Multiple trait genomic evaluation of conception rate in Holsteins

I. Aguilar,*†[1] I. Misztal,* S. Tsuruta,* G. R. Wiggans,‡ and T. J. Lawlor§
*Animal and Dairy Science Department, University of Georgia, Athens 30602
†Instituto Nacional de Investigación Agropecuaria – INIA Las Brujas, Canelones 90200, Uruguay
‡ARS, USDA, Beltsville, MD 20705-2350
§Holstein Association USA Inc., Brattleboro, VT 05302-0808

ABSTRACT

A national data set of artificial inseminations in US Holsteins was used to obtain genetic evaluations for conception rate (CR). The objective of this study was to investigate the feasibility and resulting accuracy from using all available phenotypic, pedigree, and genomic information. Evaluations were performed by regular BLUP or by BLUP with the traditional pedigree and genomic relationships combined in a unified single-step procedure (SSP). Genetic parameters of CR in the first 3 parities were estimated with data from New York State only. Heritability estimates were around 2% and genetic correlations between CR in different parities were >0.73. The R^2 obtained with the SSP were almost twice as large as those achieved with regular BLUP. Computing the SSP took 2 h, and it was 33% slower than a regular BLUP. A multiple-trait evaluation of CR using the SSP is both possible and advantageous.
Key words: BLUP, genomic selection, fertility, genetic evaluation

INTRODUCTION

A worldwide decline of fertility in Holsteins has created a need for more accurate evaluation of fertility traits. The fertility complex in dairy cattle can be broken down into several traits (González-Recio et al., 2005; Jamrozik et al., 2005). Typical fertility traits such as non-return rate and days open have their advantages and disadvantages (Huang et al., 2007). An important trait in the complex of fertility traits is conception rate (**CR**), which is defined as the probability of a successful outcome of individual breeding services (Averill et al., 2004; Huang et al., 2007). Treating each service separately allows for the specific effects influencing each breeding service to be accounted for.

Because of low heritability of CR (González-Recio et al., 2005; González-Recio et al., 2006; Tsuruta et al.,

2009), accuracies of bull EBV for CR are usually low. However, accuracy can be improved using all available services from each parity. Records from later parities will lead to increased accuracy for older animals, which will contribute to higher accuracies of the younger relatives. Furthermore, accuracy can be boosted by utilizing genomic information (Veerkamp and Beerda, 2007).

The simplest and most efficient way to use genomic information is with a single-step procedure (**SSP**; Aguilar et al., 2010). In this approach, genomic information is used to enhance relationship information and no changes to the model are required. The goals of this study were to estimate genetic parameters of CR in the first 3 parities, run a national evaluation with and without genomic information, and estimate gains in accuracy from the inclusion of genomic information.

MATERIALS AND METHODS

Data

Holstein service records in the first, second, and third parities were obtained from the Animal Improvement Programs Laboratory, ARS, USDA (Beltsville, MD). Records from breedings between 2002 and 2008 were used. Breeding information before 2002 was scarce. Data editing followed criteria presented by (Kuhn et al., 2008). Only AI services were used and DIM at insemination was required to be between 30 and 365 d. Success of insemination was determined via several reproductive criteria (heat detection, natural service, AI, and pregnancy diagnosis), as well as the presence of a subsequent calving. Service sires were restricted to Holstein bulls. Variance components were estimated using a subset of the data from only New York State. All available insemination records were used in the national genetic evaluation. A summary of both data sets is in Table 1.

