Open Access
Asian Australas. J. Anim. Sci.
Vol. 28, No. 11 : 1551-1557 November 2015
http://dx.doi.org/10.5713/ajas.15.0287
www.ajas.info
pISSN 1011-2367 eISSN 1976-5517

Genome-wide Association Study (GWAS) and Its Application for Improving the Genomic Estimated Breeding Values (GEBV) of the Berkshire Pork Quality Traits

Young-Sup Lee[a], Hyeonsoo Jeong[a], Mengistie Taye[1], Hyeon Jeong Kim[2], Sojeong Ka[1], Youn-Chul Ryu[3,*], and Seoae Cho[2,*]
Department of Natural Science, Interdisciplinary Program in Bioinformatics, Seoul National University, Seoul 151-747, Korea

ABSTRACT: The missing heritability has been a major problem in the analysis of best linear unbiased prediction (BLUP). We introduced the traditional genome-wide association study (GWAS) into the BLUP to improve the heritability estimation. We analyzed eight pork quality traits of the Berkshire breeds using GWAS and BLUP. GWAS detects the putative quantitative trait loci regions given traits. The single nucleotide polymorphisms (SNPs) were obtained using GWAS results with p value <0.01. BLUP analyzed with significant SNPs was much more accurate than that with total genotyped SNPs in terms of narrow-sense heritability. It implies that genomic estimated breeding values (GEBVs) of pork quality traits can be calculated by BLUP via GWAS. The GWAS model was the linear regression using PLINK and BLUP model was the G-BLUP and SNP-GBLUP. The SNP-GBLUP uses SNP-SNP relationship matrix. The BLUP analysis using preprocessing of GWAS can be one of the possible alternatives of solving the missing heritability problem and it can provide alternative BLUP method which can find more accurate GEBVs. (**Key Words:** Best Linear Unbiased Prediction, Genome Wide Association Study, Missing Heritability Problem, Sherman-Morrison-Woodbury Lemma, Single Nucleotide Polymorphism–Genomic Best Linear Unbiased Prediction, Berkshire Pigs)

INTRODUCTION

Pork is the most widely consumed meat, accounting for 50% of daily meat protein intake, globally (Davis and Lin, 2005). Genetic selection using best linear unbiased prediction (BLUP) methodologies, so far, have resulted in a number of successes in improving different pork quality parameters (Leeds, 2005; Sellier, 1998). Various studies have been devoted to the estimation of genetic parameters for pork quality traits to use in selection programs (Leeds, 2005). C.R. Henderson around 1950 developed the mixed model equations involving BLUP (Henderson, 1975; Jiang, 1997).

Genomic selection aims at making selection decisions based on breeding values predicted using genome-wide marker data (Meuwissen et al., 2001). There are two general categories of BLUP methods: GRM-based genomic-best linear unbiased prediction (G-BLUP), and single nucleotide polymorphism (SNP)-best linear unbiased prediction (SNP-BLUP). Genomic relationship matrix (GRM) exploits the elements of the realized proportion of the genome that two individuals share (Legarra et al., 2009; Goddard et al., 2011). The big compromise of G-BLUP and SNP-GBLUP is single nucleotide polymorphism-genomic best linear unbiased prediction (SNP-GBLUP) (Lee et al., 2014b). The breeding

* Corresponding Authors: Youn-Chul Ryu. Tel: +82-64-754-3332, Fax: +82-64-754-3332, E-mail: ycryu@jejunu.ac.kr / Seoae Cho. Tel: +82-2-876-8820, Fax: +82-2-876-8827, E-mail: seoae@cnkgenomics.com
[1] Department of Agricultural Biotechnology, Animal Biotechnology, and Research Institute for Agriculture and Life Sciences, Seoul National University, Seoul 151-921, Korea.
[2] C&K Genomics Inc., Seoul 151-919, Korea.
[3] Division of Biotechnology, Sustainable Agriculture Research Institute, Jeju National University, Jeju 690-756, Korea.
[a] These authors equally contributed and should be regarded as co-first authors.
Submitted Mar. 31, 2015; Revised Jun. 9, 2015; Accepted Jun. 24, 2015

values of the SNP-GBLUP are nearly identical to the G-BLUP. However, it predicts the SNP effects of the given traits. It is an approach unlike the single step BLUP (SS-BLUP) which finds the SNP effects iteratively (Fernando et al., 2014).

