Inclusion of herdmate data improves genomic prediction for milk-production and feed-efficiency traits within North American dairy herds

N. E. Schultz* and K. A. Weigel

Department of Dairy Science, University of Wisconsin, Madison 53706

ABSTRACT

Genomic data are widely available in the dairy industry and provide a cost-effective means of predicting genetic merit to inform selection decisions and increase genetic gains. As more dairy farms adopt genomic selection practices, dairy producers will soon have genomic data available on all of the animals within their herds. This is a very rich, but currently underused, source of information. Herdmates provide an excellent indication of how a selection candidate's genetics will perform within a given herd, noting that herdmates often include close relatives that share a similar environment. The study objective was to evaluate the utility of incorporating herdmate data into genomic predictions in a data set composed of 3,303 Holsteins from one herd in Canada and 6 herds throughout the United States. Within-herd prediction accuracy was assessed for milk-production and feed-efficiency traits determined from genomic best linear unbiased prediction under 4 different scenarios. Scenario 1 did not include herdmates in the training population. Scenarios 2 through 4 included herdmates in the training population, and scenarios 3 and 4 also included modeling of herd-specific marker effects. Leave-one-out cross validation was used to maximize the number of herdmates in the training population in scenarios 2 through 4, while maintaining constant training population size with scenario 1. Results from the present study reveal the importance of incorporating herdmate data into genomic evaluations. Inclusion of herdmates in the training population improved mean within-herd prediction accuracy for milk-production traits ($\pm$ standard error) by 0.08 ± 0.03 (milk yield), 0.07 ± 0.03 (fat percentage), and 0.05 ± 0.01 (protein percentage) and feed-efficiency traits by 0.07 ± 0.02 (milk energy), 0.03 ± 0.02 (DMI), and 0.08 ± 0.01 (metabolic body weight). Modeling herd-specific marker effects further improved mean within-

herd prediction accuracy for milk yield and energy by 0.03 ± 0.01 and 0.02 ± 0.01 , respectively. Herds with higher within-herd heritability and low genomic correlation with the remaining herds benefitted most from the inclusion of herdmate data.

Key words: genomic prediction, genotype-by-environment interaction, dairy herd

INTRODUCTION

Routine genotyping of all heifer calves within a dairy herd is a common and cost-effective means of predicting genetic merit to inform selection and culling decisions. Genomic selection increases the rate of genetic gain through increased accuracy of EBV, as compared with pedigree-based approaches that ignore Mendelian segregation within families (Meuwissen et al., 2001; Wiggans et al., 2017). Dairy producers can test (genotype) their calves, obtain predictions of genetic merit derived from an established reference population, and cull young animals with poor genetic potential; this can reduce feed costs during the rearing period and genetically improve the herd by propagating animals with superior genetic merit. As more dairy farms adopt genomic selection practices, dairy producers will soon have genomic data available on all of the animals within their herds. This is a very rich, but currently underused, source of information. Herdmates provide an excellent indication of how a selection candidate's genetics will perform within a given herd, noting that herdmates often include close relatives that share a similar environment and management system.

The goal of a national genetic evaluation system is to identify sires (and, more recently, cows) whose offspring will perform well, on average, across a wide range of environments. Environmental conditions and herd management practices are considered as "nuisance factors," and their effects are minimized or eliminated by spreading offspring across many different herds and environments in a progeny testing program. Genetic differences that could allow some animals to perform well under certain conditions, such as feeding a total mixed ration versus relying on rotational grazing, or in

Received April 17, 2019.

Accepted August 5, 2019.

*Corresponding author: nichol.schultz@gmail.com

the presence or absence of heat stress (or heat abatement devices), are ignored. This approach is in contrast to that of plant breeders, in which the performance and resilience of various crop varieties are optimized for local soil types and growing season conditions. The current approach also conflicts with an individual dairy producer's goal of breeding cattle that will perform well under the housing, nutrition, and management conditions present on their specific farms.

