



## A whole-genome association analysis of noncompensatory fertility in Holstein bulls

M. Blaschek,\* A. Kaya,† N. Zwald,† E. Memili,‡ and B. W. Kirkpatrick\*<sup>1</sup>

\*University of Wisconsin-Madison, 1675 Observatory Drive, Madison 53706

†Alta Genetics, Watertown, WI 53094

‡Department of Animal and Dairy Sciences, Mississippi State University, Mississippi State 39762

### ABSTRACT

Increasing fertility in dairy cattle is an important goal. Male infertility represents a part of the overall infertility in dairy cattle and can be partitioned into compensatory and noncompensatory components, where compensatory refers to infertility that can be overcome by increasing sperm number and noncompensatory infertility represents the remainder, presumably due to molecular and genomic defects. Through estimation of single nucleotide polymorphism (SNP) association with noncompensatory bull fertility, it is possible to identify regions of the genome influential to this trait. Use of this information in selection can allow for an increase in cattle fertility, resulting in economic benefits. In this study, high-density SNP genotypes and noncompensatory fertility data from 795 Holstein sires were used to examine SNP associations with fertility. A Bayes B analysis was performed to develop information for genomic selection and to identify genomic regions associated with noncompensatory fertility. A cross-validation approach was used to assess the effectiveness of the models within the original set of 795 bulls. Correlations of predicted and observed fertility values were approximately 0.145 in cross-validation.

**Key words:** cattle, fertility, gene, polymorphism

### INTRODUCTION

The last 50 yr has seen extensive use of AI and genetic evaluation technology in the dairy industry that has contributed to genetic improvement with regard to milk yield and other traits (Hansen, 2000). At the same time, nutrition and management of dairy cattle for milk yield has improved. However, these changes have been associated with negative effects on fertility, as documented by several researchers (Macmillan et al., 1996; Butler, 2000; Royal et al., 2000; Roche et al., 2000; Lucy, 2001; DeJarnette et al., 2004). Although

this phenomenon is attributable to the genetics and management of the dairy female, improvement in overall reproductive efficiency could be gained by improving male fertility as well.

Previous studies have documented contributions of service sire to reproductive performance in dairy cattle. Weller and Ron (1992) studied conception rate in Israeli Holstein field data and reported variance component estimates for service sire that were from one-third to two-thirds of the magnitude of sire of dam variance. Averill et al. (2004) examined conception rate using Holstein field data from California and observed variance component estimates for service sire that were approximately one-quarter of the additive genetic variance. Jamrozik et al. (2005), in a study using Canadian Holstein field data, concluded that service sire was a significant source of variation relative to additive genetic variation, given a service sire variance that was approximately 3.5 to 5% of that observed for maternal genetic variance. To the extent that service sire variation in conception rate is genetic, the opportunity exists to contribute to improved dairy cattle fertility by improving male fertility. In the context of AI, the specific aspect of fertility being considered is noncompensatory or uncompensable fertility, meaning variation in the ability of a unit of semen to lead to conception after eliminating variation that can be compensated for by an increase in sperm number. The current study was undertaken to identify genomic regions and genetic markers associated with noncompensatory fertility through a whole-genome association analysis. Thus, this study focuses on fertility as affected by molecular and genomic defects impairing the function of sperm and the ability to produce viable offspring that cannot be overcome by increasing sperm count (Saacke et al., 2000; Braundmeier and Miller, 2001; Peddinti et al., 2008). Additionally, the potential for genomic selection for noncompensatory fertility was examined.

### MATERIALS AND METHODS

Phenotypic data were provided by Alta Genetics Inc. (Watertown, WI). Animals from which phenotypic data were obtained were from the Alta Genetics (Alta

Received August 16, 2010.

Accepted May 7, 2010.

<sup>1</sup>Corresponding author: bwkirkpa@wisc.edu

Advantage) progeny-testing program, consisting of 180 dairy herds with an average of 850 milking cows per herd. Reproductive performance was regularly recorded from each herd involved in the program. As a part of this program, periodic calf parentage verification was undertaken to ensure the veracity of the data. With regard to the fertility data (i.e., conception rate), cows were routinely monitored for pregnancy based on diagnosis by veterinary palpation or ultrasonic evaluation.