Genome-wide association study (GWAS) uses genetic variants with traits of interest and it can estimate the p-value of each SNP in a given complex trait (Bolormaa et al., 2013; Lee et al., 2014a). For the analysis of each SNPs' P-value, we used the PLINK linear regression model (Purcell et al., 2007). Then we used the SNP-GBLUP (Lee et al., 2014b).

The missing heritability has been the problem which must be solved in GWAS and BLUP analyses (Eichler et al., 2010). It indicates that the narrow-sense heritability cannot be achieved satisfactorily in complex diseases and traits with a complex inheritance such as human height (Eichler et al., 2010; Yang et al., 2010). BLUP traditionally uses total genotyped SNPs. However, it has not yet solved the missing heritability problem. We tried to combine the GWAS and BLUP method to complement it. We analyzed the BLUP by using only SNPs with p-value under 0.01 in GWAS.

MATERIALS AND METHODS

Ethics statement

The study protocol and the standard operating procedures (No. 2009-077, C-grade) of Berkshire pigs were reviewed and approved by National Institute of Animal Science's Institutional Animal Care and Use Committee.

Data preparation

We used data from 702 (365 male, 204 female, 133 castrated male) Berkshire pigs. Animals were raised with the same commercial diet from the Dasan experimental farm in Namwon, Korea. The genomic DNAs of 702 individuals were genotyped using Illumina Porcine 60 K SNP Beadchip (Illumina, San Diego, CA, USA) following the standard protocol. A total number of 44,345 genotyped SNPs were filtered using quality-control processes with MAF (minor allele frequency (MAF) (<0.05), Hardy-Weinberg equilibrium (HWE) (p<0.001) and missing data (>0.01 missing) which resulted 36,896 autosomal SNPs.

A total of 8 meat quality traits were used for the analysis. The traits included carcass weight (CWT), back fat thickness (BF), intramuscular fat content (fat), protein content, Shear force (SF), water holding capacity (WHC) and color (L* and A*). Carcass weight was measured immediately after slaughter. BF and color were measured from the longissimus dorsi muscle between 10th and 11th rib. Intramuscular fat content was measured using chemical fat extraction procedures. WHC (%) was measured as a difference between moisture content (%) and expressible water (EW; %). General indication of lightness and degree of green-redness of meat color were measured referred to Minolta L (MC_L, Commission Internationale de I'Eclairage [CIE] L* color space) and Minolta A (MC_A, CIE a* color space), respectively. Shear force was measured using the Warner-Bratzler shear force meter (G-R Elec. Mfg. Co., Manhattan, New York, USA). In each sex group (365 male, 204 female, and 133 castrated male), we standardized the values to z-score separately for GWAS.

Data analysis

Linear regression GWAS: We used the linear regression model in PLINK software (additive option) for the genome-wide association (GWA) analysis with the sex adjusted data. The P-value less than the stringent level of 0.01 was selected for genome-wide significant autosomal SNPs.

The BLUP solution and SNP-GBLUP: The mixed model used to estimate the breeding values includes BLUP and best linear unbiased estimation. These models estimate the fixed effects such as sex and predicts the random effects such as SNPs for a given quantitative phenotype. The solution of the model usually can be found by using the maximum likelihood estimation (MLE) of the probability density function (pdf) of the model. The mixed model and its solution used are presented as follows:

$$y = Xb+Zu+e \quad (1)$$

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z+G_u^{-1} \end{bmatrix}\begin{pmatrix} b \\ u \end{pmatrix}=\begin{pmatrix} X'R^{-1}Y \\ Z'R^{-1}Y \end{pmatrix} \quad (2)$$

Where y is the vector of phenotypic values, X and Z are the design matrices, **b** and **u** are vectors of fixed and random effects, respectively. The random effects and residual errors are assumed to be normally distributed. These multivariate normal distributions usually are notated as $u \sim MVN(0,G_u)$ and $e \sim MVN(0,R)$ where MVN are denoted as multivariate normal distribution.

