



## Short communication: Heterosis by environment and genotype by environment interactions for protein yield in Danish Jerseys

E. Norberg,<sup>\*1</sup> P. Madsen,<sup>\*</sup> G. Su,<sup>\*</sup> J. E. Pryce,<sup>††</sup> J. Jensen,<sup>\*</sup> and M. Kargo<sup>\*§</sup>

<sup>\*</sup>Department of Molecular Biology and Genetics, Faculty of Science and Technology, Aarhus University, PO Box 50, DK-8830 Tjele, Denmark

<sup>†</sup>Biosciences Research Division, Department of Environment and Primary Industries Victoria, 5 Ring Road, Bundoora, VIC 3083, Australia

<sup>‡</sup>Dairy Futures Cooperative Research Centre, 5 Ring Road, Bundoora, VIC 3083, Australia

<sup>§</sup>Knowledge Centre for Agriculture, Agro Food Park 15, 8200 Aarhus N, Denmark

### ABSTRACT

Crossing of lines or strains within and between breeds has been demonstrated to be beneficial for dairy cattle performance. However, even within breed, differences between strains may also give rise to heterosis. A key question is whether an interaction exists between heterosis and environment ( $H \times E$ ) that is independent of genotype by environment ( $G \times E$ ) interactions. In this study,  $H \times E$  and  $G \times E$  interactions were estimated in a population of approximately 300,000 Danish Jersey cows. The cows were a mixture of pure Danish Jerseys and crosses of US and Danish Jerseys. The phenotype studied was protein yield. A reaction norm model where the unknown environmental covariates are inferred simultaneously with the other parameters in the model was used to analyze the data. When  $H \times E$ , but not  $G \times E$ , was included in the model, heterosis was estimated to be 3.8% for the intermediate environmental level. However, when both  $H \times E$  and  $G \times E$  were included in the model, heterosis was estimated to be 4.1% for the intermediate environmental level. Furthermore, when only  $H \times E$  was included in the model, the regression on the unknown environmental covariate was estimated to be 0.15, interpreted as meaning that an increase of average herd-year protein yield by 1 kg of protein led to an increase in heterosis of 0.15 kg above the average heterosis for a first-cross cow. When both  $H \times E$  and  $G \times E$  were included in the model, the regression on the unknown environmental covariate was not significantly different from zero, meaning that heterosis was similar in all environments investigated. The genetic correlation of protein yields for different environmental levels ranged from 0.72 to 0.93, which was significantly different from unity, indicating that  $G \times E$  exist for protein yield.

**Key words:** crossing of strains, heterosis, genotype by environment, protein yield

### Short Communication

Crossing of lines or strains within breeds has been demonstrated to be beneficial for dairy cattle performance. It is systematically used in breeding programs for pig and poultry, but despite being less common in dairy systems, significant advantages of crossbreeding exist (Lopez-Villalobos et al., 2000; Heins et al., 2006; Walsh et al., 2007; Sørensen et al., 2008). Also, studies have shown that expression of heterosis is dependent on the environment; Barlow (1981) and Penasa et al. (2010) stated that heterosis is largest under suboptimal production environments, whereas Bryant et al. (2007) and Kargo et al. (2012) concluded that heterosis was largest in intermediate environments. The studies by Bryant et al. (2007) and Kargo et al. (2012) on dairy cattle used average herd production level to define the environment (i.e., by grouping animals according to herd production level). However, defining environmental groups before analysis may have some disadvantages, depending on the precision of grouping. In general, such grouping will include a level of randomness due to estimation errors and due to the grouping being dependent on other factors in the model to be used. In addition, lack of connectedness between groups can be a problem. Using a Bayesian approach, Su et al. (2006) developed a reaction norm model with unknown environmental covariates, where the environmental covariates were inferred simultaneously with the other parameters in the model. Using simulated data and herd  $\times$  year as environmental value, they demonstrated this method to be superior to traditional reaction norm models [e.g., Kolmodin et al. (2002) and Calus et al. (2002)], where environmental covariates are calculated before the analysis as simple functions of the data to be analyzed. Su et al. (2009) extended the reaction norm model with unknown environmental covariate to include breed heterozygosity  $\times$  environment ( $H \times E$ ) interactions.

