Variance Components and Animal Rankings for Milk Yield and Fat Yield
in a Multibreed Dairy Cattle using Genomic-Polygenic, Genomic, and Polygenic Models

Bodin Wongpom\textsuperscript{a}, Skorn Koonawoottritriron\textsuperscript{a,}\textsuperscript{*,} Mauricio A. Elzo\textsuperscript{b} and Thanathip Suwanasopee\textsuperscript{a}

\textsuperscript{a} Department of Animal Science, Kasetsart University, Bangkok 10900, Thailand.
\textsuperscript{b} Department of Animal Sciences, University of Florida, Gainesville, FL 32611-0910, USA.

---

\* Corresponding author: Skorn Koonawoottritriron; E-mail: agrskk@ku.ac.th
Abstract

The aims of this study were to estimate variance ratios and to compare animal rankings of genomic-polygenic (GP), genomic (G), and polygenic (P) models for milk yield and fat yield in a multibreed dairy cattle population in Central Thailand. The dataset contained monthly records of milk yield (MY) and fat yield (FY) from 600 first-lactation cows from 56 farms in Central Thailand. The mixed model contained herd-year-season, Holstein fraction, heterozygosity of the cow and age at first calving as fixed effects (all models). Random effects were single nucleotide polymorphisms (SNP) (GP and G models), animal polygenic (GP and P models) and residual. GP heritability estimates were higher for MY (0.38) and FY (0.41) than for the P model (0.28 for MY and 0.30 for FY). The fraction of the additive genetic variance explained by the SNP markers were 50% for MY and 48% for FY.

Rank correlations between GP and G were the highest for both MY and FY (0.99; P<0.01). Rank correlations between G and P were the lowest for MY (0.89; P<0.01) and FY (0.73; P<0.01). SNP from GeneSeek Genomic Profiler Low-Density (GGP-LD) 9k BeadChip not only accounted for a sizeable fraction of the additive genomic variance for MY and FY, but they also yielded animal genomic estimated breeding values (EBV) whose ranking was highly correlated with rankings of both genomic-polygenic and polygenic EBV. These results indicated that GGP-LD 9k and likely higher density genotyping chips would help improve accuracy of prediction and selection in Central Thailand.

Keywords: Dairy; Breeding; Milk production; Tropics; Genomics

Introduction

Thailand is tropical and humid throughout the country during most of the year. The climate of Thailand is under the influence of seasonal monsoon weather. This country is approximately 513,000 km² of which 46.53% is devoted to agriculture, 31.84% to non-agricultural endeavors, and 21.63% to forests (Office of Agricultural Economics, 2014). The central region of Thailand has the largest concentration of dairy farms in an area comprising 26,900 km². This region contains approximately 61 % (145,578) of dairy cows, 62 % (10,111) of dairy farms, and 64% (1,921,057 kg per day) of the milk produced in the country (Department of Livestock Development, 2015; Office of Agricultural Economics, 2015). Thai government policies in support of dairy farmer efforts to increase milk production resulted in importation Holstein semen and high-percent Holstein sires by the Dairy Farming Promotion Organization (DPO) and the Department of Livestock Development (DLD). The dairy cattle population in Central Thailand is multibreed. Its largest breed component is Holstein (H), accompanied by various fractions of other breeds (Brown Swiss, Jersey, Red Dane, Brahman, Red Sindhi, and Sahiwal). Parents in this population are chosen among purebred and crossbred Holstein animals resulting in a multibreed population composed of individuals with a variety of breed compositions (Koonawootrittriron et al., 2009). Currently, 90% of the population is over 75% H with
small fractions of other breeds. This makes the structure of the Thai dairy population is substantially different from dairy populations in other countries.

The main economically important traits in the Thai dairy business are milk yield (MY) and fat yield (FY) because they are associated directly with the amount of income received by dairy farmers. Milk composition such as FY has been used as an indicator of milk quality and considered for price determination of raw milk. Dairy genetic improvement programs targeting economically important traits have been used to select the best animals as parents of the next generation (Bourdon, 2000). The traditional genetic improvement for MY and FY in dairy cattle requires pedigree and phenotype information from individual animals. Traditional estimated breeding values (EBV) for animals are computed using best linear unbiased prediction (BLUP) procedures that utilize phenotype and pedigree information from all animals (Goddard and Hayes, 2007; Goddard et al., 2010; Vitezica et al., 2011).