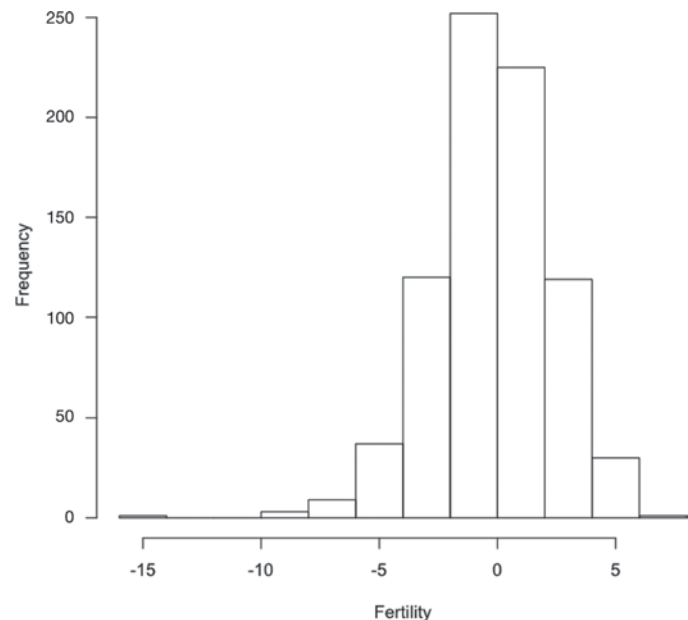
Fertility was defined as the outcome of each breeding event adjusted for environmental effects (i.e., herd, year and month, parity, cow, days in milk, sire proven status) as described by Peddinti et al. (2008). Fertility was then calculated for each bull by taking the percentage deviation of its conception from the average conception of all bulls, using only those bulls with at least 100 breedings (Figure 1). The number of breedings per bull ranged from 100 to 12,306, with a median of 711. The BLUP of bull evaluation from the model was then back-transformed to the probability scale. Fertility predictions and parameter estimates were derived using Probit.F90 software (Chang et al., 2004). Semen used by Alta Genetics as part of this study passed the minimum company threshold for viability, motility, and percentage normal sperm, and the number of sperm cells per unit of semen used for breeding was adequate to overcome compensatory defects.

Genotypes were obtained for 797 bulls and 94 paternal grandsires, using the Illumina 50K Bovine SNP microarray (Illumina, San Diego, CA). Two bulls were eliminated because of potential errors in genotyping (different animal identification numbers assigned to essentially identical genotypes). Markers were reduced in number from 54,001 by eliminating SNP with a minor allele frequency of  $<0.05$  (VanRaden et al., 2009) or complete linkage disequilibrium with other SNP, leaving 38,416 available markers. Total Alta Advantage breeding records available from 1,884 sires was 1,309,934 (Alta Advantage database). The 797 bulls used had 726,041 insemination records with an average of 913 breedings, ranging from 102 to 12,306. RandomForest, a package available within the statistical program R, was used to predict missing genotypes (R 2.7.2, The R Development Core Team, Auckland, New Zealand). Missing values were replaced by the most common genotype at that particular SNP. In this data set, however, missing values accounted for only 0.33% of all genotypes.

