Genotype by environment interactions on culling rates and 305-day milk yield of Holstein cows in 3 US regions

S. Tsuruta,* D. A. L. Lourenco,* I. Misztal,* and T. J. Lawlor†
*Animal and Dairy Science Department, University of Georgia, Athens 30602
†Holstein Association USA Inc., Brattleboro, VT 05301

ABSTRACT

The objective of this study was to investigate genotype by environment interactions for culling rates and milk production in large and small dairy herds in 3 US regions, using genotypes, pedigree, and phenotypes. Single nucleotide polymorphism (SNP) marker variances were also estimated in these different environments. Culling rates including cow mortality were based on 6 Dairy Herd Improvement termination codes reported by dairy producers. Separate data sets for culling rates and 305-d milk yield were created for large and small dairy herds in the US regions of the Southeast (SE), Southwest (SW), and Northeast (NE) for the first 3 lactation cows that calved between 1999 and 2008. Genomic information from 42,503 SNP markers on 34,506 bulls was included in the analysis to predict genomic estimated breeding value (GEBV) of culling rates and 305-d milk yield with a single-step genomic BLUP using a bivariate threshold-linear model. Cow replacement rates in large SE and NE herds were higher. Heritability estimates of culling rates ranged from 0.03 to 0.11, but the differences were small between large and small herds and among the 3 US regions. Genetic correlations between culling rates and 305-d milk yield were medium to high for cows sold for poor production and reproduction problems. Correlations of GEBV for culling rates among the 3 US regions ranged from 0.34 to 0.92 and were lower between the SW and the other regions, especially in small herds. Correlations of GEBV between large and small herds ranged from 0.44 to 0.90 and were lower in the SW. These results indicate genotype by environment interactions of cow culling rate between the US regions and between large and small herds. Correlations of top 30 SNP marker effects for culling rates between 2 US regions ranged from 0.67 to 0.98. High correlations of top SNP marker effects on culling reasons between the US regions and between large and small herds suggest that major markers can be useful for selection in different environments. The SNP variance shown in a marker gene segment on chromosome 14 was strongly associated with milk production in large and small herds in the NE but not in the SE and SW. Marker genes on chromosome 14 also showed a strong association with cow culling rates due to poor production and mortality in large herds in the NE.

Key words: genotype by environment interaction, culling rate, milk production, US Holstein

INTRODUCTION

A single-step genomic BLUP (ssGBLUP) can predict not only genomic EBV (GEBV) for economically important traits in livestock using pedigree, genotypes, and phenotypes but also SNP marker effects or direct genomic values (DGV; Aguilar et al., 2010; Wang et al., 2012). To investigate the existence of genotype by environment interaction, correlation of DGV or SNP marker effects in different environments, as well as GEBV, can be used. These interactions may be identified more accurately using DGV or SNP effects because genotypes can contain more genetic information than pedigree by including Mendelian sampling.

Gene expression for performance in dairy cattle can differ depending on environmental conditions such as season or climate, age, feed or nutrient, and management (Kolmodin et al., 2002; Windig et al., 2005). Usually, performance in dairy cattle is a composite trait (e.g., milk yield is composed of water, fat, protein, lactose, and minerals). Gene expressions for these components are partly controlled by environment and may interact with each other. Survival traits such as productive life and stayability are also a composite trait. By analyzing culling reasons such as sold for dairy purposes, poor production, reproductive problems, or mastitis or dead, using DHI termination codes and genotypes, we may be able to separate the composite survival trait into several traits (culling reasons) because the high lon-
gevity of dairy cows can be influenced by producer’s culling decision based on milk production levels, cow health, milk market prices in the region, and genes, which can interact with environments. In addition, genotype by environment interactions for cow longevity may be strongly or weakly revealed using genomic information than using only pedigree information. It may result in improving productive life and longevity for cows more efficiently. Tsuruta et al. (2014) reported that heritability estimates for cow mortality and 305-d milk yield in Holsteins and genetic correlations between those 2 traits were different in 3 US regions (Southeast, Southwest, and Northeast). If correlations of GEBV for traits among the 3 US regions are not high (e.g., <0.9), it indicates that the genotype by environment interaction should exist and different breeding programs may be required in the 3 regions. Nowadays, culling rates including cow mortality reported in DHI can be likewise analyzed using genotypes although results from those traits may be difficult to interpret because the trait definition may differ across farms and regions. And main culling reasons such as “sold for dairy purpose” or “poor milk production” can be different in each farm, based on management and selection strategies.

The objectives of this research were to investigate genotype by environment interactions by comparing differences of GEBV and SNP marker variances of cow culling rates and milk production for US Holsteins in large and small herds across US regions, using the ssG-BLUP with genotypes, pedigree, and phenotypes.