A possible genotype  $\times$  environment ( $G \times E$ ) effect has already been reported in the Danish Jersey population (Kargo et al., 2012). Those authors observed that the genetic variation of milk yield was almost twice as

Received November 8, 2013.

Accepted March 25, 2014.

<sup>1</sup>Corresponding author: [elise.norberg@agrsci.dk](mailto:elise.norberg@agrsci.dk)

large for cows in herds with high production levels compared with cows in herds with low production levels. In addition to  $G \times E$ ,  $H \times E$  interactions may also exist; therefore, a logical progression from the study by Kargo et al. (2012) is to investigate if  $G \times E$  and  $H \times E$  both exist in this population. For that reason, the reaction norm model developed by Su et al. (2006, 2009) was modified to allow for simultaneous estimation of  $G \times E$  and  $H \times E$  effects.

The aim of the present study was to estimate  $H \times E$  and  $G \times E$  effects simultaneously using a reaction norm model. Cows identified to be of Danish Jersey ancestry and crosses between Danish and US Jerseys were used. The 2 lines differ significantly in breeding values for most traits (Knowledge Centre for Agriculture, 2013) and are, therefore, considered to be 2 strains of the Jersey breed.

The data was identical to those used by Kargo et al. (2012). In brief, records on 305-d protein yield (**PY**) from 312,859 first-lactation Danish Jersey cows with varying contributions of American (US) Jersey genes from 1,746 herds and born between 1990 and 2006 were included in the analysis. Only cows with more than 45 DIM and a calving age between 18 and 40 mo were retained. Contributions of genes (of Danish and US origin) were traced as far back as possible using the Danish cattle database (Bundgaard and Høj, 2000). Danish founders were assumed to have 100% Danish genes. Contributions of imported males and females were included in the Danish cattle database based on pedigree information from the country of origin. The contribution of US Jersey genes increased from slightly below 10% for cows born in 1990 to around 40% for cows born in 2001. For cows born from 2001 onwards, the proportion of US Jersey genes has been slightly below 40%. Throughout the period, genes not belonging to original Danish or US Jersey lines were below 5%, and decreasing toward zero. To avoid the anomalies arising from another population of Jerseys, animals with more than 12.5% New Zealand Jersey genes were omitted from the study, and animals with less than 12.5% New Zealand Jersey genes were considered to be Danish Jersey, as this contribution was found to originate from an early import in the 1960s. The degree of heterozygosity was calculated as  $\text{sbp}_{(\text{line1})} \times \text{dbp}_{(\text{line2})} + \text{sbp}_{(\text{line2})} \times \text{dbp}_{(\text{line1})}$ , where sbp is the proportion of ancestry of a given sire line and dbp is the proportion of ancestry of the dam line. Lines 1 and 2 are Danish and US Jerseys, respectively. This corresponds to an assumption of heterosis being caused by dominance effects. The pedigree was traced as far back as possible in the Danish cattle database (Bundgaard and Høj, 2000), and the pedigree file for the full data set included 485,265 animals. All

analyses in this study were carried out using the DMU software package (Madsen and Jensen, 2013).

