Genomic-polygenic and polygenic evaluations for milk yield and fat percentage using random regression models with Legendre polynomials in a Thai multibreed dairy population

Danai Jattawa[a], Mauricio A. Elzo[b], Skorn Koonawootrittriron[a*], and Thanathip Suwanasopee[a]

[a]Department of Animal Science, Kasetsart University, Bangkok 10900, Thailand

[b]Department of Animal Sciences, University of Florida, Gainesville, FL 32611-0910, USA

[*] Corresponding author: Department of Animal Science, Faculty of Agriculture, Kasetsart University, Bangkok 10900, Thailand; Tel: +66 2 5791120; Fax: +66 2 5791120; Email: agrskk@ku.ac.th (Skorn Koonawootrittriron)

Abstract

The objectives of this research were to compare estimates of variance components, genetic parameters, prediction accuracies, and rankings of animals for 305-d milk yield (305-d MY) and 305-d fat percentage (305-d FP) from random regression genomic-polygenic (RRGM) and random regression polygenic (RRPM) models. In addition, RRGM and RRPM prediction accuracies and rankings were compared with those from a standard cumulative 305-d genomic-polygenic model (SCGM). The dataset contained first-lactation monthly test-day records (69,029 for MY and 29,878 for FY) from 7,206 Holstein-upgraded cows located in 761 Thai farms. Genotypic data included 74,144 actual and imputed SNP from 1,661 animals. Variance components and genetic parameters were estimated using REML procedures. The RRGM and RRPM included contemporary group (herd-year-season), calving age, heterosis, and third-order Legendre population regression coefficients. Random effects were animal additive genetic third-order Legendre regression coefficients, permanent environment third-order Legendre regression coefficients, and residual. The SCGM contained contemporary group (herd-year-season), calving age and heterosis as fixed effects, and additive genetic and residual as random effects. The RRGM yielded higher additive genetic variances and heritabilities for 305-d MY and 305-d FP than RRPM, whereas correlations between MY and FY were similar in both models. The highest prediction accuracies for both traits were for RRGM, followed by RRPM, and the lowest ones were from SCGM. Similarly, the highest rank correlations were between animal EBV for 305-d MY and 305-d FP from RRGM and RRPM, followed by those between RRGM and SCGM, and the lowest ones were between RRPM and SCGM. The higher heritability estimates and higher prediction accuracies for RRGM than for RRPM and SCGM indicated that higher selection responses for 305-d MY and 305-d FP may be achieved in this Thai dairy

population by utilizing a random-regression model and genotypic information in addition to phenotypes and pedigree.

Key words: Dairy cattle, Multibreed, Genomic, Single-step, Random regression model

1. Introduction

Random regression models (RRM; Schaeffer and Dekkers, 1994; Jamrozik and Schaeffer, 1997) are the method of choice for genetic evaluation with test-day phenotypic records in dairy cattle. Advantages of RRM over standard cumulative 305-d models include more precise accounting of environmental factors affecting milk production throughout the lactation (Ptak and Schaeffer, 1993; Schaeffer et al., 2000), and in some cases inclusion of animals with incomplete lactations in genetic evaluations (Jensen, 2001). Dairy genetic evaluations for 305-d MY with RRM were found to be more accurate than with standard cumulative 305-d models (Schaeffer et al., 2000; Santos et al., 2014a, b). The advantages of RRM over 305-d models led to their wide utilization for national dairy genetic evaluations in many countries across the world (Interbull, 2007).

The original implementation of RRM for dairy genetic evaluations utilized only test-day phenotypic records and pedigree data. Advances in genotyping technology have made information on thousands of genotypes per animal available for dairy genetic evaluations. The combination of genomic information with phenotypes and pedigree (Meuwissen et al., 2001) increased accuracy of prediction (VanRaden et al., 2009; Wiggans et al., 2011; Thomasen et al., 2012; Přibyl et al., 2014) and rate of selection progress for dairy traits in cattle populations (de Roos et al., 2011; Buch et al., 2012). Several genomic evaluation approaches have been developed and implemented to date. The first implementation of a

national genomic evaluation in dairy cattle utilized a multi-step approach (VanRaden, 2008). However, this approach is somewhat complex and difficult to implement, especially in multiple-trait model and RRM (Misztal et al., 2013; Silva et al., 2014). Thus, a single-step approach was subsequently developed that was easier to implement and more accurate for genomic evaluation than multi-step procedures (Misztal et al., 2009, 2013; Aguilar et al., 2010). Single-step genomic-polygenic EBV for milk and fat yield with a standard cumulative 305-d model yielded prediction accuracies that were, on the average, 7.2%, higher than from a polygenic model in the Holstein-upgraded Thai population (Jattawa et al., 2015). However, evaluation of animals in this population with either polygenic or single-step genomic-polygenic RRM has yet to be done. This action is crucial for the development of a national dairy cattle genomic evaluation program in Thailand. Thus, the objectives of this research were: 1) to estimate variance components and genetic parameters for 305-d milk yield and 305-d fat percentage using random regression single-step genomic-polygenic and polygenic models, and 2) to compare prediction accuracies and rankings of animals for 305-d milk yield and 305-d fat percentage from random regression single-step genomic-polygenic and polygenic models, and also with prediction accuracies and rankings from a standard cumulative 305-d genomic-polygenic model in the Holstein-upgraded dairy cattle population in Thailand.

2. Materials and methods

2.1. Animals, datasets, and traits

Animals in the dataset belonged to the Holstein-upgraded Thai dairy population. The dataset included 7,206 first-lactation cows that were the progeny of 933 sires and 6,145 dams. Animals in this population were produced through upgrading from various breeds (Brahman,

Jersey, Brown Swiss, Red Dane, Red Sindhi, Sahiwal and Thai Native) to Holstein (Koonawootrittriron et al., 2009). Approximately 90% of cows, 93% of sires, and 78% of dams were 75% Holstein or higher.

Cows were from 761 farms located across five regions in Thailand (North, Northeastern, Western, Central, and Southern). Cows had their first calving between 1997 and 2014. Phenotypic records were collected once a month starting on the fifth day after calving until completion of lactation. Only cows that had their first test-day record before 40 days and had at least 4 test-day records were used. The last test-day record used here was the eleventh record (collected between 296 d and 340 d in milk). A total of 69,029 monthly test-day records from 7,206 cows that met these criteria were used in this research.

Two separate phenotypic datasets were prepared for genetic evaluations with the random regression and the standard cumulative 305-d model. Random regression models utilized a phenotypic dataset with monthly test-day records of 69,029 milk yield (MY) and 29,878 fat percentages (FP). The standard cumulative 305-d model used a phenotypic dataset with accumulated 305-d milk yields (305-d MY) and average 305-d fat percentages (305-d FP) computed using the collected monthly test-day records. The 305-d MY records were computed using the test interval method (Sargent et al., 1968; Koonawootrittriron et al., 2001). Numbers of records, means, and SD per trait for each dataset are shown in Table 1.

2.2. Genotypic data

Tissue samples (blood and semen) were collected from 2,661 animals (89 sires and 2,572 cows). All sires had daughters with pedigree and phenotypes and all cows had pedigree and phenotypes. The tissue samples were DNA extracted using a MasterPure™ DNA Purification Kit (Epicentre®, Madison, WI, USA). A NanoDrop™ 2000 Spectrophotometer

106 (Thermo Fisher Scientific Inc., Wilmington, DE, USA) was used to assess the quality of the
107 extracted DNA. A DNA sample was considered acceptable if it had a concentration higher
108 than 15 ng/µl and an absorbance ratio (i.e., absorbance at 260 nm divided by absorbance at
109 280 nm) of approximately 1.8. Acceptable DNA samples were sent to GeneSeek (GeneSeek
110 Inc., Lincoln, NE, USA) for genotyping with genomic profiler chips (1,412 with GGP9K,
111 570 with GGP20K, 540 with GGP26K, and 139 with GGP80K). Numbers of SNP genotypes
112 per chip were 8,590 for the GGP9K, 19,616 for the GGP20K, 25,979 for the GGP26K, and
113 76,694 for the GGP80K. Animals genotyped with GGP9K, GGP20K, and GGP26K chips
114 were imputed to GGP80K using FImpute 2.2 (Sargolzaei et al., 2014). Actual and imputed
115 SNP genotypes with minor allele frequencies lower than 0.04 (n = 2,375) or call rates lower
116 than 0.9 (n = 175) were removed. The resulting genotype file after these edits contained
117 74,144 actual and imputed SNP markers.
118
119 *2.3 Estimation of variance and covariance components*
120 Estimates of variance and covariance components for MY and FP were obtained using
121 bivariate random regression genomic-polygenic (RRGM) and random regression polygenic
122 models (RRPM). The RRGM was a single-step model (Misztal et al., 2009; Aguilar et al.,
123 2010) that utilized phenotypic, genotypic, and pedigree information, whereas the RRPM
124 utilized only phenotypic and pedigree information. Contemporary groups for RRGM and
125 RRPM were defined as herd-year-seasons because of the extremely low number of cows
126 within herd-test-day subclasses (1 or 2). This resulted in a total of 2208 contemporary groups
127 with a minimum size of 4 cows and a maximum size of 36 cows per contemporary group. In
128 matrix notation, the RRGM and RRPM can be described as follows:

