Abstract W93

Linkage Disequilibrium in a Thai Dairy Cattle Population with Different Holstein Fractions

Thawee Laodim*, Skorn Koonawootrittriron*, Mauricio A. Elzo† and Thanathip Suwanasopee*

***Department of Animal Science, Kasetsart University, Bangkok 10900, Thailand**

†Department of Animal Sciences, University of Florida, Gainesville, FL 32611-0910, USA

SUMMARY

Linkage disequilibrium (**LD**) is important for gene mapping, accuracy of genomic prediction, and understanding of recombination biology in dairy cattle populations. The level of **LD** can vary among populations depending on their genetic structure, selection and recombination rates. *The objective of this study was to estimate and compare levels of **LD** in dairy cattle with different Holstein fractions under tropical conditions.* Blood samples of 2,643 dairy cattle (89 bulls and 2,554 cows) from 304 farms located in Central, Northern, Northeastern, Western and Southern Thailand were extracted for DNA. The DNA samples were genotyped with one of four GeneSeek Genomic Profiler BeadChips (9K, 20K, 26K, or 80K). Only SNPs from autosomes in common among the four chips were considered. In addition, SNPs with a minor allele frequency (**MAF**) lower than 0.01 and a call rate lower than 90% were excluded. This resulted in a set of 7,123 SNPs used in this study. Animals were classified into seven groups based on their Holstein fraction (HF): HF < 75%, 75% ≤ HF < 80%, 80% ≤ HF < 85%, 85% ≤ HF < 90%, 90% ≤ HF < 95%, 95% ≤ HF < 100%, and purebred HF. Distribution of MAF and estimation of **LD** were done using Haploview software. All HF groups had similar patterns of MAF across autosomes (fraction of SNPs increased with an increase in **MAF**). However, means of **MAF** across autosomes differed among HF groups and it tended to decrease with an increase in H fraction (from 0.376 for HF < 75% to 0.362 for purebred HF). Conversely, the mean r^2 across autosomes tended to increase as HF increased from 0.081 for HF < 75% to 0.109 for purebred HF. *Results from this study will be useful for genome wide association studies and for genomic prediction and selection of crossbred Holstein cattle in tropical regions.*

INTRODUCTION

The extent and distribution of linkage disequilibrium (**LD**) are important to study SNPs variation in the entire genome of dairy cattle. The **LD** is both a fundamental tool for gene mapping and an important criterion to select SNPs that are located in close proximity to genes of biological relevance in genome-wide association studies (Fortes *et al*., 2010). The studies of **LD** increase our understanding of recombination biology in dairy cattle populations. A level of **LD** larger than 0.2 is required to achieve reasonable accuracy of genomic selection. The **LD** is a non-random association between allele of SNPs at two sites. The extent of **LD** is measured by r^2, which is the squared correlation of the alleles at two sites. The level of r^2 ranges from 0 (no **LD**) to 1 (perfect **LD**), and it can vary among populations depending on their genetic structure, selection and recombination rates. Most dairy cattle studies reported the extent of **LD** in purebred dairy populations, which is likely to be different from crossbred or multibreed dairy populations especially those raised in tropical regions. *Thus the objective of this study was to estimate and compare levels of **LD** in dairy cattle with different Holstein fractions under Thai tropical conditions.* The results will increase our understanding of genetic variation of multibreed dairy cattle, which may be useful for genetic improvement programs under the tropical environments.

MATERIALS AND METHODS

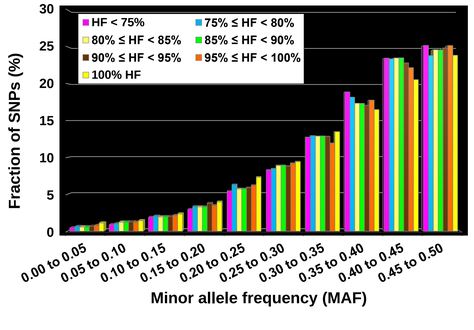
Animals and Blood samples: Animals in this study were 2,643 multibreed dairy cattle (89 bulls and 2,554 cows) born from 1994 to 2011 and raised in 304 farms located in Central (1,138 cattle from 88 farms), Northern (765 cattle from 88 farms), Northeast (255 cattle from 59 farms), Western (177 cattle from 1 farm) and Southern Thailand (308 cattle from 68 farms). These multibreed dairy cattle were the outcome of matings within and between *Bos taurus* and *Bos indicus* cattle. Cattle breeds present in this population were Brahman, Brown Swiss, Jersey, Holstein, Red Dane, Red Sindhi, and Sahiwal. Individual animals had Holstein fractions (HF) that ranged from 25% to 100%. Blood samples were taken from each animal and DNA was extracted from each sample using the protocol of Master Pure™ DNA Purification Kit (Epicentre®, USA).

