Journal of Animal Breeding and Genomics
pISSN:1226-5543 eISSN:2586-4297
JABG. 2021 December, 5(4):235-242
https://doi.org/10.12972/jabng.20210023
RESEARCH ARTICLE
OPEN ACCESS

Linkage Disequilibrium Analysis of Hanwoo in Gyeonggi Region using Hanwoo SNP Chip

Gwang Hyeon Lee [1], Hong Sik Kong [1,2,3,*]

[1]Department of Applied Biotechnology, The Graduate School of Hankyong National University, Anseong 17579, Korea
[2]Genomic Informatics Center, Hankyong National University, Anseong 17579, Korea
[3]Gyeonggi Regional Research Center, Hankyong National University, Anseong 17579, Korea

ABSTRACT

This study collected basic information for improving the Hanwoo cattle in the Gyeonggi region. Linkage disequilibrium (LD) between single nucleotide polymorphisms (SNP) markers in autosomes was estimated by analyzing the Hanwoo raised in the Gyeonggi region using the Hanwoo SNP 50K BeadChip. The Hanwoo tail hair samples used in this study were collected from 827 Hanwoo cattle in the Gyeonggi region and were subjected to SNP Chip analysis. Furthermore, 52,195 SNPs were obtained from the analysis. Quality control was performed to remove unnecessary SNPs, and 41,605 SNPs were obtained. The total genome length was 2,500.01 Mb, with chromosome 25 (42.65 Mb) being the shortest and chromosome 1 (158.09 Mb) being the longest. The r^2 value was 0.231 for the SNP distance between 0 and 50 Kb and was 0.065 for distance between 150 and 200 Kb. Thus, the closer the distance between the SNPs, the higher the r^2 value. Genetic improvement has been conducted for approximately 100 years in Angus and Holstein breeds. However, for the Hanwoo, genetic improvement has been conducted for approximately 40 years (Jo et al., 2012). In addition, the selection intensity for Hanwoo genetic improvement is lower than that of other varieties. This study confirmed that Hanwoo in the Gyeonggi region had no significant difference in LD. However, it is necessary to test cow to prevent the occurrence of inbreeding and reduction in genetic diversity caused by a preference for specific KPN.

Key words: Hanwoo, SNP chip, Linkage Disequilibrium, Gyeonggi

INTRODUCTION

The selection of economically important traits in livestock has been regularly analyzed using phenotypic data. Artificial insemination has been practiced using superior genetic quality bulls since 1960. And from 1987, the Korean proven bull No (KPN) selection has been carried out through progeny testing, resulting in substantial genetic gains (NIAS, 2012). Furthermore, recent advances in genomics have made it possible to analyze the genome genetic structure (Bovine HapMap, 2009). Currently, a technique that utilizes genomic information to evaluate the genetic ability of each animal using a single nucleotide polymorphism (SNP) chip is being highly commercialized. Whole-genome SNP chips are widely used to study genetic diversity and are regularly applied in animal breeding (McKay et al., 2008; Ben et al., 2015). Genomic selection (GS) studies and quantitative trait loci (QTL) search using genotype information for approximately 50,000 SNP markers in the genome are being conducted for domestic and international cattle breed using the 50 K Illumina Bovine Beadchip (Sved, 1971, Hayes, 2008, VanRaden et al.,

***Corresponding author:** Hong Sik Kong
Department of Applied Biotechnology, The Graduate School of Hankyong National University, Anseong 17579, Korea
Genomic Informatics Center, Hankyong National University, Anseong 17579, Korea
Gyeonggi Regional Research Center, Hankyong National University, Anseong 17579, Korea
Tel: +82-31-670-5334 Fax: 0504-842-4615 E-mail: Kebinkhs@hknu.ac.kr

Received: 10 December, 2021 **Revised:** 22 December, 2021 **Accepted:** 22 December, 2021

2009).

