,3=braak,4=yours

Correlations and regressions of national EBV on DGV

Trait	Method					
	BLUP1	BLUP2	BayesA	BayesA-P	FastBayesB	BayesB
Milk yield						
Correlation	0.68	0.69	0.70	0.72	0.67	0.65
Regression	1.31	0.78	0.99	1.49	1.30	0.98
Mean bias (kg)	-31.2	18.1	-4.6	-165	4.40	-6.60
SD of bias	147.8	146.8	141.4	144.6	148.7	149.6
Fat yield						
Correlation	0.65	0.67	0.68	0.62	0.65	0.58
Regression	1.32	0.80	0.89	1.09	1.35	0.58
Mean bias (kg)	-3.82	-0.48	-1.01	-318	3.80	-0.50
SD of bias	4.89	4.74	4.53	4.92	4.88	5.73

Correlations and regressions of national EBV on DGV

Trait	Method					
	BLUP1	BLUP2	BayesA	BayesA-P	FastBayesB	BayesB
Calving interval						
Correlation	0.70	0.71	0.66	0.69	0.70	0.49
Regression	1.18	0.75	0.77	1.29	1.19	0.30
Mean bias (d)	4.02	0.28	0.11	-41.5	3.93	0.23
SD of bias (d)	2.35	2.43	2.55	2.42	2.36	4.71
Survival						
Correlation	0.58	0.59	0.56	0.57	0.57	0.57
Regression	1.22	0.76	0.93	1.48	1.04	1.10
Mean bias (%)	-1.60	-0.45	-0.45	2.53	1.63	-0.61
Mean of bias (%)	1.46	1.47	1.47	1.49	1.46	1.70

Variables at various gene frequencies

	High Freq			Medium Freq			Low Freq		
	BLUP1	Bayes A	Bayes B	BLUP1	BayesA	BayesB	BLUP1	Bayes A	BayesB
Mean SNP	0.021	0.010	0.018	0.009	0.002	-0.010	-0.004	-0.011	-0.010
SNP VAR	1.333	0.667	1.081	1.333	0.670	1.081	1.333	0.747	1.197
SYD	0.556	0.019	0.021	0.073	0.004	-0.012	-0.107	-0.022	-0.012
WT	0.192	0.528	0.876	0.121	0.517	0.878	0.038	0.506	0.878

Conclusions

- Programmes were easy to use and have performed well in handling real data
- The accuracies from the linear and non-linear methods are similar, but it seems BayesA has the best predictive ability in the dataset analysed

Conclusions

SAC

- The relative information contributed by SNPs at various allele frequencies to DGV seems to differ for various evaluation methods

- Possibly this could affect changes in allele frequencies when selecting on DGV from the different evaluation methods and it should be studied

Acknowledgement

SAC

- Contributions of other scientists in club-ware greatly appreciated:
- R. Thompson
- J. Woolliams
- E. Mantysaari
- H. Jorjani
- E. Bongcam-Rudloff and
- DairyCo for funding
- ICBF for the data

SAC

Success through Knowledge