Abstract T87

Genomic-Polygenic Evaluation for Milk Yield and Fat Yield in a Multibreed Dairy Cattle Population in Central Thailand

Bodin Wongpom*, 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

Milk yield (**MY**) and fat yield (**FY**) are economically important traits for Thai dairy business. Genetic prediction for **MY** and **FY** in Thailand uses only pedigree and phenotypic information. Combining SNP genotypes of individual animals with pedigree and phenotypes would be expected to increase accuracy of genetic predictions and speed up selection progress. The objectives of this study were to estimate the fraction of the genetic variance accounted by 8,257 SNP from GeneSeek GGP-LD BeadChip and to compare the rankings of animals evaluated with a genomic-polygenic (**GP**), genomic (**G**), and polygenic (**P**) model for **MY** and **FY**. The dataset consisted of first-lactation **MY** and **FY** records from 600 cows from 56 farms in Central Thailand collected from 2000 to 2013. The mixed model contained herd-year-season, Holstein fraction and age at first calving as fixed effects (all model). Random effects were SNP genomic (**GP** and **G**), animal polygenic (**GP** and P) and residual. Variance components were estimated using GS3 software (option VCE; **GP** and **P**). Additive genetic predictions were computed with GS3 (option BLUP) for all models. The fraction of additive genetic variances explained by the 8,257 SNP from GGP-LD and computed with the GP model were 46% for **MY** and 45% for **FY**. Heritability estimates with the **GP** model were higher (0.37 for **MY** and 0.40 for **FY**) than those with the P model (0.28 for **MY** and 0.30 for **FY**). Rank correlations between **GP** and **G** model were the highest (0.99 for both **MY** and **FY**; P<0.0001), followed by correlations between **GP** and **P** models (0.91 for **MY** and 0.75 for **FY**; P<0.0001), and the lowest correlations were between **G** and **P** models (0.89 for **MY** and 0.73 for **FY**; P<0.0001). ***Thus, SNP from GeneSeek GGP-LD not only accounted for a sizeable fraction of the additive genetic variance for MY and FY, but they also yielded animal genomic EBV whose ranking was highly correlated with rankings of both genomic-polygenic and polygenic EBV. These results indicated that utilization of GGP-LD, and perhaps higher density genotyping chips, would be advantageous for genomic-polygenic evaluation and selection in Central Thailand.***

INTRODUCTION

Milk yield (**MY**) and fat yield (**FY**) are economically important traits in the dairy business. Both traits are used for price determination of raw milk in Thailand. Genetic improvement for **MY** and **FY** in dairy cattle require unselected 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. However, genetic improvement for dairy traits require high accuracies of prediction. Animals can now be genotyped to obtain information on single nucleotide polymorphisms (SNPs) throughout the entire genome. These SNP are expected to be associated with economically important dairy traits. The combination of marker SNPs, phenotypes, and pedigree information can be collectively used to improve the accuracy of genetic predictions. ***Thus, the objectives of this study were to estimate fraction of the genetic variance accounted by GeneSeek GGP-LD 9k BeadChip and to compare the rankings of animals evaluated with a genomic-polygenic, genomic, and polygenic models for milk yield and fat yield***.

MATERIALS AND METHODS

Data, Animals and Traits: The dataset consisted of monthly test-day milk yield (**MY**) and fat yield (**FY**) records from 600 first-lactation cows in 56 farms located in Central Thailand collected from 2000 to 2013. *Breeds of dairy cattle represented in this multibreed dairy population were Holstein, Brahman, Jersey, Red Dane, Red Sindhi, Sahiwal, and Thai Native.* The percentage of Holstein in this population ranged from 46.87 to 100 %. Test-day **MY** and **FY** were measured and collected from each individual cow once a month after calving until dry off. Monthly milk samples of cows were sent to a laboratory for milk quality analysis.

Blood sampling and single nucleotide polymorphisms (SNPs) genotyping: Blood samples were taken from each cow to extract genomic DNA using a MasterPureTM DNA Purification Kit (Epicentre®, Madison, WI, USA). The genomic DNA of each sample was prepared for SNP genotyping with the GeneSeek Genomic Profiler low-density 9k BeadChip (GeneSeek, Lincoln, NE, USA). Only SNPs that were on the 29 autosomes, had a known map position, a call rate ≥ 90%, and a minor allele frequency ≥ 0.01 (2%) were selected. Finally, 8,257 SNPs were used in this study. The number of SNPs per chromosomes ranged from 148 SNPs on the 28th chromosome to 530 SNPs on the 1st chromosome (**Figure 1**).

Genomic-polygenic and polygenic variance components and variance ratios: Genomic-polygenic and polygenic variance components for **MY** and **FY** were estimated using Markov Chain Monte Carlo (MCMC; option VCE) procedure of the GS3 software (Legarra et al., 2010). The fixed effects were herd-year-season, Holstein fraction and age at first calving, and the random effects were SNPs (only for the genomic-polygenic variance component model), animal and residual.