Linkage disequilibrium (LD) is a non-random association between alleles at different loci within a population (Weir and Ott, 1997). LD estimation is used to expand breeding and individual selection effects by estimating the genetic recombination possibility within a population. Thus, it is used to validate the QTL search and GS (McRae et al., 2007; Qanbari et al., 2010). In addition, the LD structure in a population is critical for explaining and applying the results of the genome-wide association studies (GWAS) and GS economic traits improvement (Goddard et al., 2009; Habier, 2010). Thus, recent studies explained the relation between the factors (GS, QTL, GWAS, etc.) and LD. Most studies have analyzed the Hanwoo population managed by national institutions such as the Livestock Improvement Main Center and the National Institute of Animal Science. Differences occur in Hanwoo populations raised in different provinces of Korea. Therefore, this study collected basic information for improving Hanwoo in the Gyeonggi region by estimating the LD between SNP markers in autosomes and analyzing the Hanwoo raised in the Gyeonggi region using the Hanwoo SNP 50K BeadChip.

MATERIALS & METHODS

1. DNA sample

Tail hair samples collected from 827 Hanwoo raised in the Gyeonggi region were subjected to SNP Chip analysis. The hanwoo traceability data were collected through the Korea Animal Improvement Association. The study was approved by the Hankyong National University Animal Ethics Committee (No.2018-1).

2. Genotyping

The Hanwoo 50K SNP Analysis BeadChip was used to analyze genotype using DNA extracted from the tail hair root. Final report file was output using GenomeStudio 2.0 software (Illumina Inc, 2016) in order to convert the obtained genotype information into a Plink version 1.9 (Purcell et al., 2007) analysis format, ped and map files were created after the PLINK formatting process using the Perl language. As a results, 52,195 SNPs were obtained through the analysis. The quality control (QC) performed using the Plink version1.9 (Purcell et al., 2007) removed unnecessary SNPs. SNPs with call rates < 0.90 (806), minor allele frequency (MAF) < 0.05 (9,475), and those deviating from Hardy–Weinberg equilibrium (HWE) P-value < 1×10^{-7} (309) were excluded from the analysis. The final dataset consisted of 41,605 SNPs (Table 1).

Table 1. Number of SNPs after removal of unnecessary markers from hanwoo SNP data

Category	Threshold	No. of marker
Total Genotyped SNP marker		52,195
SNP marker with missing genotyping	> 0.1	806
Minor allele frequency	< 0.05	9,475
Hardy-Weinberg Equilibrium	< 0.0000001	309
Selected SNP markers		41,605

Table 2. Number of SNPs and genetic information of each autosome in Gyeonggi hanwoo

NO. Chromosome	No. of SNPs	MAF	Total length (Mb)	Length Standard deviation	Average Interval size	Interval Standard deviation
1	2,598	0.269	158.09	46.02	0.061	0.055
2	2,177	0.269	136.66	41.59	0.063	0.067
3	2,054	0.267	121.14	34.85	0.059	0.061
4	1,931	0.272	120.01	35.13	0.062	0.053
5	1,672	0.269	121.08	36.95	0.072	0.070
6	2,565	0.265	118.98	31.14	0.046	0.056
7	2,021	0.277	112.38	30.43	0.056	0.064
8	1,792	0.264	113.01	32.26	0.063	0.054
9	1,635	0.262	105.46	31.03	0.065	0.061
10	1,875	0.267	104.17	32.88	0.056	0.092
11	1,722	0.269	107.18	32.50	0.062	0.059
12	1,251	0.267	90.83	26.20	0.073	0.117
13	1,321	0.271	83.86	24.53	0.064	0.057
14	1,827	0.254	83.15	23.15	0.046	0.046
15	1,325	0.267	84.73	23.64	0.064	0.062
16	1,244	0.270	81.37	23.41	0.065	0.068
17	1,231	0.261	74.85	22.66	0.061	0.066
18	1,053	0.265	65.16	19.27	0.062	0.062
19	1,126	0.271	63.54	17.99	0.056	0.053
20	1,244	0.265	71.59	20.94	0.058	0.052
21	1,155	0.278	71.10	20.00	0.062	0.070
22	974	0.264	61.22	18.52	0.063	0.053
23	922	0.279	52.10	14.89	0.057	0.055
24	1,001	0.269	62.10	18.36	0.062	0.053
25	790	0.282	42.65	12.69	0.054	0.045
26	829	0.269	50.95	14.48	0.062	0.047
27	733	0.278	45.33	12.73	0.062	0.061
28	728	0.267	46.18	13.69	0.064	0.056
29	809	0.273	51.10	14.62	0.063	0.065
Total	41,605	0.268	2500.01			

