Estimated Genetic Variance Explained by Single Nucleotide Polymorphisms of Different Minor Allele Frequencies for Carcass Traits in Japanese Black Cattle

Shinichiro Ogawa[1]*, Hirokazu Matsuda[1], Yukio Taniguchi[1], Toshio Watanabe[2], Yoshikazu Sugimoto[3], Hiroaki Iwaisaki[1]

[1]Graduate School of Agriculture, Kyoto University, Kyoto, Japan
[2]National Livestock Breeding Center, Nishigo, Fukushima, Japan
[3]Shirakawa Institute of Animal Genetics, Japan Livestock Technology Association, Nishigo, Fukushima, Japan
Email: *sogawa@kais.kyoto-u.ac.jp

Abstract

Japanese Black cattle are a beef breed and well known to excel in carcass quality, but the details of genetic architectures for carcass traits in beef breeds including this breed are still poorly understood. The objective of this study was to estimate the degree of additive genetic variance explained by single nucleotide polymorphism (SNP) marker groups with different levels of minor allele frequency (MAF) for marbling score and carcass weight in Japanese Black cattle. Phenotypic data on 872 fattened steers with the genotype information about 40,000 autosomal SNPs were analyzed using two different statistical models: one considering only SNPs selected based on MAF (model 1) and the other also considering all remaining SNPs as the different term (model 2). All available SNPs were classified into 10 groups based on their MAFs. For both traits, the estimated proportions of additive genetic variance explained by SNPs selected based on their MAFs using model 1 were always higher than the estimated ones using model 2. For carcass weight, relatively high values of the proportion of the additive genetic variance were estimated when using SNPs with MAFs which were in the ranges of 0.20 to 0.25 and 0.25 to 0.30, which may be partly due to the three previously-reported quantitative trait loci candidate regions. The results could have provided some information on the genetic architecture for the carcass traits in Japanese Black cattle, although its validity may be limited, mainly due to the sample size and the use of simpler statistical models in this study.

---

*Corresponding author.

Keywords

Additive Genetic Variance, Carcass Trait, Japanese Black Cattle, Minor Allele Frequency, Single Nucleotide Polymorphism

1. Introduction

Japanese Black cattle are the main breed in the Japanese native cattle, so-called Wagyu. This breed is well known to be a beef one and excel in "shimofuri" in Japanese, or marbling. Degree of marbling and carcass weight are the most economically important carcass traits in this breed. Heritability estimates of 6 representative carcass traits including the two traits in the Japanese population of this breed are moderate to high [1], which were obtained using pedigree information. For this breed, 3 candidate regions of major gene quantitative trait loci (QTLs) for carcass weight have been reported (called *CW*-1, *CW*-2 and *CW*-3 [2]-[5]) and Nishimura *et al.* (2012) [5] suggested that one-third of additive genetic variance for this trait is explained by these 3 QTLs. Ogawa *et al.* (2014) [6] stated that degree of marbling in this breed is likely to be controlled by only many polygenes. However, the whole genetic architectures for these carcass traits are still poorly understood.

Genomic evaluation (GE) and selection (GS), which were proposed by Meuwissen *et al.* (2001) [7], exploit a model including the effects of a large number of genome-wide DNA markers such as high-density single nucleotide polymorphisms (SNPs). The simultaneous use of high-density SNP markers is expected to capture the whole amount of additive genetic variance of the trait of interest, being based on the assumption that each of all QTLs is in the status of linkage disequilibrium (LD) with one or more of those SNPs. When there is fully high LD between a QTL and a SNP, the frequency of a causative allele at the QTL and that of the allele at the SNP are considered to be almost the same. Therefore, SNP markers with different minor allele frequencies (MAFs) may capture a different part of additive genetic variance.

The objective of this study was, using two different statistical models, to investigate the degrees of additive genetic variance explained by genome-wide, high-density SNP marker groups which had different levels of MAF for marbling score and carcass weight in Japanese Black cattle.

2. Materials and Methods

2.1. Ethics Statement

Animal care and use was according to the protocol approved by the Shirakawa Institute of Animal Genetics Animal Care and Use Committee, Nishigo, Japan (ACUCH21-1).

