Original Article

Genome-wide association study for lactation characteristics, milk yield and age at first calving in a Thai multibreed dairy cattle population

Pimchanok Yodklaew,[a, b] Skorn Koonawootrittriron,[a, b, *] Mauricio A. Elzo,[b, c] Thanathip Suwanasopee,[a, b] Thawee Laodim[a, b]

[a] Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand
[b] Animal Breeding and Genetics Consortium of Kasetsart University (ABG − KU), 50 Ngamwongwan, Chatuchak, Bangkok 10900, Thailand
[c] Department of Animal Sciences, University of Florida, Gainesville, FL 32611-0910, USA

ARTICLE INFO

ABSTRACT

A genome-wide association study was performed for milk yield per lactation (MY), initial yield (IY), peak yield (PY), persistency (PS) and age at first calving (AFC) in a Thai multibreed dairy cattle population. The dataset contained 1305 first-lactation cows raised on 188 farms located in Central, Northeastern and Southern Thailand. Cows were genotyped with GeneSeek Genomic Profiler low-density bead chips (8810 single nucleotide polymorphism [SNP]; $n = 1255$) and with high-density bead chips (76,883 SNP; $n = 50$). The single SNP association analyses utilized 8096 SNPs in common between the low and high density GeneSeek chips. The mixed model contained the fixed effects of contemporary group, fraction of non-Holstein breeds, age at first calving and gene content, and the random effects of animal and residual. Computations were done with the QXPAK.5 software. The number of SNPs associated with MY, IY, PY, PS and AFC at the significant threshold level of $p < 0.00001$ were 75, 102, 145, 74 and 24, respectively. Of the 366 SNP markers significantly associated with the studied traits, 54 (14.75%) were associated with two traits and 312 (85.25%) with only one trait, and all but one of the 54 SNPs associated with two traits affected MY and lactation characteristics. Genetic improvement of Thai dairy cows for lactation characteristics, milk yield and age at first calving could be aided by selecting animals with the SNP markers found to be highly associated with genes influencing these traits.

Introduction

Milk yield is an economically important trait of dairy production systems and it is a source of nutrients for mammals as it contains high-value proteins, fat, carbohydrates, vitamins and minerals (Jenness, 1988). Milk yield per lactation (MY) is affected by age at first calving (AFC) and it is associated with lactation characteristics such as initial yield (IY), peak yield (PY) and persistency (PS) (Seangjun et al., 2009). All of these traits have been included in the genetic evaluation of animals in the Thai Holstein (H)-Other (O) multibreed dairy cattle population in Thailand since 2008 (Koonawootrittriron et al., 2008). However, estimated breeding value (EBV) accuracies of animals for these traits have been low

because of the small size of the population ($n = 5835$ first-lactation cows in the 2015 genetic evaluation). Inclusion of genotypic information from single nucleotide polymorphism (SNP) markers from commercially available low density and medium density chips (for example, GGP-LD and GGP-HD; GeneSeek; Lincoln, NE, USA) may help increase the accuracy of EBV in this population. More importantly, associations between MY, IY, PY, PS, AFC and SNP markers in these chips can be used to identify genes in close proximity to significant SNP markers to help understand their biological effect on these traits. This information will be useful for the design of genotyping chips that include biologically relevant SNP markers. Genome-wide association studies (GWAS) have been used to identify important genomic regions influencing complex traits (Hayes and Goddard, 2010; Meredith et al., 2012), to identify markers that would increase EBV accuracy, and to increase understanding of genetic influences on economically important dairy traits (Pryce et al., 2010). However, most GWAS have involved

* Corresponding author. Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand.
E-mail address: agrskk@ku.ac.th (S. Koonawootrittriron).

purebred dairy populations. Conversely, most dairy cattle in Thailand (91%) are Holstein crossbreds produced through a grading-up process (Ritsawai et al., 2014). Further, environmental conditions in Thailand differ from other populations because of geography, natural resources, agricultural activity, culture and especially hot and humid weather conditions (Yodklaew et al., 2013). Genetic markers important in other populations may not necessarily be relevant under Thai conditions. Thus, the objectives of this study were: 1) to conduct GWAS to identify SNP markers associated with MY, IY, PY, PS and AFC in a Thai multibreed dairy cattle population using a set of SNP markers in common between the GGP-LD and GGP-HD chips; and 2) to identify genes in the NCBI database that are located nearest to SNP markers found to be significantly associated with these traits.

Materials and methods

Animals, traits and management

Animals in this Thai multibreed dairy cattle population were produced by upgrading of various breeds and crossbred groups to Holstein as described by Koonawootrittriron et al. (2009). Breeds represented in various fractions in the population are Holstein, Thai Native, Sahiwal, Red Dane, Red Sindhi, Brahman, Brown Swiss and Jersey (Ritsawai et al., 2014). The initial crossbreeding of native cattle to various breeds and the subsequent upgrading to Holstein aimed at improving milk production. The Holstein percentage ranged from 38% to 100%; 90% of cows in the population were 87.5% Holstein and above. All animals in the population had some Holstein fraction (86.8 ± 11.9% Holstein, range 12.5–100% Holstein). Thus, the Holstein fraction was utilized to define genetic groups and accounted for population stratification in data analyses here.

Phenotypes and SNP genotypes were from 1305 first-lactation cows raised on 188 farms located in Central (797 cows), Northeastern (223 cows) and Southern Thailand (285 cows). These cows were the progeny of 291 sires and 1192 dams and there were 12,835 animals from four generations in the pedigree file. Pedigree and performance data collected by staff from the Dairy Farming Promotion Organization of Thailand were edited for erroneous information and the remaining data were used for data analyses.

Traits were milk yield per lactation (MY), initial yield (IY), peak yield (PY), persistency (PS) and age at first calving (AFC). The lactation characteristics (IY, PY and PS) and MY of individual cows were estimated using a gamma function with monthly test-day records (Wood, 1967). The gamma function was $y_t = at^b\,e^{-ct}$, where y_t is the daily milk yield at time t, a is the average daily milk yield at the beginning of the lactation, b is the rate of increase in daily milk yield before the peak yield and c is the rate of decrease in daily milk yield after the peak yield. Traits were computed as follows: IY $= a$, PY $= a(b/c)^b\,e^{-b}$, PS $= c^{-(b+1)}$, MY $= \sum_{t=1}^{t=305} y_t$, and AFC = number of days from birth to first calving (Seangjun et al., 2009). Descriptive statistics for all traits are shown in Table 1.

