



Identification of Superior Single Nucleotide Polymorphisms (SNP) Combinations Related to Economic Traits by Genotype Matrix Mapping (GMM) in Hanwoo (Korean Cattle)

Yoonseok Lee¹, Dong-Yep Oh², Yong-Won Lee³, Jung-Sou Yeo² and Jea-Young Lee^{3,*}

¹ Charmpoom Hanwoo Research Institution, Gyeongsan-si, Gyeongbuk 712-210, Korea

ABSTRACT : It is important to identify genetic interactions related to human diseases or animal traits. Many linear statistical models have been reported but they did not consider genetic interactions. Genotype matrix mapping (GMM) has been developed to identify genetic interactions. This study uses the GMM method to detect superior SNP combinations of the CCDC158 gene that influences average daily gain, marbling score, cold carcass weight and longissimus muscle dorsi area traits in Hanwoo. We evaluated the statistical significance of the major SNP combinations selected by implementing the permutation test of the F-measure. The effect of g.34425+102 A>T (AA), g.8778G>A (GG) and g.4102+36T>G (GT) SNP combinations produced higher performance of average daily gain, marbling score, cold carcass weight and the longissimus muscle dorsi area traits than the effect of a single SNP. GMM is a fast and reliable method for multiple SNP analysis with potential application in marker-assisted selection. GMM may prospectively be used for genetic assessment of quantitative traits after further development. (**Key Words :** Economic Traits, Genotype Matrix Mapping, Single Nucleotide Polymorphisms, Hanwoo, CCDC158 Gene)

INTRODUCTION

Genetic interactions are important factors to complex biological traits, because most biological phenotypes result from the complex interplay of many genes and environmental factors. Detection of genes responsible for economic traits in livestock has been widely practiced. Several methods have been proposed to analyze genetic interactions in beef cattle (Casas et al., 2006). A typical approach is to fit a multiple regression model and, thus, relating the trait values to marker genotypes. So far, identification of presumptive genetic interaction using the multiple regression models has been demonstrated only with inter-crossed populations. The multifactor dimensionality reduction method is also a commonly used approach. It uses nonparametric methods and is most efficiently used with case-control data (Ritchi et al., 2001;

Chung et al., 2006). Most traits of economic importance in livestock are multi-factorial, *i.e.*, influenced by multiple genes and their interactions with environmental factors. Generally, models used to test the effects of genes on traits have been based on parametric methods, such as general linear models or the animal model (Henderson, 1976). However, model building may be cumbersome and over parameterization problems can arise when multiple factors, *e.g.*, multiple genes and their interaction effects, are taken into account (Ritchi et al., 2001). These complex interactions make identification of individual genetic and genetic interactions difficult. A new genetic interaction approach, termed genotype matrix mapping (GMM), is introduced to reveal genetic interactions and interaction-interaction relationships in complex traits in family data and in various genetic backgrounds (Sachiko et al., 2007).

In this study, we utilize the GMM method to detect superior SNP combinations that influence economic traits of Hanwoo. The CCDC158 gene is significantly associated with growth and carcass traits (Lee et al., 2008; 2009; 2010). Thus, we try to identify interaction effects of SNPs of the CCDC158 gene for average daily gain (ADG), marbling score (MS), carcass cold weight (CWT) and longissimus muscle dorsi area (LMA) economic traits.

* Corresponding Author : Jea-Young Lee. Tel: +82-53-810-2326, Fax: +82-53-810-4615, E-mail: jlee@yu.ac.kr

² School of Biotechnology, Yeungnam University, Gyeongsan-si, Gyeongbuk 712-749, Korea.

³ Department of Statistics, Yeungnam University, Gyeongsan-si, Gyeongbuk 712-749, Korea.

Received April 15, 2011; Accepted August 19, 2011

MATERIALS AND METHODS

Animals and phenotypes

The Hanwoo population ($n = 476$) was reared under the progeny-testing program of the National Livestock Research Institute (NLRI) of Korea. The pedigree record of 476 steers was produced from 50 sires collected by the Korean Animal Improvement Association (Seoul, Korea). Steers were fed under the tightly controlled conditions of the feeding program in the Daekwanryeong and Namwon branches. The animals were born between the spring of 1998 and autumn of 2002. All steers were slaughtered in the spring of 2002 to autumn of 2004. They were castrated at six months of age and four animals were raised per pen (4 m×8 m). They were fed with concentrates consisting of 15% crude protein (CP)/71% very digestible nutrients (TDN) for a period of 60 to 90 days after six months of age; 15% CP/71% TDN for a period of 180 days; and 13% CP/72% TDN for a period of 90 to 120 days of self-feeding. Roughage was offered *ad libitum*, and steers had free access to fresh water throughout the entire period. Average daily gain measured the difference between the six month and 24 month weight divided by 18 (i.e., the difference between 6 and 24 months). Cold carcass weight was measured following a 24-h chill. The mean and standard deviation of average daily gain, cold carcass weight, longissimus muscle dorsi area and marbling score were 0.75 ± 0.089 kg, 316.75 ± 34.459 kg, 75.30 ± 8.114 cm² and 5.61 ± 4.176 , respectively.