Two different reaction norm models were used to analyze the data in this study. The first included only an  $H \times E$  effect and the second included  $G \times E$  in addition to the  $H \times E$  effect. The properties of the models are described in detail by Su et al. (2006) and Su et al. (2009). The model estimates the unknown environmental covariates simultaneously, with all other effects in the model and the error in the estimation of these environmental effects then being integrated out of the posterior distributions. The estimated environmental covariates are based on herd-year solutions for protein yield. Following Su et al. (2006) and Su et al. (2009) the 2 reaction norm models used in this study were as follows:

- (1) A model including only  $H \times E$  effects

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{W}\mathbf{t} + \mathbf{h}\mathbf{r} + \mathbf{h}_t\mathbf{r}_t + \mathbf{Z}_0\mathbf{a}_0 + \mathbf{e};$$

- (2) A model including both  $H \times E$  effects and  $G \times E$  effects

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{W}\mathbf{t} + \mathbf{h}\mathbf{r} + \mathbf{h}_t\mathbf{r}_t + \mathbf{Z}_0\mathbf{a}_0 + \mathbf{Z}_t\mathbf{a}_t + \mathbf{e}.$$

In the 2 models, **y** is the vector of first-lactation records of PY; **b** is the vector of fixed effects of year, age at first calving (in months), calving month and calving interval (as a covariable); **t** is the vector of random herd  $\times$  year effects; **r** is a fixed regression coefficient of PY on line heterozygosity to estimate an overall heterosis effect; **r<sub>t</sub>** is a fixed regression coefficient of PY on line heterozygosity  $\times$  herd  $\times$  year effect, which estimates deviation from the overall heterosis due to production level, where the herd  $\times$  year effect is used as a proxy for the true but unknown environmental covariable; **a<sub>0</sub>** is the vector of intercepts; and **a<sub>t</sub>** is the vector of slopes of additive genetic reaction norms in model 2; **a<sub>0</sub>** in model 1 is the vector of additive genetic effects and **e** is the vector of residual effects. In this model, **X**, **W**, and **Z<sub>0</sub>** are known incidence matrices and **h** is the vector of heterozygosity. However, **h<sub>t</sub>** and **Z<sub>t</sub>** are unknown. The element *i* of vector **h<sub>t</sub>** is the product of heterozygosity (an element of **h**) and environmental value (an element of **t**) associated with observation *i*. The row *i* of matrix **Z<sub>t</sub>** has only 1 nonzero element, which is equal to the environmental value associating with observation *i*. Similarly each row of **Z<sub>0</sub>** is zero except for a value of 1 in the same position as the nonzero element in **Z<sub>t</sub>**.

The reaction norm model was implemented applying a Gibbs sampling approach, which is briefly described here based on model 2. The approach includes 2 steps

to sample location parameters. The first step is conditional on  $\mathbf{t}$ , so that  $\mathbf{h}_t$  and  $\mathbf{Z}_t$  can be treated as known. This leads to

$$\mathbf{y}_\theta = \mathbf{y} - \mathbf{Wt} = \mathbf{Xb} + \mathbf{hr} + \mathbf{h}_t\mathbf{r}_t + \mathbf{Z}_0\mathbf{a}_0 + \mathbf{Z}_t\mathbf{a}_t + \mathbf{e},$$

where  $\mathbf{y}_\theta$  is the vector of the observations corrected for herd  $\times$  year effects. Because all incidence matrices and vectors in the above model are known, location parameter  $\boldsymbol{\theta} = (\mathbf{b}, \mathbf{r}, \mathbf{r}_t, \mathbf{a}_0, \mathbf{a}_t)$  can be sampled using standard Gibbs sampling for a model assuming normal distributions.

The second step is conditional on  $\boldsymbol{\theta}$ . Thus,

$$\begin{aligned} \mathbf{y}_t = \mathbf{y} - \mathbf{Xb} - \mathbf{hr} - \mathbf{Z}_0\mathbf{a}_0 &= \mathbf{Wt} + \mathbf{h}_t\mathbf{r}_t \\ &+ \mathbf{Z}_t\mathbf{a}_t + \mathbf{e} = \mathbf{W}^*\mathbf{t} + \mathbf{e}. \end{aligned}$$

In this step,  $\mathbf{W}^*$  is the incidence matrix obtained by replacing the nonzero element in row  $i$  of matrix  $\mathbf{W}$  with  $1 + h_i\mathbf{r} + \mathbf{z}_{0i}\mathbf{a}_t$ , where  $h_i$  is the heterozygosity of the individual with record  $i$  and  $\mathbf{z}_{0i}$  is the  $i$ th row of matrix  $\mathbf{Z}_0$ . Therefore, location parameters  $\mathbf{t}$  can be sampled using standard Gibbs sampling procedure for normally distributed data.

