

Association between Genomic SNPs and Dairy Production Traits in Thai Multibreed Dairy Cattle

Pimchanok Yodklaew*, Skorn Koonawootrittriron*, Mauricio A. Elzo† and Thanathip Suwanasopee*

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

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



SUMMARY

Genomic chips containing large number of single nucleotide polymorphisms (SNPs) have been widely used to study marker-trait associations in dairy cattle. Identification of genes associated with economically relevant traits close to significant SNPs are subsequently identified in genomic databases. The objective of this study was to find SNPs and genes that were significantly associated with milk yield (MY), initial yield (IY), peak yield (PY), day to peak (DP), persistency (PS) and age at first calving (AFC) in a Thai Holstein-Other multibreed dairy population. The dataset contained 2,409 first-lactation records from 295 dairy farms located in Central, Northern, Western and Southern Thailand that were collected between 1997 and 2014. Animals were genotyped with one of four GeneSeek Genomic Profiler BeadChips (9K, 20K, 26K, or 80K). Thus, a set of 7,357 SNPs in common among the four chips were used in this study. Association between SNPs and traits were analyzed using QXPAK.5 software. The mixed model included herd-year-season, Holstein fraction, age at first calving and SNPs as fixed effects, and animal and residual as random effects. The number of significant SNPs associated with MY, IY, PY, DP, PS and AFC were 649, 491, 627, 477, 527, and 381 at $P < 0.05$, and 342, 301, 215, 259 and 156 at $P < 0.01$, respectively. All significant SNPs were checked for their association to 5,320 genes in the NCBI database obtained using Map2NCBI (R package). No gene was related to all traits. However, eight genes were associated with five traits at $P < 0.05$ (*C3H1orf87*, *LOC100848063*, *NR5A2*, *PCDH15*, *WVO30*, *NCAM1*, *GLI2*, and *LOC784126*), and seven genes were associated with four traits at $P < 0.01$ (*NR5A2*, *KCNIP1*, *INSC*, *LOC784126*, *EEF1E1*, *ATRNL1* and *LOC100294923*). Results from this research emphasized the need to validate SNP-dairy trait associations under Thai tropical environmental conditions to optimize the benefits of genomic selection.

Genotypic data: Blood samples were collected from all cows and then DNA was extracted using a MasterPure™ DNA Purification Kit (Epicentre®, Madison, WI, USA). The Quality of DNA samples was measured using a Thermo Scientific NanoDrop™ 2000 Spectrophotometer. All animals were genotyped with one of four GeneSeek Genomic Profiler BeadChips (9K, 20K, 26K, or 80K; GeneSeek, Lincoln, NE, USA). Only SNPs from autosomes in common among the four chips were considered. SNPs with call rates lower than 90% and/or minor allele frequency less than 0.01 were removed. Thus, the final set of SNP markers included in the study was 7,357.



Association Analysis: Single marker analyses were performed using QXPAK.5 software. The mixed model analysis considered herd-year-season, Holstein fraction, age at first calving, and SNP marker as fixed effects. Random effects were animal and residual. All significant SNPs were checked for their association with genes in the NCBI database using the R package Map2NCBI (Hanna et al., 2014). The SNPs and gene interaction networks with all traits were visualized using Cytoscape3.1.0 software (Shannon et al., 2003). The false discovery rate (FDR) was calculated as $(P(1-s)/s(1-P))$, when P is the P-value and s is a number of significant SNP divided by number of total SNP (Bolormaa et al., 2011).

Table 1 Description of milk yield (MY), initial yield (IY), peak yield (PY), days to peak (DP), persistency (PS) and age at first calving (AFC)

Trait	Number	Mean	SD	Min.	Max.
MY (kg)	2,409	4,380.94	1,099.87	1,104.00	9,074.00
IY (kg)	1,689	10.49	6.89	1.00	39.00
PY (kg)	1,759	18.14	4.94	4.00	39.00
DP (day)	1,727	53.24	38.90	1.00	246.00
PS	1,645	6.88	1.05	2.80	18.20
AFC (month)	2,392	31.85	6.67	19.00	65.00