The identical solution of MLE of mixed model is the generalized least squares (GLS). To compare the estimated breeding values (EBV) of the total SNPs with trimmed SNPs (unadjusted cutoff p-value 0.01), we used the G-BLUP which adopts the genomic relationship matrix (GRM) with total pruned SNPs (36,896 SNPs) and SNP-GBLUP which utilizes the SNP-SNP relationship matrix with trimmed SNPs (Lee et al., 2014b). The GRM was obtained using R package "rrBLUP" (Endelman, 2011). For the GRM of the trimmed SNP's analysis, we used the highly significant SNPs (p<0.01). The GLS solution is as follows:

$$\hat{b} = (X'V^{-1}X)^{-1}X'V^{-1}y \quad and \quad V = ZG_uZ' + R \quad (3)$$

$$\hat{u} = \mathrm{G}_u \mathrm{Z}' V^{-1} (y - X\hat{b}) \quad (4)$$

Where $\hat{b}$ is the estimated fixed effects and $\hat{u}$ is the estimated random effects.

The SNP-SNP relationship matrix and its inverse: The inverse of the SNP-SNP relationship matrix is depicted as in the below (Lee et al., 2014b):

$$G_u{}^{-1} = Z^T G^{-1} Z \quad (5)$$

Where G matrix is the genomic relationship matrix. To calculate the G_u matrix, i.e., SNP-SNP relationship matrix, we applied the Sherman-Morrison-Woodbury lemma (Sherman and Morrison, 1950; Woodbury, 1950).

$$(\mathrm{A} + \mathrm{YG}Z^*)^{-1} = A^{-1} - A^{-1}Y(G^{-1} + Z^*A^{-1}Y)^{-1}Z^*A^{-1} \quad (6)$$

Where G and A are both be invertible, and $A+YGZ^*$ are invertible if and only if $G^{-1} + Z^*A^{-1}Y$ are invertible. This formula reduces the computation time to calculate the SNP-SNP relationship matrix. We used A as identity matrix (I matrix) and the formula we used was as follows:

$$(\mathrm{I} + G_u^{-1})^{-1} = (\mathrm{I} + Z^T G^{-1} Z)^{-1} = \mathrm{I} - Z^T (G + ZZ^T)^{-1} Z \quad (7)$$

RESULTS

We first performed GWAS and identified the significant SNPs (p<0.01) associated with the phenotypic traits of interest (Table 1). The 859 (MC_L), 1,028 (CWT), 2,014 (Protein), 1,478 (BF), 2,580 (SF), 3,659 (Fat), 5,830 (WHC) and 3,210 (MC_A) SNPs were extracted and finally involved in the BLUP analysis.

In general, the results of SNP-GBLUP analyzed with trimmed SNPs mentioned above have higher heritability than those of G-BLUP with total SNPs as shown in Table 1. On the contrary, SF, fat, WHC and MC_A cases did not achieve satisfactory results and showed a small increase of heritability. We considered that the reason was failing to find the appropriate number of SNPs. The traits may have more quantitative trait loci (QTL) regions than those predicted as p-value <0.01. Because the most important part of our analysis was the number of SNPs, we regarded that the criteria of P-value in GWAS can be modified in the cases of SF, fat, WHC, and MC_A for a better performance of BLUP.

Figure 1 shows the plot of genomic estimated breeding value (GEBV) and phenotypic values for quality traits. In the plot, the colored ones refer to the trimmed SNPs' cases while the black dots refer to the total SNPs. It shows that the slopes of the colored ones (trimmed SNPs' cases) were higher than those of the black ones, which indicated the higher heritability and better performances in trimmed cases. Figure 2 and 3 show the Manhattan plot of –log 10 of absolute values of SNP effects across chromosomes. The plots indicate the aggregates of SNPs and SNP effects on each chromosome. The aggregates may imply the putative QTL regions. Specifically, Figure 2 shows the MC_L, CWT, protein, BF traits cases and Figure 3 shows the SF, fat, WHC and MC_A traits cases which showed the great and small increases of heritability, respectively. Figure 4 indicates K-means clustering (K = 4) of the phenotypic values and BLUP results of the analyzed Berkshire eight pork quality based on the 1st and 2nd discriminant functions. We used the R package "fpc" (Hennig, 2010). The plot of the trimmed SNPs' was closer than that of the total SNPs when compared with the plot of phenotypic values. These kinds of plots can assist the breeders in selecting better-performed Berkshire pigs.

