



Linkage Disequilibrium and Effective Population Size in Hanwoo Korean Cattle

S. H. Lee*, Y. M. Cho¹, D. Lim², H. C. Kim, B. H. Choi², H. S. Park², O. H. Kim³,
S. Kim¹, T. H. Kim², D. Yoon⁴ and S. K. Hong

Hanwoo Experiment Station, National Institute of Animal Science, RDA, PyeongChang 232-950 Korea

ABSTRACT : This study presents a linkage disequilibrium (LD) analysis and effective population size (N_e) for the entire Hanwoo Korean cattle genome, which is the first LD map and effective population size estimate ever calculated for this breed. A panel of 4,525 markers was used in the final LD analysis. The pairwise r^2 statistic of SNPs up to 50 Mb apart across the genome was estimated. A mean value of $r^2 = 0.23$ was observed in pairwise distances of <25 kb and dropped to 0.1 at 40 to 60 kb, which is similar to the average inter-marker distance used in this study. The proportion of SNPs in useful LD ($r^2 \geq 0.25$) was 20% for the distance of 10 and 20 kb between SNPs. Analyses of past effective population size estimates based on direct estimates of recombination rates from SNP data demonstrated that a decline in effective population size to $N_e = 98.1$ occurred up to three generations ago. (**Key Words :** Linkage Disequilibrium (LD), Effective Population Size (N_e), Hanwoo)

INTRODUCTION

The Hanwoo breed of cattle is native to Korea, and their history as a draft animal dates back at least 5,000 years. One popular theory is that the Hanwoo breed originated from a cross between zebu and taurine cattle that migrated to the Korean peninsula through North China and Mongolia (Han, 1996). Alternatively, the Hanwoo may have originated as a hybrid between the aurochs, *Bos primigenius* and zebu (Lee et al., 2002). More recently, Yoon et al. (2005) investigated the genetic diversity of 19 cattle breeds including Hanwoo using 11 microsatellite and mitochondrial DNA sequences. The genetic structure of the Hanwoo breed was completely different from zebu and was also distinct from the European and African taurine breeds. These results suggest that the Hanwoo breed did not

originate as a hybrid of taurine and zebu cattle, but rather originated independently from the European and African breeds.

Linkage disequilibrium (LD) is defined as non-random associations between alleles at different loci within a population. Information about LD has become important in the context of QTL fine mapping and population parameter estimation (Tenesa et al., 2007). The extent of LD across genomic regions, using LD information between markers and QTLs, is an important parameter to determine the statistical power of genome wide association studies with single nucleotide polymorphism (SNPs) (Meuwissen et al., 2001). In natural populations, LD is affected by genetic drift, migration, selection, mutation and recombination (Hedrick, 1987). The HapMap project (The international HapMap Consortium, 2005) has demonstrated that the underlying LD structure of the human genome can be defined in terms of discrete blocks (coldspots; low recombination regions) separated by hotspots (high recombination regions). Average LD decay with increasing distance between markers appears to vary according to population structure and parameters such as effective population size (Khatkar et al., 2007; McKay et al., 2007). In Australian dairy cattle, the average LD (r^2) decayed to 0.3 within 50 kb genomic regions (Khatkar et al., 2007). In other cattle breeds (Japanese Black, Angus, Brahmin and

* Corresponding Author : SeungHwan Lee. Tel: +82-33-330-0717, Fax: +82-33-330-0660, E-mail: slee46@korea.kr

¹ Research Policy Bureau, Rural Development Administration, Suwon, 441-706, Korea.

² Animal Genomics and Bioinformatics Division, National Institute of Animal Science, RDA, Suwon, 441-706, Korea.

³ Department of Animal Science, Kon-Kuk University, ChungJu, 380-701, Korea.

⁴ Department of Animal Science, Kyung-Pook National University, Sangju, 742-711, Korea.

Received June 1, 2011; Accepted September 22, 2011

Holstein) the average LD declined to 0.2 within 100 kb genomic regions (McKay et al., 2007).

