Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Accuracy of genomic prediction in a simulated pig populations

Jihyun Son[1], Hyeonjong Kang[1], Jayoung Kim[1], Jong-Eun Park[1,2],Deukhwan Lee[1].
[1] Graduate school of future convergence technology, Hankyong National University, Anseong, Korea
[2]Dept. of Animal science & Biotechnology, Seoul National University, Seoul, Korea,

ABSTRACT: Genomic prediction has become the promising paradigm in animal breeding programs. This study was compared prediction accuracies of genomic prediction in terms of selection strategy and candidate phenotypes using ridge regression BLUP. Simulated data mimicking pig genome was used for genomic prediction. The accuracy of traditional EBV estimated using BLUP. However the averages of accuracies between phenotypic and EBV selection were similar. The accuracy of genomic prediction for 1000 individual was calculated by 5-fold cross validation. In terms of selection strategy, accuracies of EBV-selection were **0.81** and **0.84** for phenotype and EBV as candidate phenotype. On the other hand, accuracies of phenotypic selection were **0.75** and **0.73** for phenotype and EBV. As reported in previous research, genomic prediction showed higher **(0.13-0.27)** accuracy than traditional BLUP method. In this study, EBV-selection with EBV as response variable had most accurate prediction within genomic prediction.
Keywords: Genomic prediction; Simulation; Ridge regression BLUP; Single nucleotide polymorphism (SNP)

Introduction

Genomic prediction has become the promising paradigm in animal breeding programs based on previous marker-assisted selection. Genomic prediction is a method based on total markers with LD that explores all QTL in the genome (Meuwissen et al. (2001)). The breeding value (BV) of genomic prediction is estimated by the sum of the effects of marker alleles or haplotypes covering the whole genome and its accuracy could be as high as 0.85 (Meuwissen et al. (2001)). The BV calculated from the estimated effects of markers is often called genomic estimated breeding value (GEBV) (Brito et al. (2011)). The accuracy of GEBV affected by: 1) the level of LD between markers and QTL; 2) the number of breeding stock in the training set (TS); 3) the heritability of the trait and 4) the dispersion of QTL effects. (Hayes et al. (2009)). This study was compared prediction accuracies of genomic prediction in terms of selection-strategy and candidate phenotypes using ridge regression BLUP (rrBLUP) method (Piepho (2009)). Two compared selection-strategy was phenotypic and and EBV selection. And candidate phenotypes were raw phenotype and EBV.

Materials and Methods

Simulation. Using the QMsim software (Sargolzaei and Schenkel (2009)), populations were simulated based on forward-in-time process 50k SNP markers, and 786 QTL across the 18 pig autosomes. In the first step, 1000 of the historical generations with a gradual decrease in population size from 800 to 100 were simulated. In second step, two recent generation sets were simulated by EBV and phenotypic selection with gradual increase. The breeding values were estimated by BLUP, using Henderson's mixed linear equations for an individual animal model, considering the true additive genetic variance.

In order to reproduce the current pig SNP panel (position-defined autosomal SNPs only), we set the initial value of the genomic parameter as in the following Table 1. By recording the genotype of the recent 4 generations, genotype of 2,000 animals was generated.

Table 1. Parameters of the simulation process

Simulation step and parameters	Value
Historical generations (HG)	
Initial generation (size)	0 (800)
Last generation (size)	1000 (100)
Recent expanded generation (RG)	
Number of founder males from HG	50
Number of founder females from HG	50
Number of generations	10
Number of offspring per dam	10
Ratio of male	0.5
Mating system	Random
Replacement ratio for males	0.5
Replacement ratio for females	0.5
Selection/culling design	Phenotype / EBV
BV estimation method	BLUP animal model
Heritability of the trait	0.3
Phenotypic variance	1
Genome	
Number of chromosomes	18
Total length	2453cM
Number of markers	48245
Marker distribution	Evenly spaced
Number of QTL	786
QTL distribution	Random
MAF for markers	Equal
MAF for QTL	Equal
Additive allelic effect for markers	Neutral

Additive allelic effect for QTL	Gamma distribution (shape = 0.40)
Rate of recurrent mutation	2.50E-05

EBV: estimated breeding value; BV: breeding value; QTL: Quantitative trait loci, MAF: Minor allele frequency

Genomic prediction. The method applied for genomic prediction was ridge regression, which is equivalent to best liner unbiased prediction (BLUP) in the context of mixed models implemented in rrBLUP package in R (Whittaker et al. (2000); Endelman (2011)).
The basic rrBLUP formula is

$$\mathbf{y} = \mathbf{WGu} + \boldsymbol{\varepsilon}$$

where $u \sim N(0, I\sigma_u^2)$ is a vector of marker effect, W is design matrix relating lines to observations (y), G is the genotype matrix (**e.g., {aa, Aa, AA} = {–1, 0, 1}**) for biallelic single nucleotide polymorphisms (SNPs) under an additive model and ε is a residual. The BLUP solution for the marker effects can be written as

$$\hat{\mathbf{y}} = \mathbf{Z}'(\mathbf{ZZ}' + \boldsymbol{\lambda}\mathbf{I})^{-1}$$

Where $\mathbf{Z} = \mathbf{WG}$ and the ridge parameter $\boldsymbol{\lambda} = \boldsymbol{\sigma_e^2}/\boldsymbol{\sigma_u^2}$ is the ratio between the residual and marker variances (Searle et al., 2006). The accuracy of genomic prediction was calculated by 5-fold cross validation for randomly selected 1000 animal. Data processing and analysis were performed using PLINK (Purcell et al. (2007)) and R under Linux OS.