The statistical models for predicting breeding values assume 2 components: the additive genetic effects and the environmental effects sum to an individual animal's phenotype. This approach ignores gene-by-environment interactions, meaning the expression of an individual's genetic potential is currently assumed to be independent of the environmental conditions to which the animal was exposed. In a genomic analysis, this means that the estimated effects of genetic markers are assumed to be constant across all environments. Carrying out genetic improvement at the national level using this one-size-fits-all approach has provided tremendous genetic progress in dairy cattle and other livestock species; however, it may not be well suited for selecting the optimal animals for a targeted environment or production system.

Several studies have investigated and reported on the existence of genotype-by-environment interactions in dairy cattle. Strandberg et al. (2009) and Tiezzi et al. (2017) both used sire models to examine interactions with environmental covariates using reaction norm models. Yao et al. (2017) explicitly modeled constant and across-environment (geographic region) marker effects to examine genotype-by-environment effects. However, the extent to which phenotypic variation reflects herd-specific marker effects is largely unknown. It is known that sire effects often vary by herd, which implies the existence of herd-specific marker effects.

The objective of our study was to test the predictive ability of models that incorporate herdmate data and explicitly model genotype-by-herd interactions for prediction of within-herd performance for milk-production and feed-efficiency traits in North American herds.

MATERIALS AND METHODS

Phenotypic Data

The data used in this study consisted of 3,303 Holstein cows from one herd in Canada (216 cows) and 6 herds located throughout the United States (Florida, Iowa, Michigan, Maryland, and 2 Wisconsin herds with 331, 908, 253, 483, 341, and 771 cows, respectively). More detailed information about the data sources can be found in the study by Tempelman et al. (2015).

Phenotypic data included 6 milk-production and feed-efficiency traits: milk yield (**MY**), fat %, protein %, milk energy (**MKE**), DMI, and metabolic BW. Phenotypes consisted of averaged data collected during a 6-wk period 50 to 200 d postpartum. Data from an individual cow's most recent parity were used in the analysis. Milk energy was calculated based on the gross energy per kilogram in fat, protein, and lactose. Metabolic BW was computed as BW to the 0.75 power. Each trait was adjusted for known fixed and random factors using a linear mixed-effects model. Fixed effects included herd-year-season at calving, DIM, and both linear and quadratic effects of the continuous variable parity. Experimental ration group was included as a random effect. The residuals for each cow were used as adjusted phenotypes in the following genomic evaluation.

Genotypic Data

The SNP genotypes were available for all 3,303 cows. Genotypes were obtained from the Council on Dairy Cattle Breeding (Bowie, MD) Holstein genotype database. Raw genotypes represented various low- and medium-density arrays, and missing genotypes were imputed to higher density using genotype information from bulls and cows in the Council on Dairy Cattle Breeding database (Wiggans et al., 2016). Any remaining missing genotypes (about 2%) were filled in using rounded allele frequencies from the current US Holstein population. Genotypes at each locus were coded as 0, 1, or 2 copies of the minor allele. A total of 56,575 SNP per cow with a minor allele frequency greater than 5% were available for the genomic evaluation analyses.

Statistical Models

We considered 4 scenarios for genomic evaluation. Scenario 1 excluded herdmates of the test animal in the training population. Scenario 2 included herdmates of the test animal in the training population. Both scenarios 1 and 2 assumed marker effects were constant across environments. Scenario 3 included herdmates of the test animal in the training population and allowed marker-by-herd interactions. Similar to scenario 3, scenario 4 included herdmates in the training population; however, we considered a 2-step approach to modeling marker-by-herd interactions that would allow post hoc adjustment of national EBV, where in step 1 phenotypes (preadjusted for known fixed effects) within each herd are regressed on EBV generated under scenario 1. The regression coefficient from step 1 represents a herd-specific scaling factor (expansion or shrinkage adjustment of the EBV). The EBV is the predicted