SNP genotyping

Genomic DNA from white blood cells was extracted using the phenol-chloroform method (Sambrook et al., 2001). In this study, we genotyped 19 polymorphic SNP of coiled-coil domain containing 158 (Gene ID 534614), as described by Lee et al. (2010). Primers for amplification and extension were designed for single-base extension for genotyping of polymorphic sites (SBE) (Vreeland et al., 2002). Primer extension reactions were conducted using the SNaPshot ddNTP Primer Extension Kit (Applied Biosystems, Foster City, CA, USA). One unit of shrimp alkaline phosphatase (SAP) was added to the reaction mixture, which was incubated for 1 h at 37°C, followed by 15 min at 72°C for enzyme inactivation, to clean the primer extension reaction. DNA samples containing extension products and Genescan 120 LIZ size standard solution were added to HiDi formamide (Applied Biosystems, Foster City, CA, USA) in accordance with the manufacturer's recommendations. The mixture was incubated for 5 min at 95°C, followed by 5 min on ice, after which electrophoresis was conducted using the ABI PRISM 3130XL Genetic Analyzer. Results were analyzed using GeneMapper v4.0

software (Applied Biosystems, Foster City, CA, USA).

Genotype matrix mapping

In GMM, each marker is given a matrix in which each of the total number of alleles for the marker in the tested population is represented by intersecting lines and rows. Genetic interactions are estimated and compared via virtual networks generated among the locus matrixes. Any type of population, including unrelated individuals, family data and mapping populations for linkage analysis, can be used for GMM, providing that there is no population structure within the tested data set. The number of alleles in the population should be determined for every marker before analysis.

We have the following GMM procedure (Sachiko et al., 2007):

Construct a list of locus combinations. Each locus combination in the list is depicted on the genotype matrix of the marker.

We used the F-measure to evaluate the significance of the locus combinations;

Total set S , which consists of N individuals, is divided into two non-overlapping subclasses S_0 and S_1 based on their marker genotypes. Here, we have defined S_1 as the samples that have the specific genotype pattern to be evaluated, and S_0 as the samples that do not.

Next, the mean square among classes is calculated as follows:

$$MSA = \sum_{j=0}^1 |S_j| (\mu_j - \mu)^2 = |S_0| (\mu_0 - \mu)^2 + |S_1| (\mu_1 - \mu)^2$$

Here, μ , μ_0 , and μ_1 are the means of the phenotype values in S , S_0 , and S_1 , respectively. $|X|$ is the number of individuals in X .

The mean square within each class is defined as follows:

$$MSW = \frac{\sum_{j=0}^1 \sum_{S_i \in S_j} (P_i - \mu_j)^2}{N - 2} = \frac{\sum_{S_i \in S_0} (P_i - \mu_0)^2 + \sum_{S_i \in S_1} (P_i - \mu_1)^2}{N - 2}$$

Here, P_i is the phenotype value of the i -th individual S_i . The F-measure is obtained by dividing MSA by MSW

$$F = \frac{MSA}{MSW} = \frac{(N - 2) \left[|S_0| (\mu_0 - \mu)^2 + |S_1| (\mu_1 - \mu)^2 \right]}{\sum_{S_i \in S_0} (P_i - \mu_0)^2 + \sum_{S_i \in S_1} (P_i - \mu_1)^2}$$

F indicates the bias of the distribution of phenotype values in the two subclasses. If the distribution of the phenotype in the two subclasses differs, one can conclude that the condition (i.e., the pattern of marker genotypes)

used for sample division was associated with the phenotype of interest. In such cases, the *F*-measure value is large.

Significant superior locus combinations that have large *F*-measure values are searched incrementally. During the searching procedure, the maximum *F*-measure obtained is a new *F*, and the value of the new *F* is updated when a better combination whose *F*-measure is higher than that of the current one is found. After obtaining *F*-measures based on phenotype, we need to determine statistical significance for the critical value of *F*. An empirical 100 (1-P) percentile obtained by 10,000 repetitions of the permutation process was referred to as an estimated critical value of the whole genome significance level of *P*. Thus, a permutation test is performed to determine empirical significance of *F*-measures by applying the GMM method. We obtain major combinations of the upper 15 *F*-measures using this procedure.