DISCUSSION

The features of the genome-wide association study

In the field of livestock science and animal breeding, mapping of QTL has been widely used to detect genetic variation responsible for economically important traits. However, due to the low density of markers and the confidence interval of QTL mapping studies, it has been difficult to identify genetic variation affecting complex traits (Soller et al., 2006). GWAS, also known as common-variant association study, typically focuses on the association between genomic variants and phenotypic traits especially developed in human disease study (Feero et al., 2010). GWAS has been extended for use in domestic animal

Table 1. The table of the fixed effects (male, female, and castrated male), heritability and number of SNPs used

SNP-GBLUP	MC_L	CWT	Protein	BF	SF	Fat	WHC	MC_A
# SNPs	859	1,028	2,014	1,478	2,580	3,659	5,830	3,210
Male	48.73	86.31	24.00	25.26	2.89	2.80	59.29	6.15
Female	48.15	86.00	24.00	23.03	3.14	2.41	57.84	6.10
Castrated male	48.59	85.26	23.86	28.10	2.51	3.51	60.48	6.35
h^2 (trimmed)[1]	32	24	42	37	29	39	47	35
h^2 (total)[1]	6	9	26	20	20	37	43	29

SNP-GBLUP, single nucleotide polymorphism-genomic best linear unbiased prediction; MC_L, Minolta Commission Internationale de I'Eclairage L* color space; CWT, carcass weight; BF, back fat thickness; SF, Shear force; Fat, intramuscular fat content; WHC, water holding capacity; MC_A, Minolta Commission Internationale de I'Eclairage a* color space.

It shows the heritability (%) of trimmed highly significant SNPs (p<0.01) is greater than that of total SNPs' cases in all traits.

