the **genetics**society

ARTICLE

Genotype-by-environment interaction of fertility traits in Danish Holstein cattle using a single-step genomic reaction norm model

Zhe Zhang[1,2] · Morten Kargo[1,3] · Aoxing Liu[1,4] · Jørn Rind Thomasen[1,5] · Yuchun Pan[2] · Guosheng Su[1]

Abstract

Genotype-by-environment (G × E) interactions could play an important role in cattle populations, and it should be considered in breeding programmes to select the best sires for different environments. The objectives of this study were to study G × E interactions for female fertility traits in the Danish Holstein dairy cattle population using a reaction norm model (RNM), and to detect the particular genomic regions contributing to the performance of these traits and the G × E interactions. In total 4534 bulls were genotyped by an Illumina BovineSNP50 BeadChip. An RNM with a pedigree-based relationship matrix and a pedigree-genomic combined relationship matrix was used to explore the existence of G × E interactions. In the RNM, the environmental gradient (EG) was defined as herd effect. Further, the genomic regions affecting interval from calving to first insemination (ICF) and interval from first to last insemination (IFL) were detected using single-step genome-wide association study (ssGWAS). The genetic correlations between extreme EGs indicated that G × E interactions were sizable for ICF and IFL. The genomic RNM (pedigree-genomic combined relationship matrix) had higher prediction accuracy than the conventional RNM (pedigree-based relationship matrix). The top genomic regions affecting the slope of the reaction norm included immunity-related genes (*IL17*, *IL17F* and *LIF*), and growth-related genes (*MC4R* and *LEP*), while the top regions influencing the intercept of the reaction norm included fertility-related genes such as *EREG*, *AREG* and *SMAD4*. In conclusion, our findings validated the G × E interactions for fertility traits across different herds and were helpful in understanding the genetic background of G × E interactions for these traits.

✉ Guosheng Su
  guosheng.su@mbg.au.dk

[1] Department of Molecular Biology and Genetics, Center for Quantitative Genetics and Genomics, Aarhus University, Blichers Allé 20, Postboks 50, Tjele, DK 8830, Denmark

[2] Department of Animal Science, School of Agriculture and Biology, Shanghai Jiao Tong University, 200240 Shanghai, China

[3] Knowledge Center for Agriculture (SEGES), Skejby, Aarhus N 8200, Denmark

[4] College of Animal Science and Technology, China Agricultural University, Beijing 100193, China

[5] VikingGenetics, Asesentoft 8960, Denmark

Introduction

In the dairy cattle industry, fertility traits are some of the most influential components, as declining fertility prolongs the resume cycles after calving and increases veterinary costs (De Vries 2006; Schneider et al. 2005). Some previous studies have emphasized the importance of genetic evaluation and the improvement of fertility traits in spite of their low heritability (below 5%) (Liu et al. 2017; Sun et al. 2010), and more balanced selection indices including production, longevity, health and fertility have been used instead of indices which simply focused on yield. Another issue regarding selection for improved fertility is that a wide range of environments often contribute to the phenomenon of genotype-by-environment (G × E) interaction, which is defined as different performances of animals and their offspring in different environments than those where they were raised or selected (Falconer et al. 1996).

In the Danish milk production system, different herds have different conditions in terms of feed, pharmaceuticals and housing, which may further lead to G × E interactions

(Strandberg et al. 2000), and can therefore be defined as an environmental gradient (EG) to study the G × E interactions. Two models are widely used to detect G × E interactions. One is a multi-trait model, which assumes that phenotypic expressions of a trait in various environments are different traits (Kolmodin et al. 2002), and a low genetic correlation among these different traits indicates the existence of a G × E interaction. The other one is the reaction norm model (RNM) (Falconer et al. 1996), which models the trajectory of animal performance as a function of the EG, and the breeding value of an animal is therefore partitioned into an environment-independent part (intercept) and an environment-dependent (slope) part. Compared with multi-trait models, RNMs are able to explore G × E interactions in a range of continuous environments and quantify the G × E interaction at any environment within the range.

The application of an RNM often uses the known EG as a covariate. Traditionally, the covariate was postulated as the phenotypic mean in different environments, which could lead to bias due to misleading representation of EG. An alternative method which infers the unknown EG using the Bayesian method, was suggested, and more accurate estimates could be obtained (Su et al. 2006, 2009). Although the Bayesian method is able to handle the unknown EG, the analysis is time consuming, and sometimes it is difficult to get model convergence when the model becomes complicated and too many parameters need to be estimated at the same time. Calus et al. (2002) and Kolmodin et al. (2002) proposed to first estimate the EG using a conventional linear mixed model and substitute it into the RNM to study the G × E interaction, which can be viewed as partitioning the simultaneous estimation of the EG and other parameters into two steps.

Furthermore, the inclusion of genotype information in the RNM may be beneficial for the estimation of parameters and prediction of breeding values. To integrate information from genotyped and non-genotyped animals, a single-step genomic BLUP (ssGBLUP) (Christensen and Lund 2010; Misztal et al. 2009) can be used in complicated models such as random regression or RNMs. In addition, Wang et al. 2012 proposed a new method to perform genome-wide association study (GWAS) based on ssGBLUP, which is called a single-step GWAS (ssGWAS). A simulation study has shown that this method results in more accurate estimates of SNP effects compared to BayesB and the conventional GWAS methods with single-marker regression.

The objective of this study was to use the RNM with information on genomic markers and pedigrees to explore the G × E interaction of female fertility for Danish Holstein dairy cows and to map the genomic regions contributing to the fertility phenotypes across different EGs.

Materials and methods

Data

The fertility traits of Danish Holstein used in this study included number of inseminations per conception (AIS), interval from first to last insemination (IFL), non-return rate at 56 days after first insemination (NRR) and interval from calving to first insemination (ICF). Field records during the period of insemination from 2011 to 2016 were collected from 1775 herds. The records in the first three parities were used in this study.

In addition, AIS and IFL included some censored records because of no known date of confirmed successful insemination. In this study, penalty values were used to handle censored data (Liu et al, 2017; Sun et al, 2009). For censored IFL, a penalty of 40 days was added, which is the average value of Danish Holstein cows, and a penalty of 1 was added to censored AIS. The further criteria for the data editing were in accordance with Nordic routine genetic evaluation. Briefly, the criteria included age at first insemination between 270 and 900 days, age at first calving between 500 and 1100 days, AIS between 1 and 5, IFL between 0 and 365 days, ICF between 20 and 230 days, and days open (ICF + IFL) between 20 and 365 days. Records with values below the lower limit were removed, whereas records were converted to upper-limit values if they were greater than the upper limit. Herds were defined as the herds where the animals got phenotypes. Animals that moved to other herds between the first insemination and the last insemination were removed. In addition, for each trait, the herd-year levels with <20 records were merged with the nearest herd-year level, and this combination was performed once more if the number of records was still <20. Finally, the records in herd-year levels with <20 records were discarded after two combination runs. The detailed data editing procedure can be found online (https://www.nordicebv.info/).

The number of records and animals for each trait obtained after editing are listed in Table 1. The pedigree of the animals with phenotypes was traced three generations back using the Nordic Cattle database (NAV, Skejby, Denmark).

Genotypes were available for 4534 bulls in the pedigree, among which 1777 bulls have daughters with records in the dataset. The bulls were genotyped using an Illumina BovineSNP50 BeadChip (Illumina Inc., San Diego, CA, USA) version 2 (containing 54,609 single nucleotide polymorphisms (SNPs)). The SNP data were filtered by removing markers with a minor allele frequency <1%, an average GenCall score lower than 0.60, or an unknown location in the UMD 3.1 assembly. After data editing, 46,342 SNPs were included in the analysis.