Cows were artificially inseminated at the second or third observed estrous. Cows in all regions were kept in open barns and milked twice a day. The first milking was in the morning (0500 h) and the second one in the afternoon (1500 h). Feeding was based on concentrate (1 kg of concentrate per 2 kg of milk), and fresh grass (30–40 kg/d). Seasons were classified as winter (November–February), summer (March–June), and rainy (July–October).

Genotypic data

Blood samples were collected from cows for DNA extraction using a MasterPure ™ DNA Purification Kit (Epicentre®; Madison, WI, USA). The quality of the DNA samples was measured using a spectrophotometer (NanoDrop™ 2000; Thermo Scientific; Wilmington, DE, USA). A total of 1255 cows were genotyped with GeneSeek Genomic Profiler low-density bead chips (GGP-LD; 8810 SNP) and 50 cows were genotyped with GeneSeek Genomic Profiler high-density bead chips (GGP-HD; 76,883 SNP; GeneSeek, Lincoln, NE, USA). SNPs with call rates lower than 90%, minor allele frequencies lower than 0.01 and p values for a Hardy-Weinberg equilibrium lower than 0.0001 were excluded from statistical analyses. The final set of SNP markers consisted of 8096 SNPs present in both the GGP-LD and GGP-HD chips. SNP positions within chromosomes were based on the bovine genome assembly build 6.1 (http://BovineGenome.org).

Genome-wide association analysis

The QXPAK.5 software (Perez-Enciso and Misztal, 2011) was used to perform single-trait GWAS for MY, IY, PY, PS and AFC using the mixed linear model shown in equation (1):

$$y_{ijkm} = HYS_i + b_1(BF) + b_2(AFC) + SNP_j + Anim_k + e_{ijkm} \qquad (1)$$

where y_{ijkm} was is a record for MY, IY, PY, PS or AFC, HYS_i is the contemporary group defined as herd-year-season subclass (1–698), BF is the fraction of breeds other than Holstein, AFC is the age at first calving, SNP_j is the difference between the effects of alleles A and B in locus j, $Anim_k$ is the random effect of animal k and e_{ijkm} is the random residual term. Age at first calving was excluded from the model when analyzed as a trait. The significance of the association between SNP and trait phenotypes was evaluated using a likelihood ratio test. The output of the QXPAK.5 program provided a likelihood ratio and a p value for each SNP. The threshold p value used here to assign significance was 0.00001, which was more stringent than the smallest p value of 0.0001 used by Bolormaa et al. (2011).

False positives were accounted for by using the false discovery rate (FDR). The FDR controls the expected fraction of false positives (Benjamini and Hochberg, 1995) rather than the probability of a single false discovery controlled by family-wise error rate procedures such as the Bonferroni correction. The FDR was computed for each trait at the threshold p value of 0.00001 using the ratio of $p(1-s)$ divided by $s(1-p)$, where p is the defined probability threshold (0.00001) and s is the proportion of SNPs that were significant at the defined threshold (number of significant SNPs divided by total number of SNPs) as described by Bolormaa et al. (2011). Subsequently, the number of true positives was computed as the product of the number of significant SNPs at the threshold value times (1 − FDR) for each trait. The NCBI database (http://www.ncbi.nlm.nih.gov/gene/?term=gene) and the Gene Cards Human Gene Database (http://www.genecards.org/cgi-bin/carddisp.pl?gene) were used to obtain the names and functions of genes associated to the nearest significant SNP. Genes in the NCBI database located in close proximity to significant SNPs for MY, IY, PY, PS and AFC were identified using the R package Map2NCBI (Hanna and Riley, 2014). This program identified these genes using the "GetGeneList" and "MapMarkers" functions.

Table 1
Descriptive statistics for milk yield (MY), initial yield (IY), peak yield (PY), persistency (PS) and age at first calving (AFC).

Trait	n	Mean	SD	Minimum	Maximum
MY (kg)	1282	4368.37	1099.64	1413.00	9073.00
IY (kg)	862	9.79	5.76	0.99	28.87
PY (kg)	862	18.60	4.21	6.86	36.98
PS	862	6.79	0.81	5.11	11.28
AFC (mth)	1297	30.97	6.65	16.00	62.00

The top-ten significant SNPs markers located less than 2.5 kb (nearest) to the genes were used for the gene function studies of each trait (Fortes et al., 2010). Distances between SNPs lower than 2.5 kb had linkage disequilibrium values higher than 0.515 in this Thai multibreed dairy population (Laodim et al., 2015). Thus, genes associated with these top-ten SNPs (<2.5 kb) were considered as candidate genes for genetic improvement of lactation characteristics, milk yield and age at first calving in a Thai multibreed dairy cattle population.

Additionally, the genetic correlation analysis for MY, IY, PY, PS and AFC used the ASREML software (Gilmour et al., 2009).

The qqman package (Turner, 2014) was used for creating Manhattan plots from GWAS results. Manhattan plots present $-\log_{10}$ (p values) for all SNP markers evaluated across the genome. The p values were plotted in genomic order by chromosome and position within the chromosome (x-axis). The value on the y-axis represented the $-\log_{10}$ of the p value (Turner, 2014).

Results and discussion

The final set of SNP markers (8096 SNPs) covered the region of genome 2653.30 Mb and their distribution varied among the chromosomes. Chromosome 1 had most SNPs (541) while chromosome 28 had the fewest SNPs (142). The average minor allele frequency across chromosomes was 0.37, and the range was from 0.01 to 0.5. This indicated variation in the allele frequency among the SNPs in this population. Genomic variation of cattle in this multibreed dairy population differed from those reported in other countries (The Bovine HapMap Consortium, 2009; Pérez O'Brien et al., 2014; Laodim et al., 2015).

Genome-wide association studies are important for identifying genome regions that influence traits of interest (Hayes and Goddard, 2010; Meredith et al., 2012), increasing the accuracy of genomic predictions (Zhang et al., 2015) and identifying candidate genes associated with SNP markers (Fortes et al., 2010). The number of significant SNP markers associated with lactation characteristics, milk yield and age at first calving of cattle in this Thai multibreed dairy population were different (Fig. 1). The numbers of significant SNP and false discovery rates for each trait at the threshold value of $p < 0.00001$ are shown in Table 2. Table 3 shows the top-ten SNPs (the SNPs with the 10 lowest p values) that are associated with MY, IY, PY, PS and AFC, their chromosome location and their nearest genes in the NCBI database, which derived from Map2NCBI (Hanna and Riley, 2014).