129 $$y = Xb + Z_a a_a + Z_p p_a + e,$$

130 where y was a vector of MY and FP monthly test-day phenotypic records, b was a vector of
131 fixed contemporary group (herd-year-season) subclass effects, calving age regression
132 coefficient effects, heterosis regression coefficient effects, and third-order Legendre
133 population regression coefficient effects, a_a was a vector of random animal additive genetic
134 third-order Legendre regression coefficient effects, p_a was a vector of random permanent
135 environment third-order Legendre regression coefficient effects, e was a vector of residuals,
136 X, Z_a, and Z_p were incident matrices relating elements of y to elements of b, a_a, and p_a.
137 Columns of X related phenotypic records to: a) contemporary group effects through ones and
138 zeroes, b) calving age regression coefficient effects through calving ages (mo), c) heterosis
139 regression coefficient effect through animal heterozygosities (i.e., probabilities of one
140 Holstein allele and one allele from another breed in 1 locus), and d) third-order Legendre
141 population regression coefficient effects through third-order Legendre polynomials evaluated
142 at the standardized test-day of the phenotypic record. Columns in Z_a related phenotypic
143 records to elements of a_a through third-order Legendre polynomials evaluated at the
144 standardized test-day of the phenotypic record. Columns in Z_p related phenotypic records to
145 elements of p_a through third-order Legendre polynomials evaluated at the standardized test-
146 day of the phenotypic record. Legendre polynomials evaluated at the standardized test-days
147 were computed using the following expression (Kirkpatrick et al., 1990):

148 $$P_j(a_i^*) = \frac{1}{2^j}\sqrt{\frac{2j+1}{2}} \cdot \sum_{m=0}^{[j/2]} (-1)^m \binom{j}{m} \binom{2j-2m}{j} (a_i^*)^{j-2m},$$

149 where j was the order of polynomial, and a_i^* was the standardized milk test-day (range = -1
150 to 1). The a_i^* were calculated as follows:

$$a_i^* = \frac{2\,(a_i - a_{min})}{a_{max} - a_{min}} - 1$$

where a_i was days in milk at test-day i, a_{min} was the minimum number of days in milk, and a_{max} was the maximum number of days in milk in this population (i.e., $a_{max} = 340$). The third-order Legendre polynomials evaluated at the i[th] standardized milk test-day were: $P_0 = 0.7071\,(a_i^*)^0$, $P_1 = 1.2247\,(a_i^*)^1$, $P_2 = -0.7906\,(a_i^*)^0 + 2.3717\,(a_i^*)^2$, and $P_3 = -2.8062\,(a_i^*)^0 + 4.6771\,(a_i^*)^3$.

The assumptions of RRGM and RRPM were:

$$E[y] = Xb,$$

$$Var\begin{bmatrix} a \\ p \\ e \end{bmatrix} = \begin{bmatrix} C \otimes K_a & 0 & 0 \\ 0 & I \otimes K_p & 0 \\ 0 & 0 & I \otimes R_0 \end{bmatrix},$$

$$Var\,(y) = Z_a(C \otimes K_a)Z_a' + Z_p(I \otimes K_p)Z_p' + I \otimes R_0,$$

where $C = H$, the genomic-polygenic additive relationship matrix (genotypes and pedigree information) for RRGM and $C = A$, the polygenic additive relationship matrix (pedigree information only) for RRPM, matrix K_a was the 8 × 8 variance-covariance matrix among additive genetic third-order Legendre regression coefficients for MY and FP, matrix K_p was the 8 × 8 variance-covariance matrix among permanent environment third-order Legendre regression coefficients for MY and FP, matrix R_0 was the residual variance-covariance matrix for MY and FP, and ⊗ was the Kronecker product. The variance-covariance matrix of residual effects was assumed to be homogenous for all animals throughout the lactation because of the small size of the dataset.

The genomic-polygenic relationship matrix H (Legarra et al., 2009) was equal to:

$$H = \begin{bmatrix} A_{11} + A_{12}A_{22}^{-1}(G_{22} - A_{22})A_{22}^{-1}G_{21} & A_{12}A_{22}^{-1}G_{22} \\ G_{22}A_{22}^{-1}A_{21} & G_{22} \end{bmatrix},$$

where A_{11} was the submatrix of additive relationships among non-genotyped animals, A_{12} was the submatrix of additive relationships between non-genotyped and genotyped animals, A_{22}^{-1} was the inverse of the matrix of additive relationships among genotyped animals, and G_{22} was the matrix of genomic relationships among genotyped animals. Matrix $G_{22} = ZZ'/2\sum p_j\,(1-p_j)$, where p_j = frequency of allele 2 in locus j, and the elements of matrix Z for the j[th] SNP locus of the i[th] animal were defined as follows: $z_{ij} = (0 - 2p_j)$ for genotype = 11 in locus j, $z_{ij} = (1 - 2p_j)$ for genotype = 12 or 21 in locus j, and $z_{ij} = (2 - 2p_j)$ for genotype = 22 in locus j (VanRaden, 2008; Aguilar et al., 2010). Matrix G_{22} was scaled using the default restrictions imposed by program PREGSF90 from the BLUPF90 family programs (Misztal et al., 2002). These restrictions were: 1) mean of diagonal elements of submatrix G_{22} = mean of diagonal elements of submatrix A_{22}; and 2) mean of off-diagonal elements of submatrix G_{22} = mean of off-diagonal elements of submatrix A_{22}.

Variance components for RRGM and RRPM were estimated using restricted maximum likelihood (REML) procedures with an average information algorithm (program AIREMLF90; Tsuruta, 2014). The estimated 8 × 8 variances-covariance matrices of third-order additive genetic Legendre regression coefficients ($\widehat{K}_a$) and permanent environment Legendre regression coefficients ($\widehat{K}_p$), and the 2 × 2 residual variance-covariance matrix ($\hat{R}_0$) for MY and FP were used to estimate variance components and genetic parameters for each lactation day and for the complete 305-d lactation.

Estimates of variances and covariances for trait k, k = MY or FP, and lactation day i, for i = 5 to 305, were computed as follows: 1) additive genetic variances $\hat{\sigma}^2_{aki} = x'_{ki}\widehat{K}_a x_{ki}$, where x'_{ki} was a 1 × 8 vector with 4 non-zero elements for trait k (4 third-order Legendre polynomials evaluated at standardized lactation day *i*) and 4 zeroes; 2) permanent

195 environment variances $\hat{\sigma}^2_{pki} = x'_{ki}\hat{K}_p x_{ki}$; 3) phenotypic variances $\hat{\sigma}^2_{tki} =$
196 $\hat{\sigma}^2_{aki} + \hat{\sigma}^2_{pki} + \hat{\sigma}^2_{ek}$; and 4) heritabilities $\hat{h}^2_{ki} = \frac{\hat{\sigma}^2_{aki}}{\hat{\sigma}^2_{tki}}$. Estimates of covariances and
197 correlations between lactations days i and i', for i, i' = 5 to 305, for traits k and k', k,
198 k' = MY or FP, were computed as follows: 1) additive genetic covariances $\hat{\sigma}_{aki,ak'i'} =$
199 $x'_{ki}\hat{K}_a x_{k'i'}$, where $x_{k'i'}$ was an 8 × 1 vector with 4 non-zero elements for trait k' (4 third-
200 order Legendre polynomials evaluated at standardized lactation day *i'*) and 4 zeroes; 2)
201 permanent environment covariances $\hat{\sigma}_{pki,pk'i'} = x'_{ki}\hat{K}_p x_{k'i'}$; 3) phenotypic covariances
202 $\hat{\sigma}_{tki,tk'i'} = \hat{\sigma}_{aki,ak'i'} + \hat{\sigma}_{pki,pk'i'} + \hat{\sigma}_{ek,ek'}$; 4) additive genetic correlations $\hat{r}_{aki,ak'i'} =$
203 $\frac{\hat{\sigma}_{aki,ak'i'}}{(\hat{\sigma}^2_{aki} * \hat{\sigma}^2_{ak'i'})^{0.5}}$; 5) permanent environment correlations $\hat{r}_{pki,pk'i'} = \frac{\hat{\sigma}_{pki,pk'i'}}{(\hat{\sigma}^2_{pki} * \hat{\sigma}^2_{pk'i'})^{0.5}}$; and 6)
204 phenotypic correlations $\hat{r}_{tki,tk'j} = \frac{\hat{\sigma}_{tki,tk'i'}}{(\hat{\sigma}^2_{tki} * \hat{\sigma}^2_{tk'i'})^{0.5}}$.

205 The computation variances and covariances between pairs of traits (i.e., MY and
206 MY, FP and FP, and MY and FP) for lactation days 5 to 305 resulted in three 301 × 301
207 additive genetic variance-covariance submatrices, three 301 × 301 permanent
208 environment variance-covariance submatrices, and three 301 × 301 diagonal residual
209 submatrices. These submatrices were used to estimate complete 305-d lactation
210 variance-covariance matrices for MY and FP as follows: 1) 305-d additive genetic
211 variances and covariances $\hat{\sigma}_{a305d,kk'} = 1'\hat{V}_{akk'}1$, where $1'$ is a 1 × 301 vector of ones and
212 $\hat{V}_{akk'}$ is a 301 × 301 additive genetic variance-covariance matrix for trait pair kk', k ≥
213 k'; 2) 305-d permanent environment variances and covariances $\hat{\sigma}_{p305d,kk'} = 1'\hat{V}_{pkk'}1$,
214 where $1'$ is a 1 × 301 vector of ones and $\hat{V}_{pkk'}$ is a 301 × 301 permanent environment
215 variance-covariance matrix for trait pair kk', k ≥ k'; and 3) 305-d residual variances

and covariances $\hat{\sigma}_{e305d,kk'} = 1'\hat{V}_{ekk'}1$, where $1'$ is a 1 × 301 vector of ones and $\hat{V}_{ekk'}$ is a 301 × 301 diagonal residual variance-covariance matrix for trait pair kk', k ≥'. Subsequently, estimates of phenotypic variances, heritabilities, additive genetic correlations, environmental correlations, and phenotypic correlations for 305-d MY and FP were computed using the usual expressions.

