Identification of Recently Selected Mutations Driven by Artificial Selection in Hanwoo (Korean Cattle)

Dajeong Lim1, Cedric Gondro2, Hye Sun Park1, Yong Min Cho1, Han Ha Chai1, Hwan Hoo Seong1, Bo Suk Yang3, Seong Koo Hong1, Won Kyung Chang1 and Seung Hwan Lee1,3*

1 Animal Genomics and Bioinformatics Division, National Institute of Animal Science, RDA, Suwon 441-706, Korea
2 School of Environmental and Rural Science, University of New England, Armidale, NSW 2351, Australia
3 Hanwoo Experiment Station, National Institute of Animal Science, RDA, PyeongChang 232-950, Korea

ABSTRACT: Hanwoo have been subjected over the last seventy years to intensive artificial selection with the aim of improving meat production traits such as marbling and carcass weight. In this study, we performed a signature of selection analysis to identify recent positive selected regions driven by a long-term artificial selection process called a breeding program using whole genome SNP data. In order to investigate homozygous regions across the genome, we estimated iES (Integrated Extended Haplotype Homozygosity SNP) for each SNP. As a result, we identified two highly homozygous regions that seem to be strong and/or recent positive selection. Five genes (DPH5, OLFM3, S1PR1, LRRN1 and CRBN) were included in this region. To go further in the interpretation of the observed signatures of selection, we subsequently concentrated on the annotation of differentiated genes defined according to the iES value of SNPs localized close or within them. We also described the detection of the adaptive evolution at the molecular level for the genes of interest. As a result, this analysis also led to the identification of OLFM3 as having a strong signal of selection in bovine lineage. The results of this study indicate that artificial selection which might have targeted most of these genes was mainly oriented towards improvement of meat production. (Key Words: Hanwoo, Signatures of Selection, SNP)

INTRODUCTION

Modern breeds of cattle were domesticated about 10,000 years ago to produce the distinct breed characteristics for milk or meat products from natural and human artificial selection (Bradley et al., 1999). During this history of artificial selection, mutations in genes that control important characteristics, such as high milk yield in modern dairy cows, have been selected to fixation. Hanwoo (Korean cattle) have become highly specialized for meat quality undergoing strong artificial selection (Yoon et al., 2008). Hanwoo have been intensively selected for marbling (intramuscular fat) through a progeny test in a breeding program since the 1930s. As a result of artificial selection, such as the Hanwoo progeny test, the breeding value for the marbling score could increase to 0.05 standard deviation (SD) in the Hanwoo population (Lee et al., 2011). Artificial selection might also affect genomic regions controlling Hanwoo marbling. Understanding the genetic mechanism leading to phenotypic differentiation requires identification of the genome regions that have been under long-term artificial selection.

This strong artificial selection will increase the frequency of favorable alleles at the loci affecting meat quality traits in the meat production breeds. In this process a small region of the genome surrounding the mutations is also selected, resulting in a small genome region that shows reduced variation. This region of reduced variation is referred to as a signature of selection that is identified by distributions of nucleotides around favorable mutations that differ statistically from that expected purely by chance (Kim and Stephan, 2002). Many methods have been developed for detection of selection signatures from genome analyses. Most of methods are used to compare the distribution of allelic frequencies by calculating population genetics statistics such as FST (Weir et al., 2005), linkage disequilibrium (Kim and Nielsen, 2004), Tajima (Tajima, 1989), Wu’s H-test (Fay and Wu, 2000) and the integrated Haplotype Score (iHS) (Voight et al., 2006), which is a
method based on extended haplotype homozygosity (EHH) statistics (Sabeti et al., 2002). By identifying the signatures of past selection and then identifying the functional genes and mutations involved, it is possible to identify the major genetic and metabolic pathways that control important agricultural characteristics of our modern breeds. When data are available from a large number of populations by large scale SNP data, the analysis can distinguish the genetic variation between similar populations. Searches for signatures of selection have successfully revealed many genes that are important in livestock. For example, the International Bovine HapMap project described a range of breeds that has been historically selected for different phenotypic traits. Hayes also proposed to identify divergently selected regions of the genome between dairy cattle and beef cattle breeds within Bos taurus (Hayes et al., 2009). MacEachern et al. reported the results of comparison of allelic frequencies between Australian Angus and Holstein cattle (MacEachern et al., 2009).