Table 1 Descriptive statistics for female fertility traits and heritabilities (standard errors) estimated using a pedigree-based linear mixed model

Trait	No. of records	No. of animals	Mean	SD	Min	Max	Heritability
AIS	705,835	383,179	2.06	1.24	1	5	0.027 (0.001)
ICF	723,333	391,930	74.4	31.1	20	180	0.057 (0.003)
IFL	728,426	393,561	47.7	61.3	0	230	0.032 (0.002)
NRR	715,911	386,869	0.542	0.498	0	1	0.011 (0.001)

AIS number of inseminations per conception, *ICF* interval from calving to first insemination, *IFL* interval from first to last insemination, *NRR* non-return rate at 56 days after first insemination, *SD* standard deviation, *Min* minimum value, *Max* maximum value

Model

The EG defined in this study included the production environments of herds, i.e., herd effects. We used a three-step strategy to analyze the G × E interaction for different traits. This three-step approach is an extension of the two-step RNM proposed by Calus et al. (2002) and Kolmodin et al. (2002). Briefly, the first two steps were designed to estimate herd effects and weights of heterogeneous residual variances in RNM of the third step, respectively, so that the RNM was simplified. Compared with the one-step method that estimates an unknown EG and other parameters simultaneously using a Bayesian method (Su et al. 2006), the three-step approach has much less computational demand.

Step 1

Initially, a univariate animal model was used to fit observations for each trait. The model was as follows:

$$y = Xb + Ff + Za + Ec + Wpe + e \qquad (1)$$

where y is the vector of observations; b is the vector of fixed effects other than herd effect; f is the vector of fixed herd effects; a is the vector of additive genetic effects, which are assumed to follow the normal distribution $N(0, A\sigma_a^2)$ for a pedigree-based BLUP or $N(0, H\sigma_a^2)$ for an ssGBLUP model, where σ_a^2 is the additive genetic variance and A is the numerator (pedigree) relationship matrix and H is the combined pedigree-genomic relationship matrix; c is the vector of herd-year effects, which are assumed to be normally distributed with $N(0, I\sigma_c^2)$, where σ_c^2 is the variance of herd-year effects and I is the identity matrix; pe is the vector of permanent environmental effects, which are assumed to be normally distributed with $N(0, I\sigma_{pe}^2)$, where σ_{pe}^2 is the variance of permanent environmental effects; e is the random residual vector following $N(0, I\sigma_e^2)$ and σ_e^2 is the random residual variance. X, F, Z, E and W and are incidence matrices related to the fixed and random effects in the model. The fixed effects included herd, parity, year of first insemination (for AIS, IFL and NRR) or year of calving (for ICF), month of first insemination (for AIS, IFL and NRR) or month of calving (for ICF) and age at first

insemination (divided into five age groups of 270–419 days, 420–444 days, 445–469 days, 470–500 days and 500–900 days). The random effects included additive genetic effects, herd-year effects (year of first insemination for AIS, IFL and NRR; year of calving for ICF), and permanent effects.

For the pedigree-based BLUP, the numerator relationship matrix (A) was built based only on the pedigree, whereas for the ssGBLUP model, a combined pedigree-genomic relationship matrix (H) was used, and the inverse H^{-1} was calculated as follows (Aguilar et al. 2010; Christensen and Lund 2010):

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G^{-1} - A_{22} \end{bmatrix} \qquad (2)$$

where A_{22} is the subset of the numerator relationship matrix for genotyped animals. G is the blended genomic relationship matrix. G was constructed as $(1-\omega)G_0 + \omega A_{22}$, where ω is the relative weight that could explain the fraction of the genetic variance not captured by markers and was set as 0.2 according to (Gao et al. 2012); G_0 (VanRaden 2008) was constructed using Gmatrix software (Su and Madsen 2011):

$$G_0 = ZDZ' / \sum_{i=1}^{m} 2p_i(1 - p_i) \qquad (3)$$

where the elements in column i of Z are $0 - 2p_i$, $1 - 2p_i$, and $2 - 2p_i$ for genotypes A_1A_1, A_1A_2 and A_2A_2, respectively, where p_i is the allele frequency of A_2 at locus i calculated from the current data; D is a diagonal weight matrix for each SNP, and in this step, D is an identity matrix; m is the number of SNPs. At last, G was adjusted to be compatible with A_{22} according to Christensen et al. 2012. The analysis was performed using the DMUAI module implemented in the DMU package (Madsen et al. 2013).

Step 2

In an RNM, heterogeneous residual variances are usually fitted when residual variances differ among production environments (Oliveira et al. 2018; Silva et al. 2014). To account for various residual variances, the herds were divided into five groups according to the quantiles of the

herd effects estimated in step 1, namely, (0, 10%), (10, 30%), (30, 70%), (70, 90%), and (90, 100%). The setting of five groups was based on a balance between number of observations in each group and the possible difference in residual variance among different herd levels. Then, the model in step 1 was re-run, and the variance components were estimated by a Gibbs sampling approach using the RJMC module in the DMU package (Madsen et al. 2013).

Step 3

Compared with conventional linear mixed model, the RNM included genetic sensitivities represented by random regressions on the herd effects estimated in the step 1 and was analyzed using DMUAI. It was shown as follows:

$$y = Xb + Ff + Z_0 a_0 + Z_1 a_1 + Ec + Wpe + e \qquad (4)$$

where it is assumed that $\begin{bmatrix} a_0 \\ a_1 \end{bmatrix} \sim N$ $\left(0, A \otimes \begin{bmatrix} \sigma_{a_0}^2 & \sigma_{a_0 a_1} \\ \sigma_{a_0 a_1} & \sigma_{a_1}^2 \end{bmatrix} \right)$. Z_0 is an incidence matrix connecting a_0 (intercept) to y, and Z_1 is an incidence matrix containing herd effects estimated in step 1 as covariables to connect y and a_1 (slope). Other effects and incidence matrices are the same as those for linear mixed models in step 1. The random residual vector is assumed to follow $N\left(0, R\sigma_e^2\right)$, where R is a diagonal matrix with elements for observations in the ith group of herds equal to $1/w_i$, and

$$w_i = \frac{\sum_{i=1}^{5} n_i \sigma_{e_i}^2}{\sigma_{e_i}^2 \sum_{i=1}^{5} n_i} \qquad (5)$$

where n_i is the number of observations in the ith group of herds and $\sigma_{e_i}^2$ is the residual variance of the ith group estimated by RJMC. In addition, to evaluate the need to include the G × E interaction in the model, a likelihood ratio test (LRT) between the RNM and the reduced model (RM) of (4) without considering random regression on herd effects was performed based on the statistic of $D = -2 * \log$ (likelihood) for the RM, $+ 2 * \log$(likelihood) for the RNM. The P-value for the LRT was calculated as $0.5P\left[\chi_{1\,d.f.}^2 \geq D\right] + 0.5P\left[\chi_{2\,d.f.}^2 \geq D\right]$.

The heritability $\left(h_a^2\right)$ of the linear mixed model was calculated as $h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^2}$. The additive genetic variances $\left(\sigma_a^2\right)$ and heritability $\left(h_a^2\right)$ of the RNM with a particular herd effect (f) were calculated as $\sigma_a^2 = \sigma_{a_0}^2 + \sigma_{a_1}^2 f^2 + 2\sigma_{a_0 a_1} f$. The residual variance $\left(\sigma_e^{*2}\right)$ of the RNM in a particular herd group was calculated as $\sigma_e^{*2} = \frac{1}{w}\sigma_e^2$. Consequently, the

heritability $\left(h_a^2\right)$ of the RNM was calculated as $h_a^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_{pe}^2 + \sigma_e^{*2}}$. Standard errors of heritabilities and genetic correlations were calculated according to an expansion of the Taylor series (Su et al. 2007).