2.4. Animal EBV, prediction accuracies and animal rankings

Firstly, RRGM and RRPM lactation day animal EBV for MY and FY were computed for lactation days 5 to 305 as follows: $EBV_{aki} = x'_{ki}\hat{a}_{aki}$, where x'_{ki} is a 1 × 8 vector with 4 non-zero elements for trait k (4 third-order Legendre polynomial coefficients evaluated at standardized lactation day *i*) and 4 zeroes, and $\hat{a}_{ki}$ is an 8 × 1 vector of third-order Legendre regression coefficient animal EBV for trait k (k = MY or FP) and day of lactation i. Prediction error variances for each EBV_{aki} and covariances between EBV_{aki} and $EBV_{ak'i'}$ for i ≥ i' were computed as $PEV_{aki,ak'i} = x'_{ki}\widehat{PEV}_{akik'i'}x_{k'i'}$, where $\widehat{PEV}_{akik'i'}$ is the 8 × 8 submatrix of PEV for third-order Legendre regression coefficient animal EBV between trait k (k = MY or FP) and lactation day i, and trait k' (k' = MY or FP) and lactation day i'.

Secondly, RRGM and RRPM animal EBV for 305-d MY and 305-d FP and their PEV were computed as follows: 1) $EBV_{a305d,k} = 1'EBV_{ak}$, where $1'$ is a 1 × 301 vector of ones and EBV_{ak} is a 301 × 1 vector of lactation-day EBV for animal a; 2) $PEV_{a305d,k} = 1'PEV_{akk}1$, where $1'$ is a 1 × 301 vector of ones, and PEV_{akk} is a 301 × 301 matrix of PEV variances and covariances among all lactation days for trait k (k = MY or FP) within animal a. Prediction accuracies for trait k = MY or FP, animal a, were computed as

238 $\sqrt{1-\frac{PEV_{a305d,k}}{\hat{\sigma}_{a305d,kk}}}$, where $PEV_{a305d,k}$ is the PEV for trait k, and $\hat{\sigma}_{a305d,kk}$ is the estimate of the
239 additive genetic variance for trait k (k = 305-d MY or 305-d FP).

240 Lastly, animal EBV and prediction accuracies from RRGM and RRPM were also
241 compared with a standard cumulative 305-d genomic-polygenic model (SCGM). The SCGM
242 was chosen because it had the highest prediction accuracy for milk yield and fat yield among
243 standard cumulative models in this population (Jattawa et al., 2015). The SCGM included
244 contemporary group (herd-year-season) subclass, calving age regression coefficient, and
245 heterosis regression coefficient as fixed effects, and animal additive genetic and residual as
246 random effects. The SCGM animal EBV were computed using REML additive genetic and
247 residual variance components estimated using program AIREMLF90 (Tsuruta, 2014).
248 Additive genetic variance components were: var(305-d MY) = 170,400 kg^2, var(305-d FP) =
249 0.06 %^{2}, and cov(305-d MY, 305-d FP) = -20.2 kg*%. Residual variance components were:
250 var(305-d MY) = 480,710 kg^2, var(305-d FP) = 0.18 %^{2}, and cov(305-d MY, 305-d FP) = -
251 42.9 kg*%. Prediction accuracies were computed as $\sqrt{1-\frac{PEV_{ak}}{\hat{\sigma}_{ak}}}$, where PEV_{ak} was the
252 prediction error variance for animal a, trait k, and $\hat{\sigma}_{ak}$ was the estimate of the additive
253 genetic variance for trait k, k = 305-d MY or 305-d FP from SCGM.

254 Rank correlations were calculated for 305-d MY and 305-d FP EBV from RRPM,
255 RRGM, and SCGM for all animals in the population, only sires (top 5%, 15%, 25%, and all
256 sires), and only cows (top 5%, 15%, 25%, and all cows). Associations between rankings
257 from the three models within population segments and the complete population were
258 evaluated using Spearman's rank correlations (SAS CORR procedure; SAS, 2003).

259

260 **3. Results and discussion**

261 *3.1. Variance components, heritabilities and genetic correlations*

262 Estimates of variances throughout the lactation (day 5 to 305) for MY and FP from
263 RRGM and RRPM are shown in Fig. 1 for additive genetic effects, Fig. 2 for permanent
264 environmental effects, and Fig. 3 for phenotypic effects. The pattern of daily variances
265 estimated with RRGM and RRPM was similar within traits (MY or FP) throughout the
266 lactation. Additive genetic variances for MY increased during the first three months,
267 declined during the next four months, and then increased again after seven months until the
268 end of the lactation. Similar additive genetic variances were obtained for FP from the
269 beginning of the lactation until day 245, then values sharply increased until the end of the
270 lactation. Daily permanent environmental variances (Fig. 2) and phenotypic variances (Fig.
271 3) showed the same patterns for MY and FP throughout the lactation, except during the first
272 month of lactation where both variances decreased for MY, but were low and similar for FP.
273 After the first month, daily permanent environmental and phenotypic variances for both traits
274 changed little during the next eight months and then increased until the end of the lactation.
275 Substantially larger changes in estimates of daily variance components for MY and
276 FP existed during the first 45 d and the last 45 d of lactation, especially for permanent
277 environmental effects. Implausibly high additive and permanent environmental variances at
278 the beginning and end of the lactation were also reported for MY, FP, and other dairy traits
279 (fat yield, protein yield, somatic cell count) in previous studies that fitted lactation curves
280 with Legendre polynomials (López-Romero and Carabaño, 2003; López-Romero et al., 2004;
281 Strabel and Jamrozik, 2006; Bohmanova et al., 2008, 2009). Large changes of variances at
282 the boundaries of the lactation curve have been attributed to low number of records during
283 these periods (Misztal et al., 2000; Strabel et al., 2005; Bohmanova et al., 2008) and to
284 artifacts of Legendre polynomials evaluated at extremes days in milk (Misztal et al., 2000;

López-Romero et al., 2004). Lower numbers of records after day 250 of the lactation may have contributed to the implausible values of additive genetic and permanent environmental variances at the end of the lactation. Poor adjustment of the third-degree Legendre polynomial may have been responsible for the unlikely variance component values at the beginning of the lactation. Other factors that may have contributed to the poor estimates of variance components at the extremes of the lactation curve were unaccounted effects of preferential treatment, stage of gestation, and variation among shapes of lactation curves across herds (Jamrozik et al., 2001; de Roos et al., 2004; Bohmanova et al., 2008).

Heritability estimates for daily MY and FP from RRGM and RRPM are shown in Fig. 4. Heritabilities for daily MY tended to follow the same pattern as that of daily additive genetic variances, i.e., they increased from the beginning of the lactation until the ninth month, then they decreased during the tenth month of lactation. Conversely, heritabilities estimates for daily FP increased from the beginning until the end of the lactation.

The pattern of MY heritability values here was in agreement with heritability patterns obtained in Dutch Holstein (Pool et al., 2000), Polish Black and White (Strabel and Jamrozik, 2006), and Tunisian Holstein populations (Hammami et al., 2008). Opposite patterns of high heritability at the beginning and end of the lactation were reported in Finish Ayrshire (Kettunen et al., 2000) and in Spanish Holstein (López-Romero and Carabaño, 2003). Patterns with low heritability at the extremes of the lactation may be more realistic because they indicate that MY at the extremes of the lactation were more highly influenced by environmental effects than in the middle of the lactation (Strabel et al., 2005).

Estimates of additive genetic, permanent environmental, and phenotypic variances and covariances for 305-d MY and 305-d FP computed using RRGM and RRPM are shown in Table 2. Estimates of additive genetic variances and covariances for 305-d MY and 305-

d FP were larger for RRGM than for RRPM. Conversely, estimates of permanent environmental variances and covariances from RRGM were lower than those from RRPM. However, phenotypic variances and covariances estimated for 305-d MY and 305-d FP from both models were similar. This indicated that the information from 74,144 actual and imputed genotypes helped the RRGM explain more 305-d MY and 305-d FP additive genetic variation than that explained by the RRPM using only pedigree and phenotypes.

The RRGM higher additive genetic and similar phenotypic variances to RRPM resulted in higher RRGM heritabilities (0.27 for 305-d MY; 0.16 for 305-d FP) than those from RRPM (0.21 for 305-d MY; 0.12 for 305-d FP; Table 3). The heritability estimate for 305-d MY obtained here with RRGM was similar to one previously estimated in this Thai population with a cumulative 305-d genomic-polygenic model with 74,144 actual and imputed SNP genotypes (0.26; Jattawa et al., 2015). This estimate was also within the range of heritabilities obtained using genomic models in various Holstein populations from temperate environments (0.23 to 0.33; VanRaden et al., 2009; Gao et al., 2012; Karoui et al., 2012; Rodríguez-Ramilo et al., 2014; Sun et al., 2014; Tsuruta et al, 2014). However, the RRGM heritability for 305-d FP obtained here was somewhat lower than heritabilities reported in other temperate dairy populations. Sun et al. (2014) reported 305-d FP genomic heritability of 0.54 for Jersey population in USA. Genomic heritability estimates for Holstein were 0.5 in France (Karoui et al., 2012), 0.25 in Germany (Wittenburg et al, 2015), and ranged from 0.45 to 0.5 in USA (VanRaden et al., 2009; Sun et al., 2014).

Genetic, permanent environment, and phenotypic correlations between 305-d MY and 305-d FP estimated with RRGM and RRPM were all low and negative (Table 3). The estimate of RRGM additive genetic correlation was slightly higher (-0.24) than that from RRPM (-0.19), whereas estimates of permanent environmental correlations where nearly

identical (-0.31 for RRGM and -0.32 for RRPM) and phenotypic covariances were identical (-0.14) for the two models. Thus, inclusion of SNP genotypes in addition to pedigree and phenotypes in random regression models had a very small impact on additive genetic, permanent environmental, and phenotypic correlations between 305-d MY and 305-d FY in this population. The negative additive genetic correlations between 305-d MY and 305-d FP from RRGM and RRPM obtained here indicated that cows with higher MY tended to have lower FP and vice versa. The negative additive genetic correlations between 305-d MY and 305-d FP here were somewhat lower than polygenic estimates from several Holstein populations in tropical environments (-0.32 to -0.42; Boujenane, 2002; Othmane et al., 2004; Hashemi and Nayebpoor, 2008) and in temperate environments (-0.40 to -0.55; Chauhan and Hayes, 1991; Welper and Freeman, 1992; Miglior et al., 2007; Loker et al., 2012).