Milk yield

The number of SNP markers associated with MY was 75 ($p < 0.00001$; Table 2). The top-ten significant SNP markers were nearest (less than 2.5 kb) to the *UFL1*, *RALYL*, *ADAMTS8*, *CACNB2*, *HMGCS1*, *LAMA4*, *ZNF618*, *IDE*, *AGXT2L2* and *TMEM132C* genes (Table 3). The functions of these genes can be classified into five categories. First are the protein-coding genes related to enzyme groups including *UFL1* (UFM1 specific ligase 1; http://www.genecards.org/cgi-bin/carddisp.pl?gene=UFL1), *ADAMTS8* (ADAM metallopeptidase with thrombospondin type 1 motif, 8; Georgiadis et al., 1999), and *HMGCS1* (3-hydroxy-3-methylglutaryl-CoA synthase 1) which are related to ligase activities, metallopeptidase and metalloendopeptidase activity and to HMG-CoA reductase. These genes are related to important enzymes in milk production, especially HMG-CoA reductase which is associated with cholesterol synthesis in the milk yield (Parodi, 1997). Second, the gene related to RNA and nucleotide binding is *RALYL* (RALY RNA binding protein-like; http://www.genecards.org/cgi-bin/carddisp.pl?gene=UFL1). Third are genes for transport including *CACNB2* (calcium channel,

voltage-dependent, or beta 2 subunit; Meissner et al., 2011; 1055-1060.ential anticarcinogenic agents.) and *KCNA2* (potassium voltage-gated channel, shaker-related subfamily, member 2; http://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNA2) which are related to high voltage-gated calcium and potassium channel activity, especially calcium which is a critical component in milk. Fourth are genes associated with immune response including *TLR3* (toll-like receptor 3; Oviedo-Boyso et al., 2007) and *LOC100299770* (60 kDa heat shock protein; https://www.ncbi.nlm.nih.gov/gene/100299770), which in cows may be stimulated from within or outside the body and may affect milk production. The last category is genes involved in eukaryotic cellular processes such as *LOC529930* (E3 ubiquitin-protein ligase MGRN1-like) which may be related to cell proliferation in the breast for milk production (https://www.ncbi.nlm.nih.gov/gene/529930). *LOC100848744* is a protein-coding gene of unknown function.

Significant associations between these top-ten genes and MY have not been reported in other dairy populations. Interestingly, the *DGAT1* gene on chromosome 14 (associated with intestinal fat absorption and regulation of plasma triacylglycerol concentration and energy metabolism in muscle) that was reported to have a significant association with MY in several dairy populations (Jiang et al., 2010; Meredith et al., 2012; Raven et al., 2014) was not found to be close to any SNPs significantly associated with MY in this population.

The additional genes found to be associated with MY in many dairy populations are *PPARGC1A*, *TG* and *PRLR* (Strucken et al., 2011) and *ASL*, *AVP*, *BOLA-DMB*, *CACNA1D*, *CAST*, *CD14*, *CPSF1*, *DEPDC7*, *DSC2*, *DTX2*, *DZIP3*, *FAMSC*, *GOLGA4*, *HSD17B7*, *IBSP*, *MARVELD1*, *MS4A8B*, *NLRP9*, *PARM1*, *PGR*, *TDRKH*, *TSHB* and *TXN2* (Cochran et al., 2013). Of all these genes, only *PPARGC1A* and *DSC2* were identified here. However, neither the SNP marker close to the *PPARGC1A* gene nor the one close to the *DSC2* gene were significantly associated with MY in this population.

Initial yield

There were 102 SNP markers associated with IY ($p < 0.00001$; Table 2). The top-ten significant SNPs were nearest (less than 2.5 kb) to the *ZMAT4*, *REEP1*, *GRM7*, *SCFD2*, *EEF1E1*, *TDRD9*, *DYSF* and *C13H20orf196* genes and near the *TREM1* and *C9H6orf162* genes (Table 3). The functions of these genes can be classified into seven categories. The first category includes genes related to DNA and zinc ion binding (*ZMAT4* [zinc finger, matrin-type 4]). Zinc helps to control processes that permit the body to perform efficiently and assist with the maintenance of enzymes and cells (Frederickson et al., 2000). Zinc is also important to heal wounds in cows resulting from calving (Goff and Stabel, 1990). The second category is genes that are related to the nervous system (*REEP1* [receptor accessory protein 1]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=REEP1) and *GRM7* (glutamate receptor, metabotropic 7) genes, and are involved in brain function and enhancement of cell surface expression of odorant receptors (http://www.genecards.org/cgi-bin/carddisp.pl?gene=GRM7). These olfactory senses are important for stimulating appetite in cows which would affect feed intake and milk yield. The third category consists of genes involved in protein transport (*SCFD2* [sec1 family domain containing 2]), that are essential for moving oxygen and carbon dioxide, such as hemoglobin in red blood cells which are important to keep the body functioning normally (http://www.genecards.org/cgi-bin/carddisp.pl?gene=SCFD2). The fourth category includes a gene that encodes for a multifunctional protein that is present in both the cytoplasm and the nucleus (*EEF1E1* [eukaryotic translation elongation factor 1 epsilon 1]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=EEF1E1). The fifth category is
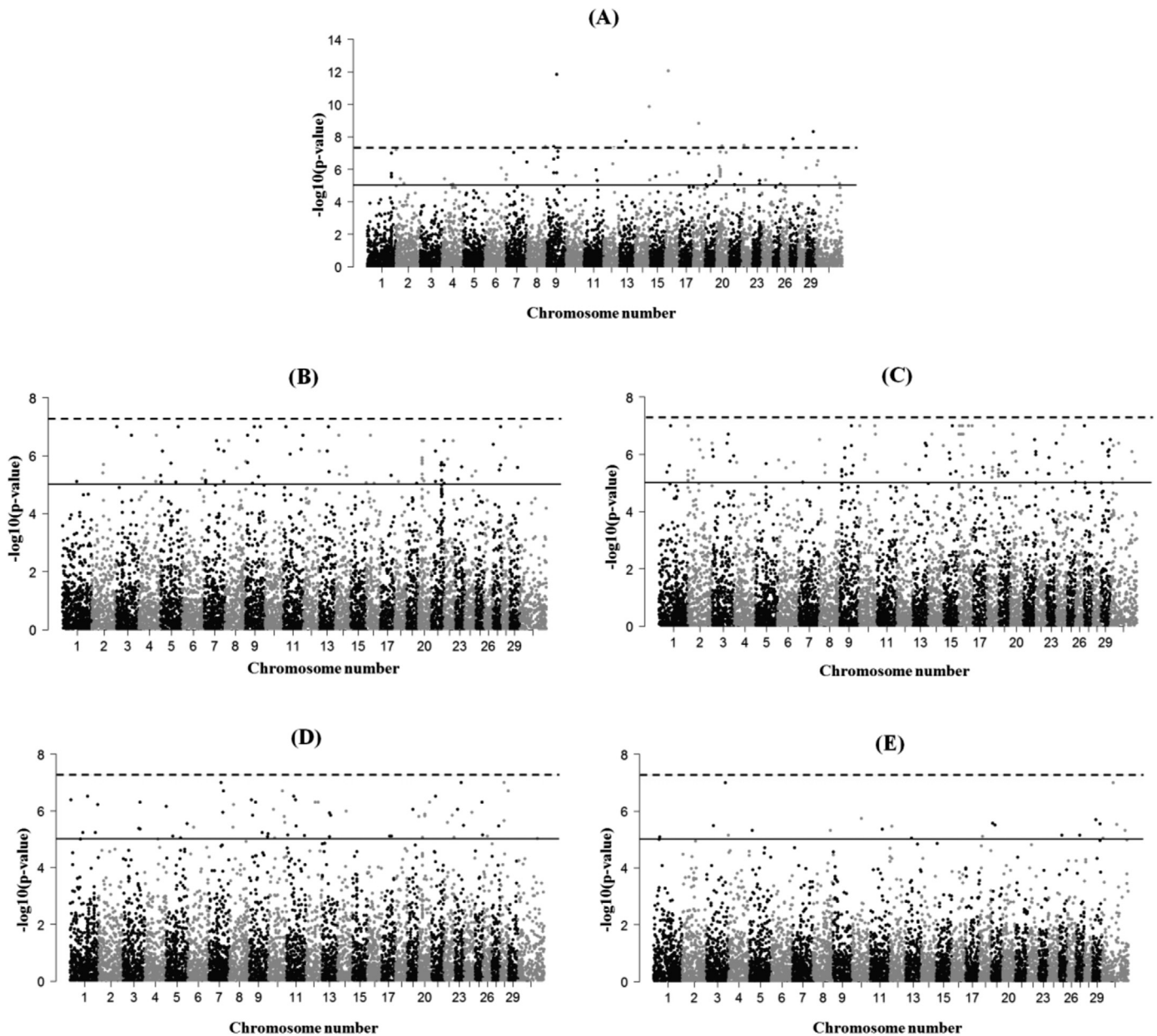
Fig. 1. Manhattan plots: (A) milk yield; (B) initial yield; (C) peak yield; (D) persistency; (E) age at first calving.