$$\mathbf{G}_0 = \begin{bmatrix} \frac{\mathbf{Z}_1 \mathbf{Z}_1'}{\sum 2p_{1j}(1-p_{1j})} & \frac{\mathbf{Z}_1 \mathbf{Z}_2'}{\sqrt{\sum 2p_{1j}(1-p_{1j})}\sqrt{\sum 2p_{2j}(1-p_{2j})}} & \dots & \frac{\mathbf{Z}_1 \mathbf{Z}_n'}{\sqrt{\sum 2p_{1j}(1-p_{1j})}\sqrt{\sum 2p_{nj}(1-p_{nj})}} \\ \frac{\mathbf{Z}_2 \mathbf{Z}_1'}{\sqrt{\sum 2p_{2j}(1-p_{2j})}\sqrt{\sum 2p_{1j}(1-p_{1j})}} & \frac{\mathbf{Z}_2 \mathbf{Z}_2'}{\sum 2p_{2j}(1-p_{2j})} & \dots & \frac{\mathbf{Z}_2 \mathbf{Z}_n'}{\sqrt{\sum 2p_{2j}(1-p_{2j})}\sqrt{\sum 2p_{nj}(1-p_{nj})}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\mathbf{Z}_n \mathbf{Z}_1'}{\sqrt{\sum 2p_{nj}(1-p_{nj})}\sqrt{\sum 2p_{1j}(1-p_{1j})}} & \frac{\mathbf{Z}_n \mathbf{Z}_2'}{\sqrt{\sum 2p_{nj}(1-p_{nj})}\sqrt{\sum 2p_{2j}(1-p_{2j})}} & \dots & \frac{\mathbf{Z}_n \mathbf{Z}_n'}{\sum 2p_{nj}(1-p_{nj})} \end{bmatrix}, \quad [4]$$

deviation from the mean phenotypic performance and assumes that a one-unit difference in EBV equates to a one-unit phenotypic difference for that trait. Fitted values from this model adjust for the expansion or shrinkage of EBV within a particular herd but do not result in the reranking of animals. In step 2 the model residuals within each herd are examined for remaining herd-specific marker effects in an effort to more accurately rank animals within a herd [i.e., the residuals (from step 1) used as the dependent variable in step 2 are computed as $y - \text{intercept} - \text{scaling factor} \times \text{EBV}$]. The herd-adjusted breeding value is equivalent to the sum of the fitted value generated by regressing phenotypes on the scenario 1 EBV (i.e., scenario 1 EBV multiplied by a herd-specific scaling factor) and the EBV generated in step 2.

Interaction Model. The interaction model used in scenario 3 assumed marker effects had 2 components: one that was common across all herds (β_0) and one that was specific for each of the herds in our study population ($\beta_1, \beta_2, \beta_3, \dots$). The equation (assuming 3 herds for ease of notation) was as follows:

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} 1\boldsymbol{\mu}_1 \\ 1\boldsymbol{\mu}_2 \\ 1\boldsymbol{\mu}_3 \end{bmatrix} + \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \beta_0 + \begin{bmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} \boldsymbol{\epsilon}_1 \\ \boldsymbol{\epsilon}_2 \\ \boldsymbol{\epsilon}_3 \end{bmatrix}, \quad [1]$$

where $\mathbf{y}_1, \mathbf{y}_2$, and $\mathbf{y}_3$ represented a vector of phenotypes and $\mathbf{X}_1, \mathbf{X}_2$, and $\mathbf{X}_3$ represented matrices of genotypes for individuals in herd 1, herd 2, and herd 3; $\boldsymbol{\mu}_1, \boldsymbol{\mu}_2$, and $\boldsymbol{\mu}_3$ were herd-specific intercepts; and $\boldsymbol{\epsilon}_1, \boldsymbol{\epsilon}_2$, and $\boldsymbol{\epsilon}_3$ represented a vector of residuals for each of the herds. Residuals were assumed to be independent [$\boldsymbol{\epsilon}_1 \sim N(\mathbf{0}, \mathbf{I}\sigma_{\epsilon_1}^2)$, $\boldsymbol{\epsilon}_2 \sim N(\mathbf{0}, \mathbf{I}\sigma_{\epsilon_2}^2)$, $\boldsymbol{\epsilon}_3 \sim N(\mathbf{0}, \mathbf{I}\sigma_{\epsilon_3}^2)$]; $\sigma_{\epsilon_1}^2, \sigma_{\epsilon_2}^2$, and $\sigma_{\epsilon_3}^2$ were herd-specific residual variances; and $\mathbf{I}$ is an identity matrix.