MAF: minor allele frequency

3. Linkage Disequilibrium (LD) analysis

LD indicated that alleles at two loci were related, and were hereditary, and could be estimated using D' or r^2, standardized for D (Lewontin et al., 1964; Hill et al., 1968). However, since LD estimation using the statistical value D' may cause overestimation when the population size is small or the allele frequency is low, the r^2 parameter, a statistic indicating the correlation of the alleles at two loci was estimated using the D' (McRae et al., 2002, Hayes, 2007). To calculate the LD between two SNP markers, analysis was performed using the –r^2 –ld-window-r^2 0 –ld-window 99999 –ld-window-kb 1000 option of the Plink version 1.9 (Purcell et al., 2007). The LD (r^2) of loci A and B on the same chromosome was calculated as follows:

$$r^2 = \frac{D^2}{(freq\ A_i \times freq\ A_j \times freq\ B_i \times freq\ B_j)}$$

where freq A_i, freq A_j, freq B_i and freq B_j are the observed frequencies of alleles A_i, A_j, B_i, and B_j, respectively. Then D was calculated as follows:

$$D = \text{freq } A_iB_i \times \text{freq } A_jB_j - \text{freq } A_iB_j \times \text{freq } A_jB_i$$

where freq A_iB_i, freq A_jB_j, freq A_iB_j, and freq A_jB_i are the observed haplotype frequencies of alleles between the two markers. LD obtained through analysis was presented numerically and graphically using the R package (The R Project for Statistical Computing, ver.4.1.2; http://www.r-project.org).

RESULTS AND DISCUSSION

In this study, 52,195 SNPs were identified by analyzing the genotypes of 827 Hanwoo cattle raised in the Gyeonggi region. LD analysis was performed using 41,605 SNP markers, excluding 10,590 SNPs, selected by the QC process. The SNPs available per chromosome are presented in Table 2. The mean MAF of 29 autosomes was 0.268. The total genome length was 2,500.01 Mb, with chromosome 25 (42.65 Mb) being the shortest and chromosome 1 (158.09 Mb) being the longest. The distance between each analyzed SNP ranged from 0.046 to 0.073 Mb, and the longest was 0.073 Mb on chromosome 12. The confirmed standard deviation showed that the distance between the SNPs was not uniform. Cho et al. (2012) reported an SNP distance range of 0.055 to 0.074 Mb and showed similar overall results.

SNPs were classified according to the MAF to confirm the distribution, and 51,080 SNPs were used, including SNPs with an allele frequency of ≤ 0.05 (Fig. 1). As a result, except for SNPs with an allele frequency of 0.05 or less, a uniform distribution across the common frequency classes was observed due to the design of the SNP chip, optimized for a uniform SNP spacing and allele frequency distribution.

In general, LD is estimated larger r^2 value by a smaller effective population size, a higher inbreeding within an analysis population, when the gene frequencies of the population are distinctly different. The degree and extent of LD in livestock breeding includes important information for marker-assisted selection and fine mapping of QTLs (Hayes et al., 2003; Du et al., 2007). Thus, it is the most appropriate method to identify SNPs or haplotypes that are significantly related to phenotypic trait variation.