Results and Discussion

Traditional BLUP. From two selection scenario, total and sire accuracies of traditional BLUP-EBV were displayed in **Table 2**. The averaged accuracies of total and sire were **0.59** and **0.57** for EBV-selection and **0.60** and **0.62** for phenotypic selection, respectively. The average, of accuracies between phenotypic and EBV selection were similar. The difference in accuracy showed weak level of significance (Paired *t* test, p = 0.06 > 0.05). These results presented the uncertain effect of EBV-selection on accuracy of EBV. Moreover, distinct increase of accuracy could not observe.

Table 2. Accuracy of EBV for two simulated data.

	Accuracy(EBV,TBV)			
	EBV-selection		**Phenotypic selection**	
Generation	**Total**	**Sire**	**Total**	**Sire**
0	0.49	NA	0.49	NA
1	0.61	0.65	0.66	0.62
2	0.65	0.62	0.60	0.72
3	0.52	0.53	0.69	0.68
4	0.59	0.64	0.65	0.58
5	0.59	0.56	0.71	0.61
6	0.62	0.46	0.62	0.68
7	0.59	0.56	0.63	0.60
8	0.61	0.55	0.57	0.59
9	0.64	0.68	0.62	0.57
10	0.53	0.50	0.58	0.70
Mean*	**0.59**	**0.57**	**0.60**	**0.62**

*: Mean of accuracies was calculated from recent 4 generation (7, 8, 9 and 10).

Genomic prediction. Accuracies of genomic prediction using rrBLUP for four different cases were calculated by 5-fold cross validation (Supplemental figure 1). In terms of selection strategy, accuracies of EBV-selection were **0.81** and **0.84** for phenotype and EBV as candidate phenotype. On the other hand, accuracies of phenotypic selection were **0.75** and **0.73** for phenotype and EBV (Table 3). This suggests EBV-selection strategy is more accurate than phenotypic selection in genomic prediction.

Table 3. Accuracy of GEBV using rrBLUP in terms of different selection strategy and candidate phenotype

	EBV-selection				**Phenotypic selection**			
Candidate phenotype	**Phenotype**		**EBV**		**Phenotype**		**EBV**	
Set	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**	**Train**	**Test**
CV1	0.84	0.81	0.85	0.82	0.80	0.77	0.75	0.73
CV2	0.82	0.81	0.84	0.86	0.79	0.75	0.75	0.73
CV3	0.84	0.84	0.84	0.85	0.78	0.77	0.75	0.73
CV4	0.84	0.82	0.85	0.83	0.79	0.73	0.75	0.73
CV5	0.84	0.77	0.85	0.81	0.80	0.72	0.75	0.74
Mean	**0.84**	**0.81**	**0.84**	**0.84**	**0.79**	**0.75**	**0.75**	**0.73**

It is common that GEBV was estimated to use phenotype, but in reality, the phenotypic data are not sufficient, the response variable GEBV will be used response for estimating GEBV. Response variable like this are actually used DYD, de-regressed proofs, national EBVs and etc. In this Experiment, using estimated EBV from simulated data as a response variable and comparing the accuracy of the difference in genome estimated breeding values. Aspect of candidate phenotype gave more accurate prediction in followed phenotype.

Conclusion

This study was compared prediction accuracies of genomic prediction in terms of selection strategy and candidate phenotypes using rrBLUP. Simulated data mimicking 1000 pig genome was used for genomic prediction. As reported in previous research, genomic prediction showed higher **(0.13-0.27)** accuracy than traditional BLUP method. In four cases of genomic prediction, EBV-selection with EBV as response variable had most accurate prediction within genomic prediction.

Acknowledgments

This work was supported by a grant from the IPET Program (No. 20093068), Ministry of agriculture, food and rural affairs, Republic of Korea.

Literature Cited

Brito, F.V., et al. (2011). Accuracy of genomic selection in simulated populations mimicking the extent of linkage disequilibrium in beef cattle.BMC genetics, 12(1), 80.

Endelman, J.B. (2011). Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome, 4(3), 250-255.

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

Hayes B.J., Bowman P.J., et al. (2009). Invited review: Genomic selection in dairy cattle: progress and challenges.J Dairy Sci, 92:433-443

Piepho, H. P. (2009). Ridge regression and extensions for genomewide selection in maize. Crop Science, 49(4), 1165-1176.

Sargolzaei, M., and Schenkel, F. S. (2009). QMSim: a large-scale genome simulator for livestock. Bioinformatics, 25(5), 680-681.

Searle, S.R., et al. (2006). Variance components. John Wiley & Sons, Hoboken, NJ.

Whittaker, J.C., et al. (2000). Marker-assisted selection using ridge regression. Genet. Res. Camb. 75:249–252.

Supplemental figure 1. Correlation (accuracy) of EBVs and predicted GEBVs in the training and test set for four different situations