**MATERIALS AND METHODS**

Termination codes and 305-d milk yields recorded in the DHI Program were obtained from USDA-Agricultural Research Service (Beltsville, MD). The following 12 termination codes were recorded in DHI farms: 1: sold for locomotion problems (LOCO), 2: sold for dairy purposes (DAIRYP), 3: sold for poor production (POORP), 4: sold for reproductive problems (REPRO), 5: sold for other reasons (OTH), 6: died (DD), 7: sold for mastitis or high SCC (MAST), 8: lactation ended with abortion (ABORT), 9: sold for udder problems, 0: still in herd (SIH), A: sold for undesirable conformation, B: sold for aggressiveness. The DHI farm can chose one of the termination codes for a cow when she is culled. Nine out of 12 termination codes were reported in the Southeast (SE: North Carolina, South Carolina, Georgia, Florida, Tennessee, Alabama, Mississippi, Arkansas, and Louisiana), Southwest (SW: Oklahoma, Texas, New Mexico, and Arizona), and Northeast (NE: Maine, New Hampshire, Vermont, Massachusetts, Rhode Island, Connecticut, and New York) regions of the United States (Table 1). We chose these 3 regions because we expect different climates (SE: hot and wet, SW: hot and dry in summer, NE: warm in summer cold and wet in winter), different managements including herd sizes, and different milk markets (SE: fluid, SW: cheese, NE: fluid and cheese; USDA-AMS, 2014). No termination codes 9, A, or B were reported in these regions. Termination codes 1, 5, and 8 were excluded from this study because events of codes 1 and 8 were too few (1% or less) and code 5 contains multiple unknown reasons for culling. Thus, 6 termination codes including 5 culling reasons and SIH were analyzed. In these 5 culling reasons, DAIRYP and POORP can be classified as voluntary culling, whereas REPRO, DD, and MAST can be classified as involuntary culling. Cows that calved from 1999 to 2008 and had at least one test-day record for the first 3 lactations in each of the 3 regions were extracted from the DHI file. Although the cause of death was unknown, the code DD was included to define cow mortality in this analysis. In each region, the data set was split into 2 groups based on the number of cows per herd-year: large herds with ≥100 cows and small herds with <100 cows. The number 100 is arbitrary, creating almost equal-size data sets for small and large herds in NE, but larger-size data sets for large herds in SW because the average herd size is large in SW, compared with the herd size in SW and NE. Selection decision may be made by different reasons in small and large herds in the 3 regions.

A bivariate threshold-linear animal model was used to estimate genetic parameters and to predict genomic EBV (GEBV) for termination codes and 305-d milk yields as repeated records in the first 3 parities. Thus, the binary traits were defined as the probability (culling rate) a cow will be culled within the first 3 parities (DAIRYP, POORP, REPRO, and MAST), as the probability (mortality) a cow will be dead within the first 3 parities (DD), and as the probability (stayability) a cow will remain in the herd within the first 3 parities (SIH). The analysis was conducted separately for large and small herds in the 3 US regions. Table 2 presents the binary trait definition for termination codes. Cows with codes DAIRYP, POORP, REPRO, and MAST, which were culled for a reason, those with DD, which had died, and those with SIH, which stayed in the herd, were categorized as 1 (yes), otherwise 0 (no). The code for SIH in this bivariate model may be confusing. Each termination code was assumed exclusive and independent events. For instance, if a cow stayed in the herd, the code is 1 (yes) for the trait SIH, otherwise (a cow was culled for some reason or died), it is 0 (no). If a cow was culled for poor production, the code is 1 (yes) for the trait POORP, otherwise (a cow stayed in the herd or was culled for other reasons), it is 0 (no). When a
cow has repeated records 1, 1, and 0 for SIH, it implies that the cow stayed in first and second parities and was culled in the third parity for some reason. When a cow has records 0 and 1 for POORP, it implies that the cow stayed in the first parity and was sold in the second parity for poor production.

In matrix notation, the bivariate trait model can be expressed as

\[ y = Xb + (Wh) + Z_1a + Z_2p + e, \]

where \( y \) is a vector of observations for a culling reason (binary: yes or no) and 305-d milk yield in the first 3 parities; \( b \) contains a vector of fixed effects of reporting month for binary traits, calving month for 305-d milk yield, age of cow nested in parity, and DIM class (in months) nested in parity and herd year; \( h \) is a vector of random herd-year effects for binary traits, \( a \) is a vector of random additive genetic effects; \( p \) is a vector of random permanent environmental effects; \( e \) is a vector of random residual effects; and \( X, W, Z_1, \) and \( Z_2 \) are incidence matrices relating the observations to \( b, h, a, \) and \( p, \) respectively. The covariance matrix \( (V) \) is defined as

\[
V = \begin{bmatrix}
\text{Var} & & & & \\
F \otimes I & 0 & 0 & 0 & \\
0 & G_0 \otimes A & 0 & 0 & \\
0 & 0 & P \otimes I & 0 & \\
0 & 0 & 0 & R \otimes I & \\
\end{bmatrix}
\]

1LOCO = sold for locomotion problems; DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; OTH = sold for other reasons; DD = died; MAST = sold for mastitis or high SCC; ABORT = lactation ended with abortion; and SIH = still in herd.

2SE = Southeast; SW = Southwest; NE = Northeast.

3Termination codes analyzed.

cow has repeated records 1, 1, and 0 for SIH, it implies that the cow stayed in first and second parities and was culled in the third parity for some reason. When a cow has records 0 and 1 for POORP, it implies that the cow stayed in the first parity and was sold in the second parity for poor production.

