Proc. Genetic Improvement of Livestock and Aquatic Animals in the Tropics: Challenge and Reward (GILAAT), Bangkok, Thailand, September 24-26, 2012.

Genomic evaluation in cattle: Experiences from UF and KU

M. A. Elzo[1], S. Koonawootrittriron[2], and T. Suwanasopee[2]
[1]University of Florida, Gainesville, USA; [2]Kasetsart University, Bangkok, Thailand

Abstract

Availability of low and high density marker chips has increased the feasibility of predicting the genetic worth of beef and dairy cattle using both phenotypes and genotypes. The objectives of this presentation are to discuss a beef cattle genomic study at the University of Florida (UF), and to present an overview of the current status of the dairy cattle genetic-genomic project being conducted in Thailand by researchers from Kasetsart University (KU) and UF with financial support from the National Science and Technology Agency, KU, and the Dairy Farming Promotion Organization of Thailand. Results from the UF genomics project showed that the fraction of additive genetic variation explained by SNP from the Illumina3K chip for four postweaning feed efficiency and weight gain traits was small (0.11 to 0.36) suggesting that a genomic-polygenic model would be the best suited to evaluate animals in the UF Angus-Brahman multibreed population. On the other hand, the primary goal of the KU dairy genomics project is the development of a dairy genetic-genomic evaluation system in Thailand. Major objectives include a DNA repository, a database system, genomic-polygenic prediction models for Thailand, associations between SNP markers and economically relevant traits, and training of graduate students. The cattle population includes 3,500 animals (250 sires and 3,250 cows) from the Central, Northeastern, and Southern regions. Sires and dams highly represented in the population will be genotyped with the Illumina BovineSNP50 chip, and remaining cows with the IlluminaLD chip, and subsequently imputed to 50K. Genomic-polygenic, genomic, and polygenic models will be used with the complete set and various subsets of markers of the BovineSNP50. Implementation of the genomic-polygenic evaluation system will utilize a single-step procedure that combines phenotypic, pedigree, and genotypic information. This evaluation system is expected to both increase prediction accuracy and rate of genetic progress under Thai environmental conditions.

Key Words: beef, cattle, dairy, evaluation, genomic, polygenic

Introduction

The availability of reasonably priced high-density marker chips for cattle has increased the feasibility of utilizing genomic and phenotypic information to obtain genetic predictions of higher accuracy than predictions obtained based on phenotypes alone. Use of genotyping chips has accelerated in the last few years with the production of low density chips (e.g., Illumina3K, IlluminaLD; Illumina, 2011a, b) that has allowed the genotyping of large number of animals at a substantially lower cost than with high density chips. Low density chips can subsequently be used in conjunction with higher density chips (e.g., IlluminaSNP50; Illumina, 2011c) to impute missing genotypes. Research was conducted at the University of Florida to determine the

fraction of additive genetic variation explained by the Illumina3K chip and to compare the ranking of animals using genomic-polygenic, genomic, and polygenic models in an Angus-Brahman multibreed population of beef cattle at the University of Florida (UF), Gainesville, USA (Elzo et al., 2012). On the other hand, a research project on the development of a dairy genetic-genomic evaluation system in Thailand was established in 2012 by researchers from Kasetsart University (KU) and UF with financial support from the National Science and Technology Development Agency of Thailand (NSTDA), KU, and the Dairy Farming Promotion Organization (DPO). Thus, the objectives of this presentation are to discuss the main results of the multibreed genomic study and to present an overview of the current status of the dairy genetic-genomic project in Thailand.

The UF Beef Genomics Project

Major objectives. 1) Estimate the fraction of the additive genetic variation explained by the Illumina Bovine3K chip for feed efficiency, growth, ultrasound, carcass, and meat palatability traits; 2) Compare the ranking of animals evaluated using genomic-polygenic, genomic and polygenic models; and 3) Assess trends in genomic-polygenic, genomic, and polygenic predictions as Brahman fraction of calves increased in a multibreed population of beef cattle ranging from 100% Angus to 100% Brahman. Only results for feed efficiency and postweaning traits will be presented here.