The development of the single-step genomic-polygenic evaluation procedure (Aguilar et al., 2010) as well as its integration into the BLUPF90 family of programs (Misztal et al., 2002) enormously facilitated the analysis and implementation of an animal random regression genomic-polygenic evaluation system in this Thai dairy population. Random regression MY and FP variance components and genetic parameters were estimated using all available test-day phenotypic, pedigree, and genotypic information from this population. The higher estimates of additive genetic variances and heritabilities for 305-d MY and 305-d FP from RRGM indicated broader additive genetic differences among individual animals, thus increasing the opportunity of selecting genetically superior animals more accurately for 305-d MY and 305-d FP than with RRPM. In particular, including genotypic information in RRGM would increase the accuracy of genetic evaluation and selection of genetically superior young bulls and cows, thus shortening generation intervals. Consequently, higher

356 rates of genetic change for 305-d MY and 305-d FP could be expected with genomic-
357 polygenic than with polygenic random regression models in this population.
358
359 *3.2. Accuracy of genomic-polygenic and polygenic EBV for 305-d milk yield and 305-d fat*
360 *percentages*
361 Fig. 5 shows the EBV accuracies for 305-d MY and 305-d FP computed with RRGM,
362 RRPM, and SCGM for all animals, sires, and cows. The RRGM had the highest mean EBV
363 accuracy for all animals (49.3% for 305-d MY and 38.6% for 305-d FP), RRPM was second
364 (45.7% for 305-d MY, and 36.1% for 305-d FP), and the least accurate was the SCGM
365 (39.5% for 305-d MY, and 30.5% for 305-d FP). Similarly, RRGM had the highest mean
366 EBV accuracy for sires (44.3% for 305-d MY and 37.2% for 305-d FP) and for cows (49.7%
367 for 305-d MY and 38.8% for 305-d FP), followed by RRPM (sires: 39.5% for 305-d MY and
368 31.3% for 305-d FP; cows: 46.2% for 305-d MY and 36.6% for 305-d FP). The lowest mean
369 EBV accuracies for sires (37.3% for 305-d MY and 30.5% for 305-d FP) and for cows (39.6%
370 for 305-d MY and 30.5% for 305-d FP) were from SCGM.
371 Higher EBV accuracies for RRGM than for RRPM (3.6% for 305-MY and 2.5% for
372 305-d FP) indicated that including genomic information in genetic evaluations increased
373 prediction accuracies over genetic evaluations based only on pedigree and phenotypic data
374 in this population. This agreed with results from previous research showing that utilization
375 of genomic information in addition to pedigree and phenotypic information to evaluate dairy
376 cattle yielded higher prediction accuracies in various dairy populations (VanRaden et al.,
377 2009; Van Doormaal et al., 2009; Wiggans et al., 2011; Su et al., 2012; Thomasen et al.,
378 2012; Bauer et al., 2014, 2015; Přibyl et al., 2014; Jattawa et al., 2015). Mean accuracies of
379 305-d MY genomic-polygenic EBV computed with single-step cumulative 305-d models

were 7.2% higher than the mean accuracy from polygenic EBV in this same Thai population (Jattawa et al., 2015). Similarly, prediction accuracy for 305-d MY from a single-step random regression genomic-polygenic model was 6.8% higher than that from random regression polygenic evaluation in a population of 1,854,275 Czech Holstein using 40,653 SNP from 2,236 genotyped sires (Bauer et al., 2015). This 6.8% increase in accuracy was higher than the value of 3.6% obtained here although the number of genotyped animals was smaller than the 2,661 animals genotyped in this Thai population. This difference was likely related to the higher level of relationships that existed in the Czech Holstein population between genotyped and non-genotyped animals (genotyped sires that had an average 240 daughters each) compared to the population here (genotyped parents had an average of 10 progenies each). A second reason may be that only 139 animals in this population had actual 80k genotypes, the rest (n = 2,522) had combinations of actual and imputed 80k genotypes. Previous studies have indicated that high levels of relationship between genotyped and non-genotype animals can improve the accuracy of genomic evaluations (Habier et al., 2010; Pszczola et al., 2012; Wu et al., 2015). Thus, increasing the fraction of genotyped animals with high-density SNP chips that are highly related to animals in the rest of the population would likely help increase genomic-polygenic prediction accuracies in future years.

Fig. 5 also shows that RRGM and RRPM yielded higher EBV accuracies for 305-d MY and 305-d FP than SCGM. On the average, RRGM EBV were 9% more accurate (9.8% for 305-d MY and 8.1% for 305-d FP) and RRPM EBV were 6% more accurate (6.2% for 305-d MY and 5.6% for 305-d FP) than SCGM EBV. These higher EBV accuracies for RRGM and RRPM than for SCGM agreed with previous studies that indicated that random regression models yielded more accurate 305-d EBV than standard cumulative 305-d models (Schaeffer et al., 2000; Santos et al., 2014a, b). The gains in accuracy from SCGM to RRGM

404 (9.8%) and from SCGM to RRPM (6.2%) for 305-d MY EBV were higher than the gain
405 obtained from polygenic cumulative 305-d to polygenic random regression models in
406 Guzerat (3.0% to 3.6%; Santos et al., 2014a, b). The higher EBV accuracies of the RRGM
407 make it the model of choice for genetic evaluation of 305-d MY and 305-d FP in the Holstein-
408 upgraded Thai population.
409
410 *3.3. Rank correlations between genomic-polygenic and polygenic EBV for 305-d milk yield*
411 *and 305-d fat percentage*
412 Table 4 shows Spearman rank correlations among all animal EBV rankings from the
413 RRGM, RRPM, and SCGM for 305-d MY and 305-d FP. The highest rank correlations were
414 between EBV from RRGM and RRPM (0.94 for 305-d MY, and 0.78 for 305-d FP), followed
415 by those between EBV from RRGM and SCGM (0.66 for 305-d MY, and 0.57 for 305-d FP),
416 and the lowest ones were those between EBV from RRPM and SCGM (0.61 for 305-d MY,
417 and 0.45 for 305-d FP). Rank correlations between animal EBV from RRGM and RRPM
418 indicated that genotypic data had little impact on EBV rankings for 305-d MY, but somewhat
419 higher impact on EBV rankings for 305-d FP. Inclusion of genomic information in dairy
420 genetic evaluations had higher impact on the accuracy of EBV for animals without
421 phenotypes than for animals with phenotypes (Schaeffer, 2006; Pollott et al., 2014; Bauer et
422 al., 2015). All cows had 305-d MY records but 3,942 cows had no 305-d FP records. The
423 lower rank correlation between RRGM and RRPM EBV for 305-d FP (0.78) than for 305-d
424 MY (0.94) was largely due to bigger changes in ranking for 305-d FP in animals without FP
425 records (mean = 2,355) compared to smaller changes in ranking for 305-d MY for these same
426 animals (mean = 1,105) because they had MY records.

427 The rank correlation between 305-d MY animal EBV from RRGM and RRPM here
428 (0.94) was higher than the value of 0.84 previously obtained in this same population between
429 animal EBV from genomic-polygenic and polygenic cumulative 305-d models (Jattawa et
430 al., 2015). The rank correlations between 305-d MY animal EBV from RRGM and SCGM
431 (0.66) and from RRPM and SCGM (0.61) here were substantially lower than the rank
432 correlation between animal EBV from polygenic random regression and cumulative 305-d
433 models (0.89) in Brazilian Guzerat (Santos et al., 2014a). This indicated that utilization of
434 genomic information in cumulative 305-d models had a higher impact on animal EBV values
435 and rankings than in random regression models in this Holstein-upgraded Thai population.
436 Rank correlations for 305-d MY and 305-d FP among RRGM, RRPM, and SCGM
437 for sires only are shown in Table 5 (top 5%, 15%, 25%, and all sires) and for cows only in
438 Table 6 (top 5%, 15%, 25%, and all cows). In addition, these two tables present percentages
439 of animals in common for 305-d MY and 305-d FP in the top 5%, 15%, and 25% of animals
440 ranked by the two models in each rank correlation. Rank correlations between EBV for sires
441 (Table 5) and for cows (Table 6) between pairs of followed the same pattern as rank
442 correlations obtained for all animals (Table 4). Rank correlations between EBV from RRGM
443 and RRPM tended to be higher across the top 5%, 15%, 25%, and all animals (0.57 to 0.94
444 for sires; 0.62 to 0.94 for cows), than those between EBV from RRGM and SCGM (0.42 to
445 0.69 for sires; 0.43 to 0.66 for cows), and those between EBV from RRPM and SCGM (0.38
446 to 0.65 for sires; 0.39 to 0.61 for cows). The top 5% of sires and cows had the lowest
447 percentages of animals in common between pairs of models, and these percentages tended to
448 increase as the fraction of sires and cows increased from 5% to 15% to 25% to 100%. The
449 highest percentages of animals in common in the top 5% were between rankings from RRGM
450 and RRPM (305-d MY: 83% for sires and 81% for cows; 305-d FP: 65% for sires and 64%

for cows). The second highest set of percentages of animals in common was the one between rankings from RRGM and SCGM (305-d MY: 58% for sires and 52% for cows; 305-d FP: 46% for sires and 45% for cows). The lowest percentages of animals in common were between rankings from RRPM and SCGM (305-d MY 54% for sires and 46% for cows; 305-d FP: 44% for sires and 40% for cows). Lower percentages of animals in common between sires and cows ranked for 305-d FP than for 305-d MY were likely the result of larger changes in 305-d FP EBV across models due to lower EBV accuracies for this trait than accuracies for 305-d MY EBV in this population. Genetic parameters, EBV accuracies, and animal rankings obtained here will help explain Thai dairy producers and stakeholders the motivation for changing the current standard cumulative polygenic model to a genomic-polygenic model based on genotypes, pedigree, and phenotypes.