In matrix notation, the bivariate trait model can be expressed as

\[ y = Xb + (Wh) + Z_1a + Z_2p + e, \]

where \( y \) is a vector of observations for a culling reason (binary: yes or no) and 305-d milk yield in the first 3 parities; \( b \) contains a vector of fixed effects of reporting month for binary traits, calving month for 305-d milk yield, age of cow nested in parity, and DIM class (in months) nested in parity and herd year; \( h \) is a vector of random herd-year effects for binary traits, \( a \) is a vector of random additive genetic effects; \( p \) is a vector of random permanent environmental effects; \( e \) is a vector of random residual effects; and \( X, W, Z_1, \) and \( Z_2 \) are incidence matrices relating the observations to \( b, h, a, \) and \( p, \) respectively. The covariance matrix \( (V) \) is defined as

\[
V = \begin{bmatrix}
\text{Var} & & & & \\
F \otimes I & 0 & 0 & 0 & \\
0 & G_0 \otimes A & 0 & 0 & \\
0 & 0 & P \otimes I & 0 & \\
0 & 0 & 0 & R \otimes I & \\
\end{bmatrix}
\]

1LOCO = sold for locomotion problems; DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; OTH = sold for other reasons; DD = died; MAST = sold for mastitis or high SCC; and SIH = still in herd.
where \( F, G_0, P, \) and \( R \) are \( 2 \times 2 \) covariance matrices for herd year, additive genetic, permanent environmental, and residual effects for each culling rate and 305-d milk yield, respectively, except for \( F \), which had herd-year variance only for binary traits, and \( A \) and \( I \) are a numerator relationship matrix and an identity matrix, respectively. The repeatability model was used because SNP variances were similar among parities in the study of cow mortality, which is one of the termination codes, by Tsuruta et al. (2014), and the selection bias in EBV can be reduced by including multiple parities in this analysis. A multiple trait model with regions as traits and random regressions on different regions and herd sizes can be used to analyze genotype by environment interactions with genotypes (Kolmodin et al., 2002; Windig et al., 2011), but to minimize computing costs and to simplify the interpretation of the results, we chose the bivariate threshold-linear model.

The SNP file contained 42,503 SNP markers, which were sequenced with the BovineSNP50 (50K) BeadChip of Illumina Inc. (San Diego, CA), for 34,506 bulls obtained from USDA-Agricultural Research Service. Parameter estimation was conducted without SNP data, whereas GEBV prediction were conducted including the SNP data, separately for each data set in large and small herds in the 3 regions. All of the genotyped bulls were used in constructing the genomic relationship matrix \( G \) and the pedigree relationship matrix \( A_{22} \) for genotyped animals. Some of the young genotyped bulls had daughters in some data sets or had no daughters in all data sets (i.e., GEBV for young genotyped bulls with no daughters can be predicted using only genomic and pedigree information so that we can select young bulls before having daughters with phenotypes). Genomic EBV of culling rates and 305-d milk yield were predicted with ssGBLUP using the same model as for parameter estimation. The POSTGSF90 program (Aguilar et al., 2010) was used to estimate SNP marker effects and those variances in windows of 30, 300, 3,000, and 34,503 adjacent SNP markers by following the methodology of Strandén and Garrick (2009) and Wang et al. (2012):

\[
\hat{u} = \frac{\delta^2}{\delta_u^2} Z' G^{-1} \hat{a},
\]

where \( \hat{u} \) is a vector of SNP marker effects, \( \hat{a} \) is a vector of animal’s additive genetic effects GEBV or DGV, and \( Z' \) and \( G^{-1} \) are matrices that relate genotypes to individuals and a genomic relationship matrix, respectively. The program also calculates DGV with DGV = \( Zu \).

The THRGIBBS1F90 program via Gibbs sampler (Tsuruta and Misztal, 2006) was used to estimate (co)variance components using cow phenotypes and pedigree information (cows, sires, and dams) in each data set and to predict GEBV for culling rates and 305-d milk yield adding genomic and pedigree information for 34,506 bulls in each model. The number of iterations to generate Gibbs samples was set to 50,000. The first 10,000 samples were discarded as burn-in. Every 10th sample was saved, and 4,000 samples in total were used to calculate posterior means and SD for parameter estimates. Visual investigation of Gibbs samples was used to determine the convergence.

RESULT AND DISCUSSION

Phenotypes

Percentages of termination codes including culling rates and cow mortality were calculated by the number of termination codes reported in the first 3 lactations (Table 1). Each culling rate could have been underestimated because not all culled cows may have been reported when DHI records were collected. The average number of records per herd-year ranged from 222 to 435 in large herds and from 32 to 33 in small herds. The average number of cows per herd-year ranged from 134 to 249 in large herds and from 18 to 22 in small herds. In these data sets, approximately 0.5% difference between culling rates can be considered statistically significant. Of the 5 analyzed termination codes for culling, the code REPRO had the highest rate in all 3 regions. The code DAIRYP had the highest culling rate for large herds in the SW. The code POORP had the highest culling rate for large herds in the SE. The code REPRO had the highest culling rate for large herds in the SE. Cow mortality with the code DD had the highest rate (8.1%) for large herds in the SE and the lowest (2.9%) for small herds in the NE. Maia et al. (2014) reported 3.0, 3.3, and 4.5% for Danish Holsteins, and Tsuruta et al. (2014) reported 2.2, 3.2, and 4.2% for US Holsteins, in first, second, and third parities, respectively. The code MAST had the highest culling rate for large herds in the SE. The code REPRO had the highest culling rate for large herds in the NE. The code SIH had high rates (>75%) for large herds in the SW and for small herds in all 3 regions. Hare et al. (2006) reported that the average survival rate to parity 3 was 50% in 1980s and 1990s. This rate was similar to 56.9 and 53.5% for large herds in the SW and for small herds in all 3 regions. The report by Norman and Walton (2013) is available on the Council on Dairy Cattle
Breeding (Beltsville, MD) website, showing that 63.1% of US Holstein cows in the DHI program stayed in the herd, 6.8 and 5.4% were sold for low production and reproduction problems, respectively, and 5.3% died. Table 1 also shows means and SD of 305-d milk yields that were standardized to 305-d mature equivalents using DHI adjustment procedures (Norman et al., 1974). The mean of 305-d milk yield for large herds in the NE was 1,001 kg higher than for small herds in the NE and 1,182 kg higher than for large herds in the SE. The large SD for 305-d milk yield (around ±2,000 kg) indicates that differences between small and large herds and among the 3 regions were not significant.