Cattle population and mating system. The UF Angus-Brahman multibreed population is composed of animals that range in breed composition from 100% Angus (A) to 100% Brahman (B). Mating is diallel involving sires and dams from six breed groups: Angus = (1.0 to 0.80 A; 0.0 to 0.20 B), ¾ A ¼ B = (0.79 to 0.60 A; 0.21 to 0.40 B), Brangus = (0.625 A; 0.375 B), ½ A ½ B = (0.59 to 0.40 A; 0.41 to 0.60 B), ¼ A ¾ B = (0.39 to 0.20 A; 0.61 to 0.80 B) and Brahman: (0.19 to 0.0 A; 0.81 to 1.00 B).

Data and traits. Only results for feed efficiency and postweaning traits are presented here. Daily feed intake and weights were collected on calves born between 2006 and 2010 at the UF Feed Efficiency Facility in Marianna, Florida. Traits were phenotypic daily residual feed intake (RFI, kg DM*day^{-1}), mean daily feed intake (DFI, kg DM*day^{-1}), mean daily feed conversion ratio (FCR, kg DM*day^{-1}/kg weight gain*day^{-1}) and postweaning gain during the 70-d feeding trial (PWG, kg).

Tissue sampling and genotyping. Semen (4 straws) or blood (10 mL) were obtained for sires. Blood samples (10 mL) were collected for dams and calves. Calf samples were collected at weaning. Genotypes were obtained at GeneSeek (GeneSeek, Lincoln, NE, USA).

Genomic-polygenic variance components. A genomic-polygenic mixed model (VanRaden, 2008; Legarra et al., 2008) was used to estimate variance components. The model had year-pen,

age of dam, sex of calf, age of calf, B fraction of calf and heterozygosity of calf as fixed effects, and animal polygenic (AP; mean zero; variance = A * additive polygenic variance; A = additive relationship matrix), additive SNP genomic effects as a function of the number of "2" alleles (AS; mean zero; variance = I * additive SNP variance) and residual effects (mean zero, I * residual variance) as random effects. Option VCE (Markov Chain Monte Carlo) of program GS3 (Legarra, 2009) was used to perform computations. Priors for the additive SNP variance were computed using the expression $\frac{\widehat{V_g}}{\sum_{i=1}^{2899} 2p_i q_i}$ (Habier et al., 2007; VanRaden, 2008; Gianola et al., 2009), where $\widehat{V_g}$ = estimate of additive polygenic variance from the polygenic model computed using ASREML, p_i = frequency of allele "1" and q_i = frequency of allele "2" in the i^{th} SNP in the Illumina3K chip.

Genomic-polygenic, genomic and polygenic predictions. Predictions were obtained using option BLUP (Gauss-Seidel iteration) of program GS3 (Legarra, 2009). Genomic-polygenic predictions were computed using the same model used for the computation of variance components, whereas the model for genomic predictions ignored polygenic effects, and the polygenic model ignored genomic effects. Genomic-polygenic predictions were equal to calf Brahman fraction * solution (Brahman – Angus) + calf additive genomic value + calf additive polygenic value. Calf additive genomic value was equal to $\sum_{i=1}^{2899} w_i \widehat{SNP_i}$, w_i = number of "2" alleles in the i^{th} SNP, where $\widehat{SNP_i}$ = BLUP of SNP_i.

Genomic and polygenic variance components and variance ratios. Table 1 presents estimates of variance ratios for RFI, DFI, FCR, and PWG. Ratios of additive genomic to total additive genetic variance were low and ranged from 0.11 ± 0.09 for DFI to 0.25 ± 0.17 for FCR. Heritability ratios ranged from 0.20 ± 0.07 for RFI to 0.36 ± 0.10 for PWG.

Ranking of animals evaluated with genomic-polygenic, genomic and polygenic models. Spearman rank correlations among predictions from the genomic-polygenic, genomic, and polygenic models are shown in Table 2. Higher correlations existed between genomic-polygenic and polygenic predictions, followed by correlations between genomic-polygenic and genomic predictions, and lastly by correlations between genomic and polygenic predictions. Negative regressions of genomic-polygenic, genomic, and polygenic predictions on Brahman fraction of calf indicated that calves tended be more efficient but grew more slowly as Brahman fraction increased.