2.2. Phenotypic Data

Marbling scores and cold carcass weights of 872 Japanese Black fattened steers whose ages ranged from 15.3 to 43.0 months were used for the current analyses. These records were collected from 2000 through 2009 at two large carcass markets in Japan, namely Tokyo Metropolitan Central Wholesale Market and Osaka Municipal South Port Wholesale Market. Marbling score is the degree of marbling ranging from null (1) to very abundant (12), assessed on the ribeye of the carcass dissected at the sixth and the seventh rib section, according to the Japan carcass grading standards [8]. Originally, these steers were chosen based on their marbling scores to perform genome-wide association studies. Consequently, while the records of carcass weight followed a bell shape distribution, those of marbling score did a bimodal one which had a peak in the lower and the upper tails each. Mean and standard deviation was 496.6 and 48.0 kg for carcass weight and 6.8 and 3.5 point for marbling score. Pedigree information on the animals was not available in this study.

2.3. Genotype Data

DNA samples of the steers were extracted from perirenal adipose tissues. Sample DNA was quantified and genotyped using the BovineSNP50v1BeadChip. The genotype data on a total of 38,467 autosomal SNPs were used, which had the values of MAFs and genotype call rates larger than 0.01 and 0.95, respectively, and were in Har-

dy-Weinberg equilibrium (p value > 0.001). Missing genotypes were filled using "BEAGLE 3.3.2" [9]. **Figure 1** is a histogram of MAFs of the available 38,467 SNPs after missing genotypes were filled. Based on MAF, all available SNPs were classified into one of ten groups for the current analyses.

2.4. Statistical Analyses

A linear model (denoted as model 1) was used to analyze the data:

$$y = Xb + g_1 + e \tag{1}$$

where y is the vector of records, b is the vector including overall mean and macro-environmental effects, or discrete effects of carcass markets and years at slaughter and continuous effects of months of age at slaughter (linear and quadratic covariates), g_1 is the vector of genomic breeding values based on the selected n SNPs of animals, or the sums of the additive effects of the selected SNPs, e is the vector of residuals, and X is an incidence matrix relating b with y. When analyzing with model 1, the mean and the variance of y were assumed, as follows:

$$E(y) = Xb \tag{2}$$

$$V(y) = G_1\sigma_{g_1}^2 + I\sigma_e^2 \tag{3}$$

where $\sigma_{g_1}^2$ is the additive genetic variance explained by the n SNPs, σ_e^2 is the residual variance, G_1 is the genomic relationship matrix (G matrix) among animals which were constructed using the genotype data on the selected n SNPs and I is an identity matrix.

The data were also analyzed using a following linear model (model 2):

$$y = Xb + g_1 + g_2 + e \tag{4}$$

where g_1 is as defined in model 1 and g_2 is the vector of genomic breeding values based on the "remaining SNPs", or the (38,467 − n) SNPs. The mean and the variance of y were assumed, as follows:

$$E(y) = Xb \tag{5}$$

$$V(y) = G_1\sigma_{g_1}^2 + G_2\sigma_{g_2}^2 + I\sigma_e^2 \tag{6}$$

where $\sigma_{g_1}^2$ and G_1 are defined as in model 1, $\sigma_{g_2}^2$ is the additive genetic variance captured by the (38,467 − n) SNPs, G_2 is the G matrix constructed using the genotype data on the (38,467 − n) SNPs.
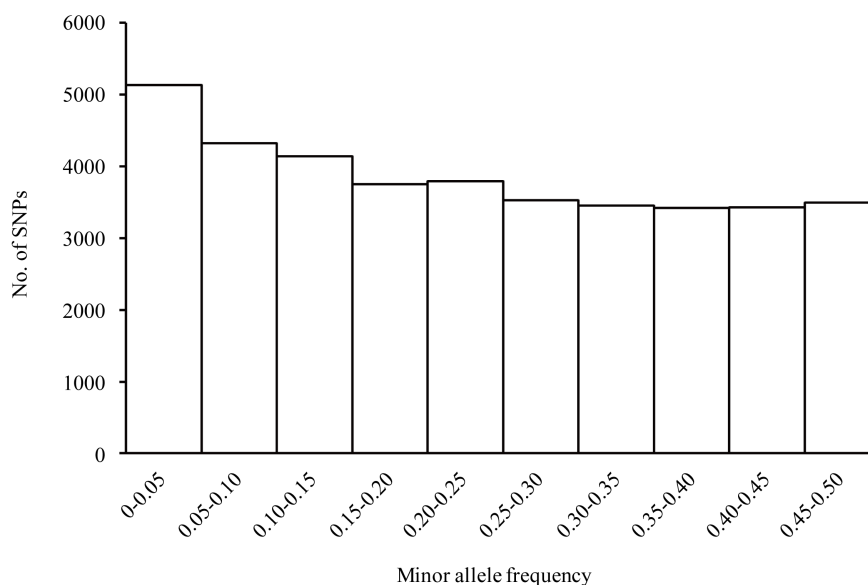