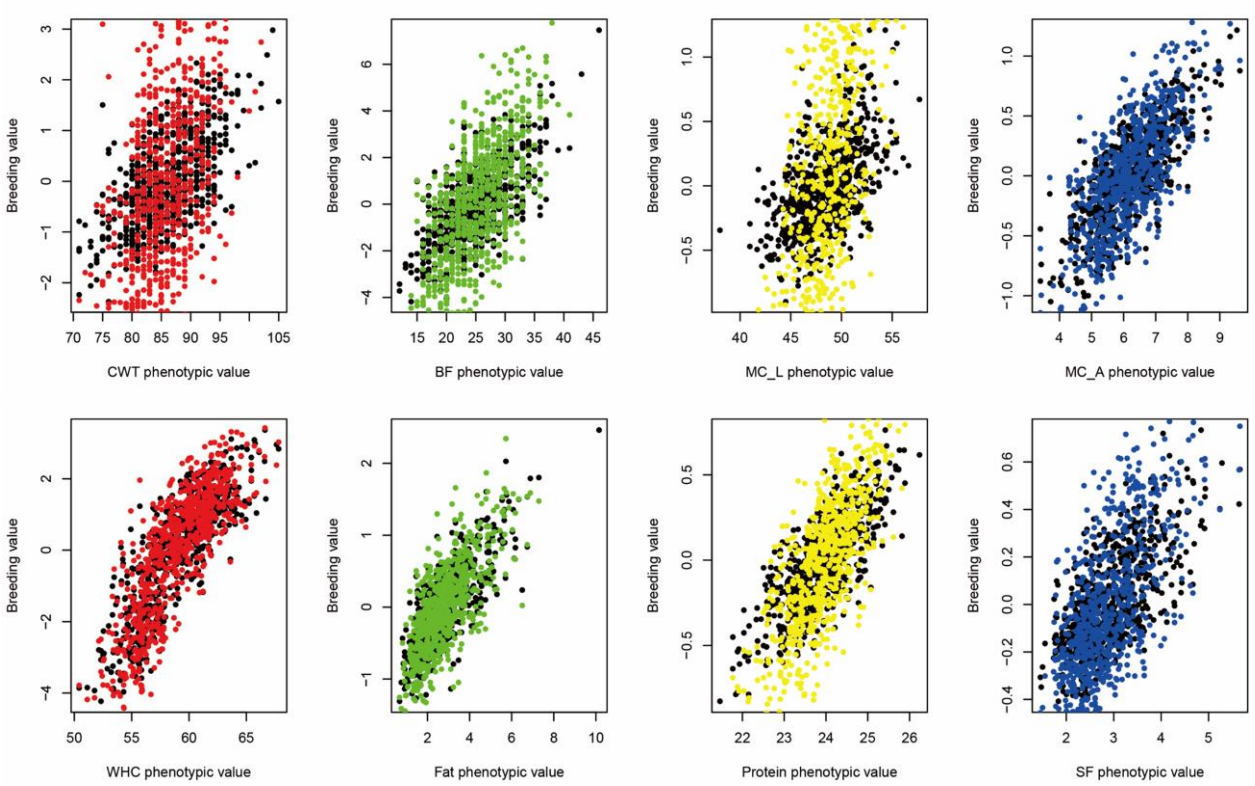

Figure 1. Plot of Berkshire genomic estimated breeding values (GEBVs) against to the phenotypic values. Black spots refer to total SNPs' cases and colored spots refer to the trimmed SNPs' cases. Because the slopes of colored ones were higher than black ones, the genomic estimated breeding values (GEBVs) of the trimmed cases can be more accurate than those of total SNPs' in terms of heritability. CWT, carcass weight; BF, back fat thickness; MC_L, Minolta L Commission Internationale de I'Eclairage L* color space; MC_A, Minolta Commission Internationale de I'Eclairage a* color space; WHC, water holding capacity; Fat, intramuscular fat content; SF, Shear force; SNP, single nucleotide polymorphism.

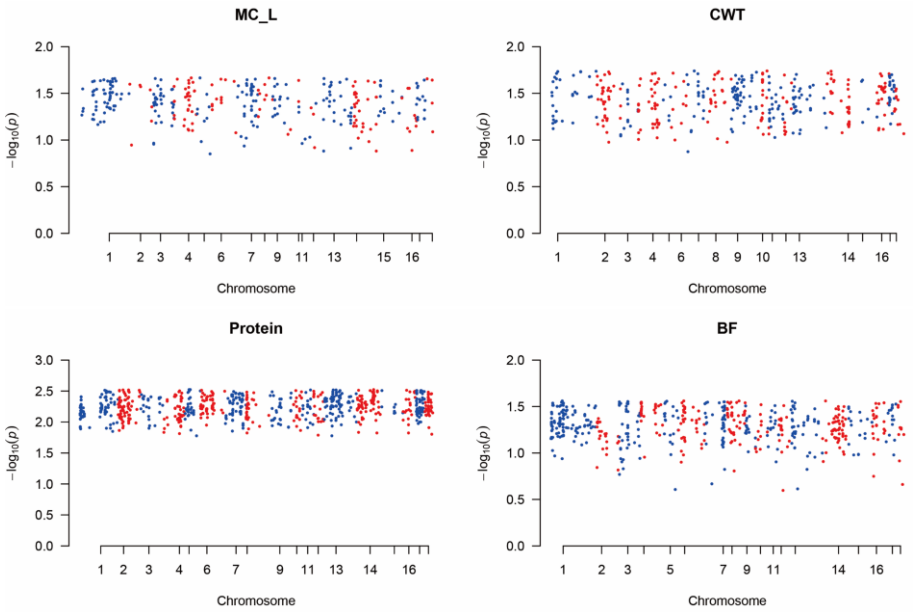

Figure 2. The Manhattan plot of –log10 of absolute values of SNP effects across chromosomes. It indicates the aggregates of the SNPs and SNP effects as predicted in GWAS. Each dot can represent the SNPs in the putative quantitative trait loci (QTL) regions. The method was single nucleotide polymorphism-genomic best linear unbiased prediction (SNP-GBLUP). SNP-GBLUP can predict the SNP effects. GWAS, genome wide association study; MC_L, Minolta L Commission Internationale de I'Eclairage L* color space; CWT, carcass weight; BF, back fat thickness.