Predicted SNP values. Small values were predicted for all SNP for all traits supporting the usual assumption that quantitative traits are determined by many genes of small effect. Table 3 contains the number and fraction of standardized predicted SNP values obtained using the genomic-polygenic model.

Conclusions. Genomic to total genetic variance ratios were low (and mostly lower than those obtained with the Illumina BovineSNP50 elsewhere). Thus, the Illumina3K chip should be used

together with higher density chips such as the Illumina BovineSNP50 to obtain predictions based on phenotypes and actual and imputed genotypes. If only genotypes from the Illumina3K chip were available, then a genomic-polygenic model should be used.

The KU Dairy Genomics Project

Major objectives. 1) Construct a DNA repository from tissue samples (semen, blood) from all animals in the Thai multibreed dairy cattle population; 2) Develop a reference dairy population for genomic evaluation; 3) Develop genomic prediction models and procedures appropriate for Thailand; 4) Determine association between SNP markers and economically relevant traits; 5) Implement a genetic evaluation system that combines phenotypes, genotypes, and pedigree; 6) Construct a database system to store phenotypes, genotypes, pedigree, and economic information; 7) Train graduate students in genetics and genomics.

Cattle population. The dairy cattle population considered over the 3 year period of this project will include 250 sires and 3,250 cows for a total of 3,500 animals. Cows will come from 240 farms (Figure 1) located in the Central region (70%; 168 farms), Northeastern region (14%; 34 farms), and Southern region (16%; 38 farms). Table 4 shows the number of sires and cows present in 2012.

Data and traits. Data for this project will come primarily from animals belonging to the DPO population (Table 4). In addition, farms with available individual cow phenotypes from the Muaklek Dairy Cooperative Limited (MDLC) will be considered in years 2 and 3. Traits to be analyzed include milk yield, fat percentage, protein percentage, total solids percentage, solids not fat percentage, initial milk yield, peak milk yield, days to peak, persistency, lactation length, somatic cell count, calving age, and physical traits (e.g., hip height, udder height, and teat length).

Tissue sampling and genotyping. Tissue samples collected from sires include semen (4 straws) and blood (10 mL) and 10 mL of blood are collected from cows. Tissue samples are kept in a repository at KU. It is projected that a total of 3,500 sires, dams, and cows will be genotyped over a period of 3 years. Genotyping will be done with IlluminaLD, Illumina BovineSNP50, and Illumina HD chips (Illumina, 2011b, c, d). For budgetary reasons, sires and dams highly represented in the pedigree will be genotyped with Illumina BovineSNP50 and the remaining cows will be genotyped with the IlluminaLD chip. In addition, some highly represented sires and dams may be genotyped with the IlluminaHD chip. Genotyping will be done at GeneSeek or other suitable laboratory.

Genomic-polygenic variance components and variance ratios. A genomic-polygenic model will be utilized to compute variance components and variance ratios (VanRaden, 2008; Legarra, 2009). Additive genomic to total genetic will be computed to determine the fraction of the

additive genetic variance explained by the markers in the IlluminaLD and Illumina BovineSNP50 chips. Animals genotyped with the IlluminaLD chip will be imputed to 50K using program Findhap (VanRaden, 2011).

Genomic-polygenic, genomic and polygenic predictions. Predictions will be obtained using genomic-polygenic, genomic, and genomic models applied to animals with actual and imputed markers from the BovineSNP50 chip. Training and validation datasets will be defined to evaluate the predictive ability of the 3 models. One alternative would be to consider a training dataset that included genotypic data from animals with their first lactation until 2013 and a validation dataset with cows that had their first lactation in 2014. Predictive abilities of all models for all traits will be evaluated using correlations between predicted values and phenotypes (Legarra et al., 2008). In addition, the predictive ability of subsets of markers relevant under Thai production conditions will be compared to the predictive ability of the complete set of markers of the BovineSNP50 chip. Association between animal rankings across models will be analyzed using Spearman rank correlations. The implementation of the genomic-polygenic evaluation system will utilize a single-step procedure that combines phenotypic, pedigree, and genotypic information (Aguilar et al., 2010).