The aim of this study is to perform a test for selection signatures within the Hanwoo population that share a similar phenotype and to detect divergently selected genomic regions. Analysis of within-population selection signatures indicate that at least some mutations, which have been differentially selected in Hanwoo, are still segregated within its population. Finally, positional candidate genes are determined in proximity to the genomic positions showing the most significant indication of selection.

**MATERIALS AND METHODS**

**Animals and genotype assay**

Carcass data and DNA samples for QTL analysis were obtained from 266 Hanwoo steers descending from 66 sires and unrelated dams (2 to 10 progeny per sire) from two NIAS experimental stations, Dae-Kwan-Ryoung and Nam-Won. Genomic DNA for genotyping assays was extracted from a blood sample and SeoLin Bioscience (Seoul, Korea) performed the SNP genotyping using the Affymetrix MegAllele GeneChip Bovine Mapping 10K SNP array (Affymetrix Inc., 2006). Three hundred steers were genotyped but 34 steers failed to genotype due to low DNA quality from phenol and chloroform contamination. Genotype data were received on 8,344 SNP and all those SNP were physically mapped to a chromosome (in bp) using the bovine genome sequence (Btau-3.1).

**Analysis of SNP statistics**

Genotypes were tested for Hardy-Weinberg equilibrium (HWE) to identify possible typing errors using a chi-square test in R/SNPPassoc Package (R Development Core Team). SNP not in HWE (p<0.05), monomorphic SNPs and minor allele frequency (<1%) were removed in this study. Finally, a total of 8,344 SNPs, genotype data were received on 4,522 SNP.

**Extended haplotype homozygosity (EHH)**

The counting algorithm of Tang et al. (2007) was implemented for identifying differential extended haplotype homozygosity regions within Hanwoo (Korean cattle). A haplotype can be identified by patterns of SNPs. Haplotype maps can be used to determine complex genetic variations of inherited diseases or complex traits. For the proportion of homologous individuals, EHHS_{ij}, at the ith and jth SNP were calculated in two steps. First, for each SNP, EHHS_{i,j} between SNP, and incrementally distant flanking SNP, were calculated until EHHS_{i,j} k<0; this was performed for both j>i and j<i.

\[
EHHS_{\text{gene},ij} = \sum_{i=1}^{n} \frac{I_{k,(hap_{i})}(1 \text{ if } hap_{i} = hap_{j})}{\sum_{i=1}^{n} I_{i,(alle)}(1 \text{ if } allei = allej)}
\]

Second, the extended haplotype homozygosity of SNP, was calculated: iES_{i} = (EHHS_{i,j}) for i,j,k for the region 3’ of i (or i,j,k for the region 5’ of i).

\[
iES_{i} = \sum_{j=1}^{k} \frac{(EHHS_{i,j+1} + EHHS_{i,j})(Posj - Posj-1)}{2}
\]

Extended haplotype homozygous regions were plotted based on the standardized log-ratio of iES, within Hanwoo. We calculated a standardized integrated extended haplotype homozygosity (iES(z)) value to identify significant regions of positive selection (p = 0.0001, z = 3.5).

**Evolutionary analysis of genes in signatures of selection (SoS)**

We identified genes within the signature of selection and obtained orthologous genes for the four species Homo sapiens, Mus musculus, Sus scrofa, Gallus gallus from the Ensembl Compara database (Flicek et al., 2008), which reports pairwise conserved synteny relations based on nucleotide alignments. Protein sequences of the orthologous genes were aligned with ClustalW. The protein sequence alignment and the corresponding coding sequences were converted into codon alignment using the pal2nal program (http://coot.embl.de/pal2nal). We obtained the orthologous sequences information for five genes in candidate regions.