Model

Conception rates in the first 3 parities were considered as correlated traits. Parameter estimates and

Table 1. Descriptive summary of national and New York data by parity[1]

Item	National			New York State		
	Parity 1	Parity 2	Parity 3	Parity 1	Parity 2	Parity 3
Insemination records (n)	3,025,115	2,033,086	945,870	165,159	116,494	55,038
Herd-year (n)	14,581	14,322	12,203	862	855	745
Cows (n)	1,186,451	790,354	380,776	67,083	46,248	22,634
Conception rate (%)	33.0	31.0	30.7	35.4	32.8	33.2

[1]Pedigree animals (n): national = 2,489,119 and New York State = 132,623.

predictions of EBV were obtained using a multiple-trait linear model. To quantify the benefit of using all 3 parities in a multiple-trait linear model, first-parity service records were also analyzed with a single-trait linear model. Fixed effects in the model included contemporary group defined by herd-year of calving, month of service, age at calving, and days to service after calving. Random effects were service sire ($\mathbf{s}$), additive genetic ($\mathbf{a}$), permanent environmental ($\mathbf{p}$) and residual ($\mathbf{e}$). The (co)variance structure was

$$\text{var}\begin{bmatrix}\mathbf{a}\\\mathbf{p}\\\mathbf{s}\\\mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{G}\otimes\mathbf{A} & \mathbf{0} & \mathbf{0} & \mathbf{0}\\\mathbf{0} & \mathbf{P}\otimes\mathbf{I} & \mathbf{0} & \mathbf{0}\\\mathbf{0} & \mathbf{0} & \mathbf{S}\otimes\mathbf{I} & \mathbf{0}\\\mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{R}\otimes\mathbf{I}\end{bmatrix},$$

where $\mathbf{A}$ is the numerator relationship matrix; $\mathbf{G}$, $\mathbf{P}$, and $\mathbf{S}$ are 3×3 (co)variance matrices for additive genetic, permanent, and service sire effects, respectively; $\mathbf{I}$ are identity matrices, and $\mathbf{R}$ is a 3×3 diagonal matrix of residual variances.

Variance Components

Parameters were estimated by the GIBBS2F90 program (Misztal et al., 2002) via a Bayesian approach using Gibbs sampling. Genomic data were not included for variance component estimation. Of a total of 100,000 samples, the first 10,000 were discarded as a burn-in, and every tenth sample was retained to calculate posterior means and standard deviations of heritability and correlation estimates.

Genetic Evaluation

Genetic evaluations were computed using a modified version of BLUP90IOD (Tsuruta et al., 2001; Aguilar et al., 2010). Approximate accuracies were calculated using ACCF90 (Misztal et al., 2002). Deregressed evaluations ($\mathbf{DD}$) were obtained from EBV and approximate accuracies (VanRaden et al., 2009). A subset of records

up to 2005 was used to assess the accuracy of prediction of breeding values. Two sets of EBV were obtained. The first set used a genetic evaluation with the regular numerator relationship matrix (parent average; $\mathbf{PA}$). The second set used a modified relationship matrix that accounts for genetic relationships via pedigree and genomic data and predicts EBV with an SSP (Misztal et al., 2009). In SSP, the $\mathbf{H}$ matrix replaces the numerator relationship matrix ($\mathbf{A}$) with the following inverse (Aguilar et al., 2010; Christensen and Lund, 2010):

$$\mathbf{H}^{-1} = \mathbf{A}^{-1} + \begin{bmatrix}\mathbf{0} & \mathbf{0}\\\mathbf{0} & \mathbf{G}^{-1}-\mathbf{A}_{22}^{-1}\end{bmatrix},$$

where $\mathbf{H}$ is a modified relationship matrix incorporating genomic information, as described by Legarra et al. (2009); $\mathbf{G}$ is a genomic relationship matrix (VanRaden, 2008); and $\mathbf{A}_{22}$ is the pedigree-based relationship matrix for genotyped animals.