Figure 1. Histogram of minor allele frequencies of all SNPs available in this study.

According to VanRaden (2008) [10], the G matrix G_1 was constructed, as follows:

$$G_1 = (M_1 - 2P_1)(M_1 - 2P_1)' / \sum_{i=1}^{n} 2p_i(1-p_i) + I10^{-4} \tag{7}$$

where M_1 is the matrix whose row elements include the number of minor alleles at each locus of the selected n SNPs of each animal, P_1 is the matrix whose row elements contain the MAF at each locus of the selected n SNPs, and p_i is the MAF at the ith SNP locus. To always make positive definite, $I10^{-4}$ was added to G_1 in its construction. A total of 11 different G_1 matrices were used, including the one constructed using all available SNPs and 10 matrices constructed using only a part of all available SNPs which were selected based on their MAFs.

Correlation coefficients were calculated between the elements of G_1 constructed using all available SNPs and those of G_1 using only the selected n SNPs. Single regression coefficients were also calculated, where the independent and dependent variables were the elements of the former and the latter G_1, respectively. The correlation and single regression coefficients were calculated for diagonal and upper-triangular elements separately.

All the parameters in the two models were estimated via the Bayesian framework using Gibbs sampling in "BLR" package under R environment [11] [12]. In the analyses, the term g_2 in model 2 was converted into $(M_2 - 2P_2)u_2$ and treated as a ridge regression term where u_2 is the vector of the allele substitution effects of the $(38,467 - n)$ SNPs, because only one G matrix could be considered at one time. The prior distributions for b were a flat distribution, and those for g_1 and u_2 were multivariate normal ones with mean vector $\mathbf{0}$ and co-variance matrix $G_1\sigma_{g_1}^2$ and mean vector $\mathbf{0}$ and co-variance matrix $I\sigma_{u_2}^2$, respectively. As prior distributions for $\sigma_{g_1}^2$, $\sigma_{u_2}^2$ and σ_e^2, independent scaled inverted chi-square distributions were used with degree of belief and scale parameters of −2 and 0, respectively, assuming that there was no prior information. A single chain of 110,000 samples was run, and the first 10,000 samples were discarded as burn-in. Samples obtained after the burn-in period were used with thinning rate of 10.

The $\sigma_{g_2}^2$ was estimated calculating $\sum_{i=1}^{38467-n} 2p_i(1-p_i)\sigma_{u_2}^2$. Phenotypic variance ($\sigma_y^2$) was estimated computing, from each sample, the value of $\sigma_{g_1}^2 + \sigma_e^2$ when using model 1 and that of $\sigma_{g_1}^2 + \sigma_{g_2}^2 + \sigma_e^2$ when using model 2. The following two values (h_1^2 and h_2^2) were also calculated:

$$h_1^2 = \sigma_{g_1}^2 / \sigma_y^2 \tag{8}$$

$$h_2^2 = (\sigma_{g_1}^2 + \sigma_{g_2}^2) / \sigma_y^2 \tag{9}$$

Posterior probability densities of the variance components were computed with the R function "density" using samples obtained, and the estimated h_1^2 and h_2^2 were the modes of the corresponding posterior probability densities.

For both traits, $\hat{g}_1$ obtained with G_1 constructed using all available SNPs in model 1 was compared with the three different predicted values, or $\hat{g}_1$ obtained with G_1 constructed using the selected n SNPs in model 1, that in model 2 and $\hat{g}_1 + \hat{g}_2$ in model 2, calculating correlation and single regression coefficients. In the regressions, the values of the independent variable were always the elements of $\hat{g}_1$ obtained with G_1 constructed using all available SNPs.