Association between individual SNPs and ADG, CWT, LMA and MS were determined by the mixed effect model, treating "sire" as a random effect; "age" at slaughter was also included in the model as a covariate in the SPSS statistics v19.1 package. We used a single SNP model. Then, the genotype combination effects were tested in the mixed effect model. In addition, statistically significant superior SNP combinations were evaluated by the permutation test (Good et al., 2000), because *F*-measures obtained by GMM did not calculate their theoretical significant levels (critical value or *p*-value). Ten thousand repetitions of the permutation process were used for the critical value. We examined significance levels (*p*-value) of the major SNP combinations selected by implementing the permutation test of the *F*-measure. This had not been obtained by Sachiko et al. (2007).

RESULTS

Lee (2009) selected 6 polymorphic SNPs, which were haplotype-tagging SNP, and six SNPs, g.4102+36T>G, g.8778G>A, g.11500-117C>G, g.32330-48A>G, g.34425+102A>T, g.66995-169insdelC, and economic traits (ADG, CWT, LMA and MS) were used for the genotype mapping matrix approach. Using the GMM algorithm, we tentatively extracted the significant SNP combinations; the critical value of $p < 0.001$ was used to determine statistical significance for the critical values of the *F*-measure. Four kinds of beef main economic traits average daily gain (ADG), cold carcass weight (CWT), longissimus dorsi muscle area (LMA), and marbling score (MS) are applied to the GMM with least squares means and standard errors (SE), and the top 15 major SNP combinations were selected (Tables 1-4). The GMM analysis obtained 15 superior SNP combinations with least squares means, standard errors (SE) and permutation *p* values for ADG (Table 1). Similar

analysis and results from CWT, LMA and MS by GMM are presented in Table 2-4 respectively. Table 1-4 showed all five superior SNP combinations out of the top 15 that were commonly related to ADG, CWT, LMA and MS. They were g.32330-48A>G (AA) g.8778G>A (GG), g.34425+102A>T (AA) g.8778G>A (GG) g.4102+36T>G (GT), g.66995-169insdelC (ins) g.34425+102A>T (AA) g.8778G>A (GG), g.66995-169insdelC (ins) g.34425+102A>T (AA) g.32330-48A>G (AA) and g.66995-169insdelC (CC) g.34425+102A>T (AA) g.32330-48A>G (AA) combinations (Table 5).

The four economic traits and five major SNP combinations groups were selected in Table 5 with least squares means, standard errors (SE) and statistical significance values (*p*-value) of the SNP genotypes for ADG, CWT, LMA and MS respectively (Table 5). These superior SNP combinations for comprehensive economic traits are used in large-scale Hanwoo. These selected five combinations of superior SNPs from each of the economic traits were used to determine whether they were statistically significant (Table 6). Table 5 was selected to investigate the influence of SNP combinations on economic traits of beef; these were tested. The selected all five SNP combinations of ADG and CWT were statistically significant ($p < 0.01$) but g.66995-169insdelC (CC), g.32330-48A>G (AA), g.8778G>A (GG) SNP genotype combinations and g.66995-169insdelC (CC), g.34425+102A>T (AA), g.8778G>A (GG) SNP genotype combinations and g.66995-169insdelC (CC), g.34425+102A>T (AA), g.32330-48A>G (AA) SNP genotype combinations were not significant or negatively significant for MS. Their *T* statistic (*p*-value) were -2.30 (0.022), -1.29 (0.198) and -1.74 (0.084) respectively. By the way, g.34425+102 A>T (AA), g.8778G>A(GG) and g.4102+36 T>G (GT) SNP genotype combination was very significant with $p < 0.001$. Table 6 presented least-square means, and standard error of g.34425+102 A>T (AA), g.8778G>A (GG) and g.4102+36 T>G (GT) SNP combinations. *T*- and permutation test are also represented the best among SNP combinations. For g.34425+102A>T (AA), g.8778G>A (GG) and g.4102+36T>G (GT) SNP genotype combinations, the single and combination effects were obtained by applying the GMM method to ADG, CWT, LMA and MS respectively. The SNP genotype combinations produced superior performance of four economic traits when compared to the total means of the Hanwoo population. In Table 6, the AAGGGT genotype group Mean±SE revealed 0.83±0.013 for AGT, 348.72±0.367 for CWT, 77.78±0.514 for LMA and 7.35±0.526 for MS. Those scores are superior to the 476 sample means and standard deviations (in materials and methods section), 0.75±0.089 kg, 316.75±34.459 kg, 75.30±8.114 cm² and 5.61±4.176, for ADG, CWT, LMA and MS respectively. Statistical tests were also significant ($p < 0.001$).