Genetic Parameter Estimates

Table 3 presents heritability and repeatability estimates for 305-d milk yields and culling rates in all 3 regions. Heritability estimates for 305-d milk yields ranged from 0.19 in SE to 0.25 in NE. The difference in large and small herds was small. Heritability estimates for culling reasons were similar in different environments, ranging from 0.03 to 0.11. Highest heritability estimates were for MAST: 0.08, 0.06 and 0.08 in large SE, SW, and NE herds and 0.10, 0.11, and 0.10 in small SE, SW and NE herds, respectively. Maia et al. (2014) estimated low heritability 0.024 for mortality in Danish Holstein cows, compared with 0.04 to 0.05 in Table 3. Tokuhisa et al. (2014) reported heritability estimates from 0.05 to 0.08 in the first to third parities. Heritability estimates (0.05 to 0.09) for SIH were similar to those (0.05 to 0.85 and 0.10) for productive life in US Holsteins reported by VanRadjen and Klaaske (1993) and Tsuruta et al. (2005), respectively. Hudson and Van Vleck (1981) also estimated the heritability of 0.05 for stayability at 60 mo of age in Holsteins.

Genetic, permanent environmental, and residual correlations for between 305-d milk yield and culling rates are shown in Table 4. The environmental correlations close to zero imply that no relationship between milk production and culling rates. However, negative genetic correlations of 305-d milk yield with DAIRYP (−0.14 to −0.32) may indicate that higher producing cows can produce daughters who may stay in the herd longer (i.e., positive genetic correlations with SIH). Negative correlations with POORP and REPRO indicate that these culled cows could have had low milk production. Correlations of 305-d milk yield with DD were positive but low. Culled cows with genetically higher milk production were more likely dead in 3 regions. Tokuhisa et al. (2014) estimated similar genetic correlations between milk yield and cow mortality, ranging from 0.0 to 0.3 in first to third parities, using different models. Correlations of 305-d milk yield with MAST were low. Genetic correlations between 305-d milk yield and SIH ranged from 0.24 to 0.41, and environmental correlations were all positive, indicating that higher producing cows tend to stay longer in the herd. Hudson and Van Vleck (1981) estimated the similar genetic correlation of 0.25 between milk production and stayability at 60 mo of age. Although culling rates for large herds in the SW were phenotypically lower than in SE and NE as shown in Table 1, no large differences were found in genetic correlations with 305-d milk yield among the 3 regions. A reason for low culling rates for large herds in SW may be related to the milk market in that region, where more milk is sold for cheese, called class III (USDA-AMS, 2014), that is, dairy producers may not sell cows primarily based on milk production levels, compared with dairy producers for large herds in SE and NE. However, those cows in the SW were not genetically different from those in the other regions.

<table>
<thead>
<tr>
<th>Trait</th>
<th>Large</th>
<th>Small</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>SW</td>
</tr>
<tr>
<td>305-d milk</td>
<td>h^2</td>
<td>r^2</td>
</tr>
<tr>
<td>DAIRYP</td>
<td>0.05</td>
<td>0.08</td>
</tr>
<tr>
<td>POORP</td>
<td>0.07</td>
<td>0.12</td>
</tr>
<tr>
<td>REPRO</td>
<td>0.04</td>
<td>0.07</td>
</tr>
<tr>
<td>DD</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>MAST</td>
<td>0.08</td>
<td>0.15</td>
</tr>
<tr>
<td>SIH</td>
<td>0.08</td>
<td>0.09</td>
</tr>
</tbody>
</table>

1DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; DD = died; MAST = sold for mastitis or high SCC; and SIH = still in herd.

*SE = Southeast; SW = Southwest; NE = Northeast.*

Table 3. Heritability (h^2) and repeatability (r^2) estimates for 305-d milk yield and termination codes in large and small herds in the 3 US regions
Table 5 shows correlations of GEBV among 3 regions for 305-d milk yield and culling rates for 34,506 genotyped bulls. The correlations ranged from 0.34 for POORP in small herds to 0.96 for SIH in large herds. The correlations for all traits except for SIH indicate genotype by environment interactions among the 3 US regions and between large and small herds. The correlations of GEBV between NE and other 2 regions were 0.05 to 0.11 lower than those between SE and SW except for 305-d milk yield and SIH in large herds, whereas the correlations between SW and other 2 regions were 0.06 to 0.12 lower than those between SE and NE in small herds. Table 6 shows correlations between GEBV for 34,506 genotyped bulls in large and small herds in the 3 regions for 305-d milk yield and culling rates. The correlations except for DAIRYP in the SW were lower than those in the SE and NE. These correlations in Tables 5 and 6 were also calculated with DGV instead of GEBV; however, the results from DGV were not shown here because they were essentially identical to the correlations for GEBV and are thus not included here.