RESULTS AND DISCUSSION

The number of significant SNPs associated with MY was 649 at $P < 0.05$, and 342 at $P < 0.01$. The most significant SNPs for MY were found in chromosome 4 at $P < 0.05$, and in chromosomes 9 and 11 at $P < 0.01$. Significant SNPs were associated with 604 genes at $P < 0.05$ and 322 genes at $P < 0.01$. An SNP marker in chromosome X that presented a strong association ($P < 6.51 \times 10^{-14}$) with MY was located within gene *LOC532022*. Genes close to significant SNP in this study were different from genes found in other dairy populations. Interestingly, *DGAT1*, a gene known for its association with MY in other dairy populations (Minozzi et al., 2013) showed no significance in this Thai multibreed population.

The number of significant SNPs associated with IY, PY, DP and PS were 491, 627, 477, 527 at $P < 0.05$, and 225, 301, 215, 259 at $P < 0.01$, respectively. The significant SNP with the strongest associations were located on chromosomes 19, 5, X and 27 for IY, PY, DP and PS, respectively. Significant SNPs were associated with 449, 576, 441 and 487 genes at $P < 0.05$ and 210, 285, 204 and 247 genes at $P < 0.01$ for IY, PY, DP and PS, respectively. The SNP with the strongest associations with lactation characteristics were within or close to the *KDM3A* gene for IY, the *LOC100300697* gene for PY, the *MAMLD1* gene for DP, and the *RARB* gene for PS.

The SNP with the strongest association with IY was located within the *KDM3A* gene, which is involved in hormone-dependent transcriptional activation, spermatogenesis, and obesity resistance through regulation of metabolic genes. The SNP close to pseudo gene *LOC100300697* had a strong association with PY. While, the SNP close to the *MAMLD1* gene was highly associated with DP. This gene encodes a mastermind-like domain containing protein which may function as a transcriptional co-activator. Lastly, the SNP with the strongest association with PS was the *RARB* gene that encodes for retinoic acid receptor beta.

The number of significant SNPs associated with AFC were 381 at $P < 0.05$ and 186 at $P < 0.01$. The most significantly SNPs associated with AFC were found in chromosomes 2 and 5 at $P < 0.05$, and in chromosome 19 at $P < 0.01$. Significant SNPs were within or close to 359 genes at $P < 0.05$ and 154 genes at $P < 0.01$. The SNP marker that had the strongest association with AFC ($P < 2.78 \times 10^{-9}$) was located within the *LOC527335* gene in chromosome 11.

The significant SNPs found here were mostly different from those found in other dairy populations. This may be have been due to population differences because most dairy cattle in Thailand were multibreed, whereas other studies were conducted in single-breed populations. Raven et al. (2014) found different numbers of significant SNPs and low fraction of SNP in common between Holstein and Jersey for milk, fat, and protein yields, and fat and protein percentages.

Table 2 False discovery rate of milk yield (MY), initial yield (IY), peak yield (PY), days to peak (DP), persistency (PS) and age at first calving (AFC)

P-Value	MY	IY	PY	DP	PS	AFC
0.05	0.54	0.74	0.56	0.76	0.68	0.96
0.01	0.21	0.32	0.24	0.34	0.28	0.47

The associations between SNPs markers and traits found here indicated that genomic improvement for MY, IY, PY, DP, PS, and AFC in this Thai multibreed dairy cattle could be done efficiently, particularly if SNPs were located within important functional genes. Thus, further studies aiming at functional genes, gene associations, and gene pathways will be necessary to help explain the biological mechanism of relevant genes and gene expression under Thai tropical production conditions.