3. Results and Discussion

Results of correlation and single regression for the elements of the two different G matrices are shown in **Table 1**. For diagonal elements, the value of correlation coefficient was 0.94 and highest when selecting the SNP groups with MAF bins of 0.20 - 0.25 and 0.25 - 0.30 and became lower when selecting SNPs with MAFs higher than 0.30 or lower than 0.20. The value of single regression coefficient was increased monotonically from 0.61 when selecting SNPs with MAFs in a MAF bin of 0.45 - 0.50 to 3.66 when selecting SNPs with MAFs in a MAF bin of 0 - 0.05. For upper-triangular elements, correlation coefficients were around 0.96 and relatively high when selecting the SNP groups with MAF bins of 0.15 - 0.20 to 0.45 - 0.50. Single regression coefficients were approximately around 1, or 0.99 to 1.06, and relatively high for the SNP groups with MAF bins of 0.15 - 0.20 to 0.45 - 0.50. These results could be mainly due to the distribution of SNP genotypes in the current population and may indicate the possibility of SNPs in different MAF groups to capture a different part of additive genetic variance for a given trait.

Table 1. Correlation and single regression coefficients for the elements of genomic relationship matrix constructed using all available SNPs and those constructed using selected SNPs based on their minor allele frequencies (MAFs).

MAF bin	No. of SNPs	Correlation coefficient		Single regression coefficient	
		Diagonal element	Upper-triangular element	Diagonal element	Upper-triangular element
0 - 0.05	5134	0.72	0.54	3.66	0.39
0.05 - 0.10	4325	0.84	0.84	2.08	0.75
0.10 - 0.15	4145	0.89	0.92	1.63	0.90
0.15 - 0.20	3750	0.92	0.95	1.28	0.99
0.20 - 0.25	3794	0.94	0.96	1.07	1.05
0.25 - 0.30	3526	0.94	0.96	0.84	1.02
0.30 - 0.35	3455	0.91	0.97	0.78	1.03
0.35 - 0.40	3419	0.87	0.96	0.69	1.05
0.40 - 0.45	3427	0.84	0.96	0.65	1.06
0.45 - 0.50	3492	0.79	0.96	0.61	1.05

Figure 2 illustrates the estimates of h_1^2 and h_2^2 for marbling score and carcass weight which were obtained using the two models and the different G matrices. For model 1 with G_1 constructed using all available SNPs, the estimates of h_1^2 were 0.67 and 0.55 for marbling score and carcass weight, respectively. These are the estimates of narrow sense heritability for the two traits, and the levels of the values are similar to those estimated previously using pedigree information in the Japanese Black population [1]. Therefore, it is likely that the use of all available SNPs, or about 40 thousand SNPs, captured a large proportion of additive genetic variance for each of the two traits.

When model 2 was used, the estimated h_2^2 were almost the same as the corresponding values of h_1^2 estimated with model 1 employing G_1 constructed using all available SNPs for both traits, but the estimated values of h_1^2 with model 2 using G_1 constructed using the selected SNPs were decreased from those with model 1 using the corresponding G_1 matrices. The estimated σ_y^2 and σ_e^2 were nearly the same with those estimated using model 1 with G_1 constructed using all available SNPs, respectively (data not shown). These mean that $\sigma_{g_1}^2$ estimated with model 1 using G_1 constructed using all available SNPs was likely to be partitioned into two different components. When analyzing with model 1 employing G_1 constructed using only the SNPs in a particular MAF group, while the estimated σ_y^2 was almost the same with that obtained using model 1 with G_1 constructed using all available SNPs, h_1^2 estimates were higher than the corresponding values of h_1^2 obtained using model 2 for both traits. This could partly be caused by the degree of LD in the whole-genome level which would be equal to or higher than those in other popular beef cattle breeds [6], resulting that the selected SNPs explained a more amount of additive genetic variance when the remaining SNPs were not considered in the analyses.

For both traits, the estimates of h_1^2 with model 2 were obviously changed, when different SNPs were selected and used. When selecting SNPs with MAFs in the middle ranges (MAF bin of 0.20 - 0.25 and 0.25 - 0.30), h_1^2 was estimated to be relatively high only for carcass weight. Nishimura *et al.* (2012) [5] reported the frequency of the critical SNP or haplotype in each of the three associated regions, *CW*-1, *CW*-2 and *CW*-3, showing that these frequencies were very similar with those of SNPs on the relevant 3 chromosomes in the BovineSNP50 Beadchip that showed the strongest association with carcass weight.