Currently, single nucleotide polymorphisms (SNP) across the genome of individual animals can be determined from approximately 7,000 SNPs up to more than 777,000 SNPs (Zhang et al., 2012). These SNPs provide an additional source of information to predict the genetic worth of an animal. The combination of genotyped SNPs, phenotype, and pedigree for animal evaluation has resulted in improved accuracies of prediction (Schaeffer, 2006; Lillehammer et al., 2010; Hayes et al., 2009). This research intends to assess the advantage of utilizing SNP information for genetic prediction with and without pedigree information in multibreed dairy cattle from Central Thailand.

Thus, the objectives of this study were: 1) to estimate variance components, genetic parameters, and fractions of genetic variances accounted by GeneSeek Genomic Profile Low-Density (GGP-LD) BeadChip relative to total genetic variances; and 2) to compare animal rankings using predicted values from genomic-polygenic, genomic, and polygenic models for milk yield and fat yield in a multibreed dairy cattle population in Central Thailand.

Materials and Methods

Animal care

This research was approved by ethics committee of the Kasetsart University Institutional Animal Care and Use Committee and animal care and use under the ethical review board of the Office of National Research Council of Thailand (ID: ACKU60-AGR-0090).

Data, traits and animal management

The dataset consisted of monthly test-day MY and FY from 600 first-lactation multibreed dairy cows that calved between 2000 and 2013 in 56 Central Thailand farms. These cows were the progeny of 198 sires and 547 dams. Cows, sires, and dams in the Thai population were either purebred H or H crossbred with various amounts of genes from other breeds. Breeds represented in the Thai dairy population were Holstein, Brahman, Jersey, Red Dane, Red Sindhi, Sahiwal, and Thai
Native. H fraction of animals in this population ranged from 46.87% to 100%, and 91% of all animals were at least 75% H.

Cows were provided a concentrate diet for milk production at a rate of 1 kg of feed (16% protein) per 2 kg of milk. Roughage consisted mainly of fresh grasses such as Guinea grass (*Penicum maximum*), Ruzi grass (*Brachiaria ruziziensis*), Napier grass (*Pennisetum purpureum*) and Para grass (*Brachiaria mutica*). Fresh grass was limited during the dry season (November to June) due to dwindling supplies of stored and underground water. Farmers also fed rice straw and other agricultural by-products (corn cobs, cassava leaves, corn silage) when available. Cows were kept in open barns and milked twice a day; once in the morning (5 to 6 a.m.) and once in the afternoon (2 to 3 p.m.). Farmers used a bucket or a pipeline system for milking. Raw milk was collected in bulk tanks and transported to a dairy cooperative or to a private milk collection center after milking twice a day (morning and afternoon).

Traits were 305-d first lactation milk yield (MY, kg) and 305-d fat yield (FY, kg). Test-day MY and milk samples were taken from each individual cow once a month from calving to drying off. Milk samples were analyzed for fat percentage and other quality traits. Monthly test-day fat yields were computed as the product of fat percentage times test-day milk yield. Monthly test-day milk yields and fat yields were used to compute MY and FY using the test-interval method (Sargent et al., 1968; Koonawootrittriron et al., 2001). Contemporary groups were defined as herd-year-season of first calving. Calving age was defined as the number of months between birth date and calving date.

**Blood sampling and single nucleotide polymorphisms (SNPs)**

Blood samples were collected from the caudal vein (9 ml) and transported from farm to laboratory at the Faculty of Agriculture of Kasetsart University in Bangkok, Thailand. Genomic DNA was extracted from blood samples using a MasterPure™ DNA Purification Kit (Epicentre®, Madison, WI, USA). The concentration and purity of DNA per sample was measured using a NanoDrop 2000 (Thermo Fisher Science Inc., Wilmington, DE, USA). DNA purity ratios of absorbance at 260 nm and 280 nm ranged from 1.8 to 2.0. DNA concentration ranged from 9 to 645 ng/µl. DNA subsamples were dried using a Freeze dry machine in about 10 to 12 hr. The dried DNA samples were prepared and transported by airmail from Kasetsart University to GeneSeek (GeneSeek; 4665 Innovation Drive, Suite 120, Lincoln, NE 68521, USA; Tel. 1-402/435-0665 402/435-0665 Fax: 1-402/435-0664) for genotyping using a GGP-LD. Only SNPs from the 29 autosomes with call rates ≥ 90% and minor allele frequencies ≥ 0.01 were used in this research (n = 8,257). The number of SNPs per chromosome ranged from 148 in chromosome 28 to 530 in chromosome 1 (Figure 1).