The genomic relationship matrix ($\mathbf{G}$) was created as

$$\mathbf{G} = \frac{\mathbf{ZZ}'}{k},$$

where $\mathbf{Z}$ is an incidence matrix for each SNP genotype, with elements

$$z_{ij} = \begin{cases}0 - 2p_j & \text{if homozygous 11}\\1 - 2p_j & \text{if heterozygous 12 or 21,}\\2 - 2p_j & \text{if homozygous 22}\end{cases}$$

for animal i and SNP genotype j with an assumed allele frequency of $p_j = 0.5$ for all SNP alleles. The scaling parameter k was defined as

$$k = 2\sum p_j(1-p_j).$$

Predictions from the 2 methods were compared by the regressions

Table 2. Estimates of posterior means and standard deviations for genetic parameters for conception rate in the first 3 parities[1]

Trait	CR 1	CR 2	CR 3
CR 1	0.018 ± 0.002	0.877 ± 0.045	0.732 ± 0.047
CR 2	0.288 ± 0.083	0.022 ± 0.002	0.808 ± 0.103
CR 3	0.162 ± 0.084	0.326 ± 0.07	0.016 ± 0.005

[1]Heritability estimates $\pm$ standard deviations on the diagonal, genetic, and permanent correlations above and below the diagonal, respectively. CR 1, CR 2, and CR 3 are the conception rates in the first, second, and third parity, respectively.

$$DD = \mu + \delta EBV_{05} + e,$$

where DD were deregressed evaluations from 154 genotyped bulls without daughter records in 2005 but with daughter records in 2009 that were computed with complete data but without genomic information; μ is a mean; δ is a regression coefficient; EBV_{05} are breeding values based on insemination records up to 2005; and e is residual error. The EBV_{05} were either PA (PA_{05}) or based on SSP (SSP-EBV_{05}).

RESULTS AND DISCUSSION

Table 2 contains estimates of genetic parameters in the first 3 parities. Heritability estimates of each parity were close to 2%, and genetic correlation between parities 1 and 2 was 0.877, between parities 2 and 3 was 0.808, and between parities 1 and 3 was 0.732.

The coefficients of determination and the coefficients of regression for bulls with no daughters in 2005 and at least 50 daughters in 2009 are presented in Table 3. The coefficients of determination obtained with PA in 2005 were low because of limited data and the low heritability of CR. In the available data set, few breeding records were before 2003. Luan et al. (2009) assessed the accuracy of genomic selection in Norwegian Red Cattle and observed a strong relationship between accuracy and heritability of the trait; traits with low heritability had EBV with lower accuracy and greater bias. In the present study, the inclusion of genomic information

(SSP-EBV) resulted in approximately doubling of coefficients of determination than when using relationships from pedigree information only (PA). The improvement of coefficients of determination corresponded to approximately 10 additional effective daughter contributions. Regression of DD_{2009} on SSP-EBV_{05} resulted in a higher coefficient δ than PA_{05}, indicating less bias SSP. Higher coefficients of determination could be achieved with data over many generations and by adding heifer insemination data.

To investigate the increase in accuracy from using a 3-parity multiple-trait model, the analyses were repeated for first-parity service records only. Results comparing SSP and PA using single-trait or multiple-trait models are in Table 4. Use of multiple parity records results in a much higher coefficient of determination and a less biased δ.

Combining a genomic-based relationship matrix with the pedigree-based relationship and the use of a multiple-trait model tripled the accuracy of estimated breeding values for CR in the first parity. This study shows the importance of using all available information (genomic markers and multiple parities) to achieve improved predictability of breeding values for a low heritable trait, such as CR. A simple repeatability model with service records from several parities could also be applied with some loss of accuracy. This topic remains to be addressed in further studies. Alternative methods and use of models incorporating genomic information, such as in VanRaden et al. (2009), could not be assessed in this study due to the limited time frame of our breeding services data. Further research on assessing the increase in accuracies of EBV, as opposed to DD, with inclusion of genomic data by an SSP would be desirable. One reason why we observed an increase in accuracy is that the genomic data provides another valuable source of information. This can be seen by inspection of the diagonals of the inverse of the pedigree-based ($\mathbf{A}$) and genomic-based ($\mathbf{G}$) relationship matrices. Diagonal elements of $\mathbf{A}^{-1}$ for bulls are close to $2 + n/2$, where n is the number of daughters, whereas the corresponding elements of $\mathbf{G}^{-1}$ are $2 + x/2$, where $x > n$. In that case, $x - n$ may be regarded as the additional number

Table 3. Coefficients of determination and coefficients of regression (δ) of daughter deviation on EBV using a single-step approach (SSP-EBV_{05}) or parent average (PA_{05})

Trait[1]	SSP-EBV_{05}		PA_{05}	
	R^2	δ	R^2	δ
CR 1	0.15	0.84	0.07	0.72
CR 2	0.13	0.81	0.06	0.66
CR 3	0.10	0.96	0.05	0.82

[1]CR 1, CR 2, and CR 3 are conception rates in the first, second, and third parity, respectively.