Table 1. Significant SNP combinations of top 15 F-measures, least square means and standard errors from average daily gain using GMM in Hanwoo

SNP combinations	Genotype	N ¹	LS mean±SE ²	F-measure	T-value	p-value
g.66995-169insdelC g.8778G>A g.4102+36T>G	Ins/GG/GT Others	72	0.857±0.010 0.748±0.001	111.543	10.56	<0.001
g.66995-169insdelC g.32330-48A>G g.8778G>A	Ins/AA/GG	112	0.826±0.008 0.747±0.001	88.469	9.40	<0.001
g.66995-169insdelC g.11500-117C>G g.8778G>A	Ins/GG/GG Others	112	0.826±0.008 0.747±0.001	88.469	9.40	<0.001
g.66995-169insdelC g.8778G>A	Ins/GG Others	112	0.826±0.008 0.747±0.001	88.469	9.40	<0.001
g.66995-169insdelC g.32330-48A>G g.4102+36T>G	Ins/AA/GT Others	99	0.828±0.011 0.748±0.001	82.387	7.09	<0.001
g.66995-169insdelC g.34425+102A>T g.8778G>A	Ins/AA/GG Others	66	0.843±0.013 0.748±0.001	75.707	7.31	<0.001
g.34425+102A>T g.8778G>A g.4102+36T>G	AA/GG/GT Others	72	0.831±0.013 0.748±0.001	62.383	6.25	<0.001
g.11500-117C>G g.4102+36T>G	GG/GT Others	1,074	0.768±0.003 0.743±0.002	59.685	7.54	<0.001
g.66995-169insdelC g.34425+102A>T g.32330-48A>G	Ins/AA/AA Others	162	0.798±0.007 0.747±0.001	51.305	7.16	<0.001
g.11500-117C>G g.8778G>A g.4102+36T>G	GG/GG/GT Others	262	0.787±0.007 0.747±0.001	49.342	5.59	<0.001
g.32330-48A>G g.11500-117C>G g.4102+36T>G	AA/GG/GT Others	561	0.774±0.004 0.746±0.001	48.895	6.53	<0.001
g.32330-48A>G g.8778G>A	AA/GG Others	1,943	0.760±0.002 0.741±0.002	47.734	6.88	<0.001
g.66995-169insdelC g.34425+102A>T g.11500-117C>G	(ins/del)/AT/GG Others	645	0.771±0.003 0.745±0.002	46.855	6.84	<0.001
g.66995-169insdelC g.32330-48A>G	Ins/AA Others	218	0.789±0.007 0.747±0.001	45.861	6.19	<0.001
g.66995-169insdelC g.32330-48A>G g.11500-117C>G	Ins/AA/GG Others	218	0.789±0.007 0.747±0.001	45.861	6.19	<0.001

¹ Number of animal. ² Standard error.

Table 2. Significance SNP combinations of top 15 F-measures, least square means and standard errors from carcass cold weight using GMM in Hanwoo

SNP combinations	Genotype	N ¹	LS mean±SE ²	F-measure	t-value	p-value
g.32330-48A>G	AA/CG/GG	915	325.1±0.994	83.346	9.74	<0.001
g.11500-117C>G	Others		313.9±0.590			
g.8778G>A						
g.11500-117C>G	CG/GG	915	325.1±0.994	83.346	9.74	<0.001
g.8778G>A	Others		313.9±0.590			
g.32330-48A>G	AA/GG	1,943	321.4±0.747	84.243	9.18	<0.001
g.8778G>A	Others		312.0±0.697			
g.66995-169insdelC	ins/del/AA/GG	752	325.7±1.219	74.302	8.62	<0.001
g.32330-48A>G	Others		314.3±0.562			
g.8778G>A						
g.66995-169insdelC	ins/del/GG	752	325.7±1.219	74.302	8.62	<0.001
g.8778G>A	Others		314.3±0.562			
g.34425+102A>T	AA/GG/GT	72	348.7±3.670	70.348	8.39	<0.001
g.8778G>A	Others		315.8±0.516			
g.4102+36T>G						
g.66995-169insdelC	Ins/AA/GG	66	350.1±3.612	69.851	8.35	<0.001
g.34425+102A>T	Others		315.8±0.516			
g.8778G>A						
g.32330-48A>G	AA/GG/GT	408	328.7±1.974	63.405	6.72	<0.001
g.8778G>A	Others		315.0±0.525			
g.4102+36T>G						
g.8778G>A	GG/GT	408	328.7±1.974	63.405	6.72	<0.001
g.4102+36T>G	Others		315.0±0.525			
g.34425+102A>T	AA/AA/GG	137	337.2±2.255	58.910	9.53	<0.001
g.32330-48A>G	Others		315.6±0.523			
g.8778G>A						
g.34425+102A>T	AA/GG/GG	137	337.2±2.255	58.910	9.53	<0.001
g.11500-117C>G	Others		315.6±0.523			
g.8778G>A						
g.34425+102A>T	AA/GG	137	337.2±2.255	58.910	9.53	<0.001
g.8778G>A	Others		315.6±0.523			
g.66995-169insdelC	ins/del/CG/GG	382	328.2±1.760	53.715	7.33	<0.001
g.11500-117C>G	Others		315.1±0.535			
g.8778G>A						
g.66995-169insdelC	Ins/AA/AA	162	334.9±2.446	52.998	7.28	<0.001
g.34425+102A>T	Others		315.6±0.523			
g.32330-48A>G						
g.66995-169insdelC	Ins/AA/GG	112	337.6±2.750	47.174	6.87	<0.001
g.32330-48A>G	Others		315.8±0.521			
g.8778G>A						