**Genomic-polygenic and polygenic variance components and variance ratios**
Genomic-polygenic and polygenic variance components for MY and FY were estimated by using the Markov Chain Monte Carlo (MCMC) procedure of GS3 with option VCE (Legarra et al., 2010). The genomic-polygenic model included the fixed effects of herd-year-season, Holstein fraction of the cow, heterozygosity of the cow, and age at first calving. Random effects were SNP, animal and residual. The polygenic model included all the effects of the genomic-polygenic model, except for SNP random effects.

GS3 required initial values for additive polygenic variance, residual variance, and additive SNP variances. Initial values for additive polygenic variance and residual variance were estimated using an average information restricted maximum likelihood (AI-REML) procedure with a model containing only additive polygenic effects by ASREML software (Gilmour et al., 2006). Initial values for additive SNP variances were computed as the ASREML estimate of the additive polygenic variance divided by \( \sum_{i=1}^{8257} 2p_i q_i \) (Habier et al., 2007; Gianola et al., 2009), where \( p_i \) = probability of allele A, and \( q_i \) = probability of allele B.

Genomic-polygenic variance component estimates from GS3 was additive SNP variances (VSNP), additive polygenic variances (VAPO), and residual variance (VE). Genomic-polygenic variance components (additive genomic variances; VAGO, additive polygenic variances; VAPO, additive genetic variances; VGTot, phenotypic variances; PheVarGP) were computed using GS3 (option VCE). Additive genomic variances were computed as the product of \( \sum_{i=1}^{8257} 2p_i q_i \) times VSNP. Total genetic variances were computed as VAGO + VAPO. Phenotypic variances were computed as VAGO + VAPO + VE.

Polygenic variance component estimates were additive polygenic variances (VGPO) and residual variances (VE). Polygenic variance components were also computed using option VCE of the MCMC procedure of program GS3. Phenotypic variances were computed as VAPO + VE. The variance ratios and heritabilities for MY and FY were computed using the estimated variances.

Estimates and variability of the genomic-polygenic, genomic, and polygenic variance components and genetic parameters in the population were obtained as the mean and SD of their sample values across all MCMC samples (number of MCMC samples = 1,200).

**Genomic-polygenic, genomic, and polygenic predictions**

Genomic-polygenic, genomic, and polygenic predictions for MY and FY used VAGO, VAPO, and VE values computed with GS3 option BLUP (Legarra et al., 2010). Genomic-polygenic predictions were obtained using a genomic-polygenic model. Genomic predictions were obtained using a genomic model (i.e., a genomic-polygenic model without a animal random effect). Polygenic predictions were obtained using a polygenic model (i.e., a genomic-polygenic model without an SNP random effect).

Genomic-polygenic EBV (GPEBV) were computed as follows:
\[ GPEBV = (\beta_{HF})(HF) + \text{additive genomic value} + \text{additive polygenic value} \] (1)

where \( \beta_{HF} \) is regression coefficient estimates for Holstein fraction, HF is Holstein fraction of cow, additive genomic value = \( \sum_{i=1}^{8257} w_i \widetilde{SNP}_i \), \( w_i \) = number of B alleles in locus i, \( \widetilde{SNP}_i \) = BLUP of SNP, and additive polygenic value = BLUP of the animal random effect from the genomic-polygenic model.

Genomic EBV (GEBV) was computed using as follows:
\[ \text{GEBV} = (\beta_{HF})(HF) + \text{additive genomic value} \] (2)

Polygenic EBV (PEBV) was computed as follows:
\[ \text{PEBV} = (\beta_{HF})(HF) + \text{additive polygenic value} \] (3)

where \( \beta_{HF} \) is regression coefficient estimates for Holstein fraction, HF is Holstein fraction of cow, and additive polygenic value = BLUP of the animal random effect from the polygenic model.

Animal rankings of BLUP predictions from genomic-polygenic, genomic, and polygenic models were analyzed using Spearman’s rank correlations (SAS, 2003).