SNP Marker Estimates

Table 7 presents correlations of the 30, 300, and 3,000 largest, and all 42,503 SNP marker effects among the 3 US regions in large and small herds. With the

Table 4. Genetic (G), permanent environmental (PE), and residual (R) correlations between 305-d milk yield and termination codes1 in large (upper) and small (lower) herds in the 3 US regions2

<table>
<thead>
<tr>
<th>Termination code/ herd size</th>
<th>G</th>
<th>PE</th>
<th>R</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>SW</td>
<td>NE</td>
</tr>
<tr>
<td>DAIRYP Large</td>
<td>−0.14</td>
<td>−0.32</td>
<td>−0.20</td>
</tr>
<tr>
<td>Small</td>
<td>−0.35</td>
<td>−0.17</td>
<td>−0.38</td>
</tr>
<tr>
<td>POORP Large</td>
<td>−0.66</td>
<td>−0.59</td>
<td>−0.77</td>
</tr>
<tr>
<td>Small</td>
<td>−0.53</td>
<td>−0.49</td>
<td>−0.58</td>
</tr>
<tr>
<td>REPRO Large</td>
<td>−0.56</td>
<td>−0.55</td>
<td>−0.56</td>
</tr>
<tr>
<td>Small</td>
<td>−0.47</td>
<td>−0.36</td>
<td>−0.36</td>
</tr>
<tr>
<td>DD Large</td>
<td>0.08</td>
<td>0.09</td>
<td>0.21</td>
</tr>
<tr>
<td>Small</td>
<td>0.16</td>
<td>0.10</td>
<td>0.21</td>
</tr>
<tr>
<td>MAST Large</td>
<td>0.04</td>
<td>−0.07</td>
<td>0.20</td>
</tr>
<tr>
<td>Small</td>
<td>0.14</td>
<td>0.15</td>
<td>0.35</td>
</tr>
<tr>
<td>SIH Large</td>
<td>0.41</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>Small</td>
<td>0.32</td>
<td>0.24</td>
<td>0.30</td>
</tr>
</tbody>
</table>

1DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; DD = died; MAST = sold for mastitis or high SCC; SIH = still in herd.
2SE = Southeast; SW = Southwest; NE = Northeast.

Table 5. Correlations of genomic EBV for 305-d milk yield and termination codes1 among the 3 US regions2 in large and small herds

<table>
<thead>
<tr>
<th>Trait</th>
<th>Large</th>
<th>Small</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE:SW</td>
<td>SE:NE</td>
</tr>
<tr>
<td>305-d milk yield</td>
<td>0.70</td>
<td>0.76</td>
</tr>
<tr>
<td>Termination code</td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAIRYP Large</td>
<td>0.73</td>
<td>0.61</td>
</tr>
<tr>
<td>Small</td>
<td>0.75</td>
<td>0.65</td>
</tr>
<tr>
<td>POORP Large</td>
<td>0.69</td>
<td>0.72</td>
</tr>
<tr>
<td>Small</td>
<td>0.92</td>
<td>0.86</td>
</tr>
<tr>
<td>REPRO Large</td>
<td>0.88</td>
<td>0.87</td>
</tr>
<tr>
<td>Small</td>
<td>0.96</td>
<td>0.96</td>
</tr>
</tbody>
</table>

1DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; DD = died; MAST = sold for mastitis or high SCC; SIH = still in herd.
2SE = Southeast; SW = Southwest; NE = Northeast.
top group of SNP markers, the correlation could explain the similarity with which common SNP markers were identified in the different regions. The minimum number of SNP 30 was used because it was sufficient to calculate stable correlations among different environments and the number should be large enough to reflect the correlation between haplotypes. In large herds, the correlations for the top 30 marker effects were greater than 0.9 for all combinations of the 3 regions and all termination codes except for DAIRYP, implying that strong marker effects are insusceptible to some environments. Therefore, if we select cows for longevity based on major SNP markers instead of (G)EBV for specific culling reasons, we may be able to improve cow longevity more efficiently in different environments. In small herds, the correlations were up to 0.37 lower than those in large herds, with exceptions for SIH and 305-d

Table 6. Correlations of genomic EBV for 305-d milk yield and termination codes1 between large and small herds within each region2

<table>
<thead>
<tr>
<th>Trait</th>
<th>SE</th>
<th>SW</th>
<th>NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>305-d milk yield</td>
<td>0.79</td>
<td>0.59</td>
<td>0.84</td>
</tr>
<tr>
<td>Termination code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAIRYP</td>
<td>0.75</td>
<td>0.70</td>
<td>0.62</td>
</tr>
<tr>
<td>POORP</td>
<td>0.72</td>
<td>0.44</td>
<td>0.72</td>
</tr>
<tr>
<td>REPRO</td>
<td>0.66</td>
<td>0.58</td>
<td>0.69</td>
</tr>
<tr>
<td>DD</td>
<td>0.88</td>
<td>0.84</td>
<td>0.86</td>
</tr>
<tr>
<td>MAST</td>
<td>0.90</td>
<td>0.78</td>
<td>0.87</td>
</tr>
<tr>
<td>SIH</td>
<td>0.97</td>
<td>0.90</td>
<td>0.97</td>
</tr>
</tbody>
</table>

1DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; DD = died; MAST = sold for mastitis or high SCC; and SIH = still in herd.
2SE = Southeast; SW = Southwest; NE = Northeast.