The model can be represented as a genomic BLUP model, by letting

$$\mathbf{u}_0 = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \beta_0, \quad \mathbf{u}_1 = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} \beta_1, \quad \mathbf{u}_2 = \begin{bmatrix} \mathbf{0} \\ \mathbf{X}_2 \\ \mathbf{0} \end{bmatrix} \beta_2, \quad \mathbf{u}_3 = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \\ \mathbf{X}_3 \end{bmatrix} \beta_3, \quad [2]$$

and represented as

$$\mathbf{y} = \boldsymbol{\mu} + \mathbf{u}_0 + \mathbf{u}_1 + \mathbf{u}_2 + \mathbf{u}_3 + \boldsymbol{\epsilon}, \quad [3]$$

where $\mathbf{u}_0 \sim N(\mathbf{0}, \mathbf{G}_0\sigma_{g_0}^2)$ is a vector of additive genetic effects due to constant marker effects across herds and $\mathbf{u}_1, \mathbf{u}_2$, and $\mathbf{u}_3$ are vectors of herd-specific genetic effects due to deviation of marker effects for a particular herd; $\sigma_{g_0}^2$ is the common genomic variance and $\sigma_{g_1}^2, \sigma_{g_2}^2$, and $\sigma_{g_3}^2$ are herd-specific genomic variances. For example, vector $\mathbf{u}_1$ contains the herd-specific genetic effects for individuals in herd 1. Elements of vector $\mathbf{u}_1$ for individuals in herds 2 and 3 are equal to zero as indicated by Equation [2]. The sum of the across-herd genetic effects ($\mathbf{u}_0$) and herd-specific genetic effects ($\mathbf{u}_1, \mathbf{u}_2$, and $\mathbf{u}_3$) represent a herd-customized breeding value.

The marker-derived genomic relationship matrix $\mathbf{G}_0$ was computed as a multipopulation genomic relationship matrix with each herd being considered a separate population using the method described in the study by Wientjes et al. (2017) and is shown in Equation [4]. In this equation, $\mathbf{Z}_n$ represented a mean-centered genotype matrix for herd n and p_{nj} represented the minor allele frequency of the j th marker in herd n . The genomic relationship matrix for herd n ($\mathbf{G}_n$) was computed as

$$\mathbf{G}_n = \frac{\mathbf{Z}_n \mathbf{Z}_n'}{\sum 2p_{nj}(1-p_{nj})}, \quad [5]$$

where $\mathbf{Z}_n$ represented a mean-centered genotype matrix for herd n and p_{nj} represented the minor allele frequency of the j th marker in herd n . A classical $\mathbf{G}$ matrix $\{[\mathbf{Z}\mathbf{Z}'/\Sigma 2p(1-p)], p = \text{allele frequency in entire study population}\}$ was also considered to estimate the common genomic variance component instead of the multipopulation $\mathbf{G}$ matrix. In this study population, the 2 approaches resulted in almost identical estimation of breeding values for each trait (correlation >0.98) and very similar within-herd prediction accuracies (correlation >0.999999 , mean difference <0.0001), see Supplemental Tables S1–S3 (<https://doi.org/10.3168/jds.2019-16820>).

Across-Herd Model. The across-herd model used in scenarios 1, 2, and step 1 of scenario 4 assumed marker effects were common across all herds (β_0) and could be obtained by setting $\beta_1 = \beta_2 = \beta_3 = 0$ from the interaction model. The equations became

$$\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix} = \begin{bmatrix} \mathbf{1}\mu_1 \\ \mathbf{1}\mu_2 \\ \mathbf{1}\mu_3 \end{bmatrix} + \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \beta_0 + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}, \quad [6]$$

and, when represented as a genomic BLUP model,

$$\mathbf{u}_0 = \begin{bmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \mathbf{X}_3 \end{bmatrix} \beta_0, \quad \mathbf{y} = \boldsymbol{\mu} + \mathbf{u}_0 + \boldsymbol{\epsilon}. \quad [7]$$