SNPs genotype: The DNA samples were genotyped with one of four GeneSeek Genomic Profiler BeadChips (GeneSeek, Lincoln, NE), included GGP-9K (8,810 SNPs), GGP-20K (17,820 SNPs), GGP-26K (26,000 SNPs) or GGP-80K (78,000 SNPs). Only SNPs from autosomes in common among the four chips were considered. In addition, SNPs with a minor allele frequency (**MAF**) lower than 0.01 and a call rate lower than 90% were excluded. This resulted in a set of 7,123 SNPs used in this study.

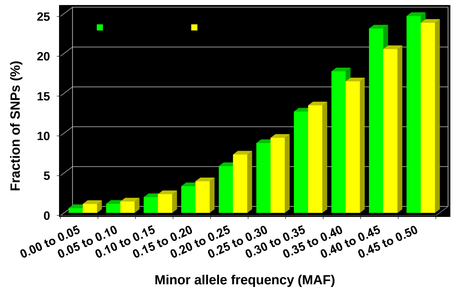
Data Analysis: Animals were classified into seven breed groups based on their Holstein fraction (HF): HF < 75% (142 cattle), 75% ≤ HF < 80% (123 cattle), 80% ≤ HF < 85% (281 cattle), 85% ≤ HF < 90% (543 cattle), 90% ≤ HF < 95% (1,077cattle), 95% ≤ HF < 100% (424 cattle), and purebred HF (53 cattle). The fraction of SNPs across autosomes was plotted against their **MAF** values for all breed groups. The patterns of **LD** were analyzed and visualized using Haploview software (Barrett *et al*., 2005). The correlation between a pair of loci (r^2) was used to measure **LD**. Only SNPs that had a P-value of Hardy–Weinberg equilibrium larger than 0.0001 and had a distance between SNPs lower than 1 Mb were included in the estimation of **LD**. Means of r^2 values were estimated for each group, and then they were compared using Bonferroni t-tests.

RESULTS AND DISCUSSION

Distribution of MAF across autosomes: All HF groups had similar patterns of MAF across autosomes (fraction of SNPs increased with an increase in MAF). Results here showed that the distributions of MAF in purebred and crossbred dairy cattle in Thailand were similar. The distribution of **MAF** in this Thai multibreed dairy cattle population was comparable to those reported for various *Bos taurus* cattle breeds (e.g., Holstein, Brown Swiss and Fleckvieh), while it had the opposite trend to distributions from *Bos indicus* cattle (e.g., Brahman, Gir and Nelore (Pérez O'Brien *et al*., 2014). This difference may be a reflection of the ancestral divergence between *Bos taurus* and *Bos indicus* (sub) species. Generally, *Bos taurus* breeds had higher **MAF** means than **MAF** from *Bos indicus* breeds (Salomon-Torres *et al*., 2013).



The means of MAF across autosomes differed among HF groups and it tended to decrease with an increase in HF fraction (from 0.376 for HF < 75% to 0.362 for purebred HF). Difference of **MAF** indicated that HF crossbreds had higher allele variation than HF purebreds. This may be due to the presence of alleles from other breeds within HF crossbreds (some animals could have had alleles from three or more other breeds). This additional genetic variation within this Thai multibreed dairy cattle population could be a reason for the difference between the dairy cattle population in Thailand and populations from other countries.



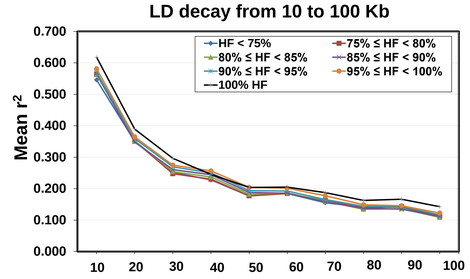
The extent and distribution of LD: The mean r^2 across autosomes (up to 1 Mb) of animals in each HF group was different (P < 0.05; **Table 1**). The purebred HF had the highest mean r^2 (0.109) and the crossbred HF < 75% had the lowest mean r^2 (0.081). Further, the mean r^2 across autosomes within Thai multibreed dairy cattle tended to increase as HF increased. This might be an indication that genomic selection for purebred HF would be somewhat more accurate than for crossbred HF cattle in this multibreed population. In addition, a high level of **LD** between markers is usually considered an indicator that these markers have not been separated by recombination, and have the same allele frequency (Ardlie *et al.,* 2002). However, all HF groups had similar patterns of **LD** distribution across autosomes, where **LD** decreased as the distance between SNPs increased. This reflected an increase in recombination rates between pairs of SNPs separated by longer distances compared to SNPs that were closer together that were rarely separated during meiosis. Hence the small distance between SNPs (2,500 bp) used as a criterion to select SNPs near the region of a gene for **GWAS** (Fortes *et al.*, 2010).