The dataset was first analyzed using a single-trait animal model. Fixed effects were herd-year-season, Holstein fraction and age at first calving. Random effects were animal and residual. Additive polygenic variance and residual variance were estimated using an average information restricted maximum likelihood (AI-REML) procedure of ASREML software (Gilmour et al., 2006). These estimates were then used as initial values for the estimation of genomic and polygenic variance components using GS3 software (Legarra et al., 2010).

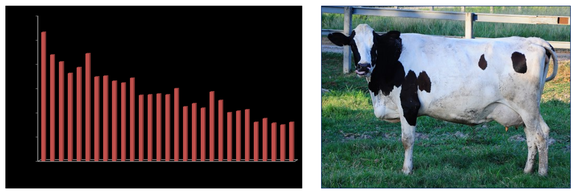

Figure1 Number of SNPs on each bovine chromosomes

Genomic-polygenic variance components: The additive SNP variances (VSNP), additive polygenic variances (VAPO), and residual variance (VE) estimates were used to compute genomic-polygenic variance components: additive genomic variances (VAGO), additive polygenic variances (VAPO), additive genetic variances (VGTot), and phenotypic variances (PheVarGP) using GS3 software (option VCE). The additive genomic variances were computed as the product of $\sum_{i=1}^{N} 2p_iq_i * VSNP_i$, the total genetic variances were computed as VAGO + VAPO, and the phenotypic variances were computed as VAGO + VAPO + VE.

Polygenic variance components: Polygenic variances were additive genetic (VAPO), residual (VE), and phenotypic (PheVarP = VAPO + VE).

Genomic-polygenic, genomic, and polygenic predictions: Genomic-polygenic, genomic, and polygenic predictions for **MY** and **FY** were computed with GS3 software (Legarra et al., 2010; option BLUP; Gauss-Seidel iteration; convergence criterion = 10^{-4}) using **VAGO**, **VAPO**, and **VE** estimates. Genomic-polygenic predictions were computed using a model with both genomic and polygenic effects, whereas genomic predictions were computed using a model with genomic effects only, and polygenic predictions were obtained using a model with polygenic effects only. Thus, ***1) Genomic-polygenic EBV (GPEBV) were computed as the sum of*** (β_{HF})***(HF) + additive genomic value + additive polygenic value,*** where β_{HF} = regression coefficient estimate for Holstein fraction and HF = Holstein fraction of cow; ***2) Genomic EBV (GEBV) were computed as the sum of*** (β_{HF})***(HF) + additive genomic value,*** where additive genomic value was from the genomic model; and ***3) Polygenic EBV (PEBV) were computed as the sum of*** (β_{HF})***(HF) + additive polygenic value,*** where additive polygenic values came from the polygenic model. Rankings of animals were analyzed using Spearman's rank correlations.

RESULTS AND DISCUSSION

Genomic-polygenic and polygenic variance components for **MY** and **FY** are presented in **Table 1**. Variance ratios and heritability from genomic-polygenic and polygenic models are presented in **Table 2**. The heritability estimates from genomic-polygenic models were 0.37 for **MY** and 0.40 for **FY**. The heritability estimates from polygenic models were 0.28 for **MY** and 0.30 for **FY**. *Genetic variance and heritability estimates from genomic-polygenic models were higher than from polygenic models.* Higher heritability estimates for **MY** and **FY** were obtained in Holstein-Friesian populations in temperate environments (Veerkamp et al., 2010; Haile-Mariam et al., 2013). Holstein-Friesian heritabilities for **MY** and **FY** from genomic-polygenic models were slightly higher than estimates from polygenic models in Australia (Haile-Mariam et al., 2013), but slightly lower in the Netherlands (Veerkamp et al., 2010). Differences in statistical models, number of SNPs, linkage disequilibria, and population structure may have contributed to these differences. *The variance ratios between additive genomic variances and total genetic variances from genomic-polygenic model were 46 % for **MY** and 45 % for **FY**.* ***These percentages indicate that the genomic information of 8,257 SNPs captured a large fraction of the total genetic variation for MY and FY in this multibreed dairy cattle population in Central Thailand.***

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

Variance components[1]	Trait	
	MY (kg^2)	FY (kg^2)
VAGO	111,866 (62,722)	196.96 (126.39)
VAPO	139,287 (82,947)	265.42 (186.74)
VGTot	251,154 (90,728)	462.39 (198.04)
PheVarGP	659,001 (49,369)	1,123.91 (105.98)
VGPO	182,975 (98,441)	337.15 (224.43)
PheVarP	644,212 (51,362)	1,091.98 (106.12)

[1] VAGO = additive genomic variances, VAPO = additive polygenic variances, VGTot = total genetic variances, PheVarGP = phenotypic variances from genomic-polygenic models, VGPO = additive genetic variances from polygenic models, PheVarP = phenotypic variances from polygenic models.