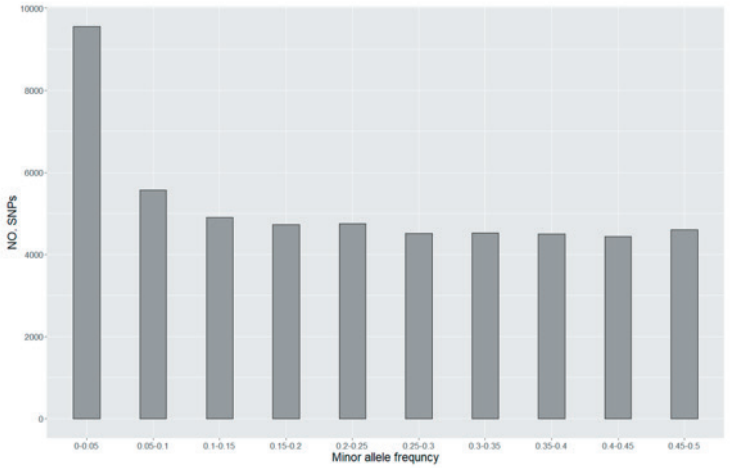

Figure 1. Minor allele frequency (MAF) of SNPs

Table 2 shows the distances of pair-wise LD that were binned into 50 Kb intervals. The average r^2 and the number of SNP pairs were calculated for each interval. A total of 766,856 SNPs pairs were evenly distributed, and the closer the distance between the SNPs, the higher was the r^2 value. The number of SNP pairs was the lowest at 36,201 pairs when the SNP distance was between 950 and 1,000 Kb, and the highest at 44,566 pairs for distance between 50 and 100 Kb. The r^2 value was 0.231 for SNP distance between 0 and 50 Kb and was 0.065 for SNP distance between 150 and 200 Kb, and Fig. 2 shows the decay of LD. A similar study performed by Cho et al. (2012) and Li and Kim (2015) found that the LD decayed at less than 200 Kb, and the r^2 value decreased after 200 kb. In addition, $r^2 > 0.2$ of adjacent SNPs were 34.83% and 18.24% for a distance of 0-50 Kb and 50-100 Kb, respectively. Marques et al. (2008) extended the range of $r^2 > 0.2$ in Holstein chromosome 14 to 100 Kb, and Sargolzaei et al. (2008) reported that the r^2 value of Holsteins in Canada and North America was 0.59 for a distance of 0-100 Kb. McKay et al. (2007) reported r^2 values of 0.55, 0.05, 0.41, 0.47, 0.61, 0.58, 0.53, and 0.28 for a distance of 0-100 Kb in Angus, Charolais, Brahman, Dutch black and white dairy cattle, Holstein, Japanese black, Limousin, and Nelore breeds, respectively. The r^2 value of the other breeds was higher than the r^2 value of Hanwoo estimated in this study. Meuwissen et al. (2001) reported that the LD r^2 value should exceed 0.2 for the genomic estimated breeding value (GEBV) accuracy to reach 85%, and Ardlie et al. (2002) pointed that for the r^2 value to provide sufficient power for GWAS it should be greater than 0.3. In the case of breeds such as Angus and Holstein, genetic improvement has been carried out for approximately 100 years, whereas the improvement of Hanwoo has been carried out for a relatively short period, approximately 40 years (Jo et al., 2012). In addition, the selection intensity for Hanwoo genetic improvement is relatively low than of other varieties.

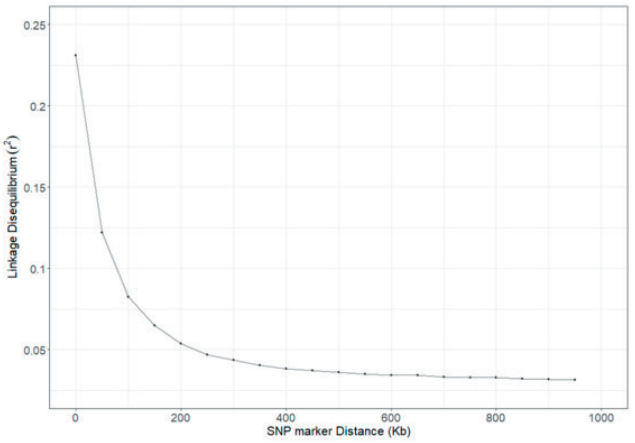

Figure 2. Linkage Disequilibrium (r2) average according to average genetic distance of Gyeonggi hanwoo.