In this study, the frequencies corresponding alleles of SNPs in the chip associated with *CW-1* and *CW-2* were 0.799 and 0.229, respectively. So these two SNPs were the SNPs in the MAF range of 0.20 to 0.25, giving a high possibility that the additive genetic variance due to the two regions, *CW*-1 and *CW*-2, was captured by using the SNPs in MAF range of 0.20 to 0.25. Therefore, why relatively high values of h_1^2 with model 2 was estimated when selecting SNPs with MAFs in the middle ranges could be partly due to the presence of candidate regions of QTLs for carcass weight, *CW*-1 and *CW*-2.
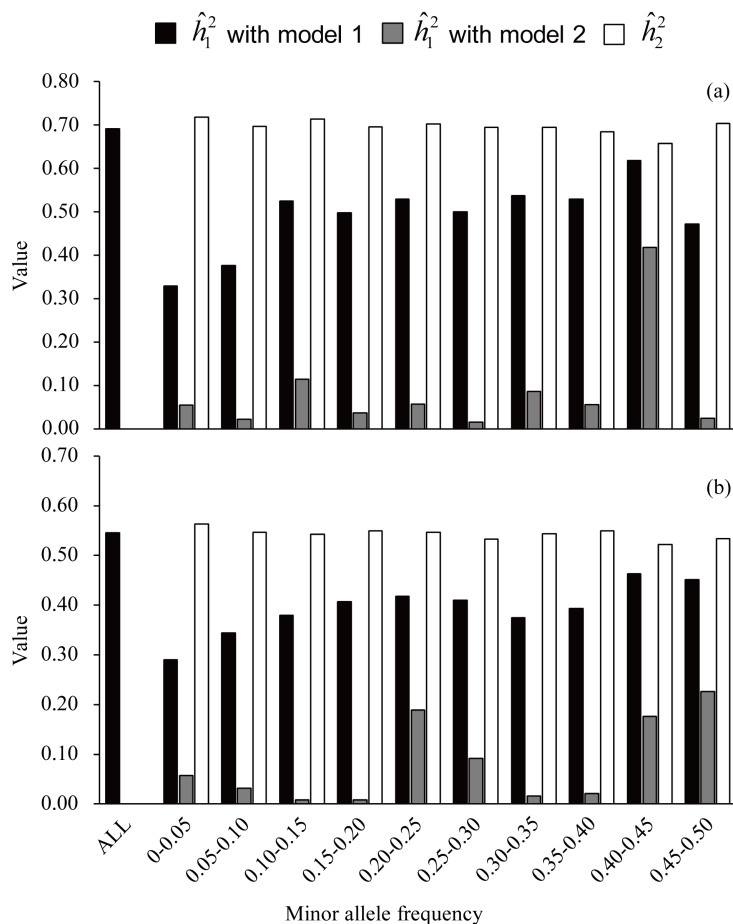
Figure 2. Changes in estimates of h_1^2 and h_2^2 for marbling score (a) and carcass weight (b): the proportion of additive genetic variance explained by the n SNPs selected based on their minor allele frequencies; h_2^2: the proportion of the sum of additive genetic variances explained by the n SNPs and the remaining SNPs with model 2.

Figure 3 and **Figure 4** show the changes in the correlation and regression coefficients, comparing $\hat{g}_1$ obtained with G_1 constructed using all available SNPs with the three different predicted values. For $\hat{g}_1 + \hat{g}_2$, values of the correlation were always higher than 0.99 for both traits. Regression coefficients were in the range of 0.99 to 1.04 for marbling score and 0.99 to 1.03 for carcass weight. However, for the $\hat{g}_1$ only, correlation and regression coefficients were decreased from those for $\hat{g}_1 + \hat{g}_2$. The degrees of decreases in both correlation and regressing coefficients were different, when SNPs of different MAF groups were selected. For $\hat{g}_1$ obtained with G_1 constructed using the selected SNPs, values of correlation and regression coefficients were somewhat higher than those for $\hat{g}_1$ with model 2 but lower than those when comparing with $\hat{g}_1 + \hat{g}_2$ in both traits. These results would be in line with the findings in variance component estimation.