Table 2
Number of significant single nucleotide polymorphisms (SNPs) and false discovery rate (FDR) for milk yield (MY), initial yield (IY), peak yield (PY), persistency (PS) and age at first calving (AFC) at the significance threshold value ($p < 0.00001$).

Trait	Number of significant SNP markers	FDR (%)
MY	75	0.11
IY	102	0.08
PY	145	0.06
PS	74	0.11
AFC	24	0.34

formation of new muscle fibers (regeneration) and in inflammation (*DYSF* [dysferlin]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=DYSF). The seventh category contains genes involved in the stimulation of neutrophil and monocyte-mediated inflammatory responses (*TREM1* [triggering receptor expressed on myeloid cells 1]) that protect cows against infections and diseases that may occur during the postpartum period (Moyes et al., 2009). Other genes including *C13H20orf196* and *C9H6orf162* have unknown functions.

Peak yield

In total, 145 SNP markers were associated with PY ($p < 0.00001$; Table 2). The top-ten significant SNPs were nearest (less than 2.5 kb) to the *RGL1*, *CCDC141*, *DPY19L1*, *LOC787329*, *C13H20orf94*, *BRP44*, *LOC100847667*, *MNS1*, *KCNH5* and *OR1J2* genes (Table 3).

related to ATP-binding RNA helicase which plays a central role during spermatogenesis by repressing transposable elements and preventing their mobilization, which is essential for the germline integrity (*TDRD9* [tudor domain containing 9]; Han and Penagaricano, 2016). The sixth category is involved in the

Table 3
Top[a]-ten single nucleotide polymorphisms (SNPs) associated with milk yield (MY), initial yield (IY), peak yield (PY), persistency (PS) and age at first calving (AFC).