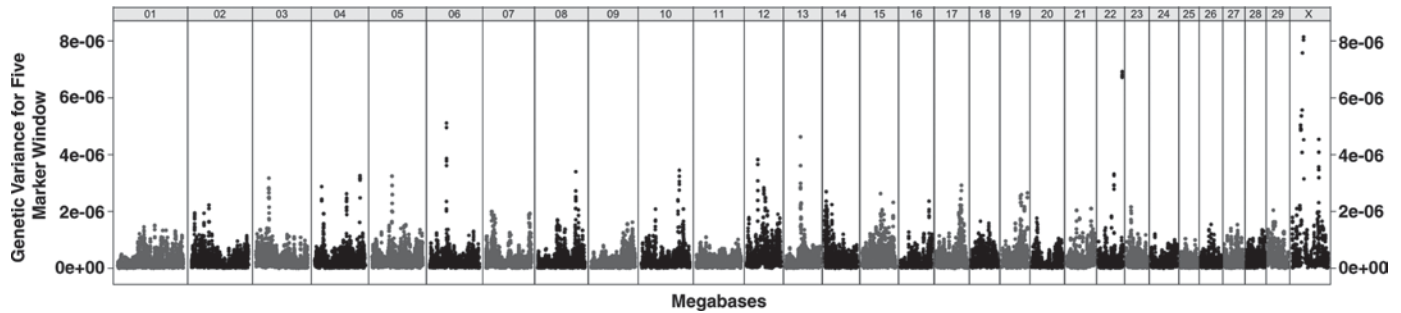
The SNP associations with noncompensatory fertility were estimated with a weighted Bayes B analysis (Meuwissen et al., 2001) that used a value of 0.5 for  $p$ , where  $1 - p$  is the proportion of loci in the genome assumed to contribute to genetic variation for the trait. The value of  $p$  was estimated in a preliminary Bayes Cp

analysis of the data (D. Habier, Department of Animal Science and Center for Integrated Animal Genomics, Iowa State University, Ames, personal communication). In the Bayes Cp analysis, a common variance is assumed for all SNP loci, but  $p$  is treated as an unknown to be estimated from the data. Weights were calculated as described in Garrick et al. (2009) based on the formula for deregressed information and assuming values of 0.03 for heritability and 0.2 for the proportion of genetic variance not accounted for by markers. All analyses were conducted using version 0.9.1101 of the Gensel program (<http://biggs.ansci.iastate.edu>). The chain length was 41,000 cycles with 1,000 cycles as burn-in. Every hundredth replicate was used in the computation of posterior estimates. For purposes of QTL mapping (i.e., identification of genomic regions associated with noncompensatory fertility), genetic variance in a moving 5-marker window was assessed by summing the genetic variances estimated for consecutive SNP.

Prediction equations for noncompensatory fertility were developed using the SNP effect estimates from the Bayes B analysis. To assess effectiveness of the prediction equations developed by the Bayesian analysis, cross-validation was conducted by subdividing the data set into 5 different groups of training and testing data comprising approximately 80 and 20% of the data, respectively. Each animal was allocated once to a testing data set. Models were fit to the individual training data sets using a Bayes B analysis, and the resulting models



**Figure 1.** Distribution of noncompensatory fertility data. Percentage deviation in conception rate from the group average is shown on the x-axis and count of bulls within the phenotype category is shown on the y-axis.



**Figure 2.** Results of whole-genome association analysis as indicated by combined genetic variance associated with adjacent SNP within a moving 5-marker window. Color version available in the online PDF.

were used to predict noncompensatory fertility genomic breeding values for animals in the corresponding testing data sets. Correspondence between predicted and actual noncompensatory fertility values was examined by calculation of correlation.

## RESULTS

Examination of the phenotypic data indicated that the data followed a normal distribution, but one bull had an extreme phenotypic value that might be considered an outlier (Figure 1). The bull in question had an extremely low fertility value that was approximately 2 standard deviations below that of the next lowest bull in the data set. Consequently, in subsequent SNP analyses, data were analyzed both with and without this individual included. The individual in question was included in analyses of parameter estimation, and heritability of fertility was estimated at  $0.03 \pm 0.001$  on the liability scale for service sire. The estimate of  $p$  from the Bayes Cp analysis was 0.50 and 0.32 for the data, including or excluding the potential outlier. Given the relatively flat and overlapping posterior distributions of  $p$  from the 2 analyses, a value of 0.5 was used in all subsequent analyses.