Using the genotype data of high-density SNPs and fitting the model similar to model 2, Abdollahi-Arpanahi *et al.* (2014) [13] studied how markers with different MAFs differ in their abilities to explain additive genetic variability for three production traits in chicken. Then, they stated that it was difficult to draw clear conclusion from the obtained results, which was a setting similar to the current study. While it was assumed with model 2 that there was no covariance structure between the selected and the remaining SNPs, there may be the possibility to exist the high degree of LDs between the SNPs used to construct G_1 and the ones used to construct G_2. In our analyses, the values of h_1^2 estimated with model 1 were always higher than those of h_1^2 estimated with model 2 for both traits, when using the same G_1 (**Figure 2**), and the similar result were observed also in Abdollahi-Arpanahi *et al.* (2014) [13]. This fact would be a certain evidence of the presence of LD structures be-
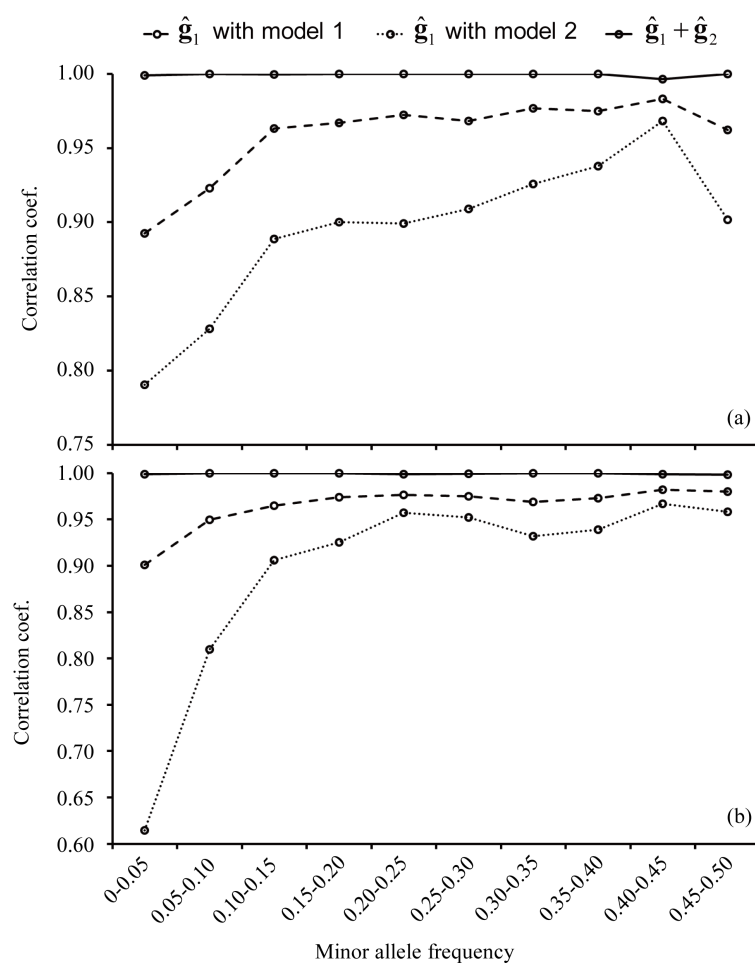
Figure 3. Changes in correlation coefficients between the elements of $\hat{g}_1$ obtained with G_1 constructed using all available SNPs in model 1 and the three different predicted values, or the elements of $\hat{g}_1$ obtained with G_1 constructed using the selected n SNPs in model 1 (dashed line), those of G_1 in model 2 (dotted line) and those of $\hat{g}_1 + \hat{g}_2$ in model 2 (solid line) for marbling score (a) and carcass weight (b).

tween SNPs of different MAF groups, and in such a situation, using model 2 could give less plausible results. A smaller sample population size often causes a spurious LD structure in the population, which would make the interpretation of the results more difficult. This problem could be mitigated by using more samples.

In this study, for both traits, the estimated proportions of additive genetic variance explained by SNPs selected based on their MAFs using model 1 were always higher than the estimated ones using model 2. For carcass weight, relatively high values of the proportion of the additive genetic variance were estimated when using SNPs with MAFs which were in the ranges of 0.20 to 0.25 and 0.25 to 0.30, which may be partly due to two of three previously-reported QTL candidate regions. The results could have provided some information on the genetic architecture for the carcass traits in Japanese Black cattle, although its validity may be limited, mainly due to the sample size and the use of simpler statistical models in this study. There will be other sources to characterize each of SNP markers (e.g., genome position information, gene function information and so on), and then these could give a chance to analyze with a different way to partition all available SNPs.