Results and Discussion

Genomic-polygenic and polygenic variance components for MY and FY are presented in Table 1. Additive genomic variances from the GP model were 215.44 (SD = 127.96) kg\(^2\) for FY and 122,763 (SD = 70,598) kg\(^2\) for MY. Additive polygenic variances from the GP model were 252.86 (SD = 192.14) kg\(^2\) for FY and 129,064 (SD = 87,924) kg\(^2\) for MY. Total genetic variances from the GP model were 468.31 (SD = 203.84) kg\(^2\) for FY and 251,828 (SD = 97,589) kg\(^2\) for MY. Phenotypic variances from the GP model were 1,131.92 (SD = 111.83) kg\(^2\) for FY and 659,615 (SD = 51,580) kg\(^2\) for MY. Additive genetic variances from the P model were 335.17 (SD = 210.10) kg\(^2\) for FY and 183,452 (SD = 100,431) kg\(^2\) for MY. Phenotypic variances from the P model were 1,097.02 (SD = 108.82) kg\(^2\) for FY and 645,573 (SD = 50,627) kg\(^2\) for MY.

Variance ratios and heritability from GP and P are shown in Table 2. The GP heritability estimates were 0.38 for MY and 0.41 for FY. The P heritability estimates were 0.28 for MY and 0.30 for FY. These P heritabilities here were within the range of 0.29 to 0.41 for MY, and 0.29 to 0.39 for FY in Dutch and Nordic Holstein Friesian populations (Stoop et al., 2008; Schopen et al., 2009; Gao et al., 2012). The P heritabilities for MY and FY here were lower than in the UK Holstein Friesian population (Pollott, 2009; Eaglen et al., 2013). The GP heritability estimates were lower than the range of 0.41 to 0.80 for MY, and 0.42 to 0.77 for FY in Dutch and Australian Holstein Friesian populations (Veerkamp et al., 2010; Veerkamp et al., 2011; Haile-Mariam et al., 2013). The GP variances and heritabilities were higher than those of the P model. GP heritabilities for MY and FY
were slightly higher than estimates from P models in Australia (Haile-Mariam et al., 2013), but slightly lower than in the Netherlands (Veerkamp et al., 2010). Differences between variances and heritability estimates in the Thai population versus those from other countries may be related to the number of SNPs in the analyses, linkage disequilibrium between SNPs and genes affecting MY and FY in each population, the size of reference populations, statistical models used in each country, and population structure across countries. The Thai population is multibreed with a larger representation of H than from six other breeds (Brahman, Jersey, Red Dane, Red Sindhi, Sahiwal, and Thai Native). In fact, over 80% of the animals in the Thai population have a Holstein fraction higher than 75% (Jattawa et al., 2015; Koonawootrittriron et al., 2009; Ritsawai et al., 2014).

The ratios of additive genomic variances to total genetic variances from the GP model were high for MY (50%) and FY (48%; Table 2). These percentages indicate that the genomic information of GGP-LD 9k BeadChip captured a substantial percentage of the total genetic variation for MY and FY in this multibreed dairy cattle population in Central Thailand. The estimates of variances and heritabilities in this study indicated that it would be feasible to increase MY and FY by genetic selection under Thai tropical conditions. Further gains could be achieved by improving management practices and feeding regimes (e.g., good quality forage throughout the year) to meet the needs of cows with high genetic production ability. Milk production in Central Thailand has been significantly increased through crossbreeding programs aimed at improving native and crossbred varieties to Holstein. Both purebred and crossbreed Holstein cattle have been selected as parents of future generations. This practice of identifying the best males and females regardless of breed composition should continue to help increase MY and FY in future years.

Spearman’s rank correlations among genomic-polygenic, genomic, and polygenic predictions for MY and FY are shown in Figure 2. Rank correlations between GP and G were 0.9973 for FY and 0.9977 for MY (P<0.0001). Rank correlations between GP and P were 0.7527 for FY 0.9095 for MY (P<0.0001). Rank correlations between G and P were 0.7277 for FY and 0.8892 for MY (P<0.0001). The correlations among GP, G, and P were high for MY and FY. For MY, the rank correlation between GP and G (0.9977; P<0.0001) was stronger than correlations between GP and P (0.9095; P<0.0001) and between G and P (0.8892; P<0.0001). For FY, the correlation between GP and G (0.9973; P<0.0001) was stronger than correlations between GP and P (0.7527; P<0.0001) and between G and P (0.7277; P<0.0001). These high correlations indicated the possibility of preselecting young animals for MY and FY using the G model to reduce the generation interval and decrease evaluation costs. After cow phenotypes are collected, final genetic evaluations can be performed using the GP model.

In conclusion, the fraction of additive genomic variances from GP model were high 50% for MY and 48% for FY. Heritability estimates with the GP model were higher for MY (0.38) and for FY (0.41) than the P model MY (0.28) and for FY (0.30). The rank correlations between GP and G were stronger than rank correlations between GP and P and between G and P. This research indicated that
genomic information from the bovine chip with 8,257 SNPs could be used in addition to phenotypic
and pedigree information to improve prediction accuracies in this multibreed dairy cattle population in
Central Thailand.

