

APPLICATION OF GENOMIC SELECTION TO VIETNAMESE HOUSEHOLD DAIRY HERDS

N.N. Bang^{1,2}, B.J. Hayes¹, I.A.S. Randhawa¹, R.E. Lyons¹, J.B. Gaughan¹, N.V. Chanh³, N.X. Trach², N.D. Khang³ and D.M. McNeill¹

¹The University of Queensland, Gatton QLD 4343 and St Lucia QLD, 4067 Australia

²Vietnam National University of Agriculture, Hanoi, 131000 Vietnam

³Nong Lam University, Ho Chi Minh, 700000 Vietnam

SUMMARY

Household dairy farms (HDFs) account for most of the demand for animal breeding support in Vietnam, as they comprise 97% of the national herd. However, most do not have individual cow pedigrees or production data. Consequently, neither pedigree-based nor genomic selection (GS) methods have been used in Vietnam. The aim of this project was to establish a milk production database and assess the accuracy of GS for production traits using only a single test-day measurement (average of pm + am milking). Phenotypic data included milk yield (MILK, kg/d), milk dry matter (DM%), fat (FAT%), and protein (PRO%) contents of 345 lactating cows from 4 dairy regions, with 8 HDFs per region. The cows were genotyped using the Bovine 50K chip. GBLUP was used to estimate genomic heritability (h^2) and evaluate the accuracy of GS per trait. Moderate heritabilities and accuracies of GS were detected for FAT% ($h^2 = 0.45$, accuracy = 0.28), PRO% ($h^2 = 0.21$, accuracy = 0.23), and DM% ($h^2 = 0.18$, accuracy = 0.48). However, the heritability for MILK was very low (0.01) and the standard errors for all heritabilities and GS accuracies were high. These data suggest the potential for a single test-day to assess Vietnamese dairy cows for milk solid content, but not milk yield, using GS.

INTRODUCTION

The dairy industry in Vietnam is characterized by approximately 500,000 household dairy farms (HDFs) (Nguyen *et al.* 2016). The HDFs account for 97% of the national dairy herd (Trach 2017), and supply >80% of fresh milk production (Vinamilk 2017). Genotypes commonly used are European breeds (predominantly Holstein Friesian but also Jersey) crossed with tropically adapted breeds (Red Sindhi and Sahiwal) and local breeds (Yellow and Lai Sind) (Hayley 2010; Lam *et al.* 2010). Improving the genetic potential of dairy cows for milk production in HDFs is necessary to improve the national supply of fresh milk. However, a national breeding program for household dairy herds is yet to be officially implemented, even for basic traits such as milk yield, fat, or protein. A lack of individual cow pedigree and phenotypic data is the main reason for this.

Genomic selection (Meuwissen *et al.* 2001) is a recently proven method which is now widely used globally in dairy cow selection. This could be a suitable tool for dairy selection in Vietnam as it enables the selection of animals based on genomic markers, most commonly single nucleotide polymorphisms, or SNPs, without the need for pedigrees. However, whilst genomic selection does not require a pedigree, it does require a large number of animals with phenotypic data to allow the development of accurate prediction equations. These data are not easily obtained in Vietnamese HDFs. Such data is expensive to collect in terms of money, time, and labour, as it requires manually separating, weighing, and sampling milk from each cow at each milking time. Consequently, we aimed to estimate the genomic heritability for key milk production traits and to assess the accuracy of genomic selection on these traits using only a single test-day measurement for each.

MATERIALS AND METHODS

Phenotype data. From August to October 2017, data from 345 lactating cows located on 32 HDFs, 8 from each of four main dairy regions in Vietnam were recorded: Lam Dong – a south high-altitude province; Ho Chi Minh – a south low-altitude city; Son La – a north high-altitude province, and Ha Nam – a north low-altitude province. Each farm was visited twice to correspond with a milking in an afternoon and following morning. At these visits, individual cow age, number of lactations and days in milk were obtained by asking the farmer and/or checking their record books where possible, and tail hair was sampled from each lactating cow. The mean \pm SD obtained for age (years), number of lactations, and days in milk of these cows were 4.5 years \pm 1.7, 2.3 lactations \pm 1.4, and 191.4 days \pm 120, respectively. A single day milk yield (MILK, kg/d) for each cow was obtained by weighing and summing the afternoon and the following morning milk yields. Milk samples for each cow were also collected at each milking, analysed at the Food Chemistry Lab (Vietnam National University of Agriculture) and averaged for milk dry matter (DM%, which is the percentage of all milk constituents excluding water), milk fat (FAT%), and milk protein (PRO%) contents. These data were used to calculate the yield of milk dry matter (DM, kg/d), fat (FAT, kg/d), protein (PRO, kg/d), and energy-corrected milk (ECM, kg/d) using the equation of Tyrrell and Reid (1965).

Genotype data. Hair samples were genotyped by Neogen Australasia, The University of Queensland, Gatton. DNA was extracted from the samples using Sbeadex Livestock Kits (LGC Limited, 2017), and then genotyped using the GGP Bovine 50K chip, which assays 48,268 SNPs (Neogen GeneSeek Operations, 2018).