Inference on Variance Components

Genomic variance components were estimated from the interaction model using the full data set. The genomic heritabilities (h^2) for each trait within herd i were calculated as

$$h^2 = \frac{(\sigma_{g_i}^2 + \sigma_{g_0}^2)}{(\sigma_{g_i}^2 + \sigma_{g_0}^2 + \sigma_{\epsilon_i}^2)}. \quad [8]$$

Trait-specific genomic correlation for a single herd with the remaining herds was measured as the square root of the proportion of genetic variance within a herd explained by common marker effects using the following equation:

$$\text{cor} = \sqrt{\frac{\sigma_{g_0}^2}{(\sigma_{g_0}^2 + \sigma_{g_i}^2)}}. \quad [9]$$

Assessment of Within-Herd Prediction Accuracy

Prediction accuracy was calculated as the correlation between genomic predicted breeding values and adjusted phenotypes. An efficient leave-one-out cross validation (Cheng et al., 2017) was used to maximize the number of herdmates included in the training population for scenarios 2 through 4. While estimating breeding values within each herd, the training population size was held constant under all 4 scenarios by removing a subset of individuals from the remaining herds in scenarios 2, 3, and step 2 of scenario 4 to maintain an equivalent number of individuals comprising the scenario 1 training population. A paired sample t -test was used to test for significant differences in mean within-herd prediction accuracy among the different scenarios evaluated.

Software

The mixed model equations for generating adjusted phenotypes, estimating (co)variance components, and predicting breeding values were solved with restricted maximum likelihood (REML) method using the R package “sommer” version 3.6 (Covarrubias-Pazarán, 2016).

RESULTS

Population Stratification

The genetic substructure of the data is plotted in Figure 1 using the first and second largest eigenvectors derived from the genomic relationship matrix of all cows. Each herd is highlighted in a separate plot to illustrate the degree of clustering of individuals from a particular herd relative to the remaining herds. In a scenario where genetic similarity within a herd is similar to the level of genetic similarity between random individuals of the remaining herds, one would observe the data points for a particular herd to random overlay data points belonging to the remaining herds. Figure 1 illustrates that for some of the herds, animals within the same herd tend to cluster together, and therefore, animals from these herds are more genetically similar to one another. The 2 most geographically distant herds, the Florida and Canada herds, each cluster toward the outer edges of the plotted data points. The F_{st} values were calculated using Weir and Hill's method (Weir and Hill, 2002) to measure the degree of genetic differentiation of each herd with the remaining herds. The F_{st} is a fixation index ranging from zero to one, with zero indicating no differentiation. The F_{st} values range from 0.002 to 0.010, with the Canadian herd being the most differentiated. The F_{st} values for 10 random subsets of individuals equivalent to the herd sizes in the study population are all less than 0.0001.

Genomic Variance Component Estimation

The genomic variance component estimates for milk-production and feed-efficiency traits for each of the herds are presented in Tables 1 and 2 and Figure 2A. Within-herd heritability determined from the interaction model is reported in Table 1. Decomposition of the variance is presented in Figure 2A. Estimated genomic heritabilities for all of the traits vary among herds. The observed differences are a result of both varying residual variances among herds and varying herd-specific genomic variances (Figure 2A). Trait-specific genomic correlations of individual herds with the remaining herds are reported in Table 2. A genomic correlation estimate less than one for an individual herd with the remaining herds indicates the presence of herd-specific

genomic variance. This herd-specific genomic variance is composed of 2 components: scaling and reranking. Scaling is a herd-specific expansion or shrinkage of the EBV that does not result in reranking of animals within the herd. In general, genomic correlations are high; however, notable differences in genomic correlation are present among herds for MY and MKE traits.

Prediction Accuracy

Table 3 presents the correlation between genomic predicted breeding values and phenotypes for 4 scenarios. For most traits, prediction accuracies are significantly higher for scenarios 2 through 4 in comparison with scenario 1, which includes the same number of indi-

Figure 1. Plot of principal components 1 and 2 for all 3,303 genotyped animals. Dark data points indicate animals from an individual featured herd (Canada, Florida, Iowa, Michigan, Maryland, Wisconsin 1, and Wisconsin 2). F_{st} = fixation index ranging from 0 to 1 that measures population differentiation.