The Gibbs sampler was run as a single chain with a length of 65,000 samples. Convergence was monitored by graphical inspection of the distributions of variance and covariance samples. The first 15,000 samples were discarded as burn-in. Every 10th sample of the remaining samples was saved, leaving 3,000 samples to estimate the features of the posterior distributions.

In general, the heterosis estimates found for PY in crosses between Danish and US Jersey agree with other estimates reported in Danish Jerseys (Sørensen et al., 2008). The parameters estimated with the 2 different reaction norm models: one including an  $\text{H} \times \text{E}$  effect and the other including both an  $\text{H} \times \text{E}$  and a  $\text{G} \times \text{E}$  effect, are presented in Table 1. The additive genetic variance is similar for both models, being slightly larger when estimated with the model that excluded a  $\text{G} \times \text{E}$  effect. This is in agreement with what was found in first-lactation Jersey cows by Campos et al. (1994). The residual variance is somewhat smaller for the model that included both  $\text{H} \times \text{E}$  and  $\text{G} \times \text{E}$  compared with the model that excluded the  $\text{G} \times \text{E}$  effect. So, the  $\text{G} \times \text{E}$  effect captures both some of the residual and some of the genetic variance estimated in a model without the  $\text{G} \times \text{E}$  effect included. Environment-independent heterosis (the intercept, or the heterosis expressed in an intermediate environment) was around 3.78% when estimated with the model including only an  $\text{H} \times \text{E}$  effect, and slightly higher when the  $\text{G} \times \text{E}$  effect was added

(4.1%). The environment-dependent heterosis (slope of the environmental covariate) was estimated to be 0.15, meaning an increase of average herd-year PY by 1 kg of protein led to an increase in heterosis of 0.15 kg above the average heterosis for a first-cross cow. When a  $\text{G} \times \text{E}$  effect was included in addition to  $\text{H} \times \text{E}$ , the regression of  $\text{H} \times \text{E}$  on the unknown environmental covariate was not significant. To make interpretation of the results from this study easier as well as to allow comparison with the study by Kargo et al. (2012), Figure 1 illustrates heterosis (in percent) for different production levels estimated with the 2 reaction norm models, in addition to the heterosis estimates obtained by Kargo et al. (2012). The selected production levels correspond to the mean of the 5 environmental groups based on herd level of PY as defined in Kargo et al. (2012). The average PY in the different environmental groups was as follows: environment 1 = 165 kg, environment 2 = 183 kg, environment 3 = 194 kg (the intermediate environment), environment 4 = 204 kg, and environment 5 = 221 kg. Thus, at these production levels, estimates of heterosis should be comparable between the 2 studies. The estimate of the environment-independent heterosis was slightly higher (3.78% compared with 3.32%) than found by Kargo et al. (2012) when estimating heterosis on the total data set using average information REML without considering differences in environment. Clearly, including a linear reaction norm on  $\text{H} \times \text{E}$  resulted in increasing estimates of heterosis as level of environment improved for PY. In Figure 1, the percentage heterosis in the different environmental groups estimated by Kargo et al. (2012) appears to exhibit a quadratic shape. However, the heterosis in PY appears to be linear. Thus, using a linear reaction norm, as we did, should be valid. However, when a  $\text{G} \times \text{E}$  effect was included in addition to the  $\text{H} \times \text{E}$  effect, the  $\text{H} \times \text{E}$  effect disappeared. The reason for this may be that the  $\text{G} \times \text{E}$  effect captured the scaling effects of increasing PY on the genetic variance. In the model without the  $\text{G} \times \text{E}$  effect, the scaling effect was captured by the  $\text{H} \times \text{E}$  term.