Quality control. R Software (R Core Team, 2016) was used for all data processing. The quality control on the genotype data removed 3,313 SNPs, which were either mitochondrial SNPs or unmapped SNPs, 5478 SNPs with call rates lower than 95%, 1980 SNPs with minor allele frequency lower than 95%, and 633 SNPs with a heterozygosity deviating ± 3 SD from the SNPs' heterozygosity mean. One sample with a call-rate less than 95% was removed, in addition to three cows with heterozygosity deviating ± 3 SD from the samples' heterozygosity mean, 14 cows from four farms with less than five lactating cows. Three cows had missing phenotypic data. The final data set for analysis comprised 323 cows from 28 farms, genotyped for 36864 SNPs.

Genomic heritability and genomic breeding values. Univariate animal linear mixed models with common environmental effects (Mrode and Thompson 2013) were fitted using the GBLUP method in the R “Sommer package” and (Covarrubias-Pazarán 2019). The matrix notation describing the model was: $y = X\beta + Z\alpha + Wc + e$, where: y was the vector of the traits observed, β was the vector of fixed effects (age, lactations, days in milk, days in milk squared), α was the vector of random additive genetic animal effects [$\alpha \sim N(0, A\sigma_a^2)$], wherein A was the genomic relationship matrix derived from the SNPs, c was the vector of the random environmental farm effect (28 farms) [$c \sim N(0, I\sigma_c^2)$], e was the vector of residual random effects [$e \sim N(0, I\sigma_e^2)$], and X , Z , and W were the incidence matrices of the fixed effects, random additive genetic effects, and random environmental effects, respectively. Animal, random environmental and residual effects were assumed to be independently distributed. Heritability (h^2) was estimated as the ratio of the additive genetic variance to total phenotypic variance [$h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_e^2)$] (Falconer and Mackay 1996).

Accuracy of genomic selection. Due to the relatively small data-set, a 10-fold cross-validation approach was applied (Kang *et al.* 2017). Briefly, the entire data set, of 323 samples, were randomly partitioned into 10 subsets of equal size. Nine were used as a training set to determine genomic estimated breeding values (GEBV) for the retained validation set (10%). This process was repeated 10 times so that each subset was used only once as the validation set. The accuracy of genomic selection for each trait was determined by: Accuracy = $r(GEBV, y)/h$, where r was the correlation between GEBV and the original phenotype (y) of each validation set and h was the square root of genomic heritability of the trait.

RESULTS AND DISCUSSION

MILK of the average household dairy cow in Vietnam was 17.87 kg/d (Table 1), which was much higher than other published survey estimates in Vietnam (14.0 – 16.0 kg/d) (Lam *et al.* 2010; Vu *et al.* 2016) but much lower than surveys in other developed countries such as Australia (22.9 kg/d) (DataGene 2017) or Asian countries such as Korea (27.8 kg/d) (Cho *et al.* 2013).

Table 1. Descriptive statistics for milk production traits in Vietnamese dairy cows

Trait	n	Mean	SD	Median	Minimum	Maximum	IQR
MILK (kg/d)	321	17.87	6.28	17.8	5.3	36.65	8.75
DM%	323	12.22	1.22	12.1	9.46	16.36	1.63
FAT%	323	3.65	0.78	3.55	1.98	5.97	0.92
PRO%	323	3.30	0.48	3.25	2.29	5.64	0.61
ECM (kg/d)	321	16.75	5.23	16.44	5.31	33.23	7.43
DM (kg/d)	321	2.14	0.67	2.11	0.75	4.26	0.88
FAT (kg/d)	321	0.63	0.20	0.61	0.14	1.31	0.28
PRO (kg/d)	321	0.58	0.19	0.56	0.19	1.20	0.27

n: number of observations; SD: standard deviation; IQR: interquartile range; MILK: milk yield; DM: milk dry matter; FAT: milk fat; PRO: milk protein; ECM: energy corrected milk.

Genomic heritability estimates (Table 2) for DM%, FAT%, PRO%, DM, FAT, and PRO in our study ranged from 0.12 (PRO) to 0.45 (FAT%), which were moderate and similar to other comparable studies that used a far greater number of cows (Kim *et al.* 2009; Toghiani 2012; Cho *et al.* 2013). These studies presented heritability for FAT% ranging from 0.15 to 0.36, PRO% from 0.07 to 0.50, FAT from 0.28 to 0.52, and PRO from 0.26 to 0.34. However, in the current study it should also be realised that the standard errors for these milk solid traits, except FAT%, were high. These high standard errors of the heritabilities could be because the single test-day measurements in our study were derived from the cows at wide ranges of lactations and days in milk, whereas the heritabilities for milk production traits change widely throughout a lactation (Kim *et al.* 2009).

The heritability for MILK in our study was also lower than expected (0.01) and with a high standard error (13 times the mean) when compared with other studies (0.15 to 0.46, Kim *et al.* 2009). The low heritability for MILK is likely due to high environmental and residual variances or measurement errors and so indicates a larger sample size would be required for a more acceptable estimate for that trait. Similarly, the low heritability for MILK was also the reason for the low heritability for ECM (0.08), as these were calculated from MILK.