As genome wide association studies become popular for the detection of QTLs, estimates of LD and population structure across cattle breeds are becoming important parameters for detecting precise QTLs. Recently, Decker et al. (2009) resolved the evolutionary relationships of extant and extinct ruminants using 16,353 animals representing 61 cattle breeds and 70 species genotyped by the Illumina BovineSNP50 BeadChip. Therefore, quantifying the extent of LD in the bovine genome will be a first step for determining the number of markers that will be sufficient for QTL mapping using LD information.

As part of a larger ongoing QTL mapping study based on LD, the goal of the present study was to estimate the LD pattern and effective population size over all autosomes in Hanwoo Korean cattle.

MATERIAL AND METHODS

DNA samples and genotype assays

DNA samples were obtained from 266 Hanwoo steers descending from 66 sires and unrelated dams (2-5 progeny number per sire) from two NIAS experimental stations, Dae-Kwan-Ryoung and Nam-Won. Genomic DNA for genotyping assays was extracted from the 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 for 8,344 SNP, but 34 steers failed to genotype due to low DNA quality from phenol and chloroform contamination. Genotyping data were received on 8,344 SNP and all those SNP were physically mapped to chromosomes (in bp) using the bovine genome sequence (Btau-3.1) (<ftp://ftp.hgsc.bcm.tmc.edu/pub/data/Btaurus/fasta/Btau20060815-freeze/>).

SNP editing

Genotypes were tested for Hardy-Weinberg equilibrium (HWE) to identify possible typing errors using chi-square test in R/SNPassoc Package (R Development Core Team). SNP not in HWE ($p < 0.0001$), monomorphic SNPs and minor allele frequency ($< 1\%$) were removed in this study.

Linkage disequilibrium analysis

The linkage disequilibrium between SNP pairs was measured using D' (Lewontin 1964) and r^2 (Hill and Robertson, 1968) on the same chromosome. The equations are defined as

$$D = \text{freq}(A_1B_1) * \text{freq}(A_2B_2) - \text{freq}(A_1B_2) * \text{freq}(A_2B_1)$$

where $\text{freq}(A_1B_1)$ is the frequency of the A_1B_1 haplotype

in the population. The D statistic is dependent on the frequencies of the individual alleles and so it is not useful for LD among multiple pairs of loci.

$$D' = \frac{|D|}{D_{\max}}$$

Where $D_{\max} = \min(\text{freq}(A_1) * \text{freq}(B_2), -1 * \text{freq}(A_2) * \text{freq}(B_1))$ if $D > 0$, else $D_{\max} = \min(\text{freq}(A_1) * \text{freq}(B_1), -1 * \text{freq}(A_2) * \text{freq}(B_2))$ if $D < 0$.

Another common measure of LD is r^2 (Hill and Robertson, 1968), which is less dependent on allele frequencies.

$$r^2 = \frac{D^2}{\text{freq}(A_1) * \text{freq}(A_2) * \text{freq}(B_1) * \text{freq}(B_2)}$$

where $\text{freq}(A_i)$ is the frequency of the A_i allele in the population.

To evaluate genome coverage for the single marker association test using the SNP set, genotype was used to estimate r^2 . For the genome-wide LD measurement, haplotypes were constructed using the Merlin program (Abecasis et al., 2002). In order to calculate LD between SNP pairs, both haplotypes, the maternally and paternally inherited haplotype were used to measure LD value between SNP pairs because the pedigree structure of this population consists of small sire half-sib families. SNP position was located across the genome using *Bos Taurus* build 3.1 map viewers at NCBI. GOLD program was used for LD measure with the constructed haplotype by Merlin (Abecasis et al., 2002).

Effective population size

The relationship between effective population size (N_e), recombination rate and LD (r^2) without mutation is summarized in $r^2 = \frac{1}{(4N_e c + 1)}$, where c is linkage map

distance in Morgans (Sved, 1971). Linkage map distance (c) was inferred based on the ratio between physical size of each chromosome and length of the corresponding linkage map (NCBI Map Viewer). Effective population size was calculated for each autosomal chromosome. In the presence

of mutation, the function of LD is $r^2 = \frac{1}{(4N_e c + 2)}$ (Tenesa et al., 2007). Using the following model:

$$r_i^2 = \frac{1}{(4N_e c + 2)} + e_i$$

where e_i is residual, effective population size (N_e) was fitted by non-linear least squares regression in R (R Development Core Team).