Table 7. Correlations of the 30, 300, and 3,000 largest, and all 42,503 SNP marker effects for 305-d milk yield and termination codes1 among the 3 US regions2 in large and small herds

<table>
<thead>
<tr>
<th>Trait/no. SNP</th>
<th>SE:SW</th>
<th>SE:NE</th>
<th>SW:NE</th>
<th>SE:SW</th>
<th>SE:NE</th>
<th>SW:NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>305-d milk yield</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.97</td>
<td>0.91</td>
<td>0.89</td>
<td>0.96</td>
<td>0.94</td>
<td>0.95</td>
</tr>
<tr>
<td>300</td>
<td>0.93</td>
<td>0.90</td>
<td>0.88</td>
<td>0.94</td>
<td>0.88</td>
<td>0.86</td>
</tr>
<tr>
<td>3,000</td>
<td>0.77</td>
<td>0.76</td>
<td>0.73</td>
<td>0.74</td>
<td>0.73</td>
<td>0.67</td>
</tr>
<tr>
<td>42,503</td>
<td>0.39</td>
<td>0.42</td>
<td>0.37</td>
<td>0.37</td>
<td>0.43</td>
<td>0.30</td>
</tr>
<tr>
<td>Termination code</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAIRYP</td>
<td>0.70</td>
<td>0.63</td>
<td>0.71</td>
<td>0.60</td>
<td>0.41</td>
<td>0.34</td>
</tr>
<tr>
<td>30</td>
<td>0.45</td>
<td>0.42</td>
<td>0.36</td>
<td>0.60</td>
<td>0.51</td>
<td>0.41</td>
</tr>
<tr>
<td>300</td>
<td>0.28</td>
<td>0.23</td>
<td>0.24</td>
<td>0.60</td>
<td>0.43</td>
<td>0.25</td>
</tr>
<tr>
<td>3,000</td>
<td>0.13</td>
<td>0.13</td>
<td>0.12</td>
<td>0.20</td>
<td>0.26</td>
<td>0.10</td>
</tr>
<tr>
<td>42,503</td>
<td>0.98</td>
<td>0.93</td>
<td>0.94</td>
<td>0.90</td>
<td>0.86</td>
<td>0.86</td>
</tr>
<tr>
<td>POORP</td>
<td>0.92</td>
<td>0.89</td>
<td>0.88</td>
<td>0.72</td>
<td>0.76</td>
<td>0.60</td>
</tr>
<tr>
<td>30</td>
<td>0.76</td>
<td>0.70</td>
<td>0.69</td>
<td>0.49</td>
<td>0.55</td>
<td>0.42</td>
</tr>
<tr>
<td>300</td>
<td>0.38</td>
<td>0.37</td>
<td>0.34</td>
<td>0.21</td>
<td>0.32</td>
<td>0.17</td>
</tr>
<tr>
<td>3,000</td>
<td>0.98</td>
<td>0.96</td>
<td>0.94</td>
<td>0.91</td>
<td>0.95</td>
<td>0.87</td>
</tr>
<tr>
<td>42,503</td>
<td>0.38</td>
<td>0.38</td>
<td>0.38</td>
<td>0.25</td>
<td>0.32</td>
<td>0.19</td>
</tr>
<tr>
<td>REPRO</td>
<td>0.91</td>
<td>0.90</td>
<td>0.91</td>
<td>0.82</td>
<td>0.81</td>
<td>0.74</td>
</tr>
<tr>
<td>30</td>
<td>0.77</td>
<td>0.73</td>
<td>0.73</td>
<td>0.56</td>
<td>0.61</td>
<td>0.48</td>
</tr>
<tr>
<td>300</td>
<td>0.39</td>
<td>0.38</td>
<td>0.38</td>
<td>0.25</td>
<td>0.32</td>
<td>0.19</td>
</tr>
<tr>
<td>3,000</td>
<td>0.96</td>
<td>0.92</td>
<td>0.89</td>
<td>0.74</td>
<td>0.71</td>
<td>0.64</td>
</tr>
<tr>
<td>DD</td>
<td>0.87</td>
<td>0.75</td>
<td>0.81</td>
<td>0.62</td>
<td>0.67</td>
<td>0.49</td>
</tr>
<tr>
<td>30</td>
<td>0.63</td>
<td>0.54</td>
<td>0.57</td>
<td>0.43</td>
<td>0.46</td>
<td>0.30</td>
</tr>
<tr>
<td>300</td>
<td>0.31</td>
<td>0.24</td>
<td>0.26</td>
<td>0.20</td>
<td>0.24</td>
<td>0.11</td>
</tr>
<tr>
<td>3,000</td>
<td>0.97</td>
<td>0.96</td>
<td>0.96</td>
<td>0.83</td>
<td>0.82</td>
<td>0.75</td>
</tr>
<tr>
<td>MAST</td>
<td>0.91</td>
<td>0.88</td>
<td>0.89</td>
<td>0.82</td>
<td>0.78</td>
<td>0.69</td>
</tr>
<tr>
<td>30</td>
<td>0.63</td>
<td>0.54</td>
<td>0.57</td>
<td>0.58</td>
<td>0.57</td>
<td>0.44</td>
</tr>
<tr>
<td>300</td>
<td>0.34</td>
<td>0.38</td>
<td>0.26</td>
<td>0.28</td>
<td>0.32</td>
<td>0.18</td>
</tr>
<tr>
<td>3,000</td>
<td>0.98</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
<td>0.96</td>
</tr>
<tr>
<td>SIH</td>
<td>0.97</td>
<td>0.94</td>
<td>0.95</td>
<td>0.94</td>
<td>0.92</td>
<td>0.91</td>
</tr>
<tr>
<td>30</td>
<td>0.90</td>
<td>0.86</td>
<td>0.87</td>
<td>0.83</td>
<td>0.82</td>
<td>0.77</td>
</tr>
<tr>
<td>300</td>
<td>0.54</td>
<td>0.53</td>
<td>0.49</td>
<td>0.46</td>
<td>0.53</td>
<td>0.34</td>
</tr>
</tbody>
</table>

1DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; DD = died; MAST = sold for mastitis or high SCC; SIH = still in herd.
2SE = Southeast; SW = Southwest; NE = Northeast.