Trait	SNP	Chromosome	Allele	MAF[2]	Position (bp)	Nearest gene	Distance (bp)	p-Value
MY	Hapmap59494-rs29020429	9	[A/G]	0.406	54,062,560	UFL1	within	1.39×10^{-12}
	ARS-USMARC-Parent-DQ846692-rs29010281	14	[A/G]	0.393	80,082,923	RALYL	within	1.38×10^{-10}
	ARS-BFGL-NGS-18010	29	[A/G]	0.338	37,101,861	ADAMTS8	within	4.94×10^{-9}
	UA-IFASA-6479	13	[T/C]	0.468	33,201,457	CACNB2	within	1.80×10^{-8}
	Hapmap56889-rs29016097	20	[G/C]	0.314	31,468,943	HMGCS1	within	3.75×10^{-8}
	Hapmap29482-BTA-146449	9	[T/A]	0.343	38,804,655	LAMA4	within	3.82×10^{-8}
	Hapmap39690-BTA-82650	8	[T/G]	0.364	104,829,510	ZNF618	within	4.02×10^{-8}
	ARS-BFGL-NGS-89510	26	[T/G]	0.339	13,936,262	IDE	within	6.19×10^{-8}
	ARS-BFGL-NGS-113876	7	[A/G]	0.330	40,644,182	AGXT2L2	within	9.36×10^{-8}
	ARS-BFGL-NGS-42950	17	[T/C]	0.358	49,531,815	TMEM132C	within	1.00×10^{-7}
IY	Hapmap46767-BTA-117224	27	[T/C]	0.402	35,486,995	ZMAT4	within	0.00×10^{-14}
	ARS-BFGL-NGS-107346	11	[T/C]	0.470	48,495,399	REEP1	within	9.34×10^{-13}
	BTB-00839354	22	[A/G]	0.349	19,129,988	GRM7	within	1.52×10^{-11}
	ARS-BFGL-NGS-92044	6	[A/G]	0.4276	70,581,495	SCFD2	within	9.89×10^{-10}
	Hapmap41022-BTA-55560	23	[A/G]	0.269	15,147,471	TREM1	600	5.70×10^{-10}
	Hapmap26693-BTA-146557	9	[A/G]	0.346	63,460,554	C9H6orf162	1767	1.63×10^{-10}
	Hapmap23842-BTA-56935	23	[G/T]	0.448	47,219,946	EEF1E1	within	2.19×10^{-8}
	ARS-BFGL-NGS-7420	21	[C/T]	0.288	70,155,265	TDRD9	within	3.11×10^{-8}
	Hapmap48102-BTA-85468	11	[C/T]	0.348	12,970,800	DYSF	within	6.35×10^{-8}
	BTB-01718516	13	[A/G]	0.396	48,318,066	C13H20orf196	within	9.53×10^{-8}
PY	ARS-BFGL-NGS-101997	16	[T/C]	0.364	66,404,619	RGL1	within	1.10×10^{-11}
	ARS-BFGL-NGS-104601	2	[A/G]	0.385	17,857,541	CCDC141	within	4.56×10^{-10}
	BTB-00193035	4	[A/G]	0.251	62,448,627	DPY19L1	within	1.92×10^{-9}
	ARS-BFGL-NGS-15071	11	[T/C]	0.373	4,373,645	LOC787329	within	8.88×10^{-9}
	ARS-BFGL-NGS-103379	13	[A/G]	0.485	3,764,223	C13H20orf94	within	1.01×10^{-9}
	ARS-BFGL-NGS-1846	11	[A/C]	0.345	93,612,322	OR1J2	1393	1.03×10^{-8}
	ARS-BFGL-NGS-67047	3	[A/G]	0.127	787371	BRP44	within	1.72×10^{-8}
	ARS-BFGL-NGS-70792	29	[T/C]	0.418	38,102,852	LOC100847667	within	2.10×10^{-8}
	BTB-01137887	10	[T/G]	0.188	54,032,992	MNS1	within	2.81×10^{-8}
	ARS-BFGL-NGS-105188	10	[C/T]	0.352	75,601,133	KCNH5	within	2.82×10^{-8}
PS	ARS-BFGL-NGS-41356	9	[T/C]	0.467	101,966,211	C9H6orf118	37	1.78×10^{-10}
	BTB-00226439	5	[T/C]	0.120	43,125,599	PTPRB	within	8.78×10^{-10}
	Hapmap48102-BTA-85468	11	[T/C]	0.348	12,970,800	DYSF	within	4.15×10^{-9}
	BTA-71670-no-rs	4	[T/C]	0.481	95,465,677	LOC100848808	within	1.01×10^{-8}
	ARS-BFGL-NGS-110084	27	[T/C]	0.319	367537	MYOM2	within	3.51×10^{-8}
	ARS-BFGL-NGS-116814	7	[A/C]	0.468	67,074,400	GRIA1	within	5.63×10^{-8}
	ARS-BFGL-NGS-31956	23	[C/G]	0.456	36,112,949	LOC100847951	within	8.93×10^{-8}
	ARS-BFGL-NGS-94323	1	[A/G]	0.350	97,746,135	SLC7A14	within	3.24×10^{-7}
	ARS-BFGL-NGS-107346	11	[T/C]	0.470	48,495,399	REEP1	within	4.46×10^{-7}
	BTB-00389059	9	[G/A]	0.497	34,943,071	NT5DC1	within	4.62×10^{-7}
AFC	ARS-BFGL-NGS-32977	X	[A/G]	0.056	65,266,220	TRPC5	within	9.51×10^{-8}
	BTB-00148619	3	[A/G]	0.260	106,442,016	ZMPSTE24	within	1.00×10^{-7}
	BTB-01137887	10	[T/G]	0.188	54,032,992	MNS1	within	1.75×10^{-6}
	ARS-BFGL-NGS-12970	29	[A/C]	0.499	40,646,639	PPP1R32	2220	2.75×10^{-6}
	Hapmap52531-rs29012092	X	[A/C]	0.140	82,446,212	CHIC1	within	2.94×10^{-6}
	ARS-BFGL-NGS-31909	19	[A/G]	0.427	20,811,074	FLOT2	within	3.13×10^{-6}
	ARS-BFGL-NGS-20636	X	[A/G]	0.068	130,773,887	MAP3K15	within	4.74×10^{-6}
	ARS-BFGL-NGS-44288	27	[C/T]	0.422	18,810,353	SLC7A2	within	7.19×10^{-6}
	ARS-BFGL-NGS-107797	25	[G/A]	0.287	15,653,583	XYLT1	within	7.20×10^{-6}
	Hapmap34824-BES10_Contig592_452	1	[A/C]	0.317	33,641,578	CADM2	within	8.09×10^{-6}

sd[2] MAF = minor allele frequency.
[a] SNP with the smallest *p*-value.

These genes can be categorized into six groups. First are genes that are associated with protein and nucleotide binding (*RGL1* [ral guanine nucleotide dissociation stimulator-like 1]; http://www. genecards.org/cgi-bin/carddisp.pl?gene=RGL1) and *CCDC141* (uncharacterized *LOC530799*; http://www.genecards.org/cgi-bin/ carddisp.pl?gene=CCDC141). Second are protein-coding genes related to transferase activity and transferring glycosyl groups (*DPY19L1* [dpy-19-like 1]; Hansen et al., 2015). Third are genes involved in pyruvate transmembrane transporter activity (*BRP44* [brain protein 44 or mitochondrial pyruvate carrier 2]), which are essential for energy balance at the peak of the lactation (https:// www.ncbi.nlm.nih.gov/gene/?term=brp44+bos+taurus). The processing of energy in the body is important to indicate how well a cow responds to feeding and management during the dry, calving, and early lactation periods (Patton et al., 2007). Fourth are genes controlling meiotic division and germ cell differentiation through regulation of pairing and recombination during meiosis (*MNS1*

[meiosis-specific nuclear structural 1]; http://www.genecards.org/ cgi-bin/carddisp.pl?gene=MNS1). Fifth are genes related to regulation of neurotransmitters and hormone release, cardiac function, and cell volume (*KCNH5* [potassium channel, voltage gated eag related subfamily h, member 5]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNH5). Sixth are genes related to the nervous system that are involved in the control of the eukaryotic cell cycle, cell division and neuronal response initiating smell perception (*OR1J2* [olfactory receptor, family 1, subfamily J, member 2]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=OR1J2).