RESULTS AND DISCUSSION

Linkage disequilibrium (LD)

The distribution of SNPs on different chromosomes is summarized in Table 1, ranging from 296 on BTA1 to 76 on BTA28. The SNPs used in this study were evenly distributed across the genome. The average SNP density varied from 0.46 to 1.06 Mb at chromosome wide.

Twenty five percent (1,873 SNP) of SNPs were monomorphic on autosomes. Twelve percent (885 SNP) of SNPs were not in Hardy-Weinberg equilibrium at the 5% significance level. A total of 912 SNPs were of unknown chromosome location and unknown SNP position. In this genome wide LD test, 3,819 SNPs (45%) including sex chromosomes (149 SNPs), were excluded. The SNP positions were determined according to the bovine genome sequence assembly version 3.1 (Btau 3.1).

According to a previous study (Arias et al., 2009), the average recombination distance in the bovine genome is

1.25 cM/Mb, but the recombination distance decreases with chromosome length, and therefore the rate of recombination increases with chromosome length. This indicates that, in general, LD will extend for shorter distances on longer chromosomes and therefore that longer chromosomes will have lower LD than shorter chromosomes. However, as shown in Table 1, certain chromosomes had higher LD than others, but there were no relationships between length of chromosome or average spacing of SNPs and extent of LD.

In this study, the average r^2 value varied from 0.07, 0.08 and 0.09 (BTA5, 9, 23 and 28, respectively) to 0.16 and 0.17 (BTA 2 and BTA29, respectively) and the average D' ranged from 0.35 (BTA18) to 0.52 (BTA10, 11). Comparing the LD structure of Hanwoo cattle with that of North American Holstein cattle, the average r^2 value in Hanwoo was 1.5 times lower than in the Holstein (Bohmanova et al., 2010).

Figure 1 shows that the r^2 value for pairs of loci were binned according to the physical distances separating the

Table 1. Number of markers, map distances and adjacent linkage disequilibrium on 29 autosomes (BTA)

BTA	No of SNP	BTA length (Mb)	SNP density (Mb)	Mean D'	SD (D')	Mean r^2	SD (r^2)
1	296	145	0.49	0.45	0.29	0.15	0.23
2	252	120	0.48	0.45	0.32	0.17	0.23
3	222	129	0.58	0.41	0.28	0.13	0.22
4	207	110	0.53	0.49	0.22	0.10	0.20
5	210	132	0.63	0.50	0.23	0.09	0.15
6	232	127	0.55	0.51	0.23	0.12	0.20
7	179	138	0.77	0.50	0.27	0.14	0.20
8	192	121	0.63	0.51	0.25	0.12	0.21
9	149	110	0.74	0.44	0.22	0.09	0.19
10	179	106	0.59	0.52	0.24	0.10	0.20
11	204	127	0.62	0.52	0.23	0.12	0.18
12	151	113	0.75	0.52	0.27	0.16	0.25
13	202	93	0.46	0.46	0.23	0.11	0.18
14	176	92	0.52	0.39	0.27	0.15	0.23
15	140	96	0.69	0.49	0.25	0.12	0.20
16	158	98	0.62	0.46	0.25	0.11	0.20
17	132	110	0.83	0.49	0.23	0.10	0.17
18	124	84	0.68	0.35	0.22	0.10	0.17
19	108	93	0.86	0.39	0.27	0.13	0.19
20	138	76	0.55	0.46	0.33	0.12	0.20
21	90	95	1.06	0.46	0.32	0.14	0.22
22	139	80	0.58	0.39	0.28	0.13	0.23
23	106	71	0.67	0.36	0.23	0.08	0.16
24	117	69	0.59	0.40	0.29	0.12	0.20
25	79	64	0.81	0.36	0.29	0.10	0.18
26	103	75	0.73	0.42	0.29	0.14	0.22
27	84	66	0.79	0.49	0.34	0.13	0.27
28	76	56	0.74	0.36	0.28	0.07	0.14
29	77	69	0.90	0.46	0.33	0.16	0.25
Mean	156	99	0.63	0.45	0.25	0.12	0.19