4. Conclusions

Similar patterns of daily variance components and heritabilities for MY and FP were obtained using random regression genomic-polygenic and polygenic models. The RRGM yielded higher estimates of genetic variances and heritabilities than RRPM estimates for both daily and cumulative 305-d MY and FP. Similarly, EBV accuracies were higher for RRGM than for RRPM, and EBV accuracies from both random regression models were higher than those from the SCGM. Considering the higher heritabilities and EBV accuracies of the RRGM than the RRPM and SCGM, selection based on RRGM animal EBV would be expected to achieve faster rates of genetic change for 305-d MY and 305-d FP than with RRPM and SCGM animal EBV in this Thai dairy population.

Conflict of interest

Authors declare that no conflicts of interest influenced this research.

Acknowledgements

The authors would like to thank the Royal Golden Jubilee Ph.D. Program (RGJ) of the Thailand Research Fund (TRF) for awarding a scholarship to the first author, the University of Florida for supporting the training of the first author as a research scholar, and the National Science and Technology Development Agency (NSTDA), Kasetsart University (KU), and the Dairy Farming Promotion Organization of Thailand (D.P.O.) for providing funding and logistic support for this research. The authors appreciate the Thai dairy farmers, dairy cooperatives, and dairy related organizations for their contribution and support of this investigation and Carlos Martinez for useful discussions on random regression models with Legendre polynomials.

References

Aguilar, I., Misztal, I., Johnson, D.L., Legarra, A., Tsuruta, S., and Lawlor, T.J., 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.

Bauer, J., Přibyl, J., and Vostrý, L., 2015. Short communication: Reliability of single-step genomic BLUP breeding values by multi-trait test-day model analysis. J. Dairy Sci. 98: 4999-5003.

Bauer, J., Vostrý, L., Přibyl, J., Svitáková, A., and Zavadilová, L., 2014. Approximation of reliability of single-step genomic breeding values for dairy cattle in the Czech Republic. Anim. Sci. Pap. Rep. 32: 301-306.

498 Bohmanova, J., Miglior, F., and Jamrozik, J., 2009. Use of test-day records beyond three
499 hundred five days for estimation of three hundred five–day breeding values for
500 production traits and somatic cell score of Canadian Holsteins. J. Dairy Sci. 92:
501 5314-5325.
502 Bohmanova, J., Miglior, F., Jamrozik, J., Misztal, I., and Sullivan, P.G., 2008. Comparison
503 of random regression models with Legendre polynomials and linear splines for
504 production traits and somatic cell score of Canadian Holstein cows. J. Dairy Sci. 91:
505 3627-3638.
506 Boujenane, I, 2002. Estimates of genetic and phenotypic parameters for milk production in
507 Moroccan Holstein-Friesian cows. Revue Élev. Méd. vét. Pays trop. 55: 63-67.
508 Buch, L.H., Sørensen, M.K., Berg, P., Pedersen, L.D., and Sørensen, A.C., 2012. Genomic
509 selection strategies in dairy cattle: Strong positive interaction between use of
510 genotypic information and intensive use of young bulls on genetic gain. J. Anim.
511 Breed. Genet. 129: 138-151.
512 Chauhan, V.P.S., and Hayes, J.F., 1991. Genetic parameters for first lactation milk
513 production and composition traits for Holsteins using multivariate restricted
514 maximum likelihood. J. Dairy Sci. 74: 603-610.
515 de Roos, A.P., Harbers, A.G., and de Jong, G., 2004. Random herd curves in a test-day
516 model for milk, fat, and protein production of dairy cattle in The Netherlands. J.
517 Dairy Sci. 87: 2693-2701.
518 de Roos, A.P.W., Schrooten, C., Veerkamp, R.F., and van Arendonk, J.A.M., 2011. Effects
519 of genomic selection on genetic improvement, inbreeding, and merit of young
520 versus proven bulls. J. Dairy Sci. 94, 1559-1567.

521 Gao, H., Christensen, O.F., Madsen, P., Nielsen, U.S., Zhang, Y., Lund, M.S., and Su, G.,
522 2012. Comparison on genomic predictions using three GBLUP methods and two
523 single-step blending methods in the Nordic Holstein population. Genet. Sel. Evol.
524 44: 8.
525 Habier, D., Tetens, J., Seefried, F.-R., Lichtner, P., and Thaller, G., 2010. The impact of
526 genetic relationship information on genomic breeding values in German Holstein
527 cattle. Genet. Sel. Evol. 42: 5.
528 Hammami, H., Rekik, B., Soyeurt, H., Gara, A.B., and Gengler, N., 2008. Genetic
529 parameters for Tunisian Holsteins using a test-day random regression model. J.
530 Dairy Sci. 91: 2118-2126.
531 Hashemi, A., and Nayebpoor, M., 2008. Estimates of genetic and phenotype parameters for
532 milk production in Iran Holstein-Friesian cows. Res. J. Biol. Sci. 3: 678-682.
533 Interbull, 2007. Interbull routine genetic evaluation for dairy production traits, August
534 2007. Available at:
535 http://www.interbull.org/web/static/mace_evaluations_archive/eval/aug07.html.
536 Jamrozik, J., and Schaeffer, L.R., 1997. Estimates of genetic parameters for a test day
537 model with random regression for yield traits of first lactation Holsteins. J. Dairy
538 Sci. 80: 762-770.
539 Jamrozik, J., Gianola, D., and Schaeffer, L.R., 2001. Bayesian estimation of genetic
540 parameters for test day records in dairy cattle using linear hierarchical models.
541 Livest. Prod. Sci. 71: 223-240.
542 Jattawa, D. Elzo, M.A., Koonawootrittriron, S, and Suwanasopee, T., 2015. Comparison of
543 genetic evaluations for milk yield and fat yield using a polygenic model and three

544 genomic-polygenic models with different sets of SNP genotypes in Thai multibreed
545 dairy cattle. Livest. Sci. 181: 58-64.
546 Jensen, J., 2001. Genetic evaluation of dairy cattle using test-day models. J. Dairy Sci. 84:
547 2803-2812.
548 Karoui, S., Carabaño, M.J., Díaz, C., and Legarra, A., 2012. Joint genomic evaluation of
549 French dairy cattle breeds using multiple-trait models. Genet. Sel. Evol. 44: 39.
550 Kettunen, A., Mäntysaari, E.A., and Pösö, J., 2000. Estimation of genetic parameters for
551 daily milk yield of primiparous Ayrshire cows by random regression test-day
552 models. Livest. Prod. Sci. 66: 251-261.
553 Kirkpatrick, M., Lofsvold, D., and Bulmer, M., 1990. Analysis of inheritance, selection and
554 evolution of growth trajectories. Genetics 124: 979-993.
555 Koonawootrittriron, S., Elzo, M.A., Thongprapi, T., 2009. Genetic trends in a Holstein ×
556 other breeds multibreed dairy population in Central Thailand. Livest. Sci. 122, 186-
557 192.
558 Koonawootrittriron, S., Elzo, M.A., Tumwasorn, S., Sintala, W., 2001. Prediction of 100-d
559 and 305-d milk yields in a multibreed dairy herd in Thailand using monthly test-day
560 records. Thai J. Agric. Sci. 34, 163-174.
561 Legarra, A., Aguilar, I., Misztal, I. 2009. A relationship matrix including full pedigree and
562 genomic information. J. Dairy Sci. 92: 4656-4663.
563 Loker, S., Bastin, C., Miglior, F., Sewalem, A., Schaeffer, L.R., Jamrozik, J., Ali, A., and
564 Osborne, V.., 2012. Genetic and environmental relationships between body
565 condition score and milk production traits in Canadian Holsteins. J. Dairy Sci. 95:
566 410–419.

López-Romero, P., and Carabaño, M.J., 2003. Comparing alternative random regression models to analyse first lactation daily milk yield data in Holstein–Friesian cattle. Livest. Prod. Sci. 82: 81-96.

López-Romero, P., Rekaya, R. and Carabaño, M.J., 2004. Bayesian comparison of test-day models under different assumptions of heterogeneity for the residual variance: the change point technique versus arbitrary intervals. J. Anim. Breed. Genet. 121: 14-25.

Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157: 1819-1829.

Miglior, F., Sewalem, A., Jamrozik, J., Bohmanova, J., Lefebvre, D.M., and Moore, R.K., 2007. Genetic Analysis of Milk Urea Nitrogen and Lactose and Their Relationships with Other Production Traits in Canadian Holstein Cattle. J. Dairy Sci. 90: 2468–2479.

Misztal, I., Aggrey, S.E., and Muir, W.M., 2013. Experiences with a single-step genome evaluation. Poult. Sci. 92: 2530-2534.

Misztal, I., Legarra, A., and Aguilar, I., 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J. Dairy Sci. 92: 4648-4655.

Misztal, I., Strabel, T., Jamrozik, J., Mäntysaari, E.A., and Meuwissen, T.H., 2000. Strategies for estimating the parameters needed for different test-day models. J. Dairy Sci. 83: 1125-1134.

Misztal, I., Tsuruta, S., Strabel, T., Auvray, B., Druet, T., Lee, D. H., 2002. BLUPF90 and related programs (BGF90). In: Proc.7th World Congr. Genet. Appl. Livest. Prod., Montpellier, France.