Table 1. Estimates ($\pm$ SE) of within-herd heritability for milk-production and feed-efficiency traits: milk yield (MY), fat %, protein %, milk energy (MKE), DMI, and metabolic BW (MBW)

Herd	MY	Fat %	Protein %	MKE	DMI	MBW
Canada	0.47 $\pm$ 0.14	0.82 $\pm$ 0.09	0.77 $\pm$ 0.09	0.40 $\pm$ 0.14	0.17 $\pm$ 0.12	0.53 $\pm$ 0.11
Florida	0.44 $\pm$ 0.12	0.51 $\pm$ 0.09	0.59 $\pm$ 0.09	0.35 $\pm$ 0.12	0.26 $\pm$ 0.11	0.57 $\pm$ 0.10
Iowa	0.17 $\pm$ 0.05	0.52 $\pm$ 0.05	0.55 $\pm$ 0.04	0.20 $\pm$ 0.05	0.29 $\pm$ 0.05	0.49 $\pm$ 0.05
Michigan	0.50 $\pm$ 0.13	0.70 $\pm$ 0.10	0.60 $\pm$ 0.10	0.44 $\pm$ 0.13	0.29 $\pm$ 0.12	0.44 $\pm$ 0.10
Maryland	0.21 $\pm$ 0.07	0.52 $\pm$ 0.06	0.56 $\pm$ 0.06	0.16 $\pm$ 0.07	0.27 $\pm$ 0.07	0.45 $\pm$ 0.06
Wisconsin 1	0.23 $\pm$ 0.11	0.51 $\pm$ 0.09	0.54 $\pm$ 0.08	0.11 $\pm$ 0.10	0.33 $\pm$ 0.11	0.46 $\pm$ 0.09
Wisconsin 2	0.15 $\pm$ 0.05	0.50 $\pm$ 0.05	0.53 $\pm$ 0.05	0.14 $\pm$ 0.05	0.21 $\pm$ 0.05	0.44 $\pm$ 0.05

viduals in the training population but does not include herdmates of the test animals. Inclusion of herdmates in the training population improved mean within-herd prediction accuracy for milk-production traits ($\pm$ SE) by 0.08 ± 0.03 (MY; $P < 0.05$), 0.07 ± 0.03 (fat %; $P < 0.05$), and 0.05 ± 0.01 (protein %; $P < 0.05$) and feed-efficiency traits by 0.07 ± 0.02 (MKE; $P < 0.05$), 0.03 ± 0.02 (DMI; NS), and 0.08 ± 0.01 (metabolic BW; $P < 0.001$).

Scenarios 2 and 3 both include herdmates in the training population; however, scenario 2 assumes marker effects are constant across herds, whereas scenario 3 allows marker effects to have a herd-specific component. Inclusion of this interaction term further improves mean within-herd prediction accuracy for MY and MKE ($\pm$ SE) by 0.03 ± 0.01 and 0.02 ± 0.01 , respectively. More notable improvements were observed for herds with higher heritabilities but low genomic correlation with the other herds.

Scenarios 3 and 4 both include a herd-specific genomic variance component; however, scenario 4 takes a 2-step approach to modeling herd-specific marker effects. In scenario 4, step 1 involves determining genomic predicted breeding values in a manner equivalent to scenario 1 (assumes constant marker effects across herds, herdmates not included in the training population). The second step involves a within-herd analysis of residuals obtained from regressing phenotypes on the step 1 genomic predicted breeding values to determine whether additional herd-specific genomic variance is present. In the case that it does exist, step 1 predicted breeding values are adjusted for herd-specific marker

effects determined in step 2. The herd adjusted breeding value is equivalent to the sum of the fitted value generated by regressing phenotypes on step 1 EBV (i.e., scenario 1 EBV multiplied by a herd-specific scaling factor, see Table 4) and the EBV generated in step 2. The 2-step approach yields very similar accuracies to scenario 3 with mean within-herd accuracy differences less than 0.02 for all traits.