Examination of genetic variance associated with 5-marker windows across the genome suggested relatively large contributions of regions on BTA3, 4, 5, 6, 8, 10, 12, 13, 14, 15, 17, 19, 22, and X (Figure 2). The relative contribution of these regions was not greatly affected by the inclusion or exclusion of the potential outlier bull so only results of analysis of the full data set are presented; the largest differences were an increase in genetic variance associated with the 109-Mb region of BTA4 and the 59-Mb region of BTA22. Twenty 5-marker windows accounting for the greatest genetic variance are listed in Table 1 along with their genomic locations (window midpoint). Complete information on SNP identifiers and estimated SNP effects are provided in Supplementary Tables 1 and 2

(available at <http://www.journalofdairyscience.org/>). Identification and chromosomal locations of SNP are also available online at [www.illumina.com/documents/products/marker\\_lists/marker\\_list\\_bovineSNP50.csv](http://www.illumina.com/documents/products/marker_lists/marker_list_bovineSNP50.csv). Supplementary Table 2 can be uploaded to the GenSel Web site ([big.s.ansci.iastate.edu](http://big.s.ansci.iastate.edu)) and used in predicting genomic breeding values from genotypes corresponding to the SNP listed in Supplementary Table 1.

Five-fold cross validation suggested that a prediction equation based on a Bayes B analysis of the full data set could be expected to have relatively low accuracy. Average Pearson product-moment correlation between predicted genomic breeding values and observed fertility values was approximately 0.145, either with the potential outlier included or excluded (Table 2).

## DISCUSSION

Several chromosomal regions potentially contributing to genetic variation in noncompensatory fertility were identified in this study. Chromosomal regions explaining the most variation in this study did not match any regions reported in a previous study using a subset of the data examined here (Feugang et al., 2009). The study conducted by Feugang et al. (2009) was broken into 2 phases: phase I searched for significant SNP in 10 high-fertility and 10 low-fertility bulls, whereas phase II tested markers found to be significant in phase I in 100 high-fertility and 101 low-fertility bulls. The small number of animals used in the Feugang et al. (2009) study, and lesser statistical power along with a liberal statistical threshold, likely accounts for the differing results between studies.

The current data were analyzed both including and excluding a potential outlier. Results of the 2 analyses were similar with regard to the SNP most commonly included in the models and regions accounting for the most genetic variance. The best SNP within the regions accounting for greatest genetic variance were the same between the 2 analyses. Predicted values based on the

**Table 1.** Location of nonoverlapping, 5-marker windows accounting for the greatest proportion of genetic variation

| BTA <sup>1</sup> | Location <sup>1</sup> | Proportion of genetic variance <sup>2</sup> (%) | Middle SNP name        | SNP identifier |
|------------------|-----------------------|---|------------------------|----------------|
| X                | 43.13                 | 0.36  | Hapmap45594-BTA-103250 | rs41614024     |
| 22               | 58.96                 | 0.31  | ARS-BFGL-NGS-113315    | rs109952523    |
| 6                | 39.71                 | 0.23  | Hapmap23201-BTC-072836 | rs42468607     |
| X                | 31.67                 | 0.22  | UA-IFASA-5646          | rs41585028     |
| 13               | 36.68                 | 0.20  | ARS-BFGL-NGS-100599    | rs111006774    |
| X                | 101.55                | 0.20  | BTA-116883             | rs41618346     |
| 12               | 29.53                 | 0.17  | ARS-BFGL-NGS-15110     | rs110380883    |
| 10               | 81.45                 | 0.15  | BTB-00437757           | rs43644054     |
| 8                | 89.91                 | 0.15  | ARS-BFGL-NGS-102412    | rs109980772    |
| 22               | 38.91                 | 0.15  | ARS-BFGL-NGS-4481      | rs110173704    |
| 4                | 109.13                | 0.14  | ARS-BFGL-NGS-33842     | rs109100643    |
| 5                | 47.38                 | 0.14  | ARS-BFGL-NGS-10812     | rs109273194    |
| 3                | 30.28                 | 0.14  | ARS-BFGL-NGS-87081     | rs109554691    |
| 17               | 58.73                 | 0.13  | ARS-BFGL-NGS-92041     | rs110411691    |
| 4                | 17.35                 | 0.13  | ARS-BFGL-NGS-20771     | rs110984522    |
| 12               | 45.01                 | 0.13  | ARS-BFGL-NGS-19223     | rs109940595    |
| 14               | 5.93                  | 0.12  | Hapmap32262-BTC-066621 | rs109118650    |
| 19               | 62.00                 | 0.12  | UA-IFASA-9390          | rs41633301     |
| 15               | 46.78                 | 0.12  | ARS-BFGL-BAC-19395     | rs110837829    |
| 4                | 76.89                 | 0.12  | ARS-BFGL-NGS-102407    | rs110968187    |

<sup>1</sup>Chromosome (BTA) assignment and location (Mb) within chromosome based on bovine genome assembly 4.0 (Bta4.0). Location denotes the midpoint of the 5-marker window.