Prediction of future performances

To evaluate and compare the prediction ability of different models, the data were divided into training and validation parts according to the birthdates of the bulls. The 20% youngest bulls (born after 01/07/2010, 355 bulls) with daughter information were regarded as the validation dataset, and the other 80% were regarded as the training dataset. In the RNM, the estimated breeding value of a particular bull obtained from the pedigree-based BLUP (EBV) or from the ssGBLUP (GEBV) was defined as

$$\widehat{a} = \widehat{a}_0 + \overline{f}\widehat{a}_1 \qquad (6)$$

where $\widehat{a}$ is the predicted breeding value for this bull; $\widehat{a}_0$ and $\widehat{a}_1$ are estimates of the intercept and slope for this bull, respectively; $\overline{f}$ is the mean effect of herds where daughters of this bull were raised. Daughter yield deviation (DYD) was used to evaluate the accuracies of the prediction. The DYD was the average performance of daughters adjusted for all fixed and non-genetic random effects as well as the dam's additive genetic effect. The accuracy of the prediction was calculated as a Pearson correlation in the form of cor(DYD, prediction). In addition, the accuracy of the prediction obtained from the RNM was compared to that calculated from the RM.

ssGWAS for ICF and IFL

ssGWAS as proposed by Wang et al. 2012 was applied to locate the genomic regions related to the intercept and slope in the RNM for ICF and IFL, as it was observed that there existed significant G × E interactions across extreme EGs, and slightly higher prediction accuracies for the RNM compared to the RM for the two traits. In a typical ssGBLUP model, it is assumed that all SNP effects have the same variance. The ssGWAS allows heterogeneous variance among different SNPs, which is realized by a weighted G-matrix, and the weights can be obtained by an iteration procedure. Thus, SNP effects and weights for ssGWAS were calculated as follows:

1. Calculate the G matrix with $D_{ii} = 1$ using (3).

2. Calculate the GEBV for all the genotyped animals using the reaction norm based ssGBLUP.

3. Calculate the SNP effects $(\widehat{u})$ based on the GEBV with $\widehat{u} = DZ'(ZDZ')^{-1}\widehat{a}$, where $\widehat{a}$ is the vector of the GEBV with elements calculated using formula (6).

Table 2 Variance of the intercept ($\sigma^2_{a_0}$), variance of the slope ($\sigma^2_{a_1}$), covariance between the intercept and slope ($\sigma_{a_0 a_1}$), permanent environmental variance (σ^2_{pe}), variance of herd-year effects (σ^2_{hy}), residual variance (σ^2_{hy}) and correlation between the intercept and slope ($r_{a_0 a_1}$), with their standard errors in parentheses estimated using an RNM

Trait	A|H	$\sigma^2_{a_0}$	$\sigma^2_{a_1}$	$\sigma_{a_0 a_1}$	σ^2_{pe}	σ^2_{hy}	σ^2_e	$r_{a_0 a_1}$
AIS	**A**	0.083 (0.025)	0.046 (0.009)	−0.054 (0.015)	0.052 (0.003)	0.052 (0.003)	1.40 (0.003)	−0.880 (0.030)
	H	0.076 (0.024)	0.044 (0.009)	−0.050 (0.014)	0.052 (0.003)	0.010 (0.0005)	1.40 (0.003)	−0.871 (0.033)
ICF	**A**	92.0 (15.8)	0.015 (0.003)	−0.849 (0.193)	53.3 (2.05)	38.7 (0.930)	728.1 (1.70)	−0.724 (0.047)
	H	95.5 (15.9)	0.016 (0.003)	−0.883 (0.195)	52.1 (2.10)	38.7 (0.929)	728.2 (1.70)	−0.725 (0.046)
IFL	**A**	48.0 (22.0)	0.052 (0.011)	−0.621 (0.462)	168.2 (7.76)	18.3 (1.03)	3383.1 (8.03)	−0.395 (0.178)
	H	44.7 (21.3)	0.050 (0.011)	−0.535 (0.446)	166.6 (7.85)	18.3 (1.03)	3383.9 (8.03)	−0.360 (0.192)
NRR	**A**	0.006 (0.002)	0.009 (0.005)	−0.005 (0.003)	0.004 (0.0004)	0.001 (0.00007)	0.235 (0.0005)	−0.709 (0.116)
	H	0.005 (0.0017)	0.009 (0.005)	−0.005 (0.003)	0.004 (0.0004)	0.001 (0.00007)	0.235 (0.0005)	−0.690 (0.122)

AIS number of inseminations per conception, *ICF* interval from calving to first insemination, *IFL* interval from first to last insemination, *NRR* non-return rate at 56 days after first insemination, *A* and *H* pedigree-based and pedigree-genomic combined matrices, respectively

4. Calculate the weight (D_{ii}) for each SNP as $D_{ii} = \hat{u}_i^2 2p_i(1 - p_i)$, where p_i is the allele frequency for the ith SNP.

5. Calculate D by normalizing the SNP weights to keep the total genetic variance constant.

6. Repeat steps 2–5.

Considering that too many iterations may cause some subjective peaks, the ssGWAS was run for three iterations (Wu et al. 2018), and in the last iteration, the SNP effects for the intercept ($\hat{u}_0$) and slope ($\hat{u}_1$) were calculated as $\hat{u}_s = DZ'(ZDZ')^{-1}\hat{a}_s$, where s equals 0 for the intercept and 1 for the slope. The percentage of genetic variances of the slope or intercept explained by a discrete window of 20 adjacent SNPs was defined as $\frac{Var(a_{st})}{\sigma^2_{a_s}} \times 100\%$, where a_{st} is the genetic value for the intercept ($s = 0$) or slope ($s = 1$) of the tth window and can be calculated as $\sum_{k=1}^{20} Z_k \hat{u}_{sk}$, where Z_k is the matrix of gene content of the kth SNP for genotyped individuals, and $\hat{u}_{sk}$ is the SNP effect of the kth SNP with the tth window for the intercept or slope; $\sigma^2_{a_s}$ is the total additive genetic variance for the intercept or slope. In addition, the remaining SNPs (<20) at the end of the chromosome were merged into the last window.

Candidate genes within the top 10 windows with the highest percentages of genetic variances of the intercept or slope were retrieved by the *BioMart* package (Durinck et al. 2005) based on UMD 3.1 assembly.

Results

Descriptive statistics

The summary statistics for each trait are listed in Table 1. A large amount of phenotypic variation was observed for each trait. Table 1 also lists the heritability of each trait estimated by the conventional linear mixed model based on pedigree information. As expected, the heritability estimates for these fertility traits were quite low, which is consistent with an earlier study (Liu et al. 2017).

### Variance and covariance components estimated by RNM

The estimates of variance components and their standard errors obtained by the RNMs based on different relationship matrices (*A* and *H*) were very similar (Table 2). Table 2 also lists the correlation coefficients between the intercept and slope for each trait. All of these correlation coefficients were negative and ranged from −0.880 to −0.360.

The heritabilities among different herd levels (in the range of mean ± 2.5 × standard deviation of herd effects) estimated by the RNM based on the *A* matrix and those from the RNM based on the *H* matrix were similar (Fig. 1). The highest heritabilities for these traits were generally observed in herds with the largest effects. The least amount of variation in heritibilities among herd levels was observed for NRR, but with large standard errors. Figure 1 also shows the discrete lines of heritabilities along the continuous EGs, which were caused by the heterogeneous residual variances assumed for five groups of herds in the model. Generally, the larger the herd effects were, the larger the residual variance was and thus the smaller weights assigned to the residual variance were, although this was not the case for NRR and ICF (Table S1).