**Conflict of Interest**

The authors declare that they have no conflict of interest.

**Acknowledgements**

The authors would like to thank the National Science and Technology Development Agency,
University and Industry Research Collaboration (NUI-RC) for giving a scholarship related to the
project P-11-00116 to the first author, Kasetsart University (Bangkok, Thailand) for partly support
through the project S-K(KS)1.58 (KURDI), the University of Florida (Gainesville, Florida, USA),
Dairy Farming Promotion Organization, dairy cooperatives, and dairy farmers in Thailand.

**References**


Department of Livestock Development. Information of farms and area in Thailand. 2015. Information
and Communications Technology Center. Ministry of Agriculture and Cooperatives,
Mar 2016

genetic relationships between calving ease, gestation length, milk production, fertility, type,

Comparison on genomic predictions using three GBLUP methods and two single-step
blending methods in the Nordic Holstein population. Genet Sel Evol. 44:8.


Manual. NSW Agriculture.


information on genome-assisted breeding values. Genetics. 77:2389-2397.


Van Arendonk, J.A.M. 2009. Genetic parameters for major milk proteins in Dutch Holstein-
2008. Genetic parameters for major milk fatty acids and milk production traits of Dutch
combining pedigree with dense SNP information on some animals. In: Proceeding of the 9th
World Conference on Genetics Applied to Livestock Production. 1-6 August 2010. Leipzig,
Germany.
pedigree-based genetic parameters for scarcely recorded traits when some animals are
Vitezica, Z.G., Aguilar, I., Misztal, I., Legarra, A. 2011. Bias is genomic predictions for
Table 1 Posterior means and posterior standard deviations of variance components for milk yield (MY) and fat yield (FY) from genomic-polygenic and polygenic models

<table>
<thead>
<tr>
<th>Variance components$^1$</th>
<th>MY (kg$^2$)</th>
<th>FY (kg$^2$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAGO</td>
<td>122,763 (70,598)</td>
<td>215.44 (127.96)</td>
</tr>
<tr>
<td>VAPO</td>
<td>129,064 (87,924)</td>
<td>252.86 (192.14)</td>
</tr>
<tr>
<td>VGTot</td>
<td>251,828 (97,589)</td>
<td>468.31 (203.84)</td>
</tr>
<tr>
<td>PheVarGP</td>
<td>659.615 (51.580)</td>
<td>1,131.92 (111.83)</td>
</tr>
<tr>
<td>VGPO</td>
<td>183,452 (100,431)</td>
<td>335.17 (210.10)</td>
</tr>
<tr>
<td>PheVarP</td>
<td>645,573 (50,627)</td>
<td>1,097.02 (108.82)</td>
</tr>
</tbody>
</table>

$^1$ VAGO = additive genomic variances, VAPO = additive polygenic variances, VGTot = total genetic variances, PheVarGP = phenotypic variances from the genomic-polygenic model, VGPO = additive genetic variances from the polygenic model, PheVarP = phenotypic variances from the polygenic model.

Table 2 Posterior means and posterior standard deviations of variance ratios and heritability for milk yield (MY) and fat yield (FY)

<table>
<thead>
<tr>
<th>Variance ratios$^1$</th>
<th>MY</th>
<th>FY</th>
</tr>
</thead>
<tbody>
<tr>
<td>VAGO / VGTot</td>
<td>0.50 (0.23)</td>
<td>0.48 (0.24)</td>
</tr>
<tr>
<td>HeritabilityGP</td>
<td>0.38 (0.13)</td>
<td>0.41 (0.16)</td>
</tr>
<tr>
<td>HeritabilityP</td>
<td>0.28 (0.14)</td>
<td>0.30 (0.18)</td>
</tr>
</tbody>
</table>

$^1$ VAGO = additive genomic variances, VGTot = total genetic variances, HeritabilityGP = heritability from genomic-polygenic model, HeritabilityP = heritability from polygenic model.
Figure 1 Number of SNPs in each bovine chromosome
Figure 2 Spearman’s rank correlation between GP and G Model of MY (A), GP and P Model of MY (B), G and P Model of MY (C), GP and G Model of FY (D), GP and P Model of FY (E), and G and G Model of FY (F). All Spearman’s rank correlations were significant (P<0.0001).