<sup>2</sup>Cumulative genetic variance as a percentage of total genetic variance for 5 consecutive SNP using data from all bulls.

models derived from data with or without the potential outlier were very similar ( $r > 0.96$ ). The bull in question had a phenotype based on 855 inseminations, only slightly above the median; consequently, his record did not have an exceptional effect on results of the analyses.

Studies by Khatib et al. (2008, 2009) have implicated the genes signal transducer and activator of transcription 5A (*STAT5A*), growth hormone receptor (*GHR*), uterine milk protein (*UTMP*), secreted phosphoprotein 1 (*SPP1*), and fibroblast growth factor 2 (*FGF2*) in contributing to variation in fertilization or early embryonic survival. Either of these functions could serve as a mechanism underlying noncompensatory fertility, so correspondence between genomic locations identified in the current study and the genomic location of these genes might be anticipated. Of these genes, only *SPP1* showed some similarity in genomic location, although its proximity to the most significant SNP on BTA6 was approximately 3 Mb (*SPP1* located at 37.51 Mb and SNP rs43463315 located at 40.72 Mb in bovine genome assembly Bta4.0). Dai et al. (2009) reported associations between a polymorphism in the follicle stimulating hormone  $\beta$ -subunit gene (*FSHB*) and sperm deformity, acrosome integrity, and nonreturn rate in beef breed bulls. The genomic location of *FSHB* on BTA15 (60.85 Mb) showed no correspondence with the most significant SNP on BTA15 in the current study.

The accuracy of predicted noncompensatory fertility, based on cross-validation analyses, was disappointingly low. Luan et al. (2009), using data from the Norwe-

gian Red cattle breed, reported a mean correlation in cross-validation analyses of 0.45 between predicted and actual breeding value for calving ease, a trait with estimated heritability similar to that reported here for noncompensatory fertility (0.03). A correlation (0.11) more comparable to that observed here was reported for a measure of clinical mastitis, which had a heritability of  $< 1\%$ . The number of observations was similar between the 2 studies, raising the question of whether the heritability estimate reported here for noncompensatory fertility overstates the true value.

Fertility traits are generally considered lowly heritable, and noncompensatory fertility of bulls is no exception. The use of semen produced by one AI organization and veterinary diagnosis of pregnancy for determination of conception rate serves to reduce 2 sources of nongenetic

**Table 2.** Cross-validation analysis of genomic predictions for noncompensatory fertility

| Cross-validation group | Correlation between predicted and observed values (outlier included) |
|------------------------|--|
| 1                      | 0.040  |
| 2                      | 0.176  |
| 3                      | 0.124  |
| 4                      | 0.144  |
| 5                      | 0.231  |
| Average                | 0.143  |

variation. Nonetheless, heritability of noncompensatory fertility in the current data set was still low, making the identification of genomic regions associated with the trait difficult.

## CONCLUSIONS

Identifying SNP that influence fertility can lead to more productive herds and economic benefits within the dairy cattle industry (Veerkamp and Beerda, 2007). A Bayesian analysis of 795 dairy bulls genotyped with 38,416 SNP was performed, leading to the development of prediction equations that may be useful in genomic selection. Given the low heritability of noncompensatory fertility, it is not surprising that the correlation between predicted and true breeding value for this trait is expected to be low ( $\sim 0.15$ ). Regions on several chromosomes that accounted for disproportionately large amounts of genetic variance were identified.

## ACKNOWLEDGMENTS

This research was supported by a Hatch grant from the Wisconsin Agricultural Experiment Station (Hatch Project WIS01333), Mississippi Agricultural and Forestry Experiment Station, and Wisconsin Department of Agriculture (ADD Program).