These genes may also increase of the number of udder cells during the PY period (Capuco et al., 2001), and affect cow response to feeding during dry, calving, and early lactation periods (Patton et al., 2007). The *LOC787329*, *C13H20orf94* and *LOC100847667* genes are protein-coding genes of unknown function. *LOC100848818* (60S ribosomal protein L9-like) is a pseudo gene in *Bos taurus* (https://www.ncbi.nlm.nih.gov/gene/100848818).

Persistency

There were 74 SNP markers associated with PS ($p < 0.00001$; Table 2). The top-ten significant SNPs were nearest (less than 2.5 kb) to the *PTPRB, DYSF, LOC100848808, MYOM2, GRIA1, LOC100847951, SLC7A14, REEP1* and *NT5DC1* genes and near the *C9H6orf118* gene (Table 3). The function of these genes can be classified into five categories. First are genes involved in cellular processes [*PTPRB* (protein tyrosine phosphatase, receptor type, B)] that are important for the maintenance and development of cells (http://www.genecards.org/cgi-bin/carddisp.pl?gene=PTPRB). Second are genes that are associated with muscle structure (*DYSF* [dysferlin, limb girdle muscular dystrophy 2B]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=DYSF) and *MYOM2* (myomesin [M-protein 2, 165 kDa]; Moreno-Sanches et al., 2010). Third are genes related to predominant excitatory neurotransmitter receptors in the mammalian brain (*GRIA1* [glutamate receptor, ionotropic, ampa 1]; Cushman et al., 2013). Fourth are genes involved in arginine transport (*SLC7A14* [solute carrier family 7, member 14]) that are important for cell division, the healing of wounds, removing ammonia from the body, immune functions, and the release of growth hormone (http://www.genecards.org/cgi-bin/carddisp.pl?gene=SLC7A14). Fifth are genes that enhance the cell surface expression of odorant receptors (*REEP1* [receptor accessory protein 1]). These genes were among the top-ten genes associated with IY. The protein-coding genes *LOC100848808, LOC100847951, NT5DC1* and *C9H6orf118* have unknown functions.

Age at first calving

In total 24 SNP markers were associated with AFC ($p < 0.00001$; Table 2). The top-ten SNPs were nearest (less than 2.5 kb) to the *TRPC5, ZMPSTE24, MNS1, CHIC1, FLOT2, MAP3K15, SLC7A2, XYLT1* and *CADM2* genes and near to the *PPP1R32* gene (Table 3). The function of these genes can be categorized into seven groups. The first group comprises genes related to cell development and ion transportation (*TRPC5* [transient receptor potential cation channel, subfamily C, member 5]; Howard et al., 2014) and a signal transduction pathway that is activated by various cell stresses and leads to apoptosis (*MAP3K15* [mitogen-activated protein kinase kinase kinase 15]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=MAP3K15). In addition, the *TRPC5* gene also influences sensory transduction, brain processing and calcium signalling (Kaczmarek, 2010). The second group encompasses protein-coding genes (*ZMPSTE24* [zinc metallopeptidase]) related to enzymes that cut other proteins, and may play a role in regulating the activity of certain genes (http://www.genecards.org/cgi-bin/carddisp.pl?gene=ZMPSTE24). The third group contains genes that control meiotic division and germ cell differentiation (*MNS1* [meiosis-specific nuclear structural 1]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=MNS1). The fourth group involves protein-coding genes related to protein binding (*FLOT2* [flotillin 2]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=FLOT2) and *PPP1R32* (protein phosphatase 1, regulatory subunit 32; http://www.genecards.org/cgi-bin/carddisp.pl?gene=PPP1R32). The fifth group is involved in the transport of the cationic amino acids (*SLC7A2* [solute carrier family 7 cationic

amino acid transporter, y+ system, member 2]; Forde et al., 2013). The sixth group is important for biosynthesis of chondroitin sulphate and dermatan sulphate proteoglycans in fibroblasts and chondrocytes (*XYLT1* [xylosyl transferase]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=XYLT1). The seventh group is important for synapse organization, providing regulated *trans*-synaptic adhesion (*CADM2* [cell adhesion molecule 2]; http://www.genecards.org/cgi-bin/carddisp.pl?gene=CADM2). The gene *CHIC1* (cysteine-rich hydrophobic domain 1) has unknown functions.

The three largest numbers of SNPs significantly associated with all traits within individual chromosomes were 32 (chromosomes 20), 28 (chromosome 9), and 18 (chromosome 16). This differed from the Brown Swiss cattle breed in Sweden, in which the greatest number of SNP associations between dairy production, fertility and conformation traits was found on chromosome 25 (Guo et al., 2012). While bovine chromosome 14 has been widely studied for its large number of QTL associated with economically important traits (Wibowo et al., 2008), only 10 SNP markers were associated with the traits in the current population.

Genes influencing more than one trait are likely to be the main source of genetic correlations between traits. Of the 366 SNP markers significantly associated with traits here, 54 (14.75%) were associated with two traits and 312 (85.25%) with only one trait, and all but one of the 54 SNPs were associated with two traits that affected MY and lactation characteristics (2 SNPs with MY and IY, 17 SNPs with MY and PY, 13 SNPs with IY and PY, 20 SNPs with IY and PS, and 1 SNP with PY and AFC). These SNP marker-trait associations were in agreement with genetic correlations estimated in this Thai population (Table 4), supporting the possible existence of pleiotropic effects of the genes associated with these SNP markers and traits in this study. Furthermore, the numbers of SNPs associated with pairs of traits were also roughly associated with the value or significance or both of the estimated genetic correlation. In particular, the estimates of genetic correlation between MY and PY (0.70; $p < 0.0001$) were associated with 17 SNPs, between IY and PY (0.42; $p < 0.0001$) with 13 SNPs and between IY and PS (-0.28; $p < 0.0001$) with 20 SNPs.

The likelihood of having the same set of significant SNPs in different populations would be rather small (Chamberlain et al., 2012) and this was supported by the fact that many of the SNPs found to be significantly associated with dairy traits in this population have not been reported elsewhere. Similarly, different sets of significant SNPs and associated genes were found to be important in other populations in temperate regions. Raven et al. (2014) found that the number of significant SNPs differed between Holstein and Jersey cows and the fraction of significant SNPs in common was low (4.7%, 6.9%, 6.3%, 38.9% and 27.3% for fat, milk, protein, and % fat and % protein, respectively). Further, no significant SNPs in common existed between these two breeds for calving interval. Differences in breed composition between the Thai population (multibreed *Bos taurus-Bos indicus* with a major Holstein component) and dairy populations in other studies (mostly purebred *Bos taurus*) as well as differences in environmental conditions (such as tropical versus temperate) likely influenced the sets of genes found to be associated with dairy traits in each case. The tropical environmental conditions in Thailand may have affected the expression of genes

Table 4
Genetic correlations between milk yield (MY), initial yield (IY), peak yield (PY), persistency (PS) and age at first calving (AFC).