The regression coefficient for phenotypes regressed on EBV should be close to one. A one-unit change in EBV should equate to a one-unit change in phenotype. Within-herd regression coefficients for EBV generated under scenarios 1 through 4 are reported in Table 4. The regression coefficients represent a herd-specific scaling (expansion or shrinkage) of the EBV. Bias (deviation from 1) is reduced with the inclusion of herd-mate information and explicit modeling of genotype $\times$ herd interactions.

Figure 2 illustrates increased gains in prediction accuracy achievable with inclusion of herd-mate data, especially for herds with a large herd-specific variance component. Gain in within-herd prediction accuracy indicates the herd-specific genomic variance is due to reranking of animals within a herd and not just herd-specific scaling (shrinkage or expansion) of the EBV. The within-herd correlation of herd-mate information unadjusted (scenario 1) and adjusted (scenario 4 – 2-step genotype $\times$ herd approach) genomic predicted breeding values is reported in Table 5. A correlation coefficient < 1 indicates the presence of herd-specific genomic variance independent of scaling effects. The 2-step approach to modeling marker-by-herd interactions

Table 2. Estimates ($\pm$ SE) of genomic correlation with other North American herds for milk-production and feed-efficiency traits: milk yield (MY), fat %, protein %, milk energy (MKE), DMI, and metabolic BW (MBW)

Herd	MY	Fat %	Protein %	MKE	DMI	MBW
Canada	0.74 $\pm$ 0.14	0.97 $\pm$ 0.08	1.00 $\pm$ 0.08	0.74 $\pm$ 0.16	1.00 $\pm$ 0.36	1.00 $\pm$ 0.12
Florida	0.64 $\pm$ 0.10	1.00 $\pm$ 0.10	0.96 $\pm$ 0.08	0.61 $\pm$ 0.12	0.85 $\pm$ 0.18	0.85 $\pm$ 0.09
Iowa	1.00 $\pm$ 0.15	0.88 $\pm$ 0.05	1.00 $\pm$ 0.04	0.75 $\pm$ 0.12	0.89 $\pm$ 0.09	1.00 $\pm$ 0.05
Michigan	0.56 $\pm$ 0.09	0.78 $\pm$ 0.07	0.88 $\pm$ 0.09	0.58 $\pm$ 0.11	1.00 $\pm$ 0.21	1.00 $\pm$ 0.12
Maryland	0.83 $\pm$ 0.15	0.90 $\pm$ 0.07	0.97 $\pm$ 0.06	0.85 $\pm$ 0.19	0.89 $\pm$ 0.13	1.00 $\pm$ 0.07
Wisconsin 1	0.81 $\pm$ 0.20	0.97 $\pm$ 0.09	1.00 $\pm$ 0.08	1.00 $\pm$ 0.46	0.74 $\pm$ 0.14	0.94 $\pm$ 0.10
Wisconsin 2	1.00 $\pm$ 0.18	0.95 $\pm$ 0.06	1.00 $\pm$ 0.05	1.00 $\pm$ 0.20	1.00 $\pm$ 0.13	1.00 $\pm$ 0.06

would facilitate post hoc adjustment of national EBV. The gains in within-herd explained phenotypic variance using this approach are also reported in Table 5 along with the improvement in ranking accuracy (measured by Spearman's correlation coefficient). Improvement in ranking accuracy is observed for herds with herd-specific genomic variance that is independent of scaling effects (i.e., herds that demonstrate gain in phenotypic variation explained by herdmate information adjusted

EBV and herds that demonstrate decreased correlation of herdmate information adjusted and unadjusted EBV). For example, the herdmate information adjusted EBV and unadjusted EBV exhibit near zero correlation ($r = -0.03$) for the milk-yield trait in the Florida herd. However, the herdmate information adjusted EBV better explains phenotypic variation within the herd, and a large gain in ranking accuracy is attained with the herdmate adjusted EBV.

Figure 2. (A) Plots of within-herd variance decomposition of milk-production and feed-efficiency traits estimated from the interaction model using the full data set. (B) Bar plots of improvement in predictive ability determined under 4 scenarios: scenario 1 = exclusion of herd mates from training population; scenario 2 = inclusion of herd mates in training population; scenario 3 = inclusion of herd mates in training population and modeling of herd-specific genetic effects; scenario 4 = inclusion of herd mates in training population and modeling of herd-specific genetic effects using 2-step approach. MY = milk yield; MKE = milk energy content; MBW = metabolic BW; CAN = Canada.