Journal of Dairy Science Vol. 98 No. 8, 2015
milk yields. These results suggest that markers related to culling reasons and mortality were more likely selected in larger herds, where cows produce more milk. Dairy producers in large herds should have more opportunities to cull cows, and their culling decisions can be more targeted or focused on the primary traits of economic importance that have a stronger relationship with the SNP markers. On the other hand, producers in smaller herds do not have the same opportunity to cull genetically inferior cows, and their culling decisions may be more involuntary. As the number of the top SNP markers increased, these correlations declined. This suggests that as more minor genes or polygenes are involved, each marker effect may be too small to estimate accurately, more genes are susceptible to some environments, or both. Table 8 presents correlations of the 30, 300, and 3,000 largest, and all 42,503 SNP marker effects between large and small herds in the 3 regions. All correlations for the top 30 SNP marker effects in the SE and NE were greater than 0.9 except for DAIRYP in the SE. Low correlations for DAIRYP (0.77), POORP (0.67), DD (0.73), and MAST (0.74) in the SW may indicate large genotype by environment interactions but could also be attributed to the relatively small data size. In Tables 7 and 8, high correlations (e.g., > 0.9) for the top 300 or 3,000 SNP marker effects may indicate more genes associated with the trait, fewer genotype by environment interactions in the trait, or both. When estimating SNP marker effects in this analysis, additive genetic effects and SNP marker effects were assumed random and normally and independently distributed, ignoring all interactions among SNP markers and with various environments. If several of these significant interactions exist and play important roles in gene expression and repression, SNP marker estimates in this study could be misleading.

Variances in the 30 adjacent SNP windows, which are approximately equivalent to 2 Mbp, were calculated as a proportion of the total additive genetic variance. Variances greater than 0.5% for 305-d milk yield (m) and the 6 termination codes (2: DAIRYP, 3: POORP, 4: REPRO, 6: DD, 7: MAST, 0: SIH) in BTA were indicated with no asterisk (>0.5%), *(>1.0%), **(>2.0%) or ***(>3.0%) in Table 9. A segment of SNP markers in BTA14 showed a strong association with 305-d milk yield in all 3 regions except for small SW herds. Cole et al. (2009) reported the highest additive variance ratios in BTA14 for milk and fat yields. Pryce et al. (2010) found more than 70 significant SNP effects on BTA14 that included the DGAT1 gene for milk production in dairy cattle (Grisart et al., 2002; Thaller et al., 2003). Streit et al. (2011) investigated a major gene (DGAT1) by polygene interaction in German Holsteins and found that the interaction for milk yield was not significant although the interactions for fat percentage and protein percentage were significant. In BTA14, SNP variances for POORP, REPRO, DD, and MAST explained more than 1% of the total additive genetic variance in large NE herds. In addition, in BTA14, SNP variances explained more than 1% for DAIRYP and POORP in small NE herds. Wibowo et al. (2008) reviewed over 40 studies of QTL on BTA14 and showed 56 QTL for milk production and some common QTL shared with SCC or mastitis in dairy cattle. Wijga et al. (2012) reported strong genomic associations with SCC on BTA4, BTA6, and BTA18 explained by 11 to 16% SNP variances. Sahana et al. (2014) reported QTL associated with mastitis on BTA6 for Holsteins. Höglund et al. (2009a) reported 24 QTL for fertility traits on 17 chromosomes. Höglund et al. (2009b) reported correlated QTL regions between fertility and milk production traits on 19

### Table 8. Correlations of the 30, 300, and 3,000 largest, and all 42,503 SNP marker effects for 305-d milk yield and termination codes between large and small herds within each region