If amino acid changes are selectively neutral (i.e., mutations that are neither advantageous or deleterious), they will be fixed at the same rate as synonymous mutations and ω ratio (dS/dD) = 1. ω values > 1 are taken to indicate that amino acid changes are accumulating at a faster rate than is acceptable under a neutral mutation model. That is to say,
the rate of amino acid changes (d_N) significantly exceeds the rate of synonymous changes (d_S) at the DNA level. The codon-based likelihood model proposed by Goldman and Yang (Goldman and Yang, 1994) is implemented in the program codeml of PAML package (Yang, 1997). The program is useful for estimating synonymous and non-synonymous substitution rates (d_N/d_S). To estimate the d_N/d_S values, model 0 (M0, null model) with a single dN/dS value for all branches of the tree and model 2 (M2, positive selection) with the branches of interest were assumed to have different d_N/d_S ratios. We obtained a log-likelihood value for each model. We tested for positive selection by comparing twice the log-likelihood difference among models (M0 vs. M2) from a chi-square distribution with n−1 df, where n is the number of branches of the phylogeny in the LRT (Yang and Bielawski, 2000). If a significant p-value is obtained, it can be concluded that the positive selection model (M2) is the favored model. Next, models of variable selective pressures among amino acid sites were used to test for the presence of sites under positive selection. The four models (M1a, M2a, M7, M8) in the CODEML program of the PAML package were tested (Lynn et al., 2005).

RESULTS AND DISCUSSION

Identification of candidate genes in the positively selected region

Figure 1 shows the plot of iES scores for one strong candidate region identified by our genome-wide scan. We extended core regions in both directions up to 1.5 cM from a core SNP (rs29012432, 41.75 cM) and annotated a subset of genes in the core region. The total distance between the first point to the left and to the right of the core SNP is from 40.93 cM to 43.87 cM of BTA3. There is a clear clustering of high values into the region where some SNPs show evidence of selection. As a result, five genes were determined as putative targets of recent artificial selection as follows: diphthine synthase 5 (DPH5), sphingosine-1-phosphate receptor 1 (S1PR1), olfactomedin 3 (OLFM3), leucine-rich repeat neuronal protein 1 (LRRN1) and protein cereblon (CRBN).

Table 1. Summary statistics of the integrated EHHS (iES) values for selection signature in candidate genes

| Chromosome | Candidate region | Closest SNP name and position (bp) | iES value |
|------------|------------------|-----------------------------------|-----------|
| BTA3       | DPH5 (DPH5 homolog (S. cerevisiae)) | rs29020061 (40,929,695) | 7.21 |
|            | S1PR1(sphingosine-1-phosphate receptor 1) | rs29018907 (41,249,954) | 7.46 |
|            | OLFM3(olfactomedin 3) | rs29018230 (41,792,758) | 9.82 |
| BTA12      | LRRN1(Leucine-rich repeat neuronal protein 1) | rs29015171 (20,796,087) | 12.55 |
|            | CRBN (Protein cereblon) | rs29017072 (21,841,651) | 11.71 |

Figure 1. Genome wide extended haplotype homozygosity (EHH) profiling to detect signature of selection in Hanwoo.
Table 2. Gene Ontology and KEGG pathway of the candidate genes showing evidence for selection signatures

| Candidate gene | GO term | KEGG pathway |
|----------------|---------|--------------|
| DPH5           | Peptidyl-diphthamide biosynthetic process | - |
| S1PR1          | Angiogenesis (GO:0001525), cell adhesion (GO:0007155), G-protein coupled receptor protein signaling pathway (GO:0007186), inhibition of adenylate cyclase activity by G-protein signaling pathway (GO:0007193), brain development (GO:0007420) | Neuroactive ligand-receptor interaction |
| OLFM3          | Eye photoreceptor cell development (GO:0042462) | - |
| LRRN1          | Integral to membrane (GO:0016021) | - |
| CRBN           | Negative regulation of protein homooligomerization (GO:0032463), negative regulation of ion transmembrane transport (GO:0034766) | - |

Evidence of positive selection between species

We next implemented a further test to study interspecific divergence between bovine and the other species against the five candidate genes. It is important to know how these genes have been positively selected along bovine lineage with different selective pressures. The evolutionary forces operating on particular genes use the ratio of non-synonymous (\(d_{NS}\)) to synonymous (\(d_{SY}\)) substitution. To study differences in selection pressures, we conducted likelihood ratio tests comparing a one-ratio model to a two-ratio alternative model. Table 3 shows the results of the likelihood ratio test using different evolutionary models. Only one gene showed significant acceleration in the \(o\)-ratio on the bovine lineage. For OLFM3, the two-ratio (\(o = 0.12\)) models detected significant positive selection in bovine lineage (Figure 2). It suggests that OLFM3 is an accelerated protein evolution driven by positive selection or a relaxation of constraints. The branch shows evidence of