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

Variance ratios[1]	Trait	
	MY	FY
VAGO / VGTot	0.46 (0.22)	0.45 (0.24)
HeritabilityGP	**0.37 (0.12)**	**0.40 (0.16)**
HeritabilityP	0.28 (0.14)	0.30 (0.18)

[1] VAGO = additive genomic variances, VGTot = total genetic variances, HeritabilityGP = heritability from genomic-polygenic model, HeritabilityP = heritability from polygenic model.

Spearman's rank correlations between genomic-polygenic, genomic, and polygenic predictions for **MY** and **FY** are presented in **Table 3**. Rank correlations between **GP** and **G** models ranged from 0.9973 to 0.9977 (P < 0.0001). Rank correlations between **GP** and **P** models ranged from 0.7527 to 0.9095 (P < 0.0001) while rank correlations between **G** and **P** models ranged from 0.7277 to 0.8892 (P < 0.0001). Rank correlations between **GP**, **G**, and **P** models were high for all traits. *For **MY**, the rank correlation between **GP** and **G** models (0.9977; P < 0.0001) was stronger* than rank correlations between the **GP** (0.9095; P < 0.0001) and **G** (0.8892; P < 0.0001) models with the **P** model. *For **FY**, the correlation between **GP** and **G** models (0.9973; P < 0.0001) was stronger* than correlations between **GP** (0.7527; P < 0.0001) and **G** (0.7277; P < 0.0001) with **P**. ***These high correlations indicated the possibility of preselecting young animals for MY and FY using a G model and later on, after phenotypes are collected, do a final genetic evaluation using a GP model.***

Correlation between predicted animal EBV and phenotypes of genomic-polygenic, genomic, and polygenic predictions for **MY** and **FY** are presented in **Table 4**. Correlations were ranged from 0.5201 to 0.6068 (P < 0.0001) for **GP** model, 0.5183 to 0.6060 (P < 0.0001) for **G** model, and 0.5046 to 0.5871 (P < 0.0001) for **P** model. *Correlations between predicted animal EBV and phenotypes from the **GP** model (0.5201 for **MY** and 0.6068 for **FY**; P < 0.0001) were higher than correlations from the **G** (0.5183 for **MY** and 0.6060 for FY; P < 0.0001) and **P** models (0.5046 for **MY** and 0.5871 for **FY**; P < 0.0001).* ***The results suggested that genomic information from the bovine chip with 8,257 SNPs could be used to combine genomic, phenotypic and pedigree information to improve prediction accuracies in this multibreed dairy cattle population in Central Thailand.***

Table 3 Spearman's rank correlation among genomic-polygenic, genomic, and polygenic predictions for milk yield (MY) and fat yield (FY)

Model[1]	Trait	
	MY	FY
GP Model, G Model	**0.9977**	**0.9973**
GP Model, P Model	0.9095	0.7527
G Model, P Model	0.8892	0.7277

[1] All Spearman's rank correlations were significant (P < 0.0001).

Table 4 Correlation between predicted animal EBV and phenotypes of genomic-polygenic, genomic, and polygenic predictions for milk yield (MY) and fat yield (FY)

Model[1]	Trait	
	MY	FY
GP Model	**0.5201**	**0.6068**
G Model	0.5183	0.6060
P Model	0.5046	0.5871

[1] All correlations were significant (P < 0.0001).

FINAL REMARKS

- ***Fractions of genomic to total genetic variances were high for MY and FY***
- ***Heritability estimates with the GP model were higher than the P model for MY and FY***
- ***Rank correlations between predicted values from GP and G models for MY and FY were higher than between GP and P models***

" Combining genomic SNPs, phenotypes, and pedigree information would help to improve accuracy of prediction for MY and FY in Central Thailand "

LITERATURE CITED

Gilmour, A.R., B.R. Cullis, S.J. Welham and R. Thompson. 2006. ASREML Reference Manual. NSW Agriculture.

Haile-Mariam, M., G.J. Nieuwhof, K.T. Beard, K.V. Konstatinov and B.J. Hayes. 2013. J. Anim. Breed. Genet. 130(1): 20-31.

Legarra, A., A. Ricard and O. Filangi. 2010. GS3 Genomic Selection Gibbs Sampling Gauss Seidel and BayesC Manual. INRA, Institut National de Recherche Agronomique-Laboratoires, Auzeville, France.

Veerkamp, R.F., H.A. Mulder and M.P.L. Calus. 2010. Proc 9th WCGALP, Leipzig, Germany.