Association between marker SNP and economically relevant traits. Standardized predicted SNP values will be plotted by their location and by chromosome number. Predictions of SNP values within and across chromosomes will be evaluated for their proximity to markers with known association to traits of economic importance. A systems biology approach will subsequently be tried to explore multiple-trait SNP associations (Fortes et al., 2010).

Final Remarks. Available phenotypic data in the contributing populations and genotypic information from animals genotyped with the Illumina50K and IlluminaLD chips (imputed to 50K) are expected to yield combined predictions of higher accuracy than current polygenic predictions. The implementation of a system that utilizes all sources of information (phenotypes, genotypes, pedigree) will increase genetic progress in the Thai dairy population by improving the ability of dairy farmers to identify the best sires and dams for their production systems under Thai environmental conditions.

Literature Cited

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. J. Dairy Sci. 93:743–752.

Elzo, M. A., G. C. Lamb, D. D. Johnson, M. G. Thomas, I. Misztal, D. O. Rae, C. A. Martinez, J. G. Wasdin and J. D. Driver. Genomic-polygenic evaluation of Angus-Brahman multibreed

cattle for feed efficiency and postweaning growth using the Illumina3k chip. J. Anim. Sci. 90:2488-2497.

Fortes, M. R. S., A. Reverter, Y. Zhang, E. Collis, S. H. Nagaraj, N. N. Jonsson, K. C. Prayaga, W. Barris, and R. J. Hawken. 2010. Association weight matrix for the genetic dissection of puberty in beef cattle. Proc. Natl. Acad. Sci. USA 107:13642–13647.

Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. Genetics 183:347-363.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177: 2389-2397.

Illumina, Inc. 2011a. GoldenGate Bovine3K Genotyping BeadChip. Illumina Data Sheet, San Diego, CA. http://www.illumina.com/Documents//products/datasheets/datasheet_bovine3k.pdf.

Illumina, Inc. 2011b. BovineLD Genotyping BeadChip. Illumina Data Sheet, San Diego, CA. http://www.illumina.com/documents/products/datasheets/datasheet_bovineLD.pdf.

Illumina, Inc. 2011c. BovineSNP50 Genotyping BeadChip. Illumina Data Sheet, San Diego, CA. http://www.illumina.com/Documents/products/datasheets/datasheet_bovine_snp5O.pdf.

Illumina, Inc. 2011d. BovineHD Genotyping BeadChip. Illumina Data Sheet, San Diego, CA. http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf

Legarra, A. 2009. GS3. INRA, Auzeville, France. http://snp.toulouse.inra.fr/~alegarra/manualgs3_2.pdf.

Legarra, A., C. Robert-Granié, E. Manfredi, and J-M. Elsen. 2008. Performance of genomic selection in mice. Genetics 180:611-618.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91:4414–4423.

VanRaden, P. M. 2011. Findhap, Version 2. USDA-ARS. http://aipl.arsusda.gov/software/findhap/.

Table 1. Posterior means and posterior standard deviations for additive genetic and genomic variance ratios

	Trait[1]			
Variance Ratios[2]	RFI	DFI	FCR	PWG
VAGO/VGTot	0.15 ± 0.12	0.11 ± 0.09	0.25 ± 0.17	0.15 ± 0.11
Heritability	0.20 ± 0.07	0.31 ± 0.09	0.21 ± 0.08	0.36 ± 0.10
HeritabilityPO	0.17 ± 0.08	0.28 ± 0.09	0.15 ± 0.07	0.32 ± 0.09

[1]RFI = residual feed intake; DFI = mean daily feed intake; FCR = mean daily feed conversion ratio; PWG = postweaning gain.

[2]VAGO = additive genomic variance; VGTot = VAGO + VAPO; HeritabilityPO = heritability from a polygenic model.