Trait	MY	IY	PY	PS
IY	0.53 ($p < 0.0001$)			
PY	0.70 ($p < 0.0001$)	0.42 ($p < 0.0001$)		
PS	0.03 ($p = 0.3352$)	-0.28 ($p < 0.0001$)	-0.06 ($p = 0.0432$)	
AFC	0.06 ($p = 0.0418$)	0.13 ($p < 0.0001$)	0.05 ($p = 0.0615$)	-0.05 ($p = 0.0725$)

controlling the syntheses of proteins and enzymes influencing dairy biological processes differently from animals in temperate environments. Natural and artificial selection pressures may have favored a somewhat different set of genes associated with lactation characteristics, MY and AFC in these dairy cattle populations. Identification and evaluation of biologically relevant sets of genes under prevalent regional environmental conditions would help increase the accuracy and effectiveness of dairy cattle selection in Thailand. Selection programs in Thailand would need to target SNPs and genes found to be associated with dairy traits under local environmental conditions. Additional research will need to be conducted to explain biological relationships between genes found to be near significant SNPs and their associated traits. Understanding the biological function of specific genes and their role in physiological pathways will help to identify and predict the impact of selecting animals with specific sets of genes on future generations under Thai environmental conditions.

Implications

The genetic improvement of Thai dairy cows for lactation characteristics, milk yield, and age at first calving could be aided by selecting animals with SNP markers found to be highly associated with genes influencing these traits. The identified SNPs could be included in new genotyping chips targeting upgraded Holstein dairy cows kept under tropical conditions not only within Thailand but also in other tropical countries. Future research would need to compare the accuracy of genomic predictions of these tropically targeted genotyping chips with the accuracy of generic, commercially available, low and high density genotyping chips.

Conflict of interest

The authors declare that they do not have any conflicts of interest.

Acknowledgements

References

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 289–300.

Bolormaa, S., Hayes, B.J., Savin, K., Hawken, R., Barendse, W., Arthur, P.F., Herd, R.M., Goddard, M.E., 2011. Genome-wide association studies for feedlot and growth traits in cattle. J. Anim. Sci. 89, 1684–1697.

Capuco, A.V., Wood, D.L., Baldwin, R., Mcleod, K., Paape, M.J., 2001. Mammary cell number, proliferation, and apoptosis during a bovine lactation: relation to milk production and effect of bST. J. Dairy Sci. 84, 2177–2187.

Chamberlain, A.J., Hayes, B.J., Savin, K., et al., 2012. Validation of single nucleotide polymorphisms associated with milk production traits in dairy cattle. J. Dairy Sci. 95, 864–875.

Cochran, S.D., Cole, J.B., Null, D.J., Hansen, P.J., 2013. Discovery of single nucleotide polymorphism in candidate genes associated with fertility and production traits in Holstein cattle. BMC Genet. 14, 49. http://dx.doi.org/10.1186/1471-2156-14-49.

Cushman, R., Miles, J.R., Rempel, L.A., McDaneld, T.G., Kuehn, L.A., Chitko-McKown, C.G., Nonneman, D., Echternkamp, S.E., 2013. Identification of an ionotropic glutamate receptor AMPA1/GRIA1 polymorphism in crossbred beef cows differing in fertility. J. Anim. Sci. 91, 2640–2646.

Forde, N., Mehta, J.P., McGettigan, P.A., Mamo, S., Bazer, F.W., Spencer, T.E., Lonergan, P., 2013. Alterations in expression of endometrial genes coding for proteins secreted into the uterine lumen during conceptus elongation in cattle. Genomics 14, 321. http://dx.doi.org/10.1186/1471-2164-14-321.

Fortes, M.R.S., Reverter, A., Zhang, Y., et al., 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. Proc. Natl. Acad. Sci. USA. 107, 13642–13647.

Frederickson, C.J., Suh, S.W., Silva, D., Frederickson, C.J., Thompson, R.B., 2000. Importance of zinc in the central nervous system: the zinc-containing neuron. J. Nutr. 130, 1471–1483.

Georgiadis, K.E., Hirohata, S., Seldin, M.F., Apte, S.S., 1999. ADAM-TS8, a novel ADAM-TS8, a novel metalloprotease of the ADAM-TS family located on mouse chromosome 9 and human chromosome 11. Genomics 62, 312–315.

Gilmour, A.R., Gogel, B.J., Cullis, B.R., Thompson, R., 2009. ASREML User Guide Release 3.0. VSN International Ltd., Hemel Hempstead, HP1 1ES, UK.

Goff, J.P., Stabel, J.R., 1990. Decreased plasma retinol, α-tocopherol, and zinc concentration during the periparturient period: effect of milk fever. J. Dairy Sci. 73, 3195–3199.

Guo, J., Hossein, J., Orjan, C., 2012. A genome-wide association study using international breeding-evaluation data identifies major loci affecting production traits and stature in the Brown Swiss cattle breed. BMC Genet. 13, 82. http://dx.doi.org/10.1186/1471-2156-13-82.

Han, Y., Penagaricano, F., 2016. Unravelling the genomic architecture of bull fertility in Holstein cattle. BMC Genet. 17, 143. http://dx.doi.org/10.1186/s12863-016-0454-6.

Hanna, L.L.H., Riley, D.G., 2014. Mapping genomic markers to closest feature using the R package Map2NCBI. Livest. Sci. 162, 59–65.

Hansen, L., Lind-Thomsen, A., Joshi, H.J., et al., 2015. A glycogene mutation map for discovery of diseases of glycosylation. Glycobiology 25, 211–224.

Hayes, B., Goddard, M., 2010. Genome-wide association and genomic selection in animal breeding. Genome 53, 876–883.

Howard, J.T., Kachman, S.D., Snelling, W.M., Pollak, E.J., Ciobanu, D.C., Kuehn, L.A., Spangler, M.L., 2014. Beef cattle body temperature during climatic stress: a genome-wide association study. Int. J. Biometeorol. 58, 1655–1672.