Genetic improvement is a technology that selects desirable livestock economic traits and enhances them to suit human needs. Performance tests or progeny tests evaluated based on phenotype and pedigree were used in the past. However, they were time-consuming and could not predict the results accurately; hence, genetic improvement using genomic information is now utilized (Kim, 2021). Genetic marker development has made it possible to improve livestock genetic ability and will become a major factor in deciphering various traits and diseases. Recently, a large-scale SNP panel (50K, 700K, HD, etc.) was commercialized, many studies have used SNP markers in Hanwoo (Lee et al., 2011, Cho et al., 2012, Li and Kim, 2015). In addition, a Hanwoo SNP chip with a Hanwoo specific marker was developed and commercialized, and the second version had been recently developed. Hanwoo improvement has been conducting

genetic ability evaluations since the 1980s to select genetically superior KPN. The frozen semen of the selected KPN is used in the artificial insemination of cattle raised across the country (Park et al., 2011; Shin et al., 2018). Genetic improvement through limited KPN shows a reduction in the genetic diversity, effective population of Hanwoo, and improvement plateau, and an increase in the inbreeding coefficient within the population (Woolliams, 2004; Weigel, 2001). A significant reduction in genetic diversity due to inbreeding or effective population reductions can lead to inbreeding depression and species extinction (Zenger et al., 2007). This study confirmed that Hanwoo in the Gyeonggi region showed no difference in LD. It is thought that the genetic diversity was retained due to the influx of various Hanwoo species nationwide. However, it is necessary to test cow to prevent the occurrence of inbreeding and a reduction in genetic diversity caused by a preference for a specific KPN. This study can be a source of basic data for the genetic improvement of Hanwoo cattle in the Gyeonggi region.

Table 3. Pairwise linkage disequilibrium (r2) for SNPs at various distance in Gyeonggi hanwoo

Distance (Kb)	No. of SNP pairs	Mean (r^2)	Median	Standard deviation (r^2)	$r^2 > 0.2$ (%)	$r^2 > 0.3$ (%)
0 - 50	37,541	0.231	0.099	0.288	34.83	26.78
50 - 100	44,566	0.122	0.045	0.190	18.24	11.87
100 - 150	42,983	0.083	0.029	0.147	10.75	6.53
150 - 200	41,510	0.065	0.023	0.124	7.49	4.24
200 - 250	40,506	0.054	0.019	0.111	5.37	3.06
250 - 300	39,561	0.047	0.017	0.101	4.13	2.41
300 - 350	38,633	0.043	0.016	0.094	3.70	2.06
350 - 400	38,693	0.040	0.015	0.090	3.08	1.80
400 - 450	37,998	0.038	0.014	0.086	2.73	1.55
450 - 500	37,688	0.037	0.014	0.086	2.62	1.60
500 - 550	37,081	0.036	0.013	0.086	2.52	1.61
550 - 600	37,137	0.035	0.012	0.084	2.50	1.54
600 - 650	36,909	0.034	0.012	0.081	2.42	1.53
650 - 700	36,906	0.034	0.012	0.083	2.31	1.56
700 - 750	36,813	0.033	0.012	0.080	2.17	1.41
750 - 800	36,616	0.033	0.012	0.080	2.23	1.43
800 - 850	36,494	0.033	0.012	0.082	2.17	1.48
850 - 900	36,482	0.032	0.012	0.079	2.12	1.38
900 - 950	36,538	0.032	0.011	0.078	2.14	1.43
950 - 1000	36,201	0.031	0.011	0.077	2.05	1.39

REFERENCES

Ardlie KG, Kruglyak L., and Seielstad M. 2002. Patterns of linkage disequilibrium in the human genome. Nat. Rev. Genet, 3: 299–309.

Ben Jemaa S, Boussaha M, Ben Mehdi M, Lee JH, Lee S-H. 2015. Genome-wide insights into population structure and genetic history of tunisian local cattle using the illumina bovinesnp50 beadchip. BMC Genomics, 16: 677.

Bovine HapMap C, Gibbs RA, Taylor JF, et al. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324: 528-532.

Chang CC, Chow CC, Tellier L.C, Vattikuti S, Purcell SM, and Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience, 4(1).

Cho CI, Lee JH, and Lee DH. 2012. Estimation of linkage disequilibrium and effective population size using whole genome single nucleotide polymorphisms in Hanwoo. J. Life Sci, 22(3): 366-372.

Du FX., Clutter AC. and Lohuis MM. 2007. Charaterizing Linkage Disequilibrium in Pig Populations. Int. J. Biol. Sci, 3: 166-178.

Goddard ME, Hayes BJ. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat Rev Genet, 10(6): 381–91.