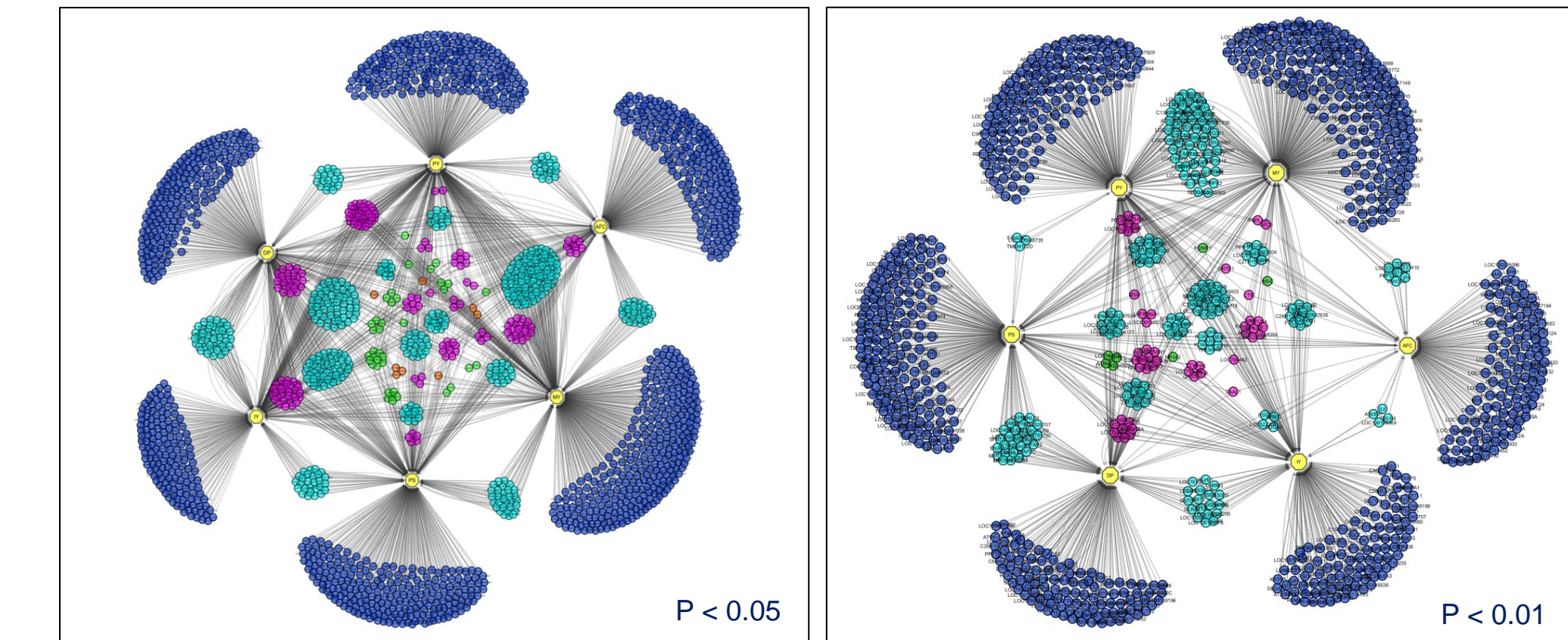
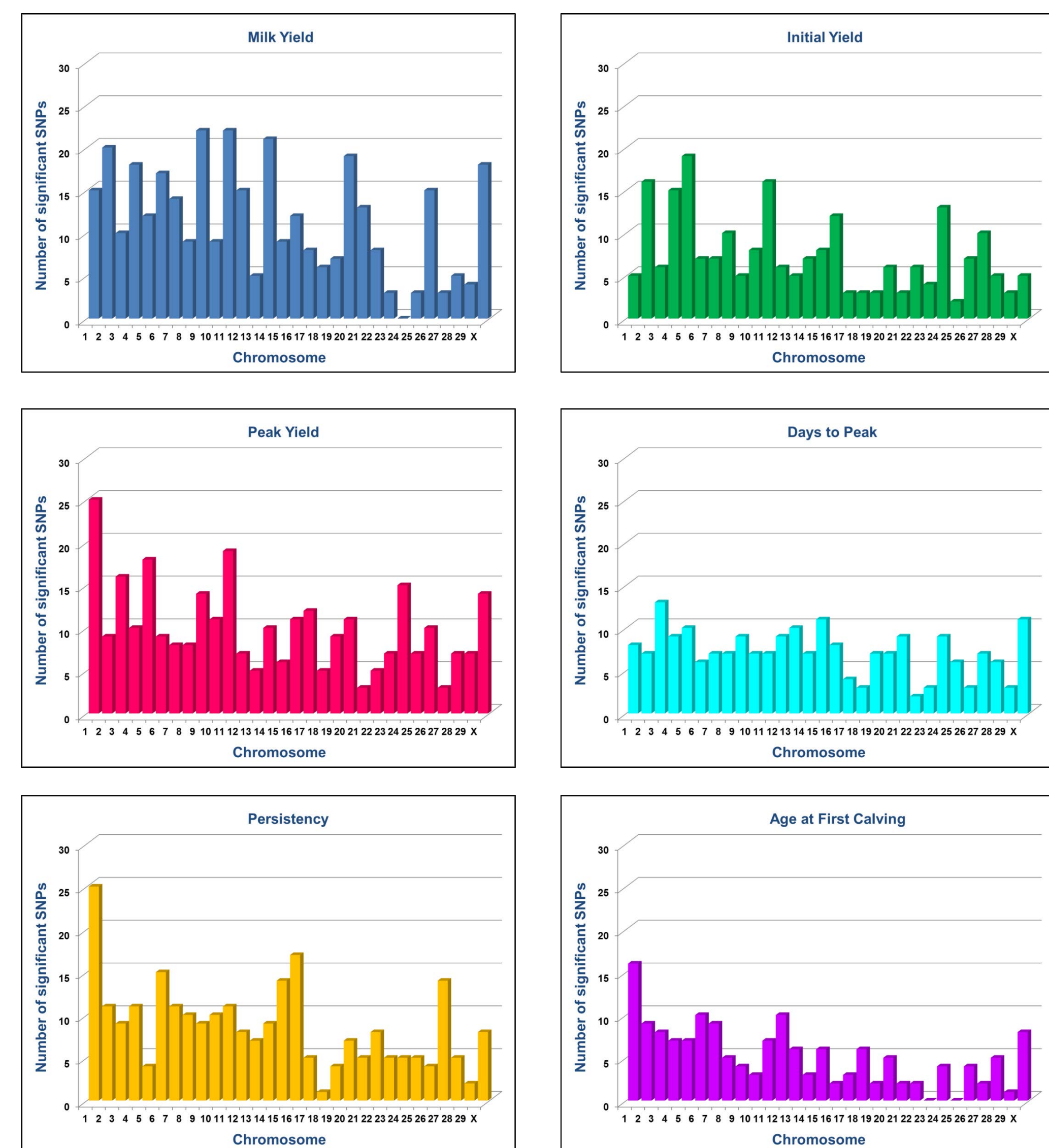