Heritabilities of PY obtained in the present study for different environmental groups as defined in Kargo et al. (2012) and genetic correlations between PY in the different groups estimated with the reaction norm model including both an  $\text{H} \times \text{E}$  effect and a  $\text{G} \times \text{E}$  effect are presented in Table 2. The heritabilities ranged from 0.21 to 0.52 and were, as expected, lowest in the group reflecting the lowest herd production level and highest in the group with the highest production level. The heritability of PY in an intermediate environment was estimated to be 0.35 and is similar to that found by Kargo et al. (2012) when estimating the heritability of PY on the same data using average information REML.

**Table 1.** Estimated parameters for protein yield from 2 different reaction norm models where the first included only a heterosis  $\times$  environment ( $H \times E$ ) effect and the second included both an  $H \times E$  effect and genotype  $\times$  environment ( $G \times E$ ) effect (with SE given in the subscripts)

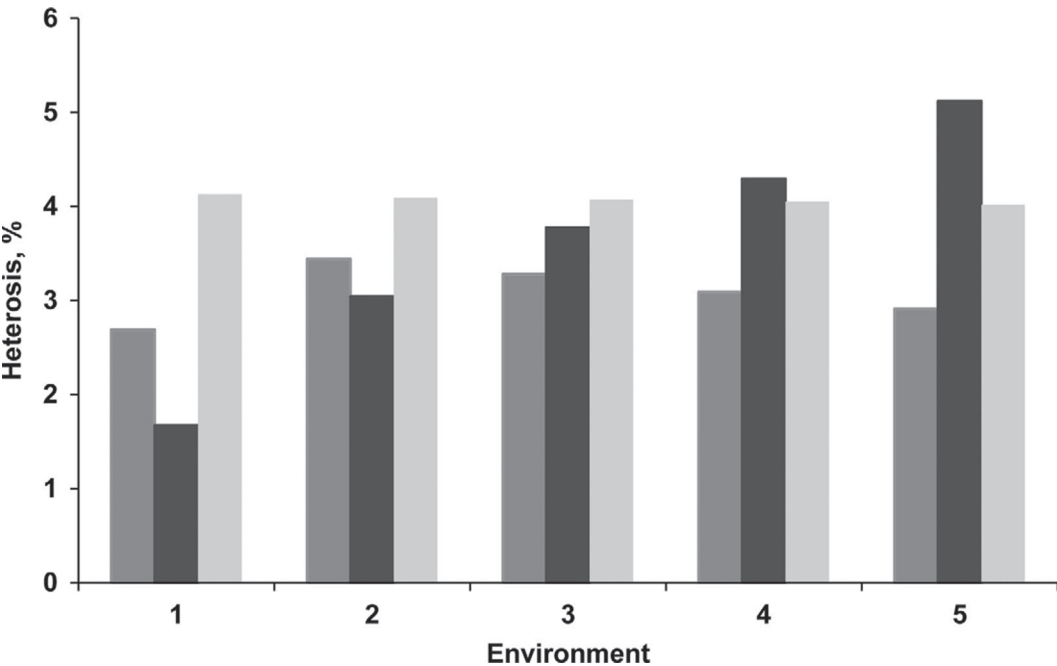
Model	$\sigma_a^2$ <sup>1</sup>	$\sigma_b^2$ <sup>2</sup>	$\sigma_{ab}$ <sup>3</sup>	$\sigma_e^2$ <sup>4</sup>	EI Het. <sup>5</sup>	ED Het. <sup>6</sup>
$H \times E$	257 <sub>4.59</sub>	—	—	477 <sub>3.10</sub>	7.34 <sub>0.43</sub>	0.15 <sub>0.01</sub>
$H \times E$ and $G \times E$	247 <sub>4.62</sub>	0.08 <sub>0.004</sub>	3.3 <sub>0.09</sub>	451 <sub>3.17</sub>	7.96 <sub>0.43</sub>	0.037 <sub>0.02</sub>

<sup>1</sup>Additive genetic variance.  
<sup>2</sup>Variance of the slope of the additive genetic reaction norm.  
<sup>3</sup>Covariance between  $\sigma_a^2$  and  $\sigma_b^2$ .  
<sup>4</sup>Residual variance.  
<sup>5</sup>Environment-independent heterosis (kg).  
<sup>6</sup>Environment-dependent heterosis (kg).