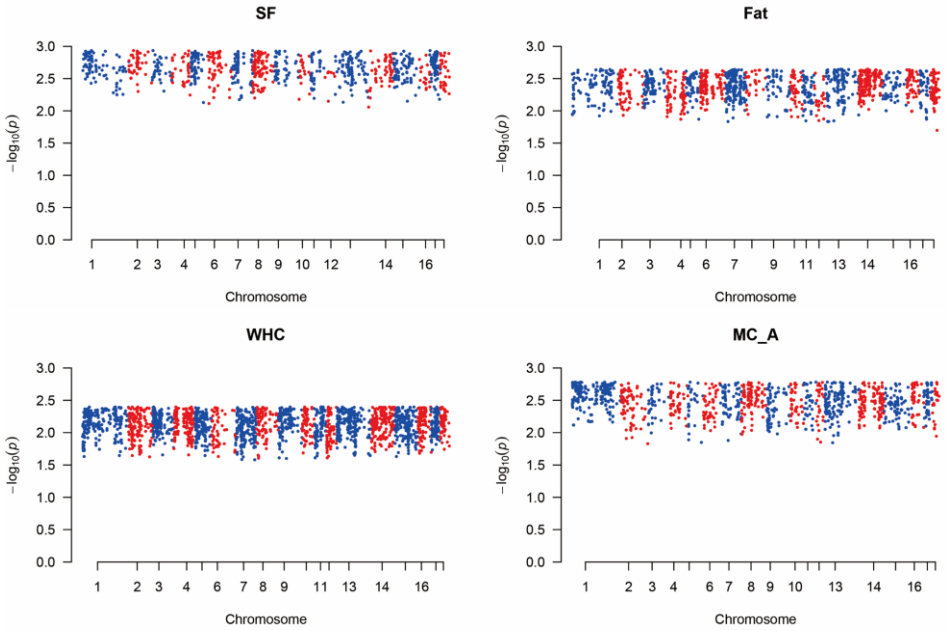

Figure 3. The Manhattan plot of –log10 of absolute values of SNP effects across chromosomes. It shows the aggregates of the SNPs and SNP effects as predicted in GWAS. Each dot can represent the SNPs in the putative QTL regions. The method was single nucleotide polymorphism-genomic best linear unbiased prediction (SNP-GBLUP). GWAS, genome wide association study; SF, Shear force; Fat, intramuscular fat content; WHC, water holding capacity; MC_A, Minolta Commission Internationale de I'Eclairage a* color space.