Habier D. 2010. More than a third of the WCGALP presentations on genomic selection. J Anim Breed Genet, 127(5): 336–7.

Hayes BJ, Visscher PE., McPartlan H. and Goddard ME. 2003. A novel multi-locus measure of linkage disequilibrium and it use to estimate past effective population size. Genome Research, 13: 635-643.

Hayes BJ. 2007. QTL mapping, MAS, and genomic selection: A short-course. Animal Breeding & Genetics Department of Animal Science. Iowa State University, IA.

Hayes BJ, Bowman PJ, Chamberlain AC., Verbyla K, and Goddard ME. 2008. Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genetic Selection Evolution, 41(1): 51.

Hill WG, Robertson A. 1968. Linkage disequilibrium in finite populations. Theor Appl Genet, 38(6): 226-31.

Illumina Inc. 2016. GenomeStudio Polyploid Genotyping Module v2.0 Software Guide. Tech Note. Illumina Inc. CA, USA.

Jo C, Cho SH, Chang J, Nam KC. 2012. Keys to production and processing of Hanwoo beef: A perspective of tradition and science. Anim Front, 2: 32–38.

Kim DH. 2021. Studies on the Genetic Evaluation of Hanwoo using the Genomic Information. Ph.D. dissertation, Jeonbuk National University, Jeonbuk, Korea.

Lewontin RC. 1964. The interaction of selection and linkage. I. General considerations; heterotic models. Genetics, 49(1): 49-67.

Marques, E., Schnabel, RD, Stothard, P, Kolbehdari, D, Wang Z, Taylor JF, and Moore SS. 2008. High density linkage disequilibrium maps of chromosome 14 in Holstein and Angus cattle. BMC Genet, 9(1): 45.

McKay SD, Schnabel RD, Murdoch BM, et al. 2007. Whole genome linkage disequilibrium maps in cattle. BMC Genet, 8: 74.

McKay SD, Schnabel RD, Murdoch BM, et al. 2008. An assessment of population structure in eight breeds of cattle using a whole genome SNP panel. BMC Genet, 9: 37.

McRae AF, McEwan JC, Dodds KG, Wilson T, Crawford AM, and Slate J. 2002. Linkage disequilibrium in domestic sheep. Genetics, 160: 1113-1122.

Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4): 1819–29.

NIAS (National Institute of Animal Science). 2012. 2011 Animal Improvement Facts and Figures. p.3-166. NIAS, Korea.

Park SR, Lee, SK, Lee KS, Shin YS, Song YH, and Lee SJ. 2011. Analysis of reproduction and breeding status in Gangwon east area. Ann. Anim. Resour. Sci, 22: 1-5.

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

Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, and Simianer H. 2010. The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet, 41: 346-356.

Sargolzaei M, Schenkel FS, Jansen GB, and Schaeffer LR. 2008. Extent of linkage disequilibrium in Holstein cattle in North America. J. Dairy Sci, 91(5): 2106-2117.

Shin EG, Lee SH, Yoon DH. 2018. Accuracy of genomic estimated breeding value with Hanwoo cows in the commercial farms. J. Agric. Life Sci, 52: 91-98.

Sved JA. 1971. Linkage disequilibrium and homozygosity of chromosome segments in finite population. Theor. Popul. Biol, 2: 125-141.

VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, and Schenkel F. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci, 92: 16-24.

Weigel KA., 2001. Controlling inbreeding in modern breeding programs. J. Dairy Sci, 84(E. Suppl.): E177–E184.

Weir BS, and Ott J. 1997. Genetic data analysis II. Trends Genet, 13(9): 379.

Woolliams JA, Gwaze GP, Meuwissen THE, et al. eds. 2004. Secondary Guidelines for Development of National Farm Animal Genetic Resources Management Plans. Management of Small Populations at Risk. Food and Agriculture Organization of the United Nations. Washington, DC.

Zenger KR, Khatkar MS, Cavanagh JAL, Hawken RJ, and Raadsma HW. 2007. Genome-wide genetic diversity of Holstein Friesian cattle reveals new insights into Australian and global population variability, including impact of selection. Animal genetics, 38: 7-14.