Table 4. Coefficients of determination and coefficients of regression (δ) of daughter deviation on EBV using a single-step approach (SSP-EBV_{05}) or parent average (PA_{05}) for first parity conception rate using single-trait or multiple-trait analysis

Model	SSP-EBV_{05}		PA_{05}	
	R^2	δ	R^2	δ
Single trait	0.07	0.86	0.02	0.57
Multiple trait	0.15	0.84	0.07	0.72

of daughter equivalents from the genomic information. The average (minimum, maximum) x for young bulls in this study was 10.7 (6.1, 82.2).

Computations were done in an Opteron 64-bit processor with a clock speed of 3.02 GHz, 64 Mbyte of memory, and a cache size of 1 Mbyte. Initial computing with BLUP took 1.5 h. Computing with the added genomic information via SSP increased the time to 2 h. Computing with multiple-trait models and the genomic information is, therefore, realistic, even if the data set is much larger.

CONCLUSIONS

Multiple trait genetic evaluation for conception rate using outcomes of all available inseminations is technically possible. Large improvements in accuracy are possible when genomic information is used. Computation with the SSP is straightforward. More accurate assessment of such an improvement would require either records over a longer period of time or a different methodology for comparisons.

ACKNOWLEDGMENTS

REFERENCES

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

Averill, T. A., R. Rekaya, and K. Weigel. 2004. Genetic analysis of male and female fertility using longitudinal binary data. J. Dairy Sci. 87:3947–3952.

Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. Genet. Sel. Evol. 42:2.

González-Recio, O., Y. M. Chang, D. Gianola, and K. A. Weigel. 2005. Number of inseminations to conception in Holstein cows using censored records and time-dependent covariates. J. Dairy Sci. 88:3655–3662.

González-Recio, O., Y. M. Chang, D. Gianola, and K. A. Weigel. 2006. Comparison of models using different censoring scenarios for days open in Spanish Holstein cows. Anim. Sci. 82:233–239.

Huang, C., I. Misztal, S. Tsuruta, and T. J. Lawlor. 2007. Methodology of evaluation for female fertility. Pages 156–160 in Proceedings of the 2007 Interbull Meeting. Interbull Bulletin 37, Dublin, Ireland.

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

Kuhn, M. T., J. L. Hutchison, and H. D. Norman. 2008. Modeling nuisance variables for prediction of service sire fertility. J. Dairy Sci. 91:2823–2835.

Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. J. Dairy Sci. 92:4656–4663.

Luan, T., J. A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T. H. E. Meuwissen. 2009. The accuracy of genomic selection in Norwegian Red cattle assessed by cross-validation. Genetics 183:1119–1126.

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

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, and D. H. Lee. 2002. BLUPF90 and related programs (BGF90). Commun. No. 28–07 in 7th World Congress on Genetics Applied to Livestock Production, Montpellier, France.

Tsuruta, S., I. Misztal, C. Huang, and T. J. Lawlor. 2009. Bivariate analysis of conception rates and test-day milk yields in Holsteins using a threshold-linear model with random regressions. J. Dairy Sci. 92:2922–2930.

Tsuruta, S., I. Misztal, and I. Stranden. 2001. Use of the preconditioned conjugate gradient algorithm as a generic solver for mixed-model equations in animal breeding applications. J. Anim. Sci. 79:1166–1172.

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

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Invited review: Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.

Veerkamp, R. F., and B. Beerda. 2007. Genetics and genomics to improve fertility in high producing dairy cows. Theriogenology 68:S266–S273.