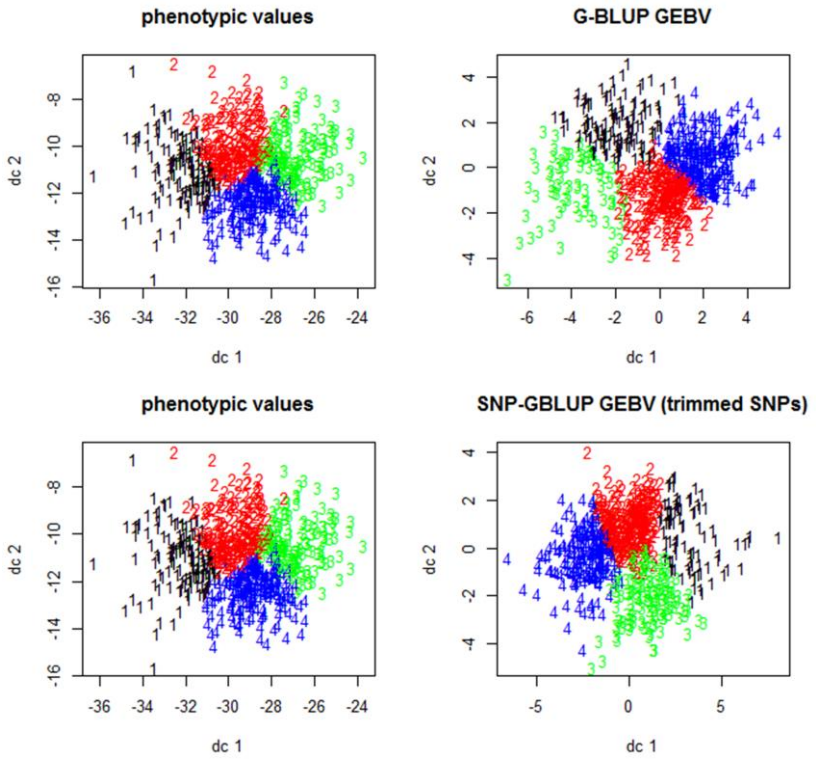

Figure 4. The plot of the 1st and the 2nd discriminant functions of Berkshire eight port quality traits and corresponding genomic estimated breeding values (GEBVs) of total SNPs (genomic-best linear unbiased prediction; G-BLUP) and trimmed SNPs (single nucleotide polymorphism-genomic best linear unbiased prediction; SNP-GBLUP). These plots represent the similarity between phenotypic values and GEBVs of the trimmed SNPs ($p<0.01$) as compared to those of the total SNPs. To classify individuals using the GEBVs can be an aid to the Berkshire breeders.

genetics since genomic sequences of livestock have become available, and the large scale of genomic variants (SNPs, Indel and CNV) were discovered. Using sequence variation, GWAS can detect the causal mutation responsible for the economic traits underlying in QTL (Zhang et al., 2012a). This was the basis of our study because GWAS could detect the putative QTL regions and it could improve the BLUP.

Genome-wide association study and its application to the best linear unbiased prediction

Zhang et al. reported that GWAS improved the accuracy of genomic selection (GS) (Zhang et al., 2014). They asserted the superiority of BLUP|GA model over G-BLUP, which used the QTL counts and obtained p-value from GWAS. BLUP|GA model requires prior knowledge about which SNPs belonged to QTL regions. We further assert that results of GWAS can contain the information about the QTL regions' SNPs. Otherwise, it can contain the information about the QTL regions via linkage disequilibrium. Specifically, the estimated heritability of MC_L and CWT were highly improved with 3 to 5 fold increase. This may arise because of the detection of putative QTL regions by GWAS. The combination of GWAS and SNP-GBLUP can make it possible to estimate the QTL-related SNP effects and GEBVs.

The number of analyzed SNPs was 2% to 10% of the total SNPs. We considered that as the number of QTL varies, the analyzable SNPs varies. Thus, we adopted the criteria as P-value of GWAS results. BLUP of trimmed SNPs were better than that of total SNPs in terms of heritability and GEBVs.

Missing heritability and trimming of single nucleotide polymorphisms

The missing heritability problem can occur in the association of the traits and genetic markers. In GWAS, the difficulty in analyzing complex diseases and genetic traits such as human height have emphasized the missing heritability problem (Manolio et al., 2009; Yang et al., 2010). Furthermore, there has been the missing heritability in the BLUP analysis. Many BLUP analyses have not fulfilled the narrow-sense heritability. Thus, the GEBVs could not be predicted accurately. The application of GWAS to the BLUP was a success in MC_L, CWT, protein, BF analyses and partly a success in MC_L, CWT, protein BF analyses. However, the number of SNPs required to predict the GEBVs better, can be a controversy.

The genomic relationship matrix, GRM statistically a variance-covariance matrix, uses the whole SNP information. On the contrary, partial GRM which uses the SNP information in part was a major concern in our study. We used the partial GRM because it can be variance-covariance matrix. The partial GRM which was constructed by using the SNP information in part ($p<0.01$ in GWAS) cannot matter because it can be a variance-covariance matrix.

IMPLICATIONS

We applied the Genome-wide Association Study (GWAS) to complement the best linear unbiased prediction (BLUP). The criteria of selected SNPs in the BLUP analysis was p-value <0.01 in GWAS. We concluded that analysis of BLUP with SNPs (p-value <0.01) had a better performance than that of total SNPs in terms of narrow-sense heritability. However, whether the criteria of p-value can predict GEBVs better, remains a controversy for the future.

CONFLICT OF INTEREST

We certify that there is no conflict of interest with any financial organization regarding the material discussed in the manuscript.

ACKNOWLEDGMENTS

This work was supported by a grant 'Researcher capacity-building project for R&D on biological resources to prepare for the Nagoya Protocol' from the National Institute of Biological Resources (NIBR), funded by the Ministry of Environment (MOE) of the Republic of Korea (NIBR No. 2013-02-071).