When comparing the heritability estimates by investigating 5 environmental groups that were approximately the same as in the study by Kargo et al. (2012), the heritabilities obtained in our study were lower in the lower production level groups and higher in the groups with a higher production level. This is likely because the analysis of Kargo et al. (2012) used predefined environmental groups, whereas in the study presented here, all parameters were estimated simultaneously, avoiding the need for prior classification of groups. Errors in estimated environmental effects will lead to downward biases in estimated regression parameters when regressing on this effect, which may explain the differences

observed in heritabilities of the extreme groups between the 2 studies.

The genetic correlations between PY in different environments were, as expected, highest between neighboring environments. The correlation between the lowest production level group (environment 1) and the others ranged from 0.72 to 0.93. The correlations between environment 2 and 3, 3 and 4, and 4 and 5 were close to unity. These results demonstrate that PY is not the same trait in herds with very low production levels, where the cow does not have the possibility to express their genetic potential. Therefore, models for genetic evaluations of production traits should include



**Figure 1.** Heterosis (%) estimated by Kargo et al. (2012) for different environmental groups based on protein production level using average information REML (AI-REML; medium gray bars) and in the current study using a reaction norm model including a heterosis  $\times$  environment effect (dark gray bars) or a reaction norm model including both a heterosis  $\times$  environment and a genotype  $\times$  environment effect (light gray bars). Estimates are related to the same groups as in the AI-REML analysis by Kargo et al. (2012).

**Table 2.** Heritabilities (on the diagonal) and genetic correlations (below the diagonal) between protein yield in 5 different environments (Env. 1–5) corresponding to the environmental groups used in Kargo et al. (2012)<sup>1</sup>

Environment	Env. 1	Env. 2	Env. 3	Env. 4	Env. 5
Env. 1	0.209 <sub>0.008</sub>				
Env. 2	0.927 <sub>0.08</sub>	0.289 <sub>0.006</sub>			
Env. 3	0.860 <sub>0.015</sub>	0.989 <sub>0.001</sub>	0.354 <sub>0.006</sub>		
Env. 4	0.802 <sub>0.021</sub>	0.968 <sub>0.001</sub>	0.995 <sub>0.001</sub>	0.416 <sub>0.006</sub>	
Env. 5	0.721 <sub>0.029</sub>	0.928 <sub>0.008</sub>	0.974 <sub>0.003</sub>	0.992 <sub>0.001</sub>	0.518 <sub>0.006</sub>

<sup>1</sup>Standard errors are given in the subscripts, estimated with a reaction norm model using Gibbs sampling.

this interaction, as this will lead to more accurate evaluations of individual animals and further enable a fair comparison of animals independent of their production environment. However, the  $G \times E$  interaction is currently not included in most genetic evaluation procedures implemented worldwide, including Denmark. If it is included in routine evaluations, a relevant question would be how to interpret and use the resulting breeding values. One strategy could be to select bulls based on their performance for specific environmental conditions (herds with different levels of management). Another strategy could be to select bulls with the flat-test slope (i.e., the bulls that are least sensitive to the environment, or “robust” across environments; Bryant et al., 2007). The results presented here also clearly support that models for genetic evaluation should take account of differences in estimates of heritability and variance in different environments. The optimum strategy is dependent on the phenotypic records available, information on environment/herds, size of the population, and the magnitude of  $G \times E$  interaction for the trait of interest. The results also clearly support that extreme herds do not obtain all benefits of the current system of genetic evaluation for PY in Denmark.