Environmental sensitivity

The genetic sensitivity to different environments for each trait was indicated by the variance of the slope (Table 2), which was statistically (P-values < 0.05) different from zero for all traits based on one-tailed t-tests. The genetic correlations between a herd at a particular quantile of herd effects (in the range of mean ± 2.5 × standard deviation of herd effects) and herd level at 5% (lowest), 55% (intermediate)
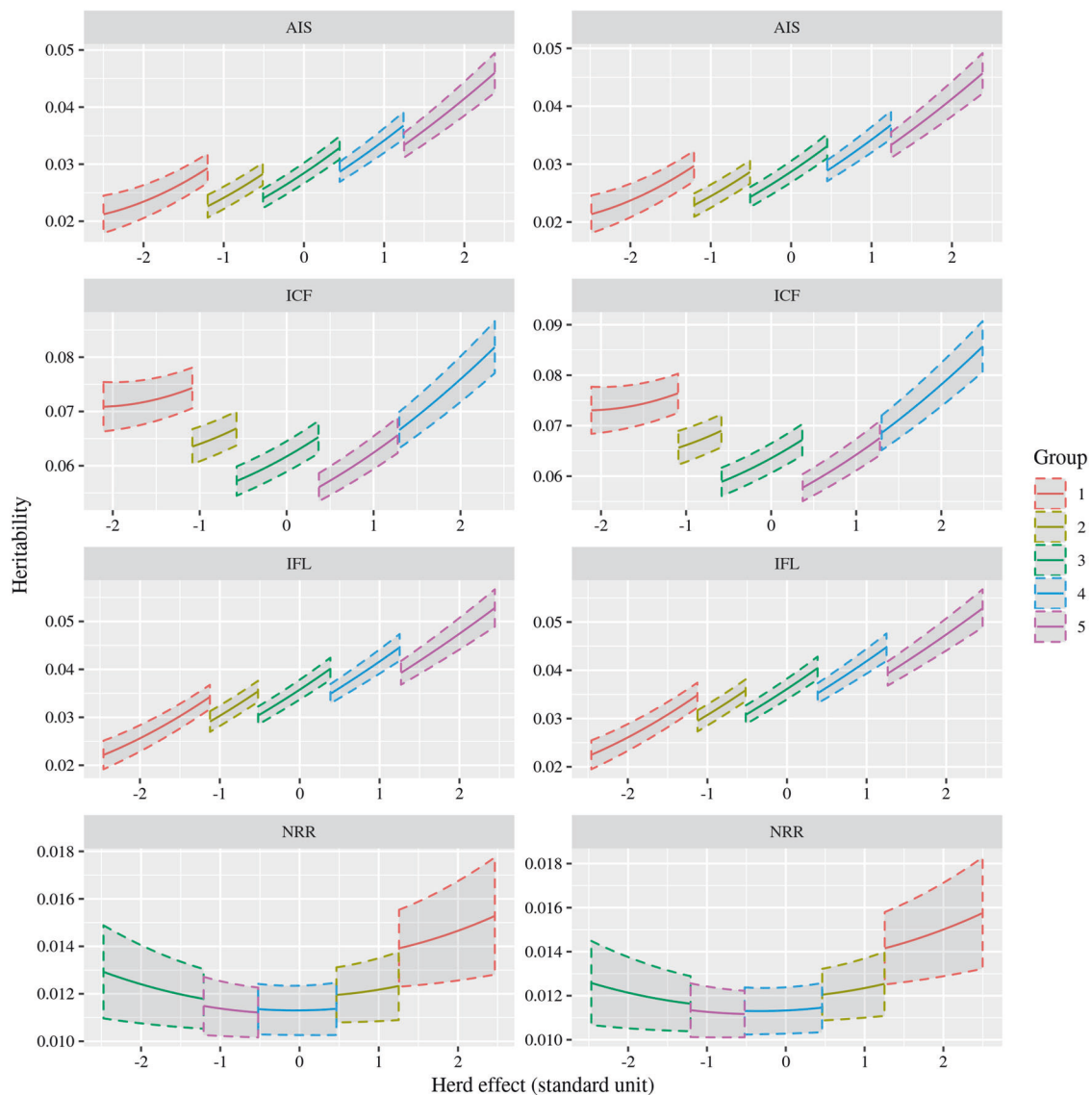
Fig. 1 Heritabilities estimated using an RNM with the **A** relationship matrix (left) and **H** relationship matrix (right) for female fertility traits: AIS (number of inseminations per conception), ICF (interval from calving to first insemination), IFL (interval from first to last insemination), and NRR (non-return rate at 56 days after first insemination). The x-axis represents the normalized herd effects with a range of mean ± 2.5 standard units. The lines with different colours show the heritabilities within different groups of heterogeneous residual variance along different EGs, and the shades represent standard errors

or 95% (highest) quantiles are shown in Fig. 2. The differences of patterns between **A** and **H** are very small for ICF and NRR, while the genetic correlations estimated by **A** matrix were slightly lower than those estimated by **H** matrix for AIS and IFL (Fig. 2). The values of genetic correlations between extreme EGs are listed in Table 3. It was observed that genetic correlations between quantiles of (1 and 99%), (5 and 95%), (10 and 90%) and (20 and 80%) of herd levels were significantly different from unity, mainly for ICF and IFL. Table 4 lists the number of overlapping animals of the top 50 sires between extreme EGs. The more extreme

between these EGs were, the lower the overlap of the top sires was, which indicates that the re-ranking of sires can happen across extreme EGs, especially for ICF.