REFERENCES

Bolormaa, S., J. E. Pryce, K. Kemper, K. Savin, B. J. Hayes, W. Barendse, Y. Zhang, C. M. Reich, B. A. Mason, and R. J. Bunch et al. 2013. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in bos taurus, bos indicus, and composite beef cattle. J. Anim. Sci. 91:3088-3104.

Davis, C. G. and B.-H. Lin. 2005. Factors affecting us pork consumption US Department of Agriculture, Economic Research Service, Washington, DC, USA.

Eichler, E. E., J. Flint, G. Gibson, A. Kong, S. M. Leal, J. H. Moore, and J. H. Nadeau. 2010. Missing heritability and strategies for finding the underlying causes of complex disease. Nat. Rev. Genet. 11:446-450.

Endelman, J. B. 2011. Ridge regression and other kernels for genomic selection with r package rrblup. Plant Genome 4:250-255.

Feero, W. G., A. E. Guttmacher. and T. A. Manolio. 2010. Genomewide association studies and assessment of the risk of disease. N. Engl. J. Med. 363:166-176.

Fernando, R. L., J. C. Dekkers, and D. J. Garrick. 2014. A class of bayesian methods to combine large numbers of genotyped and non-genotyped animals for whole-genome analyses. Genet. Sel. Evol. 46:50.

Goddard, M., B. J. Hayes, and T. H. E. Meuwissen. 2011. Using the genomic relationship matrix to predict the accuracy of genomic

selection. J. Anim. Breed. Genet. 128:409-421.

Henderson, C. R. 1975. Best linear unbiased estimation and prediction under a selection model. Biometrics 31:423-447.

Hennig, C. 2010. Fpc: Flexible procedures for clustering. R package version 2:0-3. https://cran.r-project.org/web/packages/fpc/, Accessed August 13, 2015.

Jiang, J. 1997. A derivation of blup—best linear unbiased predictor. Stat. Probabil. Lett. 32:321-324.

Lee, T., D.-H. Shin, S. Cho, H. S. Kang, S. H. Kim, H.-K. Lee, H. Kim, and K.-S. Seo. 2014a. Genome-wide association study of integrated meat quality-related traits of the duroc pig breed. Asian Australas. J. Anim. 27:303-309.

Lee, Y.-S., H.-J. Kim, S. Cho, and H. Kim. 2014b. The usage of an snp-snp relationship matrix for best linear unbiased prediction (blup) analysis using a community-based cohort study. Genome. Inform. 12:254-260.

Leeds, T. D. 2005. Pork Quality Improvement: Estimates of Genetic Parameters and Evaluation of Novel Selection Criteria. Ph.D. Thesis, The Ohio State University, Columbus, OH, USA.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656-4663.

Manolio, T. A., F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, and A. Chakravarti et al. 2009. Finding the missing heritability of complex diseases. Nature 461:747-753.

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

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. De Bakker, and M. J. Daly. 2007. Plink: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559-575.

Sellier, P. 1998. Genetics of meat and carcass traits. The Genetics of the Pig (Eds. M. Rothschild and A. Ruvinsky). CAB International Wallingford, Oxon, UK: 463-510.

Sherman, J. and W. J. Morrison. 1950. Adjustment of an inverse matrix corresponding to a change in one element of a given matrix. Ann. Math. Stat. 21:124-127.

Soller, M., S. Weigend, M. N. Romanov, J. C. M. Dekkers, and S. J. Lamont. 2006. Strategies to assess structural variation in the chicken genome and its associations with biodiversity and biological performance. Poult. Sci. 85:2061-2078.

Woodbury, M. A. 1950. Inverting modified matrices. Memorandum report, Princeton University, Princeton, NJ USA, 42:106.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, and G. W. Montgomery. 2010. Common snps explain a large proportion of the heritability for human height. Nat. Genet. 42:565-569.

Zhang, H., Z. Wang, S. Wang, and H. Li. 2012a. Progress of genome wide association study in domestic animals. J. Anim. Sci. Biotech. 3:26.

Zhang, Z., J. He, H. Zhang, P. Gao, M. Erbe, H. Simianer, and J. Li. 2014. Results of genome wide association studies improve the accuracy of genomic selection. 10th world congress on genetics applied to livestock production, The Westin Bayshore, Vancouver, BC, Canada, #695 (the poster number).