4. Conclusion

In this study, for marbling score and carcass weight in the Japanese Black cattle population, we tried to partition
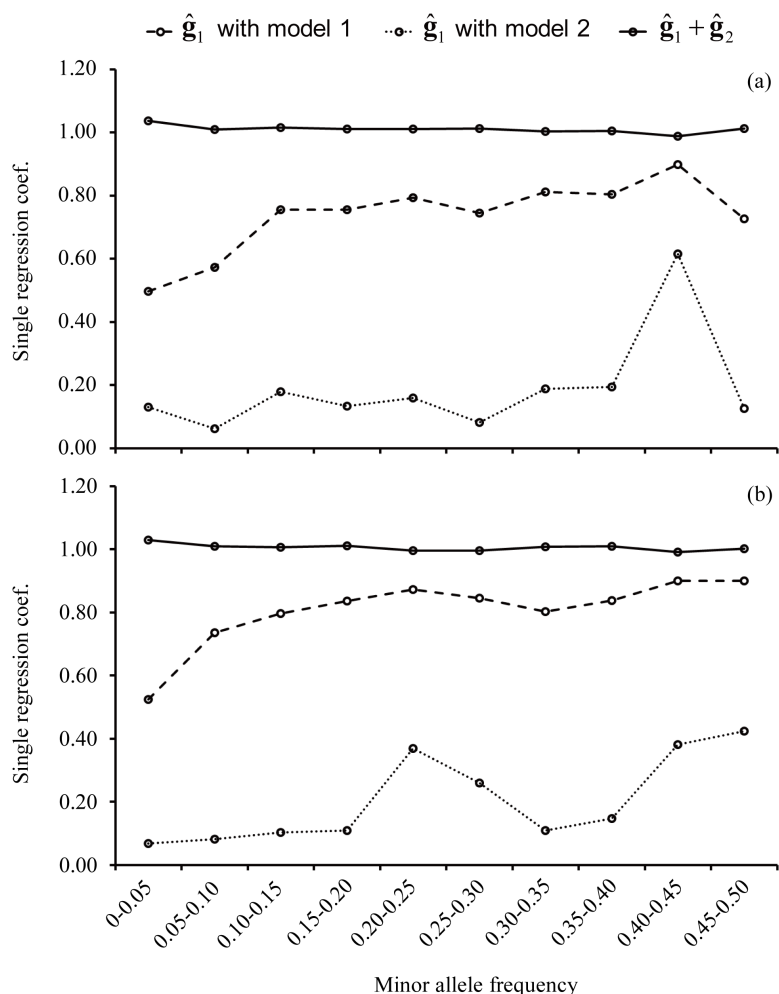
Figure 4. Changes in regression coefficients where the values of the independent variable were the elements of $\hat{\boldsymbol{g}}_1$ obtained with $\boldsymbol{G}_1$ constructed using all available SNPs and the values of the dependent variable were the elements of $\hat{\boldsymbol{g}}_1$ obtained with $\boldsymbol{G}_1$ constructed using selected n SNPs with model 1 (dashed line), those of $\hat{\boldsymbol{g}}_1$ in model 2 (dotted line), and those of $\hat{\boldsymbol{g}}_1 + \hat{\boldsymbol{g}}_2$ in model 2 (solid line) for marbling score (a) and carcass weight (b).

the additive genetic variances captured by genome-wide SNP markers into two different components, based on the information on their MAFs. Results indicated that the whole additive genetic variance captured by all available SNPs could be separately estimated as the two components. Using SNPs in different MAF ranges might explain different parts of the additive genetic variance for the carcass traits. For carcass weight, relatively high values of the proportion of the additive genetic variance were estimated when using SNPs with MAFs which were in the ranges of 0.20 to 0.25 and 0.25 to 0.30, which may be partly due to two of three previously-reported QTL candidate regions. The validity of the findings in this study is definitely necessary to examine using relevant more data and a more sophisticated statistical method.