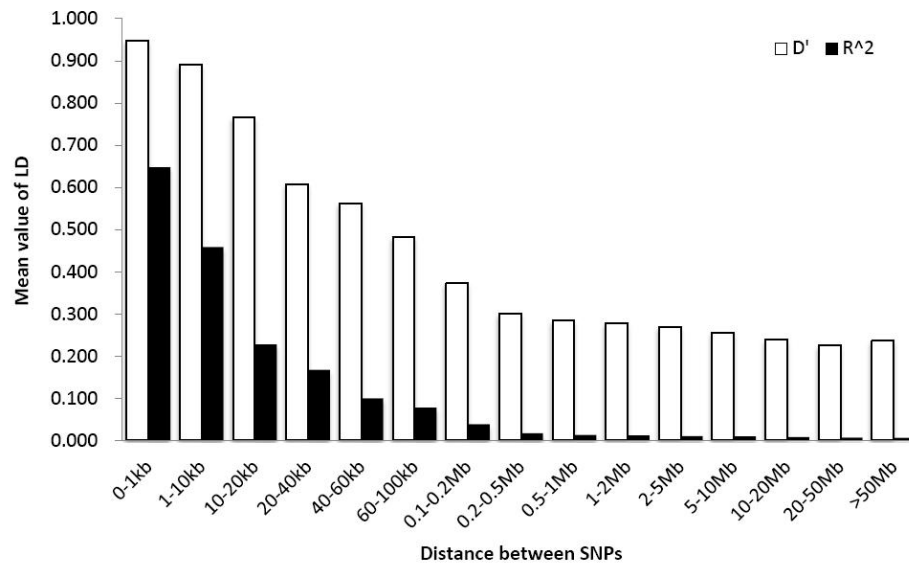


Figure 1. Mean value of linkage disequilibrium (D' and r^2) among syntenic SNP pairs over different map distance pooled over all autosomes.

loci. As shown in Figure 1, there is an inverse relationship between LD and physical distance and the r^2 value decreased by 0.1 at a locus separation of approximately 100 kb in Hanwoo. A similar study performed in eight cattle breeds by McKay et al. (2007) found that average r^2 values had fallen to 0.1 for SNPs with a distance of 500 kb.

There are similar studies measuring LD in pig and dog breeds, and both species demonstrate a LD value of 0.1 for SNPs with a distance of several Mb (Lindblad-Toh et al., 2005; Du et al., 2007). Our results suggest a shorter range of LD than has previously been reported in cattle (McKay et al., 2007). This indicates that Hanwoo may have a larger historical effective population size than other breeds and that there was a rapid recent expansion of the Hanwoo population.

Effective population size

In most studies estimating effective population sizes in animals, genetic distance c is approximated directly using physical distance (1 Mb-1 cM) (Hayes, 2008; Kim et al., 2009). Recently, the average recombination distance was

empirically measured as about 1.25 cM/Mb in the bovine genome (Arias et al., 2009).

LD value (r^2) was averaged in bins of estimated linkage distance and used to investigate changes in effective population size from 500 generations ago to the present in Hanwoo cattle (Table 2). Given the estimation of the average recombination rate in the bovine genome calculated by Arias et al. (2009), physical map distances were transformed into Morgan genetic map distances. Linkage map distance (c) in Morgans was calculated by dividing the total size of the MARC linkage map distance by the physical genomic size (Btau 4.0). The effective size of the Hanwoo population was estimated chromosome-by-chromosome using the mutation-included model as explained in M & M (Table 2; Figure 2).