Accuracy of predicted breeding values

Table S2 lists the LRT statistics for the RM and RNM, which were all statistically significant, indicating the reasonability of the RNM. The accuracies of predicted breeding values for bulls in the validation dataset obtained by different models are listed in Table 5. The highest accuracy
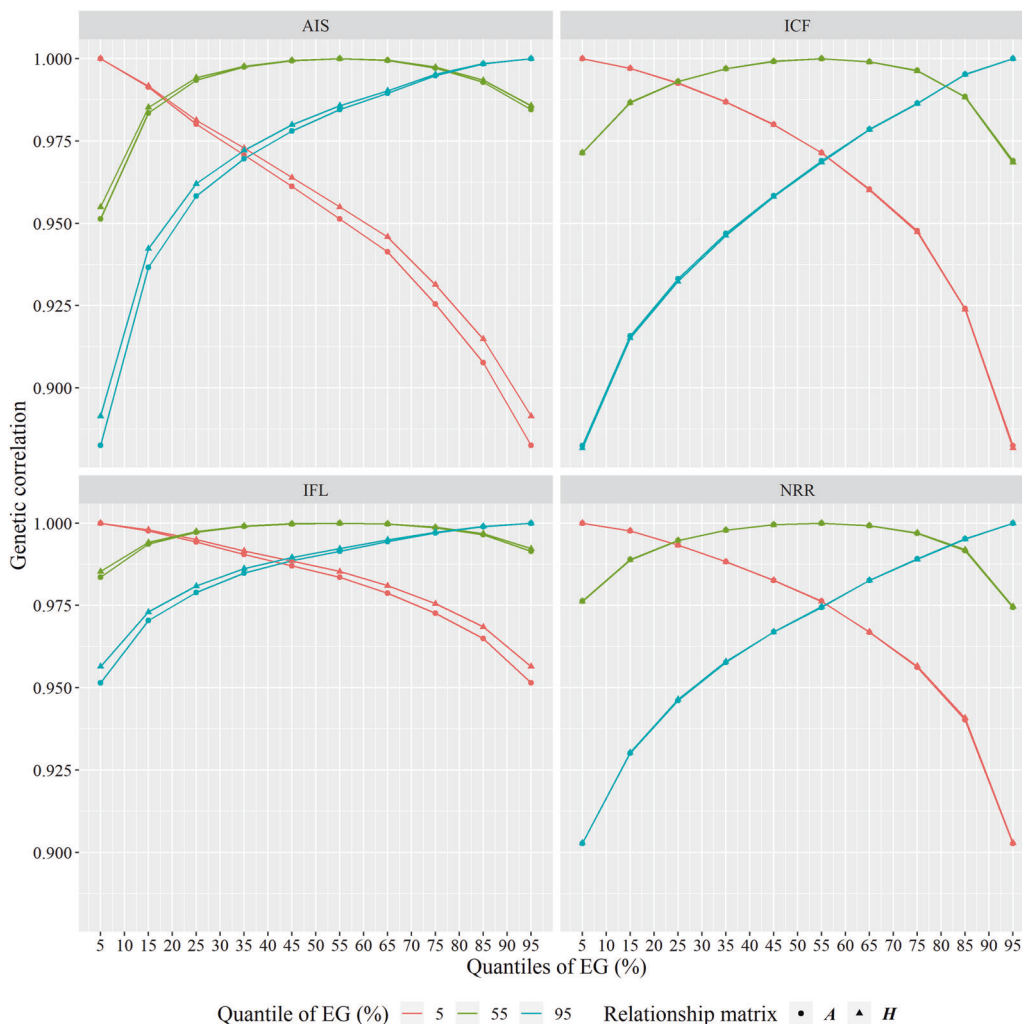
Fig. 2 The genetic correlations between performances in a herd at a particular quantile of environmental gradients (EGs) and a herd at 5, 55 or 95% quantile of EGs for AIS (number of inseminations per conception), ICF (interval from calving to first insemination), IFL (interval from first to last insemination), and NRR (non-return rate at 56 days after first insemination) obtained from an RNM with the *A* relationship matrix and *H* relationship matrix. The x-axis represents different quantiles of EGs with a range of mean ± 2.5 standard units, and y-axis represents genetic correlation

Table 3 Genetic correlations (standard errors) across extreme EGs

Trait	*A*|*H*	Genetic correlation (1 and 99%)[a]	Genetic correlation (5 and 95%)[a]	Genetic correlation (10 and 90%)[a]	Genetic correlation (20 and 80%)[a]
AIS	*A*	0.773 (0.217)	0.883 (0.103)	0.932 (0.058)	0.971 (0.024)
	H	0.793 (0.189)	0.891 (0.092)	0.938 (0.050)	0.974 (0.021)
ICF	*A*	0.794* (0.105)	0.883* (0.063)	0.928* (0.039)	0.968* (0.018)
	H	0.793* (0.104)	0.882* (0.062)	0.928* (0.039)	0.968* (0.018)
IFL	*A*	0.901* (0.051)	0.952* (0.026)	0.970* (0.016)	0.987* (0.007)
	H	0.911* (0.047)	0.956* (0.023)	0.973* (0.015)	0.988* (0.006)
NRR	*A*	0.816 (0.172)	0.903 (0.096)	0.939 (0.060)	0.974 (0.026)
	H	0.818 (0.165)	0.903 (0.093)	0.940 (0.058)	0.975 (0.024)

AIS number of inseminations per conception, *ICF* interval from calving to first insemination, *IFL* interval from first to last insemination, *NRR* non-return rate at 56 days after first insemination, *A* and *H* pedigree-based and pedigree-genomic combined matrices, respectively

[a]The genetic correlations between the 1 and 99%, 5 and 95%, 10 and 90%, and 20 and 80% quantiles of EGs; one asterisk (*) means the value deviates from unity by more than 1.645 × SE.

Table 4 The number of overlapping animals of the top 50 sires along two extreme EGs

Trait	A|H	Top 50 overlaps (1 and 99%)[a]	Top 50 overlaps (5 and 95%)[a]	Top 50 overlaps (10 and 90%)[a]	Top 50 overlaps (20 and 80%)[a]
AIS	A	26	34	36	43
	H	24	32	37	42
ICF	A	23	26	29	37
	H	19	26	33	39
IFL	A	39	42	43	46
	H	38	41	43	45
NRR	A	37	40	42	44
	H	34	38	41	46

AIS number of inseminations per conception, *ICF* interval from calving to first insemination, *IFL* interval from first to last insemination, *NRR* non-return rate at 56 days after first insemination, *A* and *H* pedigree-based and pedigree-genomic combined matrices, respectively

[a]The number of overlapping animals in the top sires in the 1 and 99%, 5 and 95%, 10 and 90%, and 20 and 80% quantiles of EGs

Table 5 The prediction accuracies for different traits using an RM and RNM

Trait	Model	Cor(EBV, DYD)[a]	Cor(GEBV, DYD)[b]
AIS	RM	0.154	0.248
	RNM	0.147	0.236
ICF	RM	0.262	0.318
	RNM	0.265	0.319
IFL	RM	0.133	0.218
	RNM	0.140	0.228
NRR	RM	0.123	0.127
	RNM	0.103	0.117

AIS number of inseminations per conception, *ICF* interval from calving to first insemination, *IFL* interval from first to last insemination, *NRR* non-return rate at 56 days after first insemination, *RM* reduced model, *RNM* reaction norm model

[a]The correlation coefficients between DYD and predicted breeding values estimated using the *A* matrix

[b]The correlation coefficients between DYD and predicted breeding values estimated using the *H* matrix

ssGWAS for ICF and IFL

The percentages of genetic variances explained by genomic windows are shown in the Manhattan plot in Fig. 3. Table S3 shows the location information of top windows for the intercept, slope and candidate genes included in those windows. Some top windows that accounted for the largest variance of the intercept were also found to be the top windows that explained the largest variance of the slope for ICF (6 windows) and IFL (1 window), but the effects of these shared windows were mainly in opposite directions because of the negative correlation coefficients between the intercept and slope for ICF (-0.725), and for IFL (-0.360). The top genomic window explained 4.34% of the variance of the intercept for ICF, which also accounted for the largest percentage (3.00%) of the variance of the slope for ICF. Compared with ICF, the percentages of genetic variances explained by genomic windows for IFL were smaller. There were two genomic regions associated with both ICF and IFL, namely, 24.04–24.90 Mb on BTA23 and 70.78–71.59 Mb on BTA17. The candidate genes in these two windows were the immunity-related genes *IL17*, *IL17F* and *LIF*.

Discussion

In this study, we explored the G × E interaction for female fertility traits using RNMs with or without genomic information and evaluated prediction accuracies for future performance. Further, we tried to locate the particular regions in the genome that probably explain the genetic background of ICF and IFL, which exhibited G × E interactions across extreme EGs. The results showed that the slope variances of the reaction norm for all the traits were statistically significant from zero, while genetic correlations across the extreme EGs could only be identified as significantly different from unity for ICF and IFL. Generally, the inclusion of genotype information improved the prediction accuracy. In addition, some candidate genes influencing intercepts and slopes of reaction norms for ICF and IFL were also identified.