Table 3. Likelihood estimates of different evolutionary models (Model0 vs Model2)

| Gene Name | \(d_{NS}/d_{SY}\) | Degree of freedom | \(\chi^2\) | p-value |
|-----------|------------------|------------------|------------|--------|
| S1PR1     | 0.05             | 1                | 1.93       | NS     |
| DPH5      | 0.09             | 1                | 0.18       | NS     |
| OLFM3     | 0.12             | 1                | 31.47      | <0.001*|
| LRRN1     | 0.03             | 1                | 5.34       | NS     |
| CRBN      | 0.05             | 1                | 1.21       | NS     |

Degree of freedom is the difference in the number of parameters between evolutionary models. \(\chi^2\) is twice the difference of log likelihood between models. p-value is the probability that two models should differ in log likelihood given the degree of freedom.
Gene expression analyses identify genes affecting a...gous regions. In addition, we...nt bovine lineage because this...

---

**REFERENCES**

Andreae, L. C., D. Peukert, A. Lumsden and J. D. Gilthorpe. 2007. Analysis of Lrrn1 expression and its relationship to neuromeric boundaries during chick neural development. Neural. Dev. 2:22.

Bluhner, M., L. Wilson-Fritch, J. Leszyk, P. G. Laustsen, S. Corvera and C. R. Kahn , 2004. Role of insulin action and cell size on protein expression patterns in adipocytes. J. Biol. Chem. 279:31902-31909.

Bradley, D. G. and E. P. Cunningham. 1999. Genetic aspects of domestication. In: The genetics of cattle (Ed. R. Fries and A. Ruvinsky) pp. 15-31. CAB Publishing, Wallingford.

Clark, A. G., S. Glanowski, R. Nielsen, P. D. Thomas, A. Kejarival, M. A. Todd, D. M. Tanenbaum, D. Civello, F. Lu, B. Murphy, et al. 2003. Inferring nonneutral evolution from human-chimp-mouse orthologous gene trios. Science 302:1960-1963.

Fay, J. C. and C. I. Wu. 2000. Hitchhiking under positive Darwinian selection. Genetics 155:1405-1413.

Flichek, P., B. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham and T. Cutts. 2008. Ensembl 2008. Nucleic Acids Res. 36 (Suppl 1):D707.

Goldman, N. and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11:725-736.

Hannun, Y. A. and L. M. Obeid. 2008. Principles of bioactive lipid signalling: lessons from sphingolipids. Nat. Rev. Mol. Cell Biol. 9:139-150.

Hayes, B. J., A. J. Chamberlain, S. Macachern, K. Savin, H. McPartlan, I. MacLeod, L. Setheruman and M. E. Goddard. 2009. A genome map of divergent artificial selection between Bos taurus dairy cattle and Bos taurus beef cattle. Anim. Genet. 40:176-184.

Kim, Y. and R. Nielsen. 2004. Linkage disequilibrium as a signature of selective sweeps. Genetics 167:1513-1524.

Kim, Y. and W. Stephan. 2002. Detecting a local signature of genetic hitchhiking along a recombining chromosome. Genetics 160:765-777.

Kimura, T., K. Sato, E. Malchinkhuu, H. Tomura, K. Tamama, A. Kuvabara, M. Murakami and F. Okajima. 2003. High-density lipoprotein stimulates endothelial cell migration and survival through sphingosine 1-phosphate and its receptors. Arterioscler Thromb. Vasc. Biol. 23:1283-1288.

Lee, K. M., S. Jo, H. Kim, J. Lee and C. S. Park. 2011. Functional modulation of AMP-activated protein kinase by cereblon. Biochim. Biophys. Acta. 1813:448-455.

Lee, S. H., J. H. van der Werf, N. K. Kim, S. H. Lee, C. Gondro, E. W. Park, S. J. Oh, J. P. Gibson and J. M. Thompson. 2011. QTL and gene expression analyses identify genes affecting carcass weight and marbling on BTA14 in Hanwoo (Korean Cattle). Mamm. Genome 22:589-601.

Lee, W. J., M. Kim, H. S. Park, H. S. Kim, M. J. Jeon, K. S. Oh, E.
H. Koh, J. C. Won, M. S. Kim and G. T. Oh. 2006. AMPK activation increases fatty acid oxidation in skeletal muscle by activating PPARα and PGC-1. Biochem. Biophys. Res. Commun. 340:291-295.