<table>
<thead>
<tr>
<th>Trait/no. SNP</th>
<th>SE</th>
<th>SW</th>
<th>NE</th>
</tr>
</thead>
<tbody>
<tr>
<td>305-d milk yield</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.99</td>
<td>0.99</td>
<td>0.98</td>
</tr>
<tr>
<td>300</td>
<td>0.97</td>
<td>0.97</td>
<td>0.97</td>
</tr>
<tr>
<td>3,000</td>
<td>0.89</td>
<td>0.89</td>
<td>0.92</td>
</tr>
<tr>
<td>All 42,503</td>
<td>0.50</td>
<td>0.50</td>
<td>0.56</td>
</tr>
<tr>
<td>Termination code</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DAIRYP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.78</td>
<td>0.77</td>
<td>0.92</td>
</tr>
<tr>
<td>300</td>
<td>0.66</td>
<td>0.59</td>
<td>0.82</td>
</tr>
<tr>
<td>3,000</td>
<td>0.43</td>
<td>0.37</td>
<td>0.60</td>
</tr>
<tr>
<td>All 42,503</td>
<td>0.18</td>
<td>0.15</td>
<td>0.26</td>
</tr>
<tr>
<td>POORP</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.96</td>
<td>0.67</td>
<td>0.98</td>
</tr>
<tr>
<td>300</td>
<td>0.94</td>
<td>0.70</td>
<td>0.96</td>
</tr>
<tr>
<td>3,000</td>
<td>0.78</td>
<td>0.46</td>
<td>0.86</td>
</tr>
<tr>
<td>All 42,503</td>
<td>0.39</td>
<td>0.21</td>
<td>0.48</td>
</tr>
<tr>
<td>REPRO</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.99</td>
<td>0.98</td>
<td>0.98</td>
</tr>
<tr>
<td>300</td>
<td>0.92</td>
<td>0.79</td>
<td>0.95</td>
</tr>
<tr>
<td>3,000</td>
<td>0.75</td>
<td>0.54</td>
<td>0.83</td>
</tr>
<tr>
<td>All 42,503</td>
<td>0.38</td>
<td>0.23</td>
<td>0.44</td>
</tr>
<tr>
<td>DD</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.98</td>
<td>0.73</td>
<td>0.97</td>
</tr>
<tr>
<td>300</td>
<td>0.85</td>
<td>0.74</td>
<td>0.89</td>
</tr>
<tr>
<td>3,000</td>
<td>0.65</td>
<td>0.55</td>
<td>0.73</td>
</tr>
<tr>
<td>All 42,503</td>
<td>0.29</td>
<td>0.25</td>
<td>0.35</td>
</tr>
<tr>
<td>MAST</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.98</td>
<td>0.74</td>
<td>0.99</td>
</tr>
<tr>
<td>300</td>
<td>0.96</td>
<td>0.74</td>
<td>0.96</td>
</tr>
<tr>
<td>3,000</td>
<td>0.75</td>
<td>0.49</td>
<td>0.88</td>
</tr>
<tr>
<td>All 42,503</td>
<td>0.36</td>
<td>0.20</td>
<td>0.49</td>
</tr>
<tr>
<td>SIH</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>0.99</td>
<td>0.96</td>
<td>0.98</td>
</tr>
<tr>
<td>300</td>
<td>0.98</td>
<td>0.89</td>
<td>0.97</td>
</tr>
<tr>
<td>3,000</td>
<td>0.94</td>
<td>0.71</td>
<td>0.93</td>
</tr>
<tr>
<td>All 42,503</td>
<td>0.62</td>
<td>0.34</td>
<td>0.60</td>
</tr>
</tbody>
</table>

1DAIRYP = sold for dairy purposes; POORP = sold for poor production; REPRO = sold for reproductive problems; DD = died; MAST = sold for mastitis or high SCC; SIH = still in herd.
2SE = Southeast; SW = Southwest; NE = Northeast.
chromosomes. Based on results in this study, no strong associations were found in the SE and SW. A large number of animals with phenotypes of culling rates may be required to estimate accurate SNP variances in a specific chromosome, even though heritability estimates for culling rates are not low compared with those in the NE. A reason why the stronger associations in BTA14 being found in the NE could be that dairy producers for large herds in the NE make culling decision based more on poor or high milk production. Inconsistent relationships with BTA14 among the regions could explain why the DGAT1 gene does not go to fixation in the Holstein breed. As shown in the USDA report (USDA-AMS, 2014), more milk for cheese (Class III) was sold in the SW. The high or low milk price in the region may affect the decision of how farmers cull a cow. A different selection criterion may have affected the gene frequencies linked to milk production in each region. If the DGAT1 gene was equally important in the same direction in all regions, the frequency of the favorable DGAT1 allele would increase. However, the lack of a consistent selection pressure on either allele across the entire population would help ensure that the different alleles of the DGAT1 gene remain at some intermediate frequency.

**CONCLUSIONS**

Culling rates including cow mortality and milk production for US Holstein cows in large and small herds in the SE, SW, and NE regions of the United States were analyzed using genotypes, pedigree information, and phenotypes. Cows with poor milk production and reproductive problems were more likely to be culled. According to genomic breeding values for genotyped bulls, larger genotype by environment interactions on culling rates seem to exist in large and small herds in the Southwest than in the other regions. Therefore, the interaction term should be included in the model to reduce the effect on breeding values. Genes responsible for culling reasons are more likely linked to milk production, especially in the NE. Different genomic segments could be responsible for culling reasons in the different regions of the United States. Major SNP markers could be used to improve cow longevity by selecting animals for specific culling reasons in different environments. Culling rates based on DHI reports may not be consistent with actual productive and reproductive performance because culling decisions are made by each dairy producer’s management strategy, which varies by producer. Therefore, these farmer-recorded traits on cow culling rates could be interpreted differently. If these traits are recorded under similar conditions, genetic evaluations of cow longevity can be beneficial for dairy producers, using termination codes with genotypes.

In future studies with genotypes, it will be important to investigate how each individual SNP interacts with a specific environment condition to predict genomic evaluations more accurately.

---

**Table 9.** Termination codes\(^1\) and 305-d milk yield (m) with SNP marker variance greater than 0.5%, >1.0% (*), >2.0% (**), or >3.0% (***). for each chromosome in large and small herds in the 3 US regions\(^2\)

<table>
<thead>
<tr>
<th>BTA</th>
<th>Large</th>
<th>Small</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SE</td>
<td>SW</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>m*, 3, 4</td>
<td>m, 4</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>17</td>
<td>M</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>M</td>
<td>6</td>
</tr>
<tr>
<td>22</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>24</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>28</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

\(^1\) = sold for dairy purposes; \(^2\) = sold for poor production; \(4\) = sold for reproductive problems; \(6\) = died; \(7\) = sold for mastitis or high SCC; \(0\) = still in herd.

\(^2\)SE = Southeast; SW = Southwest; NE = Northeast.
ACKNOWLEDGMENTS

This study was partially funded by the Holstein Association USA Inc. (Brattleboro, VT) The authors thank the USDA-Agricultural Research Service for providing DHI and genomic data.

REFERENCES