Heritabilities estimated from conventional linear mixed model

The heritabilities estimated from the conventional linear mixed model (Table 1) were similar to those from previous studies (Hou et al. 2009; Jamrozik et al. 2005; Liu et al. 2008, 2017). Several studies showed that heritability estimates for interval traits (IFL and ICF) were generally higher than those for count trait (AIS) and binary trait (NRR) for dairy cattle (Berry et al. 2013; Ghiasi et al. 2011), which was also true for our study.

(0.319) was observed between GEBV and DYD for ICF using the RNM. The inclusion of genomic information improved the prediction accuracy in both the RM and RNM. The largest improvement was observed for AIS with the RM, which increased from 0.154 to 0.248. The accuracies for ICF and IFL obtained from the RNM were slightly higher than those obtained from the RM, and the largest improvement was observed in the EBV of IFL for the RNM, which increased from 0.133 to 0.140.

Fig. 3 Proportions of the intercept and slope variances explained by each 20-SNP region for ICF (interval from calving to first insemination) and IFL (interval from first to last insemination) based on ssGWAS. The x-axis represents the chromosomes, and the y-axis shows the percentages of genetic variances
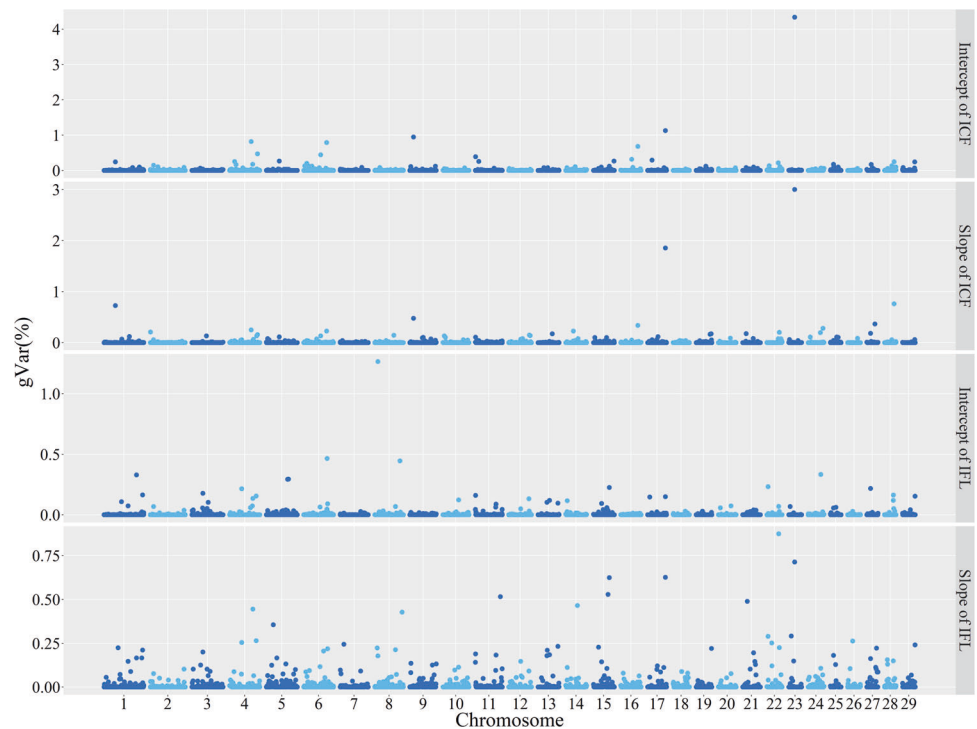
Variance components estimated with the RNM

In this study, we estimated variance components using an RNM with A or H matrices. A previous study on milk yield, dry matter and body weight of Holstein dairy cattle compared genetic variance estimates obtained from ssGBLUP with those obtained from conventional BLUP and showed that variance components estimated with the A matrix were higher than the components obtained from the H matrix, while the latter led to higher accuracy (Veerkamp et al. 2011). However, genetic variance estimates obtained from the conventional RNM and those from the genomic RNM were similar, which was consistent with the result of an earlier study on G × E interactions for total number born in pigs using an RNM (Silva et al. 2014).

Residual variances across different EGs could be different (Knap and Su 2008; Kolmodin et al. 2002). In our study, the heterogeneous residual variances were defined for different herd groups, but not at each herd level. This resulted in discontinuous estimates of heritability between herd groups, as shown in Fig. 1. We found that the residual variance components did not always increase along with the increase in herd effects for ICF and NRR. This pattern has also been observed and discussed in previous studies for G × E interactions in pigs (Silva et al. 2014) and cattle (Cardoso and Tempelman 2012). Calus et al. (2006) suggested that a higher-order RNM and alternative heteroskedastic error specifications might be used in analysis of G × E interactions.

G × E interaction

In this study, the variance of the slope in the RNM was significantly different from zero for all traits, suggesting that all traits exhibited G × E interaction to some extent (Falconer 1990). The correlations between the intercept and slope for all the traits were strongly negative. Some studies asserted that a low correlation between the intercept and slope suggested the re-ranking of animals across different environments, which means that the best animal in one environment was not necessarily the best in another environment (Santana et al. 2013; Su et al. 2006). Actually, even if the correlation between the intercept and slope is very high, the re-ranking of animals can still occur as long as the range of EG is large enough and the variance of the slope is significantly different from zero, because the trajectories of different animals' performances along the EG are not parallel. Therefore, the selection of robust animals with flat slopes and preferential intercepts in breeding practice is very important because they maintain the superior performance across different environments (Strandberg et al. 2000). In addition, environmental sensitivity must be taken into account in the selection of animals as it offers the opportunity to automatically include the environment in the breeding goal (De Jong and Bijma 2002). Furthermore, when RNM was used for a conventional breeding program, the breeding value for each herd level could be appropriate for selection of a heifer/cow, which is raised in a particular farm. However, the breeding value at average effect of

herds could be appropriate for a bull, whose daughters are raised in different farms.

Conventional RNMs and genomic RNMs led to similar genetic correlations between different EGs for ICF and NRR (Fig. 2), which was consistent with previous estimates for yearling weight trait of beef cattle (Oliveira et al. 2018). However, slightly lower genetic correlation coefficients estimated by conventional RNMs were observed for AIS and IFL, which were consistent with previous results for total number born of pigs (Silva et al. 2014). As we have mentioned above, there existed differences of variance components estimated with different relationship matrices, but genomic relationship matrix led to more accurate estimates. Even so, the differences of estimates between genomic and conventional RNMs led to the similar results of existence of G × E interaction, i.e., the genetic correlations between extreme EGs were significantly from unity for ICF and IFL using both conventional and genomic RNMs (Table 3), indicating that the re-ranking of sires may occur if their daughters are distributed across different extreme EGs, which can be observed in Table 4. It was observed that re-ranking was more evident for ICF and AIS than for IFL and NRR (Table 4). Some previous studies have reported the existence of G × E interactions in different fertility traits for cows under different environmental descriptors. For instance, ICF was found to have a high sensitivity under different calving months in Danish and Swedish Holstein dairy cattle (Ismael et al. 2016). Significant G × E interactions in different herds of conventional and organic production systems were observed for calving interval, days open and pregnant at first insemination only in the second parity of Swedish Holsteins (Sundberg et al. 2010), but no G × E interactions existed in the first parity. This suggests that the existence of G × E interactions depends not only on the different environmental descriptors but also on the different stages within lactations.