Acknowledgements

References

[1] Oyama, K. (2011) Genetic Variability of Wagyu Cattle Estimated by Statistical Approaches. *Animal Science Journal*, **82**, 367-373. http://dx.doi.org/10.1111/j.1740-0929.2011.00895.x

[2] Mizoshita, K., Watanabe, T., Hayashi, H., Kubota, C., Yamakuchi, H., Todoroki, J. and Sugimoto, Y. (2004) Quantitative Trait Loci Analysis for Growth and Carcass Traits in a Half-Sib Family of Purebred Japanese Black (Wagyu) Cattle. *Journal of Animal Science*, **82**, 3415-3420.

[3] Takasuga, A., Watanabe, T., Mizoguchi, Y., Hirano, T., Ihara, N., Takano, A., Yokouchi, K., Fujikawa, A., Chiba, K., Kobayashi, N.,Tatsuda, K., Oe, T., Furukawa-Abe, A., Fujita, T., Inoue, K., Mizoshita, K., Ogino, A. and Sugimoto, Y. (2007) Identification of Bovine QTL for Growth and Carcass Traits in Japanese Black Cattle by Replication and Identical-by-Desceny Mapping. *Mammalian Genome*, **18**, 125-136. http://dx.doi.org/10.1007/s00335-006-0096-5

[4] Setoguchi, K., Furuta, M., Hirano, T., Nagao, T., Watanabe, T., Sugimoto, Y. and Takasuga, A. (2009) Cross-Breed Comparisons Identified a Critical 591-kb Region for Bovine Carcass Weight QTL (*CW-2*) on Chromosome 6 and the Ile-442-Met Substitution in *NCAPG* as a Positional Candidate. *BMC Genetics*, **10**, 43. http://dx.doi.org/10.1186/1471-2156-10-43

[5] Nishimura, S., Watanabe, T., Mizoshita, K., Tatsuda, K., Fujita, T., Watanabe, N., Sugimoto, Y. and Takasuga, A. (2012) Genome-Wide Association Study Identified Three Major QTL for Carcass Weight Including the *PLAG1-CHCHD7*QTL for Stature in Japanese Black Cattle. *BMC Genetics*, **13**, 40. http://dx.doi.org/10.1186/1471-2156-13-40

[6] Ogawa, S., Matsuda, H., Taniguchi, Y., Watanabe, T., Nishimura, S., Sugimoto, Y. and Iwaisaki, H. (2014) Effects of Single Nucleotide Polymorphism Marker Density on Degree of Genetic Variance Explained and Genomic Evaluation for Carcass Traits in Japanese Black Beef Cattle. *BMC Genetics*, **15**, 15. http://dx.doi.org/10.1186/1471-2156-15-15

[7] Meuwissen, T.H.E., Hayes, B.J. and Goddard M.E. (2001) Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. *Genetics*, **157**, 1819-1829.

[8] Japan Meat Grading Association (JMGA) (1988) New Beef Carcass Grading Standards. JMGA, Tokyo.

[9] Browning, S.R. and Browning, B.L. (2007) Rapid and Accurate Haplotype Phasing and Missing Data Inference for Whole Genome Association Studies by Use of Localized Haplotype Clustering. *The American Journal of Human Genetics*, **81**, 1084-1097. http://dx.doi.org/10.1086/521987

[10] VanRaden, P.M. (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science*, **91**, 4414-4423. http://dx.doi.org/10.3168/jds.2007-0980

[11] De los Campos, G., Naya, H., Gianola, D., Crossa, J., Legarra, A., Manfredi, E., Weigel, K. and Cotes, J.M. (2009) Predicting Quantitative Traits with Regression Models for Dense Molecular Markers and Pedigree. *Genetics*, **182**, 375-385. http://dx.doi.org/10.1534/genetics.109.101501

[12] R Development Core Team (2011) R: A Language and Environment for Statistical Computing. R Development Core Team, Vienna.

[13] Abdollahi-Arpanahi, R., Pakdel, A., Nejati-javaremi, A., Moradi Shahrbabak, M., Morota, G., Valente, B.D., Kranis, A., Rosa, G.J.M. and Gianola, D. (2014) Dissection of Additive Genetic Variability for Quantitative Traits in Chickens Using SNP Markers. *Journal of Animal Breeding and Genetics*, **131**, 183-193. http://dx.doi.org/10.1111/jbg.12079