Table 1 Mean of MAF and r^2 across autosomes in a Thai multibreed dairy population with different Holstein fractions

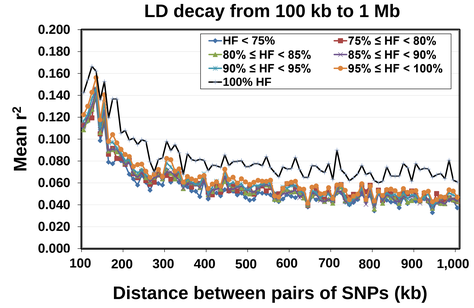
Holstein fraction	Number of cattle	Mean MAF ± SE	Mean r^2 ± SE
HF < 75%	142	0.376 ± 0.001	0.081 ± 0.001^{d}
75% ≤ HF < 80%	123	0.371 ± 0.001	0.086 ± 0.001^{cd}
80% ≤ HF < 85%	281	0.372 ± 0.001	0.085 ± 0.001^{cd}
85% ≤ HF < 90%	543	0.372 ± 0.001	0.085 ± 0.001^{cd}
90% ≤ HF < 95%	1,077	0.370 ± 0.001	0.089 ± 0.001^{bc}
95% ≤ HF < 100%	424	0.370 ± 0.001	0.093 ± 0.001^{ab}
100% HF	**53**	**0.362 ± 0.001**	**0.109 ± 0.001^{a}**

The LD decay up to 100 kb: The mean r^2 for pairs of SNPs located at distances shorter than 10 kb was 0.546 for HF < 75%, 0.564 for 75% ≤ HF < 80%, 0.563 for 80% ≤ HF < 85%, 0.566 for 85% ≤ HF < 90%, 0.570 for 90% ≤ HF < 95%, 0.581 for 95% ≤ HF < 100% and 0.618 for purebred HF. However, the mean r^2 for pairs of SNPs located from 10 to 100 kb apart declined from 0.546 to 0.111 for HF < 75%, 0.564 to 0.113 for 75% ≤ HF < 80%, 0.563 to 0.109 for 80% ≤ HF < 85%, 0.566 to 0.111 for 85% ≤ HF < 90%, 0.570 to 0.116 for 90% ≤ HF < 95%, 0.581 to 0.122 for 95% ≤ HF < 100% and 0.618 to 0.142 for purebred HF. This indicated that the mean r^2 decreased rapidly with the distance increasing from 10 to 100 kb for all breed groups. Meuwissen *et al.* (2001) simulated various levels of **LD** (r^2) and achieved an accuracy of 0.85 for genomic breeding values when r^2 = 0.2. At this threshold (r^2 = 0.2), the distance between SNPs was less than 40 kb in crossbreds HF < 95%, while it was 60 kb in crossbreds HF ≥ 95% and purebred HF. The differences among HF groups could be due to genetic structure, where animals from crossbred HF groups had a wider genetic variation than animals in the purebred HF group.

LD decay from 10 to 100 Kb



LD decay from 100 kb to 1 Mb



The LD decay up to 1 Mb: The analysis of **LD** decay for distances separated every 10 kb from 100 kb up to 1 Mb, the mean r^2 declined from 0.111 to 0.037 for HF < 75%, 0.113 to 0.043 for 75% ≤ HF < 80%, 0.109 to 0.043 for 80% ≤ HF < 85%, 0.111 to 0.043 for 85% ≤ HF < 90%, 0.116 to 0.047 for 90% ≤ HF < 95%, 0.122 to 0.047 for 95% ≤ HF < 100%, and 0.142 to 0.061 for 100% HF. The **LD** decay pattern from 100 kb to 1 Mb declined rapidly and were lower in HF crossbreds than in HF purebreds. High **LD** in HF crossbreds at shorter distances was likely related to higher ancestral relatedness. Higher level of **LD** at longer distances in purebred HF could be an indication of a recent reduction in effective population size and a stronger influence of recent positive selection (Ardlie *et al.,* 2002).

FINAL REMARKS

- ***All HF groups had similar SNP distribution by MAF***
- ***The mean MAF decreased as HF fraction increased***
- ***The mean r^2 increased as HF fraction increased***
- ***The distance between SNP at r^2 = 0.2 was longer for Purebred HF and 95% ≤ HF < 100% (60 kb) than for crossbred HF < 95% (40 kb)***

LITERATURE CITED

Ardlie, K.G., L. Kruglyak and M. Seielstad. 2002. Nat. Rev. 3: 299-309.

Barrett, J.C., B. Fry, J. Maller and M.J. Daly. 2005. Bioinformatics 21:263-265.

Fortes, M.R.S., A. Reverter, Y. Zhang, E. Collis, S.H. Nagaraj, N.N. Jonsson, K.C. Prayaga, W. Barris and R.J. Hawken. 2010. PNAS 107(31): 13642-13647.

Meuwissen, T.H.E., B.J. Hayes and M.E. Goddard. 2001. Genetics 157: 1819-1829.

Pérez O'Brien A.M., G. Mészáros, Y.T. Utsunomiya, T.S. Sonstegard, J.F. Garcia, C.P. Van Tassell, R. Carvalheiro, M.V.B. da Silva and J. Solkner. 2014. Livest. Sci. 166: 121-132.

Salomon-Torres, R., L.K. Matukumalli, C.P. Van Tassell, C. Villa-Angulo, V.M. Gonzalez-Vizcarra and R. Villa-Angulo. 2014. PLoS ONE 9(7):e103046.