Accuracy of genomic prediction

Although the existence of G × E interactions between different EGs was not pervasive for all traits, the LRT results showed that the RNM fit the data better than the RM without considering G × E interactions of all these traits (Table S2). However, the LRT simply measures the goodness of fit of the model rather than the accuracy of prediction (Su et al. 2012). Therefore, the prediction accuracies were compared between the RM and RNM, and the RNM had slightly higher prediction accuracies for ICF and IFL, which is also consistent with the existence of G × E interactions across extreme EGs. Whether the accuracy of prediction for future performance could be improved by the inclusion of the reaction norm was examined by the evaluation of genomic predictions using the RNM with A and H matrices. Compared with pedigree-based BLUP, higher accuracies were obtained with the inclusion of genotype information. In a previous study that evaluated the genomic prediction accuracy using a conventional ssGBLUP model in the same Holstein population, the genomic prediction accuracy of ICF was evaluated to be 0.24 for pedigree-based BLUP and 0.29 for ssGBLUP (Ismael et al. 2017), which is comparable to the results we obtained in this study.

Mapping of genomic regions associated with the intercept and slope of the reaction norm

This study performed ssGWAS to locate the particular regions along the genome that were associated with fertility phenotypes and environmental sensitivity. ICF and IFL were chosen to perform ssGWAS as significant G × E interactions were observed across extreme EGs for the two traits (Table 3), and slightly higher prediction accuracies for the RNM were observed compared to the RM (Table 5). Some identified genomic windows were common for both the intercept and slope for each of the two traits (Table S3), and more were found for ICF. The latter result could be because the absolute value of the correlation between the intercept and slope was high (0.725) for ICF, and low (0.360) for IFL. The overlaps of genomic windows between ICF and IFL contain some immune genes such as *IL17*, *IL17F* and *LIF*, which were proved to be related to fertility. For instance, *IL17* and *IL17F* encode interleukin 17 cytokine, and the increased level of this cytokine in plasma has a negative impact on human fertility (Ozkan et al. 2014). *LIF* encodes another interleukin cytokine, leukaemia inhibitory factor, which plays an important role in embryo implantation. It has been reported that aberrant leukaemia inhibitory factor production is linked to implantation failure (Salleh and Giribabu 2014). In addition, immunity-related genes were observed in top genomic windows for the intercept and slope of the two traits, such as *IL17*, *IL17F*, *IL17RB*, *LIF*, *CHDH*, *CD59* and *TLR4*, indicating influences of immunity on fertility in cattle. It has been reported that immunity and fertility could share some common elements (Hurley 2014). As different antimicrobials are used in different herds among different production systems (Bennedsgaard et al. 2006; Bennedsgaard et al. 2010), the pleiotropy of these immunity-related genes may play roles in the performance of fertility traits (Banos et al. 2013). Some genes in the top regions of the intercept and slope for ICF have been previously identified as contributing to cattle fertility. For instance, *TFG*, *EREG* and *AREG* are related to preovulatory follicles in cattle (Li et al. 2009), and *TPR* is linked to age at first calving in Nellore cattle (Mota et al. 2017). For IFL, the genomic regions influencing the intercept contain the *SMAD4* gene, which is necessary for early embryonic development and embryotrophic actions of

follistatin in cattle (Lee et al. 2014). The genomic windows related to the slope for both traits also contain two growth-related genes, *MC4R* and *LEP*, both of which can regulate obesity and energy expenditure (Dempfle et al. 2004; Paolini et al. 2016; Rosenbaum and Leibel 2014; Rutanen et al. 2004). The association between these genes and ICF and IFL suggests that that fertility of cattle is an energy-dependent process, which could be influenced by another energy-consuming process of growth, based on the energy allocation theory (Heino 1999).

In conclusion, genetic parameters obtained from the RNM suggested the existence of G × E interactions for all traits, indicating that the breeding value of an individual may be changed in different herds for these traits. However, genetic correlations across extreme EGs indicated significant G × E interactions for ICF and IFL. The RNM resulted in a better goodness of fit than the RM for all the traits and higher prediction accuracies for ICF and IFL than the RM. Genotype information improved the prediction accuracies of both the RM and RNM for all of the traits. Several reproduction-related, immunity-related and growth-related genes were identified in the genomic regions affecting the intercepts and slopes of the reaction norm for ICF and IFL.

Data archiving

Data supporting this paper were obtained from the commercial dairy farms in Denmark. The phenotype and genotype data are available only upon agreement with commercial breeding organizations and should be requested directly from the corresponding author.

Compliance with ethical standards

Ethical approval All phenotypic data were recorded as part of routine dairy cattle management and genetic evaluations. The DNA samples were obtained for the purpose of routine genomic evaluations in previous projects. Thus, no additional animal handling or experiment was performed specifically for this study.

Conflict of interest The authors declare that they have no conflict of interest.

References

Aguilar I, Misztal I, Johnson D, Legarra A, Tsuruta S, Lawlor T (2010) Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score1. J Dairy Sci 93(2):743–752

Banos G, Wall E, Coffey MP, Bagnall A, Gillespie S, Russell GC, et al. (2013) Identification of immune traits correlated with dairy cow health, reproduction and productivity. PLOS ONE 8(6): e65766

Bennedsgaard T, Klaas IC, Vaarst M. (2006). Outphasing of antibiotics in organic dairy production-challenges for the epidemiologist. In: Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics. Cairns, Australia

Bennedsgaard TW, Klaas IC, Vaarst M (2010) Reducing use of antimicrobials-experiences from an intervention study in organic dairy herds in Denmark. Livest Sci 131(2-3):183–192

Berry DP, Kearney J, Twomey K, Evans R (2013) Genetics of reproductive performance in seasonal calving dairy cattle production systems. Ir J Agric Food Res 52(1):1–16

Calus M, Groen A, De Jong G (2002) Genotype× environment interaction for protein yield in Dutch dairy cattle as quantified by different models. J Dairy Sci 85(11):3115–3123

Calus M, Janss L, Veerkamp R (2006) Genotype by environment interaction for somatic cell score across bulk milk somatic cell count and days in milk. J Dairy Sci 89(12):4846–4857

Cardoso F, Tempelman R (2012) Linear reaction norm models for genetic merit prediction of Angus cattle under genotype by environment interaction 1. J Anim Sci 90(7):2130–2141

Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. Genet Sel Evol 42(1):2

Christensen OF, Madsen P, Nielsen B, Ostersen T, Su G (2012) Single-step methods for genomic evaluation in pigs. Animal 6 (10):1565–1571

De Jong G, Bijma P (2002) Selection and phenotypic plasticity in evolutionary biology and animal breeding. Livest Prod Sci 78 (3):195–214

De Vries A (2006). Determinants of the cost of days open in dairy cattle. In: Proceedings of the 11th International Symposium on Veterinary Epidemiology and Economics: Cairns, Australia, pp 1114–1115

Dempfle A, Hinney A, Heinzel-Gutenbrunner M, Raab M, Geller F, Gudermann T, et al. (2004) Large quantitative effect of melanocortin-4 receptor gene mutations on body mass index. J Med Genet 41(10):795–800

Durinck S, Moreau Y, Kasprzyk A, Davis S, De Moor B, Brazma A, et al. (2005) BioMart and Bioconductor: a powerful link between

biological databases and microarray data analysis. Bioinformatics 21(16):3439–3440

Falconer D (1990) Selection in different environments: effects on environmental sensitivity (reaction norm) and on mean performance. Genet Res 56(1):57–70

Falconer DS, Mackay TF, Frankham R (1996) Introduction to quantitative genetics, 4th edn. Longman, Burnt Mill, Harlow, UK

Gao H, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, et al. (2012) Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. Genet Sel Evol 44(1):8

Ghiasi H, Pakdel A, Nejati-Javaremi A, Mehrabani-Yeganeh H, Honarvar M, González-Recio O, et al. (2011) Genetic variance components for female fertility in Iranian Holstein cows. Livest Sci 139(3):277–280