¹ Number of animal. ² Standard error.

Table 3. Significance SNP combinations of top 15 F-measures, least square means and standard errors from longissimus muscle dorsi area using GMM in Hanwoo

SNP combinations	Genotype	N ¹	LS mean±SE ²	F-measure	t-value	p value
g.32330-48A>G	AG/AG/GG	146	79.966±0.796	66.005	6.46	<0.001
g.8778G>A	Others		74.763±0.118			
g.4102+36T>G						
g.32330-48A>G	AG/GG/GG	335	77.537±0.525	42.122	5.24	<0.001
g.11500-117C>G	Others		74.719±0.120			
g.4102+36T>G						
g.32330-48A>G	AG/GG	335	77.537±0.525	42.122	5.24	<0.001
g.4102+36T>G	Others		74.719±0.120			
g.66995-169insdelC	ins/del/AG/GG	29	84.035±0.270	41.519	31.09	<0.001
g.8778G>A	Others		74.881±0.119			
g.4102+36T>G						
g.11500-117C>G	GG/AG/GG	227	78.084±0.587	40.682	5.54	<0.001
g.8778G>A	Others		74.765±0.120			
g.4102+36T>G						
g.8778G>A	AG/GG	227	78.084±0.587	40.686	5.54	<0.001
g.4102+36T>G	Others		74.765±0.120			
g.34425+102A>T	AA/AA/GT	204	78.191±0.578	38.868	6.23	<0.001
g.32330-48A>G	Others		74.779±0.120			
g.4102+36T>G						
g.66995-169insdelC	ins/del/GG/GG	74	80.054±0.493	33.778	10.26	<0.001
g.11500-117C>G	Others		74.853±0.120			
g.4102+36T>G						
g.66995-169insdelC	ins/del/GG	74	80.054±0.493	33.778	10.26	<0.001
g.4102+36T>G	Others		74.853±0.120			
g.34425+102A>T	AA/AG/GG	325	77.273±0.534	32.911	4.62	<0.001
g.32330-48A>G	Others		74.749±0.120			
g.4102+36T>G						
g.34425+102A>T	AA/AA	386	76.990±0.361	30.517	5.89	<0.001
g.32330-48A>G	Others		74.737±0.125			
g.4102+36T>G						
g.34425+102A>T	AA/AA GG	386	76.990±0.361	30.517	5.89	<0.001
g.32330-48A>G	Others		74.737±0.125			
g.11500-117C>G						
g.34425+102A>T	AA/AG/GG	217	77.719±0.603	30.228	4.76	<0.001
g.8778G>A	Others		74.793±0.120			
g.4102+36T>G						
g.34425+102A>T	AA/GG/AG	584	76.551±0.327	30.063	5.48	<0.001
g.11500-117C>G	Others		74.684±0.126			
g.8778G>A						
g.34425+102A>T	AA/AG	584	76.551±0.327	30.063	5.48	<0.001
g.8778G>A	Others		74.684±0.126			

¹ Number of animal. ² Standard error.

Table 4. Significance SNP combinations of top 15 F-measures, least square means and standard errors from marbling score using GMM in Hanwoo