Lynn, D. J., A. R. Freeman, C. Murray and D. G. Bradley. 2005. A genomics approach to the detection of positive selection in cattle: adaptive evolution of the T-cell and natural killer cell-surface protein CD2. Genetics 170:1189-1196.

MacEachern, S., B. Hayes, J. McEwan and M. Goddard. 2009. An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (Bos taurus) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in Domestic cattle. BMC Genomics 10:181.

Sabeti, P. C., D. E. Reich, J. M. Higgins, H. Z. Levine, D. J. Richter, S. F. Schaffner, S. B. Gabriel, J. V. Platko, N. J. Patterson and G. J. McDonald, et al. 2002. Detecting recent positive selection in the human genome from haplotype structure. Nature 419:832-837.

Sakai, Y., T. Fujita, T. Sasaki, S. Miyata, K. Nagata, B. E. Nichols. 2000. Measures of human population structure show positive selection in the human genome from haplotype structure. Nature 419:832-837.

Sharma, N. K., S. K. Das, A. K. Mondal, O. G. Hackney, W. S. Chu, P. A. Kern, N. Rauoli, H. J. Spencer, A. Yao-Borengasser and S. C. Elbein. 2008. Endoplasmic reticulum stress markers are associated with obesity in nondiabetic subjects. J. Clin. Endocrinol. Metab. 93:4532-4541.

Sharon, D., G. Glusman, Y. Pilpel, M. Khen, F. Gruetzner, T. Haaf and D. Lancet. 1999. Primate evolution of an olfactory receptor cluster: diversification by gene conversion and recent emergence of pseudogenes. Genomics 61:24-36.

Stone, E. M., J. H. Fingert, W. L. Alward, T. D. Nguyen, J. R. Polansky, S. L. Sunden, D. Nishimura, A. F. Clark, A. Nystuen and B. E. Nichols. 1997. Identification of a gene that causes primary open angle glaucoma. Science 275:668-670.

Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585-595.

Tang, K. K. R. Thornton, M. Stoneking. 2007. A new approach for using genome scans to detect recent positive selection in the human genome. PLoS Biol. 5:e171.

Tang, S., J. P. Perry, D. Sheehan, D. J. Buckley and P. A. Morrissey. 2001. Antioxidative effect of added tea catechins on susceptibility of cooked red meat, poultry and fish patties to lipid oxidation. Food Res. Int. 34:651-657.

Torrado, M., R. Trivedi, R. Zinovieva, I. Karanova and S. I. Tomarev. 2002. Optimedin: a novel olfactomedin-related protein that interacts with myocilin. Hum. Mol. Genet. 11:1291-1301.

Vought, B. F., S. Kudaravalli, X. Wen and J. K. Pritchard. 2006. A map of recent positive selection in the human genome. PLoS Biol. 4:e72.

Wei, Y., D. Wang and M. J. Pagliassotti. 2007. Saturated fatty acid-mediated endoplasmic reticulum stress and apoptosis are augmented by trans-10, cis-12-conjugated linoleic acid in liver cells. Mol. Cell Biochem. 303:105-113.

Weir, B. S., L. R. Cardon, A. D. Anderson, D. M. Nielsen and W. G. Hill. 2005. Measures of human population structure show heterogeneity among genomic regions. Genome Res. 15:1468-1476.

Yamada, T., S. Sasaki, S. Sukegawa, T. Miyake, T. Fujita, H. Kose, M. Morita, Y. Takahagi, H. Murakami and F. Morimatsu et al. 2010. Novel SNP in 5’ flanking region of EDG1 associated with marbling in Japanese Black beef cattle. Anim. Sci. J. 80:486-489.

Yang, Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 13:555-556.

Yang, Z. and J. P. Bielawski. 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15:496-503.

Yoon, D., Y. S. Kwon, K. Y. Lee, W. Y. Jung, S. Sasazaki, H. Mannen, J. T. Jeon and J. H. Lee. 2008. Discrimination of Korean cattle (Hanwoo) using DNA markers derived from SNPs in bovine mitochondrial and SRY genes. Asian Australas. J. Anim. Sci. 21:25-28.