Heino K (1999) Evolution of resource allocation between growth and reproduction in animals with indeterminate growth. J Evolut Biol 12(3):423–429

Hou Y, Madsen P, Labouriau R, Zhang Y, Lund M, Su G (2009) Genetic analysis of days from calving to first insemination and days open in Danish Holsteins using different models and censoring scenarios1. J Dairy Sci 92(3):1229–1239

Hurley DJ (2014). The relationship of immunity and reproduction in dairy cows. In: 25th Annual Florida Ruminant Nutrition Symposium: Cairns, Australia, p 1

Ismael A, Løvendahl P, Fogh A, Lund MS, Su G (2017) Improving genetic evaluation using a multitrait single-step genomic model for ability to resume cycling after calving, measured by activity tags in Holstein cows. J Dairy Sci 100(10):8188–8196

Ismael A, Strandberg E, Berglund B, Kargo M, Fogh A, Løvendahl P (2016) Genotype by environment interaction for the interval from calving to first insemination with regard to calving month and geographic location in Holstein cows in Denmark and Sweden. J Dairy Sci 99(7):5498–5507

Jamrozik J, Fatehi J, Kistemaker G, Schaeffer L (2005) Estimates of genetic parameters for Canadian Holstein female reproduction traits. J Dairy Sci 88(6):2199–2208

Knap P, Su G (2008) Genotype by environment interaction for litter size in pigs as quantified by reaction norms analysis. Animal 2(12):1742–1747

Kolmodin R, Strandberg E, Madsen P, Jensen J, Jorjani H (2002) Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. Acta Agric Scand, Sect A-Anim Sci 52(1):11–24

Lee K-B, Zhang K, Folger JK, Knott JG, Smith GW (2014) Evidence supporting a functional requirement of SMAD4 for bovine pre-implantation embryonic development: a potential link to embryotrophic actions of follistatin. Biol Reprod 91(3):62

Li Q, Jimenez-Krassel F, Ireland JJ, Smith GW (2009) Gene expression profiling of bovine preovulatory follicles: gonadotropin surge and prostanoid-dependent up-regulation of genes potentially linked to the ovulatory process. Reproduction 137(2):297–307

Liu A, Lund MS, Wang Y, Guo G, Dong G, Madsen P, et al. (2017) Variance components and correlations of female fertility traits in Chinese Holstein population. J Anim Sci Biotechnol 8(1):56

Liu Z, Jaitner J, Reinhardt F, Pasman E, Rensing S, Reents R (2008) Genetic evaluation of fertility traits of dairy cattle using a multiple-trait animal model. J Dairy Sci 91(11):4333–4343

Madsen P, Sørensen P, Su G, Damgaard LH, Thomsen H, Labouriau R. (2013). A user's guide to DMU-a package for analysing multivariate mixed models. Version 6; release 5.2. http://dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf.

Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. J Dairy Sci 92(9):4648–4655

Mota R, Guimarães S, Fortes M, Hayes B, Silva F, Verardo L, et al. (2017) Genome-wide association study and annotating candidate gene networks affecting age at first calving in Nellore cattle. J Anim Breed Genet 134(6):484–492

Oliveira D, Lourenco D, Tsuruta S, Misztal I, Santos D, de Araújo Neto F, et al. (2018) Reaction norm for yearling weight in beef cattle using single-step genomic evaluation. J Anim Sci 96(1):27–34

Ozkan ZS, Deveci D, Kumbak B, Simsek M, Ilhan F, Sekercioglu S, et al. (2014) What is the impact of Th1/Th2 ratio, SOCS3, IL17, and IL35 levels in unexplained infertility? J Reprod Immunol 103:53–58

Paolini B, Maltese P, Del Ciondolo I, Tavian D, Missaglia S, Ciuoli C, et al. (2016) Prevalence of mutations in LEP, LEPR, and MC4R genes in individuals with severe obesity. Genet Mol Res 15(3):n. pag

Rosenbaum M, Leibel RL (2014) The role of leptin in energy homeostasis in humans. J Endocrinol 223(1):T83–T96

Rutanen J, Pihlajamäki J, Karhapää P, Vauhkonen I, Kuusisto J, Mykkänen LM, et al. (2004) The Val103Ile polymorphism of melanocortin-4 receptor regulates energy expenditure and weight gain. Obes Res 12(7):1060–1066

Salleh N, Giribabu N (2014) Leukemia inhibitory factor: roles in embryo implantation and in nonhormonal contraception. Sci World J 2014:201514

Santana M, Eler J, Cardoso F, Albuquerque L, Ferraz J (2013) Phenotypic plasticity of composite beef cattle performance using reaction norms model with unknown covariate. Animal 7(2):202–210

Schneider MdP, Strandberg E, Ducrocq V, Roth A (2005) Survival analysis applied to genetic evaluation for female fertility in dairy cattle. J Dairy Sci 88(6):2253–2259

Silva F, Mulder H, Knol E, Lopes M, Guimarães S, Lopes P, et al. (2014) Sire evaluation for total number born in pigs using a genomic reaction norms approach 1. J Anim Sci 92(9):3825–3834

Strandberg E, Kolmodin R, Madsen P, Jensen J, Jorjani H (2000) Genotype by environment interaction in Nordic dairy cattle studied by use of reaction norms. Inter Bull 25:41–45

Su G, Lund MS, Sorensen D (2007) Selection for litter size at day five to improve litter size at weaning and piglet survival rate1. J Anim Sci 85(6):1385–1392

Su G, Madsen P (2011). User's Guide for Gmatrix. A program for computing Genomic relationship matrix. http://www.dmu.agrsci.dk/Gmatrix/Doc/Current/GmatrixV2-User-Guide.pdf.

Su G, Madsen P, Lund MS (2009) Reaction norm model with unknown environmental covariate to analyze heterosis by environment interaction. J Dairy Sci 92(5):2204–2213

Su G, Madsen P, Lund MS, Sorensen D, Korsgaard IR, Jensen J (2006) Bayesian analysis of the linear reaction norm model with unknown covariates 1. J Anim Sci 84(7):1651–1657

Su G, Madsen P, Nielsen US, Mäntysaari EA, Aamand GP, Christensen OF, et al. (2012) Genomic prediction for Nordic Red Cattle using one-step and selection index blending. J Dairy Sci 95(2):909–917

Sun C, Madsen P, Lund M, Zhang Y, Nielsen U, Su G (2010) Improvement in genetic evaluation of female fertility in dairy cattle using multiple-trait models including milk production traits 1. J Anim Sci 88(3):871–878

Sun C, Madsen P, Nielsen US, Zhang Y, Lund MS, Su G (2009) Comparison between a sire model and an animal model for

genetic evaluation of fertility traits in Danish Holstein population. J Dairy Sci 92(8):4063–4071

Sundberg T, Rydhmer L, Fikse W, Berglund B, Strandberg E (2010) Genotype by environment interaction of Swedish dairy cows in organic and conventional production systems. Acta Agric Scand Sect A 60(2):65–73

VanRaden PM (2008) Efficient methods to compute genomic predictions. J Dairy Sci 91(11):4414–4423

Veerkamp R, Mulder H, Thompson R, Calus M (2011) Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. J Dairy Sci 94 (8):4189–4197

Wang H, Misztal I, Aguilar I, Legarra A, Muir W (2012) Genome-wide association mapping including phenotypes from relatives without genotypes. Genet Res 94(2):73–83

Wu P, Yang Q, Wang K, Zhou J, Ma J, Tang Q, et al. (2018) Single step genome-wide association studies based on genotyping by sequence data reveals novel loci for the litter traits of domestic pigs. Genomics 110(3):171–179