## REFERENCES

- Barlow, R. 1981. Experimental evidence for interaction between heterosis and environment in animals. *Anim. Breed. Abstr.* 49:715–737.
- Bryant, J. R., N. López-Villalobos, J. E. Pryce, C. W. Holmes, D. L. Johnson, and D. J. Garrick. 2007. Short communication: Effects of environment on the expression of breed and heterosis effects for production traits. *J. Dairy Sci.* 90:1548–1553.
- Bundgaard, E., and S. Høj. 2000. Direct Access to the Cattle Database with Livestock Registration. Annual Report 1999. National Committee on Danish Cattle Husbandry, Aarhus, Denmark.
- Calus, M. P. L., A. F. Groen, and G. de Jong. 2002. Genotype  $\times$  environment interaction for protein yield in Dutch dairy cattle as quantified by different models. *J. Dairy Sci.* 85:3115–3123.
- Campos, M. S., C. J. Wilcox, C. M. Becerril, and A. Diz. 1994. Genetic parameters for yield and reproductive traits of Holstein and Jersey cattle in Florida. *J. Dairy Sci.* 77:867–873.
- Heins, B. J., L. B. Hansen, and A. J. Seykora. 2006. Production of pure Holsteins versus crossbreds of Holstein with Normande, Montbeliarde, and Scandinavian Red. *J. Dairy Sci.* 89:2799–2804.
- Kargo, M., P. Madsen, and E. Norberg. 2012. Is crossbreeding only beneficial in herds with low management level? *J. Dairy Sci.* 95:925–928.
- Knowledge Centre for Agriculture. 2013. Udenlandske avlsværdital. Accessed June 15, 2013. <http://www.landbrugsinfo.dk/Kvaeg/Avl/Udenlandske-avlsvaerdital/Sider/Startside.aspx>.
- Kolmodin, R., E. Strandberg, P. Madsen, J. Jensen, and H. Jorjani. 2002. Genotype by environment interaction in Nordic dairy cattle studied using reaction norms. *Acta Agric. Scand. A Anim. Sci.* 52:11–24.
- Lopez-Villalobos, N., D. J. Garrick, C. W. Holmes, H. T. Blair, and R. J. Spelman. 2000. Profitabilities of some mating systems for dairy herds in New Zealand. *J. Dairy Sci.* 83:144–153.
- Madsen, P., and J. Jensen. 2013. A user's guide to DMU: A package for analysing multivariate mixed models. Version 6, release 5.2. Danish Institute of Agricultural Sciences, Tjele, Denmark. Accessed Jan. 2012. [http://www.dmu.agrsci.dk/DMU/Doc/Current/dmuv6\\_guide.5.2.pdf](http://www.dmu.agrsci.dk/DMU/Doc/Current/dmuv6_guide.5.2.pdf).
- Penasa, M., M. De Marchi, R. Dal Zotto, G. de Jong, G. Bittante, and M. Cassandro. 2010. Heterosis effects in a black and white dairy cattle population under different production environments. *Livest. Sci.* 131:52–57.
- Sørensen, M. K., E. Norberg, J. Pedersen, and L. G. Christensen. 2008. Crossbreeding in dairy cattle: A Danish perspective. *J. Dairy Sci.* 91:4116–4128.
- Su, G., P. Madsen, and M. S. Lund. 2009. Reaction norm model with unknown environmental covariate to analyze heterosis by environment interaction. *J. Dairy Sci.* 92:2204–2213.
- Su, G., P. Madsen, M. S. Lund, D. Sørensen, I. R. Korsgaard, and J. Jensen. 2006. Bayesian analysis of the linear reaction norm model with unknown covariate. *J. Anim. Sci.* 84:1651–1657.
- Walsh, S., F. Buckley, D. P. Berry, M. Rath, K. Pierce, N. Byrne, and P. Dillon. 2007. Effects of breed, feeding system, and parity on udder health and milking characteristics. *J. Dairy Sci.* 90:5767–5779.