Jenness, R., 1988. Composition of milk. In: Wong, N.P., Jenness, R., Keeney, M., Marth, E.H. (Eds.), Fundamentals of Dairy Chemistry, pp. 1–38 (New York, NY, USA).

Jiang, L., Liu, J., Sun, D., Ma, P., Ding, X., Yu, Y., Zhang, Q., 2010. Genome wide association studies for milk production traits in Chinese Holstein population. PLoS One 5, e13661. http://dx.doi.org/10.1371/journal.pone.0013661.

Kaczmarek, J.S., 2010. TRPC5: a startling discovery. Open Drug Discov. J. 2, 103–106.

Koonawootrittriron, S., Elzo, M.A., Suwanasopee, T., Sopanarat, P., Chaimongkol, C., Sainui, T., Tongprapi, T., Ralukmun, T., 2008. D.P.O. Sire & Dam Summary 2008. Dairy Farming Promotion Organization, Ministry of Agriculture and Co-operatives of Thailand, Bangkok, Thailand.

Koonawootrittriron, S., Elzo, M.A., Thongprapai, T., 2009. Genetic trends in a Holstein×other breeds multibreed dairy population in Central Thailand. Livest. Sci. 122, 186–192.

Laodim, T., Koonawootrittriron, S., Elzo, M.A., Suwanasopee, S., 2015. Genome-wide linkage disequilibrium in a Thai multibreed dairy cattle population. Livest. Sci. 180, 27–33.

Meissner, M., Weissgerber, P., Camacho Londono, J.E., et al., 2011. Moderate calcium channel dysfunction in adult mice with inducible cardiomyocyte-specific excision of the cacnb2 gene. J. Biol. Chem. 286, 15875–15882.

Meredith, B.K., Kearney, F.J., Finlay, E.K., Bradley, D.G., Fahey, A.G., Berry, D.P., Lynn, D.J., 2012. Genome-wide associations for milk production and somatic cell score in Holstein-Friesian cattle in Ireland. BMC Genet. 13, 21. http://dx.doi.org/10.1186/1471-2156-13-21.

Moreno-Sanches, N., Rueda, J., Carabano, M.J., Reverter, A., McWulliam, S., Gonzalez, C., Diaz, C., 2010. Skeletal muscle specific genes networks in cattle. Funct. Integr. Genom. 10, 609–618.

Moyes, K.M., Drackley, J.K., Morin, D.E., Bionaz, M., Rodriguez-Zas, S.L., Everts, R.E., Lewin, H.A., Loor, J.J., 2009. Gene network and pathway analysis of bovine mammary tissue challenged with Streptococcus uberis reveals induction of cell proliferation and inhibition of PPARγ signaling as potential mechanism for the negative relationships between immune response and lipid metabolism. BMC Genom. 10, 542. http://dx.doi.org/10.1186/1471-2164-10-542.

Oviedo-Boyso, J., Valdez-Alarcon, J.J., Cajero-Juarez, M., Ochoa-Zarzosa, A., Lopez-Meza, J.E., Bravo-Patino, A., Baizabal-Aguirre, V.M., 2007. Innate immune response of bovine mammary gland to pathogenic bacteria responsible for mastitis. J. Infect. 54, 399–409.

Parodi, P.W., 1997. Cows' milk fat components as potential anticarcinogenic agents. J. Nutr. 127, 1055–1060.

Patton, J., Kenny, D.A., McNamara, S., Mee, J.F., O'Mara, F.P., Diskin, M.G., Murphy, J.J., 2007. Relationships among milk production, energy balance, plasma analytes, and reproduction in Holstein-Friesian cows. J. Dairy Sci. 90, 649–658.

Perez-Enciso, M., Misztal, I., 2011. Qxpak. 5: old mixed model solutions for new genomics problems. BMC Bioinf. 12, 202. http://dx.doi.org/10.1186/1471-2105-12-202.

Pérez O'Brien, A.M., Meszaros, G., Utsunomiya, Y.T., et al., 2014. Linkage disequilibrium levels in Bos indicus and Bos taurus cattle using medium and high density SNP chip data and different minor allele frequency distributions. Livest. Sci. 166, 121–132.

Pryce, J.E., Bolormaa, S., Chamberlain, A.J., Bowman, P.J., Savin, K., Goddard, M.E., Hays, B.J., 2010. A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. J. Dairy Sci. 93, 3331–3345.

Raven, L.A., Cocks, B.G., Hayes, B.J., 2014. Multibreed genome wide association can improve precision of mapping causative variants. BMC Genom. 15, 62. http://dx.doi.org/10.1186/1471-2164-15-62.

Ritsawai, P., Koonawootrittriron, S., Jattawa, D., Suwanasopee, T., Elzo, M.A., 2014. Fraction of cattle breeds and their influence on milk production of Thai dairy cattle. In: Proceeding of 52th Kasetsart University Annual Conference, pp. 25–32. Bangkok, Thailand.

Seangjun, A., Koonawootrittriron, S., Elzo, M.A., 2009. Characterization of lactation patterns and milk yield in a multibreed dairy cattle population. Kasetsart J. Nat. Sci. 43, 74–82.

Strucken, E.M., Bortfeldt, R.H., Koning, D.J., Brockmann, G.A., 2011. Genome-wide associations for investigating time-dependent genetic effects for milk production traits in dairy cattle. Anim. Genet. 43, 375–382.

The Bovine HapMap Consortium, 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324, 528–532.

Turner, S.D., 2014. qqman: an R Package for Visualizing GWAS Results Using Q-Q and Manhattan Plots. bioRxiv.

Wibowo, T.A., Gaskins, C.T., Newberry, R.C., Thorgaard, G.H., Michal, J.J., Jiang, Z., 2008. Genome Assembly Anchored QTL map of bovine chromosome 14. Int. J. Biol. Sci. 4, 406–414.

Wood, P.D.P., 1967. Algebraic model of the lactation curve in cattle. Nature 216, 164–165.

Yodklaew, P., Koonawootrittriron, S., Elzo, M.A., Suwanasopee, T., 2013. Impact of Sire by Region Interaction on First-lactation Traits of Dairy Cows Raised under Tropical Conditions in Thailand. ADSA-ASAS Joint Annual Meeting, Indianapolis, IN, USA. 8–12 July 2013.

Zhang, Z., Erbe, M., He, J., Ober, U., Gao, N., Zhang, H., Simianer, H., Li, J., 2015. Accuracy of whole-genome prediction using a genetic architecture-enhanced variance-covariance matrix. G3 (Bethesda) 5, 615–627.