## REFERENCES

- 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.
- Braundmeier, A. G., and D. J. Miller. 2001. The search is on: Finding accurate molecular markers of male fertility. *J. Dairy Sci.* 84:1915–1925.
- Butler, W. R. 2000. Nutritional interactions with reproductive performance in dairy cattle. *Anim. Reprod. Sci.* 60–61:449–457.
- Chang, Y. M., D. Gianola, B. Heringstad, and G. Kelmetsdal. 2004. Effects of trait definition on genetic parameter estimates and sire evaluation for clinical mastitis with threshold models. *Anim. Sci.* 79:355–363.
- Dai, L., Z. Zhao, X. Yue, X. Li, Y. Gao, J. Liu, and J. Zhang. 2009. Effects of novel single nucleotide polymorphisms of the FSH beta-subunit gene on semen quality and fertility in bulls. *Anim. Reprod. Sci.* 114:14–22.
- DeJarnette, J. M., C. E. Marshall, R. W. Lenz, D. R. Monke, W. H. Ayars, and C. G. Sattler. 2004. Sustaining the fertility of artificial-ly inseminated dairy cattle: The role of the artificial insemination industry. *J. Dairy Sci.* 87(E. Suppl.):E93–E104.
- Feugang, J. M., A. Kaya, G. P. Page, L. Chen, T. Mehta, K. Hirani, L. Nazareth, E. Topper, R. Gibbs, and E. Memili. 2009. Two-stage genome-wide association study identifies integrin beta 5 as having potential role in bull fertility. *BMC Genomics* 10:176–185.
- Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol.* 41:55–62.
- Hansen, L. B. 2000. Consequences of selection for milk yield from a geneticist's viewpoint. *J. Dairy Sci.* 83:1145–1150.
- 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.
- Khatib, H., W. Huang, X. Wang, A. H. Tran, A. B. Bindrim, V. Schutzkus, R. L. Monson, and B. S. Yandell. 2009. Single gene and gene interaction effects on fertilization and embryonic survival rates in cattle. *J. Dairy Sci.* 92:2238–2247.
- Khatib, H., C. Maltecca, R. L. Monson, V. Schutzkus, X. Wang, and J. J. Rutledge. 2008. The fibroblast growth factor 2 gene is associated with embryonic mortality in cattle. *J. Anim. Sci.* 86:2063–2067.
- 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.
- Lucy, M. C. 2001. Reproductive loss in high-producing dairy cattle: Where will it end? *J. Dairy Sci.* 84:1277–1293.
- Macmillan, K. L., I. J. Lean, and C. T. Westwood. 1996. The effects of lactation on the fertility of dairy cows. *Aust. Vet. J.* 73:141–147.
- Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Peddinti, D., B. Nanduri, A. Kaya, J. M. Feugang, S. C. Burgess, and E. Memili. 2008. Comprehensive proteomic analysis of bovine spermatozoa of varying fertility rates and identification of biomarkers associated with fertility. *BMC Syst. Biol.* 2:19–32.
- Roche, J. F., D. Mackey, and M. D. Diskin. 2000. Reproductive management of postpartum cows. *Anim. Reprod. Sci.* 60–61:703–712.
- Royal, M. D., A. O. Darwash, P. E. Flint, R. Webb, J. A. Woolliams, and G. E. Lamming. 2000. Declining fertility in dairy cattle: Changes in traditional and endocrine parameters of fertility. *Anim. Sci.* 70:487–501.
- Saacke, R. G., J. C. Dalton, S. Nadir, R. L. Nebel, and J. H. Bame. 2000. Relationship of seminal traits and insemination time to fertilization rate and embryo quality. *Anim. Reprod. Sci.* 60–61:663–677.
- 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.
- Weller, J. I., and M. Ron. 1992. Genetic analysis of fertility traits in Israeli Holsteins by linear and threshold models. *J. Dairy Sci.* 75:2541–2548.

Copyright of Journal of Dairy Science is the property of Elsevier Science and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.