SNP combinations	Genotype	N ¹	LS mean±SE ²	F-measure	t-value	p-value
g.66995-169insdelC	ins/del/AG/CG	521	6.202±0.120	63.769	6.97	<0.001
g.32330-48A>G	Others		4.748±0.063			
g.11500-117C>G						
g.34425+102A>T	AT/AG/CG	521	6.202±0.120	63.769	6.97	<0.001
g.32330-48A>G	Others		4.748±0.063			
g.11500-117C>G						
g.32330-48A>G	AG/CG	521	6.202±0.120	63.769	6.97	<0.001
g.11500-117C>G	Others		4.748±0.063			
g.32330-48A>G	AG/CG/AG	521	6.202±0.120	63.769	6.97	<0.001
g.11500-117C>G	Others		4.748±0.063			
g.8778G>A						
g.32330-48A>G	AG/CG/GT	521	6.202±0.120	63.769	6.97	<0.001
g.11500-117C>G	Others		4.748±0.063			
g.4102+36T>G						
g.66995-169insdelC	ins/del/CG/GT	852	5.853±0.158	60.427	6.80	<0.001
g.11500-117C>G	Others		4.693±0.064			
g.4102+36T>G						
g.34425+102A>T	AT/CG/GT	872	5.810±0.155	56.428	6.63	<0.001
g.11500-117C>G	Others		4.697±0.064			
g.4102+36T>G						
g.11500-117C>G	CG/GT	872	5.810±0.155	56.428	6.63	<0.001
g.4102+36T>G	Others		4.697±0.064			
g.66995-169insdelC	ins/del/CG/AG	706	5.865±0.170	49.113	6.20	<0.001
g.11500-117C>G	Others		4.739±0.064			
g.8778G>A						
g.34425+102A>T	AT/AG/GT	798	5.792±0.155	48.987	6.40	<0.001
g.32330-48A>G	Others		4.725±0.065			
g.4102+36T>G						
g.34425+102A>T	AT/AG	808	5.787±0.153	48.619	6.41	<0.001
g.32330-48A>G	Others		4.724±0.065			
g.34425+102A>T	AT/AG/AG	808	5.787±0.153	48.619	6.41	<0.001
g.32330-48A>G	Others		4.724±0.065			
g.8778G>A						
g.66995-169insdelC	ins/del/AG	894	5.711±0.142	45.858	6.36	<0.001
g.32330-48A>G	Others		4.717±0.066			
g.34425+102A>T	AT/CG/AG	726	5.813±0.166	45.191	6.01	<0.001
g.11500-117C>G	Others		4.744±0.064			
g.8778G>A						
g.11500-117C>G	CG/AG	726	5.813±0.166	45.191	6.01	<0.001
g.8778G>A	Others		4.744±0.064			

¹ Number of animal. ² Standard error.

Table 5. Level of t- and permutation test, p value, least square means and standard errors of 5 SNP combinations for 4 economic traits in Hanwoo

SNP combinations	Traits ¹	Genotype	LS mean±SE ²	Bootstrapping data		
				N ³	T-test (p-value)	Permutation test (p-value)
g.66995-169insdelC g.32330-48A>G g.8778G>A	ADG	CC/AA/GG Others	0.826±0.009 0.747±0.001	112	9.40(<0.000)	0.000
	CWT	CC/AA/GG Others	337.6±2.750 315.8±0.521	112	6.87(<0.000)	0.000
g.66995-169insdelC g.34425+102A>T g.8778G>A	LMA	CC/AA/GG Others	77.143±0.431 74.884±0.121	112	5.04(<0.000)	0.000
	MS	CC/AA/GG Others	4.089±0.359 4.952±0.061	112	-2.30(0.022)	0.000
	ADG	CC/AA/GG Others	0.843±0.013 0.748±0.001	66	7.31(<0.000)	0.000
	CWT	CC/AA/GG Others	350.1±3.672 315.8±0.516	66	8.35(<0.000)	0.000
g.34425+102A>T g.8778G>A g.4102+36T>G	LMA	CC/AA/GG Others	78.258±0.549 74.892±0.120	66	5.99(<0.000)	0.000
	MS	CC/AA/GG Others	5.546±0.543 4.919±0.061	66	1.29(0.198)	0.000
	ADG	AA/GG/GT Others	0.831±0.013 0.748±0.001	72	6.25(<0.000)	0.000
	CWT	AA/GG/GT Others	348.7±3.670 315.8±0.516	72	8.39(<0.000)	0.000
g.66995-169insdelC g.34425+102A>T g.32330-48A>G	LMA	AA/GG/GT Others	77.778±0.514 74.900±0.120	72	5.46(<0.000)	0.000
	MS	AA/GG/GT Others	7.347±0.526 4.887±0.061	72	5.30(<0.000)	0.000
	ADG	CC/AA/AA Others	0.798±0.007 0.747±0.001	162	7.16(<0.000)	0.000
	CWT	CC/AA/ AA Others	334.9±2.446 315.6±0.523	162	7.28(<0.000)	0.000
g.32330-48A>G g.8778G>A	LMA	CC/AA/ AA Others	76.290±0.411 74.891±0.122	162	3.26(0.001)	0.000
	MS	CC/AA/ AA Others	4.469±0.268 4.947±0.062	162	-1.74(0.084)	0.000
	ADG	AA/GG Others	0.760±0.002 0.741±0.002	1,943	6.88(<0.000)	0.000
	CWT	AA/GG Others	321.4±0.747 312.0±0.700	1,943	9.18(<0.000)	0.000
g.32330-48A>G g.8778G>A	LMA	AA/GG Others	75.207±0.171 74.718±0.164	1,943	2.06(0.039)	0.000
	MS	AA/GG Others	4.763±0.087 5.073±0.084	1,943	-2.55(0.011)	0.000

¹ Number of animal. ² Standard error.³ ADG = Average daily gain, CWT = Carcass cold weight, LMA = Longissimus muscle dorsi area, MS = Marbling score.