Figure 1 Genes associated with traits (yellow color) at $P < 0.05$ and $P < 0.01$ [dark blue color represents 1 trait, light blue color represents 2 traits, pink color represents 3 traits, green color represents 4 traits, and orange color represents 5 traits]

There were eight genes (*C3H1orf87*, *LOC100848063*, *NR5A2*, *PCDH15*, *WVO30*, *NCAM1*, *GLI2*, and *LOC784126*) associated with 5 traits ($P < 0.05$), and 7 genes (*NR5A2*, *KCNIP1*, *INSC*, *LOC784126*, *EEF1E1*, *ATRNL1* and *LOC100294923*) associated with 4 traits ($P < 0.01$; Figure 1). Groups of genes that had different functions affected different target organs. These genes may be correlated with other genes or receptor genes that affected traits in this study (gene function was searched at NCBI and GeneCards databases).

The groups of genes that were associated with five traits at $P < 0.05$ had important roles in embryonic development (*NR5A2*) and in embryogenesis (*GLI2*). Gene *PCDH15* is related to development and maintenance of hairlike projections (stereocilia), and gene *NCAM1* is involved in neuron-neuron adhesion, neurite fasciculation, outgrowth of neurites. Others were pseudo genes of unknown functions.

The groups of genes that associated with four traits at $P < 0.01$ were related to embryonic development (*NR5A2*), transportation of potassium (*KCNIP1*), cell proliferation and differentiation in the developing nervous system (*INSC*), encoding of a multifunctional protein localized in the cytoplasm and nucleus (*EEF1E1*), and were important for melanocortin signaling pathways that regulate energy homeostasis (*ATRNL1*).