Table 2. Spearman rank correlations for animals evaluated using genomic-polygenic, genomic, and polygenic models

	Trait[1]			
Correlation[2]	RFI	DFI	FCR	PWG
GP Model, G Model	0.65	0.62	0.66	0.74
GP Model, P Model	0.98	0.99	0.95	0.99
G Model, P Model	0.52	0.51	0.42	0.65

[1]RFI = residual feed intake; DFI = mean daily feed intake; FCR = mean daily feed conversion ratio; PWG = postweaning gain.

[2]GP Model = genomic-polygenic model; G Model = genomic model; P Model = polygenic model. All correlations were significant ($P < 0.0001$).

Table 3. Number and percentage of standardized predicted SNP values from the genomic-polygenic model

	Trait[1]							
	RFI		DFI		FCR		PWG	
SDSNP Range[2]	N	%	N	%	N	%	N	%
-0.4 to -0.5	0	0	0	0	0	0	1	0.03
-0.3 to -0.4	0	0	1	0.03	4	0.14	1	0.03
-0.2 to -0.3	8	0.28	19	0.66	60	2.07	66	2.28
-0.1 to -0.2	187	6.45	244	8.42	393	13.55	371	12.80
0 to -0.1	1204	41.53	1171	40.39	1007	34.74	998	34.43
0 to 0.1	1289	44.46	1169	40.32	1004	34.63	1010	34.84
0.1 to 0.2	202	6.97	277	9.56	379	13.07	376	12.97
0.2 to 0.3	9	0.31	18	0.62	48	1.66	72	2.48
0.3 to 0.4	0	0	0	0	4	0.14	4	0.14

[1]RFI = residual feed intake; DFI = mean daily feed intake; FCR = mean daily feed conversion ratio; PWG = postweaning gain.

[2]SDSNP = additive SNP standard deviation.

Table 4. Number of sires and dams from the DPO population in 2012

Breed group	Sires	Dams
(0.8 – 1.0) Holstein	110	1,815
(0.6 – 0.8) Holstein	22	530
(0.4 – 0.6) Holstein	10	126
(0.2 – 0.4) Holstein	0	14
(0.0 – 0.2) Holstein	8	15
Total	150	2,500

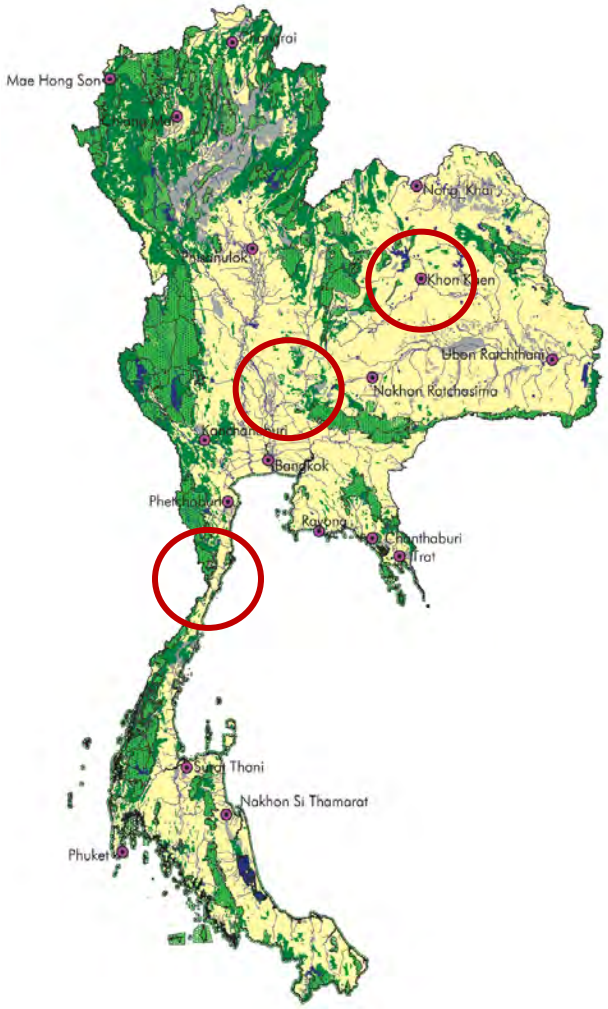

Figure 1. Location of farms participating in the KU genomics project in the Central, Northeastern, and Southern regions