The effective population size (N_e) decreased slightly over the last 100 generations in Hanwoo. However, the effective population size was consecutively falling almost double from 50 generations ago to the present (Table 2). Figure 2 shows that the change in N_e between 500 to 100 past generations was small, but that N_e fell dramatically

Table 2. Effective population size (N_e) in the past Hanwoo (Korean cattle) population

Fragment size (kb)	100	200	500	1,000	2,000	5,000	10,000	15,000
Effective population size (N_e)								
Mean	2,743.7	3,905.6	2,748.3	1,487.6	729.4	352	162.5	98.1
Median	2,362.3	3,533.2	2,712.3	1,476.5	718.2	352.2	163.1	96.3
Standard deviation	2,197.1	3,570	836.8	453.8	111.1	22.7	13.7	7.7
Minimum	625.6	1,132.9	1,259.5	925.6	555.7	306.8	137	87.3
Maximum	12,072.6	19,792.8	4,778.4	3,279.6	1,005.7	392.5	188.9	119.3
Linkage distance (cM)	0.1	0.2	0.5	1	2	5	10	15
Generations	500	250	100	50	25	10	5	3

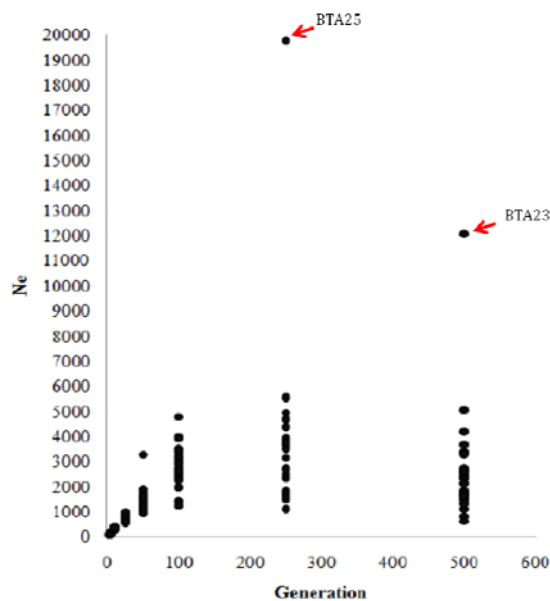


Figure 2. Effective population size from chromosome of the past Hanwoo (Korean cattle) population. Effective population size was estimated in each autosomal chromosome. The generation in the past population was calculated as $1/2c$. N_e is plotted against the past generation up to 500 generation.

during the last 100 generations to the present. For example, the mean N_e within 1 cM was 1,487, which was the same mean N_e estimated over 50 generations ago. The effective population size with the 2 Mb haplotype calculated for the previous 25 generations was 729.4. When haplotype size was 15 Mb, the mean and median N_e was 98.1 and 96, respectively, which was larger (71.7 and 70, respectively) than previous estimates for the North American Holstein population (Kim et al., 2009). However, average N_e on BTA23 and BTA25 within 0.1 cM and 0.2 cM interval was extremely higher than other chromosome. The average r^2 values on these chromosomes appear around 0.2 lower than other chromosomes within a 0.1 to 0.2 cM interval. Additionally, each of the loci on these chromosomes had similar allele frequencies to those loci on the other autosomes (data not shown). This suggests that the loci on BTA23 and BTA25 may have been clustered around one or more recombination hotspots on each of these chromosomes (McKay et al., 2007).

In this study, the N_e in Hanwoo is larger overall than in North American Holstein cattle. In particular, the N_e of Hanwoo dramatically decreased 25 generations ago; this is when official Hanwoo breeding programs were initiated. This is evidence that selection pressure affects N_e in the Hanwoo population. The first Hanwoo breeding program, or Hanwoo Gaeryang Danji (HGD), was initiated by the Ministry of Agriculture and Forestry (MAF) in 1979. The HGD program spanned eight provinces across Korea and 3,967 animals in an effort to improve growth while at

retaining or improving marbling scores (Lee, 2010).