591 Othmane, M.H., Hamouda, M.B., and Hammami, H., 2004. Multivariate animal model
592 estimates of genetic, environmental and phenotypic correlations for early lactation
593 milk yield and composition in Tunisian Holstein-Friesians. Interbull Bull 32: 129-
594 132.
595 Pollott, G.E., Charlesworth, A., and Wathes, D.C., 2014. Possibilities to improve the
596 genetic evaluation of a rare breed using limited genomic information and
597 multivariate BLUP. Animal 8: 685-694.
598 Pool, M.H., Janss, L.L., and Meuwissen, T.H., 2000. Genetic parameters of Legendre
599 polynomials for first parity lactation curves. J. Dairy Sci. 83: 2640-2649.
600 Přibyl, J., Bauer, J., Pešek, P., Přibylová, J., Vostrý, L., and Zavadilová, L., 2014. Domestic
601 and Interbull information in the single step genomic evaluation of Holstein milk
602 production. Czech J. Anim. Sci. 59: 409-415.
603 Pszczola, M., Strabel, T., Mulder, H.A., and Calus, M.P.L., 2012. Reliability of direct
604 genomic values for animals with different relationships within and to the reference
605 population. J. Dairy Sci. 95: 389-400.
606 Ptak, E., and Schaeffer, L.R., 1993. Use of test day yields for genetic evaluation of dairy
607 sires and cows. Livest. Prod. Sci. 34: 23-34.
608 Rodríguez-Ramilo, S.T., García-Cortés, L.A., González-Recio, Ó., 2014. Combining
609 genomic and genealogical information in a reproducing Kernel Hilbert spaces
610 regression Model for genome-enabled predictions in dairy cattle. PLoS ONE, 9,
611 e93424. Available at:
612 http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0093424.
613 Santos, D.J., Boison, S.A., Utsunomya, A.T., Peixoto, M.G.C.D., Tonhati, H., Sölkner, J.,
614 and da Silva, M.V., 2014a. An approach to genomic analysis of longitudinal data

using random regression. In: Proc.10th World Congr. Genet. Appl. Livest. Prod., Vancouver, Canada.

Santos, D.J.A., Peixoto, M.G.C.D., Borquis, R.R.A., Panetto, J.C.C., Faro, L.E., and Tonhati, H., 2014b. Predicting breeding values for milk yield of Guzerá (*Bos indicus*) cows using random regression models. Livest. Sci. 167: 41-50.

Sargent, F.D., Lytton, V.H., Wall Jr., O.G., 1968. Test interval method of calculating Dairy Herd Improvement Association records. J. Dairy Sci. 51, 170-179.

Sargolzaei, M., Chesnais, J.P., Schenkel, F.S., 2014. A new approach for efficient genotype imputation using information from relatives. BMC Genomics, 15, 478.

SAS, 2003. SAS OnlineDoc 9.1.3. SAS institute Inc., Cary, North Carolina, USA.

Schaeffer, L.R., and Dekkers, J.C.M., 1994. Random regressions in animal models for test-day production in dairy cattle. Proc. 5th World Congr. Genet. Appl. Livest. Prod., Guelph, XVIII: 443.

Schaeffer, L.R., 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123: 218-223.

Schaeffer, L.R., Jamrozik, J., Kistemaker, G.J., and Van Doormaal, B.J., 2000. Experience with a test-day model. J. Dairy Sci. 83: 1135-1144.

Silva, M.V.B., Santos, D.J.A., Boison, S.A., Utsunomiya, A.T.H., Carmo, A.S., Sonstegard, T.S., Cole, J.B., and Van Tassell, C.P., 2014. The development of genomics applied to dairy breeding. Livest. Sci. 166: 66-75.

Strabel, T., and Jamrozik, J., 2006. Genetic analysis of milk production traits of Polish Black and White cattle using large-scale random regression test-day models. J. Dairy Sci. 89: 3152-3163.

638 Strabel, T., Szyda, J., Ptak, E., and Jamrozik, J., 2005. Comparison of random regression
639 test-day models for Polish Black and White cattle. J. Dairy Sci. 88: 3688-3699.
640 Su, G., Madsen, P., Nielsen, U.S., Mäntysaari, E.A., Aamand, G.P., Christensen, O.F., and
641 Lund, M.S., 2012. Genomic prediction for Nordic Red cattle using one-step and
642 selection index blending. J. Dairy Sci. 95: 909-917.
643 Sun, C., VanRaden, P.M., Cole, J.B., and O'Connell, J.R. 2014. Improvement of prediction
644 ability for genomic selection of dairy cattle by including dominance effects. PLoS
645 ONE 9: e103934.
646 Thomasen, J.R., Guldbrandtsen, B., Su, G., Brøndum, R.F., and Lund, M.S., 2012.
647 Reliabilities of genomic estimated breeding values in Danish Jersey. Animal 6: 789-
648 796.
649 Tsuruta, S., 2014. Average Information REML with several options including EM-REML
650 and heterogeneous residual variances. Available at:
651 http://nce.ads.uga.edu/wiki/doku.php?id=application_programs.
652 Tsuruta, S., Misztal, I., Aguilar, I., and Lawlor, T.J., 2014. Genome wide association study
653 on cow mortality in three US regions. In: Proc.10th World Congr. Genet. Appl.
654 Livest. Prod., Vancouver, Canada.
655 Van Doormaal, B.J., Kistemaker, G.J., Sullivan, P.G., Sargolzaei, M., and Schenkel, F.S.,
656 2009. Canadian implementation of genomic evaluations. Interbull Bull. 40: 214-
657 218.
658 VanRaden, P.M., 2008. Efficient methods to compute genomic predictions. J. Dairy Sci.
659 91, 4414-4423.

660 VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D.,
661 Taylor, J.F., Schenkel, F.S., 2009. Invited review: Reliability of genomic
662 predictions for North American Holstein bulls. J. Dairy Sci. 92, 16-24.
663 Welper, R.D., and Freeman, A.E., 1992. Genetic parameters for yield traits of Holsteins,
664 including lactose and somatic cell score. J. Dairy Sci. 75: 1342–1348.
665 Wiggans, G.R., VanRaden, P.M., Cooper, T.A., 2011. The genomic evaluation system in
666 United States: Past, present, future. J. Dairy Sci. 94, 3202-3211.
667 Wittenburg, D., Melzer, N., and Reinsch, N., 2015. Genomic additive and dominance
668 variance of milk performance traits. J. Anim. Breed. Genet. 132: 3-8.
669 Wu, X., Lund, M.S., Sun, D., Zhang, Q., and Su, G., 2015. Impact of relationships between
670 test and training animals and among training animals on reliability of genomic
671 prediction. J. Anim. Breed. Genet. 132: 366-375.
672

673 **Table 1**

674 Description of datasets used for the two random regression models and the standard

675 cumulative 305-d model

Item	n	Mean	SD
Random Regression Models			
Cows	7,206		
Milk yield, kg	69,029	13.8	4.9
Fat percentage, %	29,878	3.5	0.9
Standard cumulative 305-d model			
Cows	7,206		
305-d Milk yield, kg	7,206	4,243	1,112
305-d Fat percentage, %	3,264	3.5	0.7

676

677

Table 2

Additive genetic, permanent environmental, phenotypic variances and covariances for 305-d milk yield (305-d MY) and 305-d fat percentage (305-d FP) estimated using two random regression models

Variance component	Model[a]	
	RRGM	RRPM
Additive genetic		
Var (305-d MY), kg^2	279,893.2	217,247.9
Cov (305-d MY, 305-d FP), kg%	-41.3	-24.9
Var (305-d FP), $\%^2$	0.10	0.08
Permanent environment		
Var (305-d MY), kg^2	556,455.4	612,728.6
Cov (305-d MY, 305-d FP), kg%	-72.9	-90.4
Var (305-d FP), $\%^2$	0.10	0.13
Phenotypic		
Var (305-d MY), kg^2	1,023,747.6	1,017,384.8
Cov (305-d MY, 305-d FP), kg%	-114.1	-115.2
Var (305-d FP), $\%^2$	0.66	0.66

[a] RRGM = Random regression genomic-polygenic model; RRPM = Random regression polygenic model

Table 3

Heritabilities and correlations for 305-d milk yield (305-d MY) and 305-d fat percentage (305-d FP) computed using two random regression models

Parameter	Model[a]	
	RRGM	RRPM
Heritability (305-d MY)	0.27	0.21
Heritability (305-d FP)	0.16	0.12
Additive genetic correlation (305-d MY, 305-d FP)	-0.24	-0.19
Permanent environmental correlation (305-d MY, 305-d FP)	-0.31	-0.32
Phenotypic correlation (305-d MY, 305-d FP)	-0.14	-0.14

[a] RRGM = Random regression genomic-polygenic model; RRPM = Random regression polygenic model

Table 4

Rank correlations between animal EBV for 305-d milk yield (305-d MY) and 305-d fat percentage (305-d FP) evaluated using two random regression models and a standard cumulative 305-d model

Trait	Rank correlations[a]		
	RRGM, RRPM	RRGM, SCGM	RRPM, SCGM
305-d MY	0.94	0.66	0.61
305-d FP	0.78	0.57	0.45

[a] RRGM = Random regression genomic-polygenic model; RRPM = Random regression polygenic model; SCGM = Standard cumulative 305-d genomic-polygenic model; All rank correlations were significant at $P < 0.0001$.

Table 5

Rank correlations between sire EBV for 305-d milk yield (305-d MY) and 305-d fat percentage (305-d FP) evaluated using two random regression models and a standard cumulative 305-d model

Trait	Sires[a]	Rank correlations[b]		
		RRGM, RRPM	RRGM, SCGM	RRPM, SCGM
305-d MY	top 5% (52)	0.78 (83)	0.50 (58)	0.50 (54)
	top 15% (155)	0.82 (86)	0.62 (59)	0.56 (58)
	top 25% (259)	0.88 (88)	0.63 (61)	0.64 (59)
	100%	0.94	0.69	0.65
305-d FP	top 5% (52)	0.57 (65)	0.42 (46)	0.38 (44)
	top 15% (155)	0.66 (76)	0.46 (59)	0.40 (55)
	top 25% (259)	0.74 (75)	0.48 (60)	0.52 (53)
	100%	0.82	0.58	0.47

[a] Numbers in brackets are numbers of sires in the top 5%, 15%, and 25%.