These markers could be use to improve MY, IY, PY, DP, PS, and AFC. Selection for these various markers could help dairy farmers identify the best replacement animals. Commercial enterprises and dairy farmers in Thailand could benefit from the use of genotype chips like the ones used in this study to further improve their genetic management and selection programs.

FINAL REMARKS

- GWAS analyses were used to identify SNP associations with dairy production traits in a Thai multibreed population
- *C3H1orf87*, *LOC100848063*, *NR5A2*, *PCDH15*, *WVO30*, *NCAM1*, *GLI2*, and *LOC784126* genes were associated with five traits at $P < 0.05$ and *NR5A2*, *KCNIP1*, *INSC*, *LOC784126*, *EEF1E1*, *ATRNL1* and *LOC100294923* genes were associated with four traits at $P < 0.01$
- Significant SNP here emphasized the need to validate SNP-dairy trait associations under tropical environmental conditions to optimize the benefits of genomic selection in the Thai multibreed population

LITERATURE CITED

Bolormaa S., B.J. Hayes, K. Savin, R. Hawken, W. Barendse, P.F. Arthur, R.M. Herd and M.E. Goddard. 2011. J. Anim. Sci. 89: 1684-1697.
 Hanna, L. L. H. and D. G. Riley. 2014. Livest. Sci. 162: 59-65.
 Minozzi, G., E.L. Nicolazzi, A. Stella, S. Biffani, R. Negrini, B. Lazzari, P. Ajmone-Marsan and J.L. Williams. 2013. PLoSONE 8(11): e80219.
 Perez-Enciso, M. and I. Misztal. 2011. BMC Bioinform. 12: 202.
 Raven, L.A., B.G. Cocks and B.J. Hayes. 2014. BMC Genomics 15: 62.
 Shannon, P., A. Markiel and O. Ozier. 2003. Genome Res. 13: 2498-2504.

INTRODUCTION

Genetic improvement of dairy cattle for economically important traits is necessary for dairy farmers. Dairy trait phenotypes are quantitative and complex, particularly when they come from multibreed genotypes and are expressed in a variety of environments. Best linear unbiased prediction (BLUP) procedure has been used for decades to predict animal breeding values. Genome-wide association study (GWAS) is a technology of interest to livestock genetic improvement programs. GWAS has been used to identify SNPs that are related to several complex traits in humans and animals. Genotypic data from thousands of loci obtained by using SNP arrays has increased the power to identify genes related to economically relevant phenotypes. Researchers can detect statistical associations between SNPs and phenotypes, and also determine biological functions of candidate genes that are close to significant SNPs. GWAS have usually focused on economically important traits such as production and reproduction traits in dairy cattle. However, no report on GWAS for dairy cattle in Thailand currently exists. Thus, the objectives of this research were to conduct a genome wide association study to identify SNP associated with milk yield, lactation characteristics, and age at first calving in a Thai multibreed population, and locate genes near these SNP in the NCBI database.

MATERIALS AND METHODS

Animals and Phenotypes: The complete pedigree and phenotype records of 2,409 first lactation cows were collected from 295 dairy farms located in Central (919 cows), Northern (775 cows), Northeastern (251 cows), Western (177 cows) and Southern (287 cows) of Thailand during 1997 to 2014. The main breed fraction of all animals in this multibreed dairy population was Holstein, while the other breeds were Brahman, Brown Swiss, Jersey, Red Danish, Red Sindhi, Sahiwal, Simmental and Thai Native. Seasons in Thailand were winter (November to February), summer (March to June), and rainy (July to October). All cows were artificial inseminated with selected bull's semen. Animals were kept in open-barns and milked twice a day (5 a.m. and 3 p.m.). Feeding was based on fresh grasses (*Brachiaria mutica*, *Brachiaria ruziziensis*, *Penicum maximum*, and *Pennisetum purpureum*), concentrate and minerals. During the dry seasons (winter and summer), when fresh grasses were limited, rice straw, crop residues, and agricultural by-products were provided.

Monthly test-day milk yields for individual cows were recorded from calving until drying off. These monthly records were used to estimate IY, PY, DP, PS and MY using Wood's gamma function (Wood, 1967). Age at first calving (AFC) was computed as the number of days between birth and first calving dates. Descriptive statistics of the phenotypic observations are presented in Table 1.