In addition, three SNP combination genotype groups represented higher performance than the effect of a single SNP genotype in Table 6. Therefore, we concluded g.34425+102A>T (AA), g.8778G>A (GG) and g.4102+36T>G (GT) SNP combinations, in beef complex economic traits, were decided as the best SNP combinations.

DISCUSSION

A novel method, GMM, was applied to reveal interactions between various SNP genotypes in complex traits. This study used the bootstrap sampling method (Efron et al., 1993) and generated 4,190 animals. The top 15 SNPs combinations were selected by the GMM method for ADG, CWT, LMA and MS. The permutation test was added for statistical significant in this GMM method (Tables 1-4).

The five common superior SNP combinations with beef complex economic traits were selected and statistical tests conducted (Table 5). In addition, we discovered g.34425+102A>T (AA), g.8778G>A (GG) and g.4102+36T>G (GT) SNP combinations that were best commonly affected for ADG, CWT, LMA and MS within the CCDC158 gene and screened in a large-scale Hanwoo population (n = 4,190). The other four SNP combinations did not enable the choice of complex economic traits in Table 5. In contrast, the average for the GG genotype of g.8778G>A SNP was 319.88 kg for cold carcass weight, as described by Lee et al. (2010); however, in this study, the average for genotype of g.34425+102A>T (AA), g.8778G>A (GG) and g.4102+36T>G (GT) SNP combinations was 348.72 kg. It was superior to 28.84 kg in the GG genotype of g. 8778 G>A SNP. The effect of SNP

Table 6. Least square means and standard errors (SE), single and combinations effect of superior 3 SNP for average daily gain and carcass traits in large-scale Hanwoo population

Traits ¹	SNP	Genotype	N ²	LS mean±SE	T-value	p ³
ADG	g.34425+102A>T	AA	1,200	0.745±0.003	-2.06	0.040
		Other	2,990	0.751±0.002		
	g.8778G>A	GG	1,946	0.760±0.002	6.88	0.000
		Other	2,244	0.741±0.002		
	g.4102+36T>G	GT	1,947	0.749±0.002	-0.34	0.730
		Other	2,243	0.750±0.002		
	Combinations	AAGGGT	72	0.831±0.013	6.25***	0.001
		Other		0.748±0.001		
CWT	g.34425+102A>T	AA	1,200	315.68±0.929	-0.83	0.409
		Other	2,990	316.60±0.618		
	g.8778G>A	GG	1,946	321.32±0.746	9.10	0.000
		Other	2,244	312.01±0.699		
	g.4102+36T>G	GT	1,947	314.63±0.795	-3.07	0.002
		Other	2,243	317.81±0.669		
	Combinations	AAGGGT	72	348.72±0.367	8.39***	0.001
		Other		315.77±0.516		
LMA	g.34425+102A>T	AA	1,200	75.462±0.220	2.77	0.006
		Other	2,990	74.737±0.140		
	g.8778G>A	GG	1,946	75.225±0.171	2.21	0.027
		Other	2,244	74.701±0.164		
	g.4102+36T>G	GT	1,947	74.818±0.168	-1.00	0.316
		Other	2,243	75.055±0.166		
	Combinations	AAGGGT	72	77.778±0.514	5.46***	0.001
		Other		74.895±0.120		
MS	g.34425+102A>T	AA	1,200	4.9183±0.110	-0.11	0.912
		Other	2,990	4.9331±0.073		
	g.8778G>A	GG	1,946	4.7797±0.088	-2.3	0.022
		Other	2,244	5.0584±0.084		
	g.4102+36T>G	GT	1,947	5.1177±0.092	2.89	0.004
		Other	2,243	4.7652±0.080		
	Combinations	AAGGGT	72	7.3472±0.526	5.30***	0.001
		Other		4.8866±0.060		

¹ ADG = Average daily gain, CWT = Carcass cold weight, LMA = Longissimus muscle dorsi area, MS = Marbling score.

² Number of animals. ³ Permutation p value. *** p<0.001.

combinations produced very higher performance of ADG, CWT, LMA and MS traits than did the effect of single SNP (Table 6).