[b] RRGM = Random regression genomic-polygenic model; RRPM = Random regression polygenic model; SCGM = Standard cumulative 305-d genomic-polygenic model. All rank correlations were significant at $P < 0.0001$, except for top 5% between sire EBV for 305-d MY and 305-d FP that were significant at $P < 0.005$. Numbers in brackets are percentages of sires in common in the top 5%, 15%, and 25% of sires ranked by each pair of models.

Table 6

Rank correlations between cow EBV for 305-d milk yield (305-d MY) and 305-d fat percentage (305-d FP) evaluated using two random regression models and a standard cumulative 305-d model

Trait	Cows[a]	Rank correlations[b]		
		RRGM, RRPM	RRGM, SCGM	RRPM, SCGM
305-d MY	top 5% (624)	0.81 (81)	0.45 (52)	0.40 (46)
	top 15% (1,873)	0.82 (84)	0.50 (58)	0.41 (54)
	top 25% (3,121)	0.83 (86)	0.52 (63)	0.45 (60)
	100%	0.94	0.66	0.61
305-d FP	top 5% (624)	0.62 (64)	0.43 (45)	0.39 (40)
	top 15% (1,873)	0.67 (66)	0.46 (52)	0.39 (48)
	top 25% (3,121)	0.68 (70)	0.44 (57)	0.38 (52)
	100%	0.77	0.57	0.45

[a] Numbers in brackets are numbers of cows in the top 5%, 15%, and 25%.

[b] RRGM = Random regression genomic-polygenic model; RRPM = Random regression polygenic model; SCGM = Standard cumulative 305-d genomic-polygenic model. All rank correlations were significant at $P < 0.0001$. Numbers in brackets are percentages of cows in common in the top 5%, 15%, and 25% of cows ranked by each pair of models.

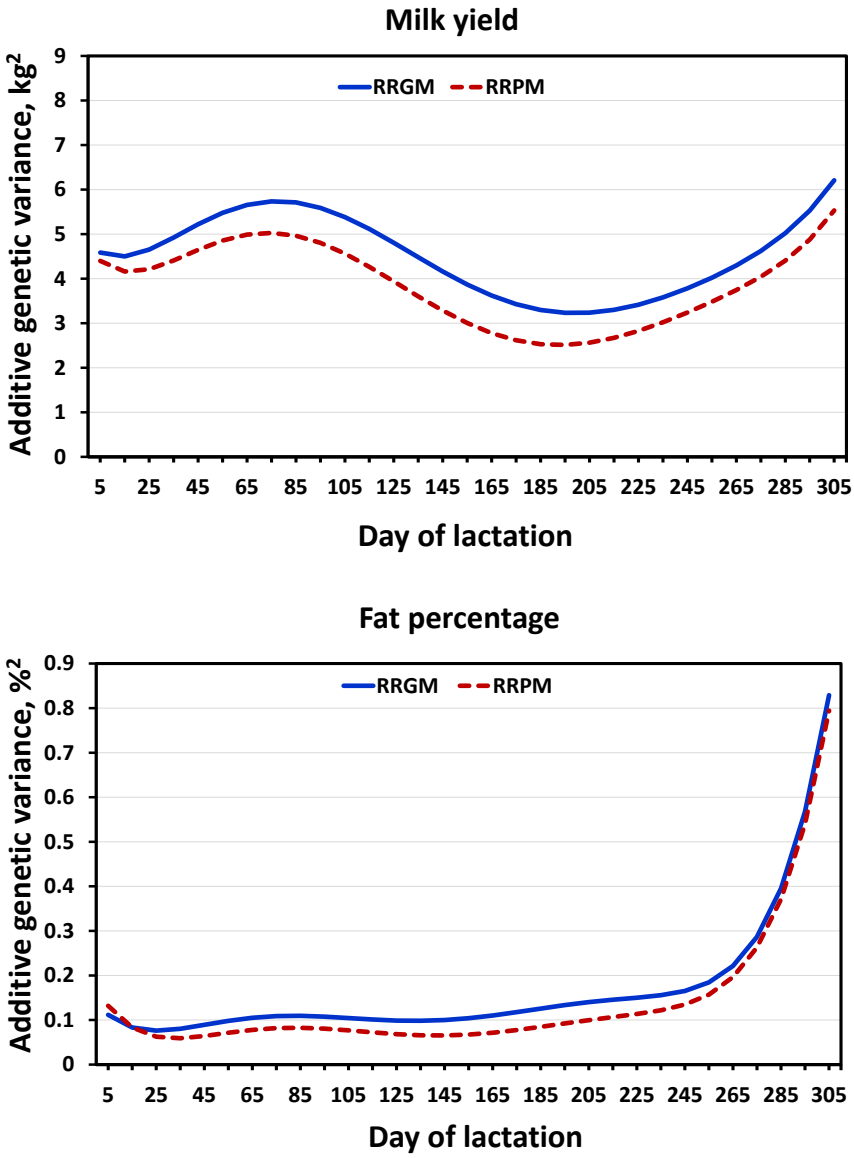

Fig. 1. Additive genetic variances for milk yield and fat percentage estimated using random regression genomic-polygenic (RRGM) and polygenic (RRPM) model

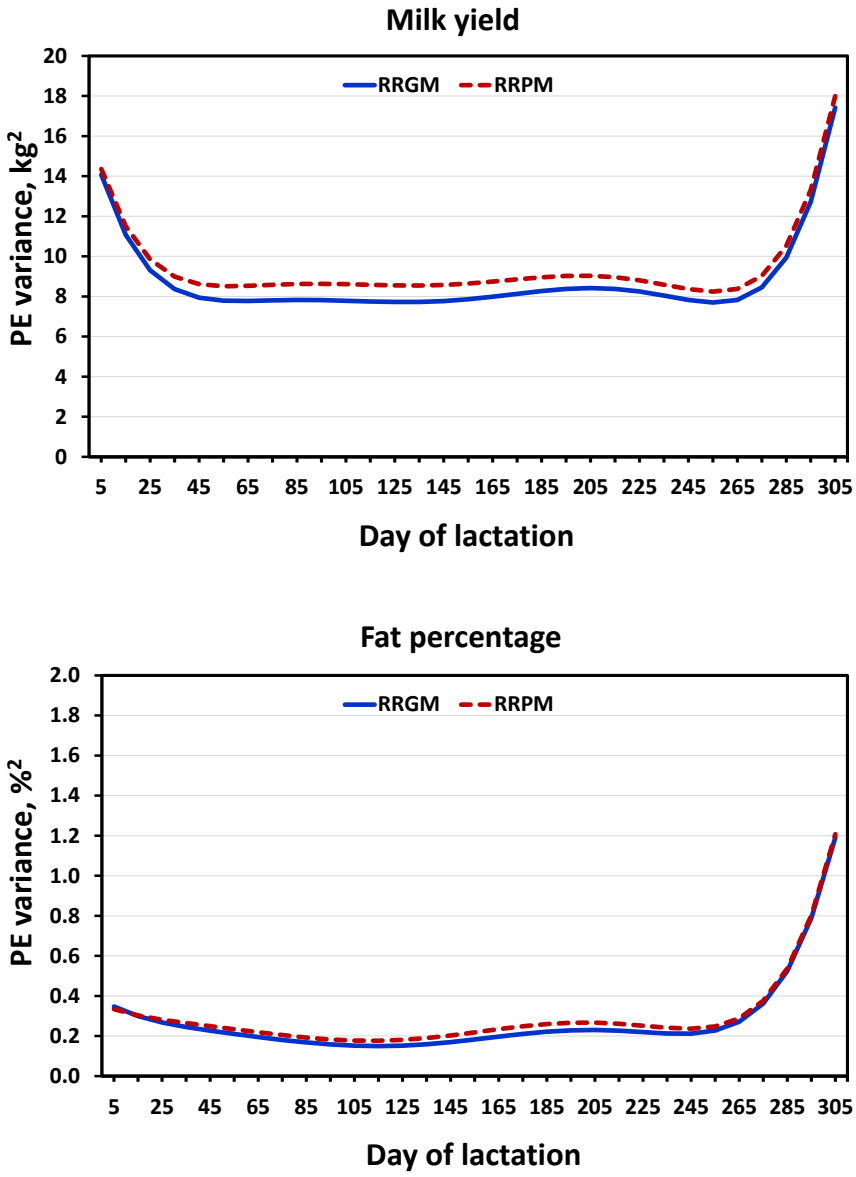

Fig. 2. Permanent environmental (PE) variances for milk yield and fat percentage estimated using random regression genomic-polygenic (RRGM) and polygenic (RRPM) models