REFERENCES

- Abecasis, G. R., S. S. Cherny, W. O. Cookson and L. R. Cardon. 2002. Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat. Genet.* 30:97-101.
- Arias, J. A., M. Keehan, P. Fisher, W. Coppieters and R. Spelman. 2009. A high density linkage map of the bovine genome. *BMC Genet.* 10:1471-2156.
- Bohmanova, J., M. Sargolzaei and F. S. Schenkel. 2010. Characteristics of linkage disequilibrium in North American Holsteins. *BMC Genomics* 11: doi:10.1186/1471-2164-11-421
- Decker, J. E., J. C. Pires, G. C. Conant, S. D. McKay, M. P. Heaton, K. Chen, A. Cooper, J. Vilki, C. M. Seabury, A. R. Caetano, G. S. Johnson, R. A. Brennenman, O. Hanotte, L. S. Eggert, P. Wiener, J. J. Kim, K. S. Kim, T. S. Sonstegard, C. P. van Tassell, H. L. Neibergs, J. C. McEwan, R. Brauning, L. L. Coutinho, M. E. Babar G. A. Wilson, M. C. McClue, M. M. Rolf, J. W. Kim, R. D. Schnabel and J. E. Taylor. 2009. Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics. *PNAS* 106:18644-18649
- Du, F. X., A. C. Clutter and M. M. Lohuis. 2007. Characterizing linkage disequilibrium in pig population. *Int. J. Biol. Sci.* 3: 166-178.
- Han, S. W. 1996. The breed of cattles. Sun-Jin publishing pp. 148-160.
- Hayes, B. J. 2008. QTL mapping, MAS and Genomic selection. Text book for Armidale summer animal breeding course pp. 6-10.
- Hedrick, P. 1987. Gametic disequilibrium measures: proceed with caution. *Genetics* 117:331-341.
- Khatkar, M. S., K. R. Zenger, M. Hobbs, R. J. Hawken, J. A. L. Cavanagh, W. Barris, A. E. McClintock, S. McClintock, P. C. Thomson, B. Tier, F. W. Nicholas and H. W. Raadsma. 2007. A primary assembly of a bovine haplotype block map based on a 15 k SNP panel genotyped in Holstein-Friesian cattle. *Genetics* 176:763-772.
- Kim, E. S. and B. W. Kirkpatrick. 2009. Linkage disequilibrium in the North American Holstein population. *Anim. Genet.* 40: 279-288.
- Lee, C. and E. J. Pollak. 2002. Genetic antagonism between body weight and milk production in beef cattle. *J. Anim. Sci.* 80: 316-321.
- Lee, S. H. 2010. Genome analysis to identify QTL and genes affecting carcass traits in Hanwoo (Korean cattle). The University of New England, PhD thesis, pp. 10-20.
- Lewontin, R. C. 1964. The interaction of selection and linkage. I. General considerations; heterotic model. *Genetics* 49:49-67.
- Lindblad-Toh, K et al. 2005. Genome sequence, comparative analysis and haplotype structure of the domestic dog. *Nature* 438:803-819.
- Hill, W. G. and A. Robertson. 1968. Linkage disequilibrium in finite populations. *Theor. Appl. Genet.* 38:226-231.
- McKay, S. D., R. D. Schnabel, B. M. Murdoch, L. K. Matukumalli, J. Aerts, W. Coppieters, D. Crews E. D. Neto, C. A. Gill, Chuan Gao, H. Mannen, P. Stothard, Z. Wang C. P. van Tassell, J. L. Williams J. F. Taylor, and S. S. Moore. 2007. Whole genome linkage disequilibrium maps in cattle. *BMC Genet.*

- 8:74. doi: 10.1186/1471-2156-8-74.
- Meuwissen, T. H. E., B. J. Hayes and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- R Development Core Team: R: a language and environment for statistical computing. 2008 R Foundation for statistical computing, Vienna, Austria.
- Sved, J. V. 1971 Linkage disequilibrium and homozygosity of chromosome segments in finite population. *Theor. Popul. Biol.* 2:125-141.
- Tenesa, A., P. Navarro, B. J. Hayes, D. L. Duffy, G. M. Clarke, M. E. Goddard and P. M. Visscher. 2007. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 17:520-526.
- The international HapMap Consortium 2005. A haplotype map of the human genome. *Nature* 437:1299-1320.
- Yoon, D. H., E. W. Park, S. H. Lee, H. K. Lee, S. J. Oh, I. C. Cheong and K. C. Hong. 2005. Assessment of genetic diversity and relationships between Korean cattle and other cattle breeds by microsatellite loci. *J. Anim. Sci. Technol. (Kor)* 47(3):341-354.