Therefore, we suggested that g.34425+102 A>T (AA), g.8778G>A (GG) and g.4102+36T>G (GT) SNP combinations were the best SNP combinations on growth and carcass traits in Hanwoo. GMM is a fast and reliable method for multiple SNP analysis of potential application in marker-assisted selection. However, effectual data size is an indispensable issue to be evaluated. One of the most probable reasons for insufficient output from the GMM analysis is the limited size of the data set. Increasing the number of individuals for data acquisition would improve the accuracy of the entire analysis. GMM may prospectively be used for genetic assessment of quantitative traits after further development.

ACKNOWLEDGEMENTS

This research was supported by a Yeungnam University research grant in 2009.

REFERENCES

- Casas, E., S. N. White, T. L. Wheeler, S. D. Shackelford, M. Koohmaraie, D. G. Riley, C. C. Chase, Jr., D. D. Johnson and T. P. L. Smith. 2006. Effect of calpastatin and u-calpain markers in beef cattle on tenderness traits. *J. Anim. Sci.* 84:520-525.
- Chung, Y. J., S. Y. Lee and T. S. Park. 2006. Multifactor dimensionality reduction in the presence of missing observations. *Journal of Korea Statistical Society, Proceedings of the Autumn Conference, Korea.* pp. 31-36.
- Churchill, G. A. and R. W. Deorge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.
- Efron, B. and R. Tibshirani. 1993. *An Introduction to the bootstrap.* Chapman & Hall/CRC.
- Good, P. 2000. *Permutation test: a practical guide to resampling method for testing hypotheses.* Springer-Verlag Berlin and Heidelberg GmbH & Co., New York.
- Henderson, C. R. 1976. A Simple method for computing the inverse of a numerator relationship matrix used in predicting of breeding values. *Biometrics* 32:69-83.
- Lee, J. Y., J. C. Kwon and J. J. Kim. 2008. Multifactor dimensionality reduction (MDR) analysis to detect single nucleotide polymorphisms associated with a carcass trait in a Hanwoo population. *Asian-Aust. J. Anim. Sci.* 21(6):784-788.
- Lee, J. Y. and J. Y. Goh. 2009. Selection of the principal genotype with genetic algorithm. *Journal of the Korean Data & Information Science Society.* 20(4):639-647.
- Lee, J. Y. and H. G. Lee. 2009. Multifactor dimensionality reduction (MDR) analysis by dummy variables. *The Korean Journal of Applied Statistics.* 22(2):435-442.
- Lee, J. Y., Y. W. Lee and J. S. Yeo. 2007. Bootstrapping of Hanwoo chromosome17 based on BMS1167 microsatellite locus. *Journal of the Korean Data & Information Science Society.* 18(1):175-184.
- Lee, Y. S. 2009. Study on the identification of candidate genes and their haplotypes that are associated with growth and carcass traits in the QTL region of BTA6 in a Hanwoo population, Ph. D. Thesis, University of Yeungnam, Gyeongbuk, Korea.
- Lee, Y. S., J. H. Lee, J. Y. Lee, J. J. Kim, H. S. Park and J. S. Yeo. 2008. Identification of candidate SNP (single nucleotide polymorphism) for growth and carcass traits related to QTL on chromosome 6 in Hanwoo (Korean cattle). *Asian-Aust. J. Anim. Sci.* 21(12):1703-1709.
- Lee, Y. S., D. Y. Oh, J. J. Kim, J. H. Lee, H. S. Park and J. S. Yeo. 2010. A single nucleotide polymorphism in LOC534614 as an unknown gene associated with body weight and cold carcass weight in Hanwoo (Korean cattle). *Asian-Aust. J. Anim. Sci.* 23(12):1543-1551.
- Ritchie, M. D., L. W. Hahn, N. Roodi, L. R. Bailey, W. D. Dupont, F. F. Parl and J. H. Moore. 2001. Multifactor-dimensionality reduction reveals high-order interactions among estrogen-metabolism genes in sporadic breast cancer. *Am. J. Hum. Genet.* 69(1):138-147.
- Sachiko, I., N. Akihiro and T. Satoshi. 2007. Genotype matrix mapping: searching for quantitative trait loci interactions in genetic variation in complex trait. *DNA Res.* 14:217-225.
- Sambrook, J. and D. W. Russell. 2001. *Molecular cloning: A laboratory manual.* Vol. 1. 3rd edition: Cold Spring Harbor. New York. pp. 6.4-6.12.
- Snelling, W. M., E. Casas, R. T. Stone, J. W. Keele, G. P. Harhay, G. L. Bennett and T. P. L. Smith. 2005. Linkage mapping bovine EST-based SNP. *BMC Genomics* 6(1):74-84.
- Vreeland, W. N., R. J. Meagher and A. E. Barron. 2002. Multiplexed, highthroughput genotyping by single base extension and end labeled free solution electrophoresis. *Anal. Chem.* 74:4328-4333.