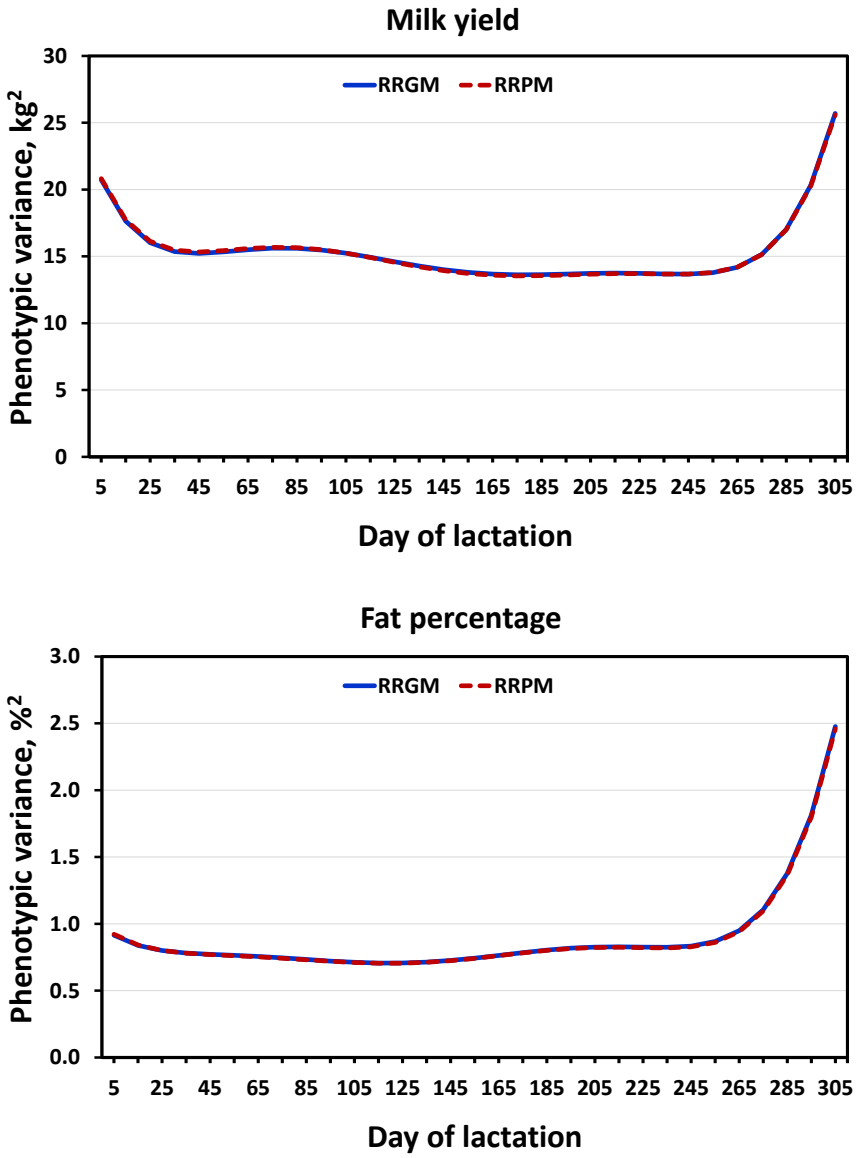

Fig. 3. Phenotypic variances for milk yield and fat percentage estimated using random regression genomic-polygenic (RRGM) and polygenic (RRPM) models

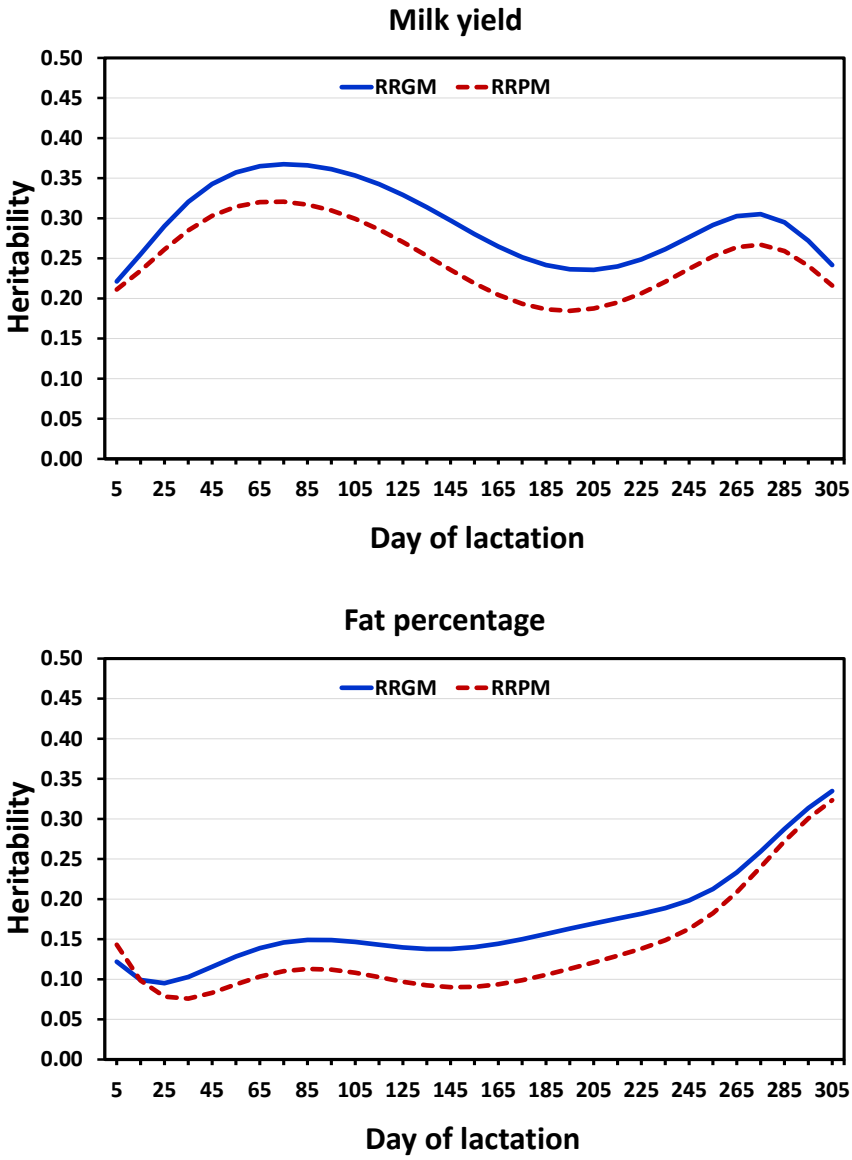

Fig. 4. Heritabilities for milk yield and fat percentage estimated using random regression genomic-polygenic (RRGM) and polygenic (RRPM) models

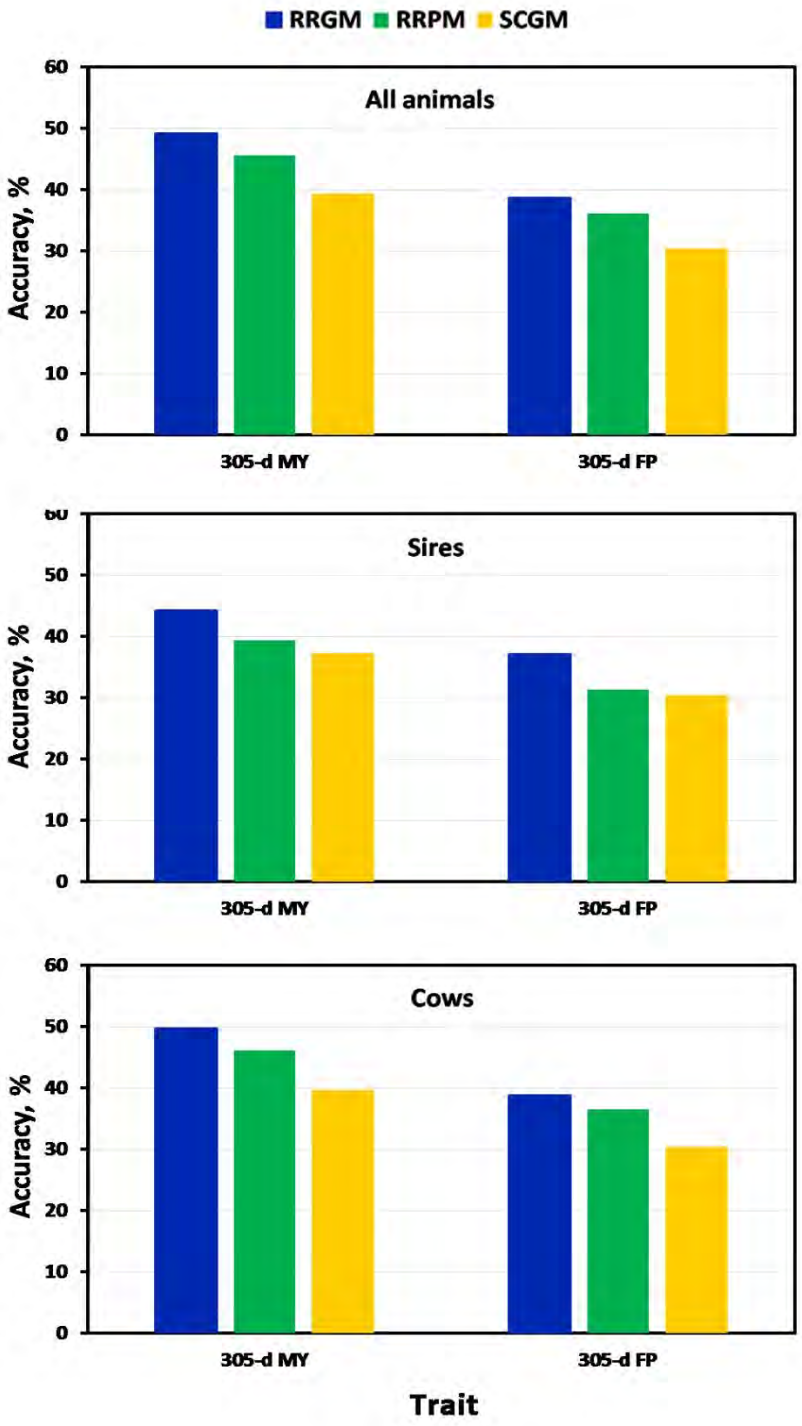

Fig. 5. Accuracy of estimated breeding values for 305-d milk yield (305-d MY) and 305-d fat percentage (305-d FP) in a Holstein-upgraded dairy cattle population using random regression genomic-polygenic (RRGM), random regression polygenic (RRPM), and standard cumulative 305-d genomic-polygenic (SCGM) models