The accuracies of GEBV from GBLUP for DM%, FAT%, and PRO% were moderate (0.23 to 0.48) and significantly different from zero, as the mean of accuracies for these traits were at least almost twice their standard errors. However, the accuracy for MILK in our study was inflated by its very low heritability (0.01) to become an unrealistically high number (>1). The accuracies of GS for other traits were unstable with moderate means (0.11 to 0.34), but with very high standard error (0.7 to 2 times the mean). To avoid inflated accuracy resulting from close family relationships between training and test animals, partitioning animal into training and validation sets should be based on family so that the highly related animals were in the same validation set (Pszczola *et al.* 2012). However, due to the lack of pedigree data, cows in our study were just randomly partitioned into training and validation sets and this could be a bias source in our GS accuracies.

Table 2. Estimates of additive genetic variance (σ_a^2), random environmental variance (σ_e^2), residual variance (σ^2), heritability (h^2), correlation between GEBV and phenotype, and accuracy of genomic selection of milk production traits using univariate models

Trait	σ_a^2	σ_e^2	σ^2	$h^2 \pm SE$	Correlation $\pm SE$	Accuracy $\pm SE$
MILK (kg/d)	0.16	9.84	18.35	0.01 \pm 0.13	0.15 \pm 0.08	1.60 \pm 0.82
DM%	0.17	0.27	0.75	0.18 \pm 0.14	0.21 \pm 0.06	0.48 \pm 0.14
FAT%	0.21	0.04	0.25	0.45 \pm 0.15	0.19 \pm 0.05	0.28 \pm 0.07
PRO%	0.03	0.02	0.12	0.21 \pm 0.14	0.11 \pm 0.05	0.23 \pm 0.12
ECM (kg/d)	1.09	7.51	12.78	0.08 \pm 0.14	0.06 \pm 0.11	0.21 \pm 0.41
DM (kg/d)	0.04	0.11	0.21	0.15 \pm 0.14	0.13 \pm 0.09	0.34 \pm 0.24
FAT (kg/d)	0.005	0.009	0.019	0.19 \pm 0.15	0.05 \pm 0.08	0.11 \pm 0.18
PRO (kg/d)	0.002	0.012	0.016	0.12 \pm 0.14	0.08 \pm 0.08	0.24 \pm 0.25

Abbreviations of traits as in Table 1; SE: standard error

CONCLUSIONS

This study suggests that genomic selection using the GGP Bovine 50K chip and a single test day measurement could potentially be applied in Vietnam to the milk solid traits DM%, FAT%, and CP%, but not to milk yield traits. However, a larger sample size is recommended to confirm these findings. The very low estimation of the heritability of MILK could give misleading results when the accuracy of genomic selection is assessed.

REFERENCES

- Cho C., Cho K., Choy Y., Choi J., Choi T., Park B. and Lee S. (2013) *J. Ani. Sci. Tec.* **55**: 7.
- Covarrubias-Pazarán, G. (2019) 'Solving Mixed Model Equations in R'. CRAN.
- DataGene (2017) 'Australian Dairy Herd Improvement Report 2016'. DataGene Limited.
- Falconer D.S. and Mackay T.F.C. (1996) 'Introduction to quantitative genetics. Longman.
- Hayley R.A.B.A. (2010) Master Thesis, The Ohio State University.
- Kang H., Zhou L., Liu J., Kang H., Zhou L. and Liu J. (2017) *Fro. Agri. Sci. Engi.* **4**: 268.
- Kim B.W., Lee D., Jeon J.T. and Lee J.G. (2009) *Asian-Aust. J. Anim.Sci.* **22**: 923.
- Lam V., Wredle E.K., Thao N.T., Man N.V. and Svennersten S.K. (2010) *Afr. J. Agr. Re.* **5**: 2668.
- LGC Limited (2017) Sbeadex livestock kit.
- Meuwissen T. H., Hayes B. J. and Goddard M. E. (2001) *Genetics* **157**:1819.
- Mrode A. R. and Thompson R. (2013) 'Linear Models for the Prediction of Animal Breeding Values' 3rd Ed. CABI.
- Neogen GeneSeek Operations (2018) GeneSeek Genomic Profiler Bovine 50K.
- Nguyen M.H., Cesaro J.D., Pham D.K., Hoang V.Q. and Duteurtre G. (2016) *J. Anim. Sci. Tech.* **61**: 12.
- Pszczola M. Strabel T. Mulder H. A. Calus M. P. L. (2012) *J. Dairy Sci.* **95**: 389.
- R Core Team (2019) 'R: A language and environment for statistical computing'.
- Toghiani S. (2012) *Archiv Tierzucht* **55**: 458.
- Trach N. X. (2017) *Improving dairy cattle health and production in Vietnam.*
- Tyrrell H. F. and Reid J. T. (1965) *J. Dairy Sci.* **48**:1215.
- Vinamilk (2017) Vinamilk invested many high-tech dairy farms.
- Vu N. H., Lambert C. and Gauly M. (2016) *Asian-Aust. J Anim. Sci.* **10**: 290.