DISCUSSION

In the near future, many dairy herds will have complete genomic data as increasing numbers of dairy farmers incorporate genomic testing as a routine management tool. This is a potentially rich and untapped source of contextual information that can be leveraged to capture genotype-by-herd interactions and tailor breeding programs to the individual dairy producer. Results from the present study reveal the importance of incorporating herdmate data into genomic evaluations. Significant gains in within-herd prediction accuracy are attainable when herdmate data are included in the training population. The inclusion of herd-specific genetic effects will improve the accuracy of selection decisions, particularly on farms where the genetic background of the herd or the environmental conditions differ from the typical herd in the national reference population.

The observed improvement in prediction accuracy may be multifactorial. It is well known that accuracy of genomic prediction is dependent on the degree of relatedness between the selection candidates and the reference population (Habier et al., 2007, Edwards et al., 2018). Schultz and Weigel (2018) demonstrated in data simulated to be similar to Holstein herd structure, improved prediction accuracy with inclusion of herdmates irrespective of the prediction method (genomic BLUP vs. a Bayesian approach). Herdmates tend to have similar genetic background, due to reliance on semen from a single AI company; consistency in the herd's breeding objectives over time; and retention of progeny within the herd. Second, improvement in prediction accuracy may be in part due to the presence and capture of gene-by-herd interactions, even when they are not explicitly modeled. With the limited number of herds included in this study, herd-specific genetic effects may have a strong influence on the across-herd marker effect esti-

Table 3. Within-herd correlation estimates ($\pm$ SE) for genomic predicted breeding values with milk-production and feed-efficiency traits determined under 4 scenarios

Item	Herd						
	Canada	Florida	Iowa	Michigan	Maryland	Wisconsin 1	Wisconsin 2
MY							
Scenario 1	0.11 $\pm$ 0.07	0.03 $\pm$ 0.06	0.14 $\pm$ 0.03	0.14 $\pm$ 0.06	0.17 $\pm$ 0.04	0.26 $\pm$ 0.05	0.21 $\pm$ 0.04
Scenario 2	0.30 $\pm$ 0.07***	0.21 $\pm$ 0.05***	0.18 $\pm$ 0.03	0.23 $\pm$ 0.06***	0.23 $\pm$ 0.04*	0.29 $\pm$ 0.05	0.18 $\pm$ 0.04
Scenario 3	0.33 $\pm$ 0.06***	0.27 $\pm$ 0.05***	0.18 $\pm$ 0.03	0.33 $\pm$ 0.06**	0.24 $\pm$ 0.04*	0.28 $\pm$ 0.05	0.18 $\pm$ 0.04
Scenario 4	0.33 $\pm$ 0.06**	0.27 $\pm$ 0.05***	0.19 $\pm$ 0.03*	0.32 $\pm$ 0.06**	0.23 $\pm$ 0.04*	0.29 $\pm$ 0.05	0.21 $\pm$ 0.04
Fat%							
Scenario 1	0.15 $\pm$ 0.07	0.34 $\pm$ 0.05	0.32 $\pm$ 0.03	0.45 $\pm$ 0.06	0.44 $\pm$ 0.04	0.44 $\pm$ 0.05	0.46 $\pm$ 0.03
Scenario 2	0.32 $\pm$ 0.06***	0.30 $\pm$ 0.05	0.47 $\pm$ 0.03***	0.52 $\pm$ 0.05**	0.50 $\pm$ 0.04**	0.47 $\pm$ 0.05	0.49 $\pm$ 0.03
Scenario 3	0.32 $\pm$ 0.06***	0.30 $\pm$ 0.05	0.47 $\pm$ 0.03***	0.52 $\pm$ 0.05*	0.50 $\pm$ 0.04**	0.46 $\pm$ 0.05	0.49 $\pm$ 0.03
Scenario 4	0.30 $\pm$ 0.07*	0.34 $\pm$ 0.05	0.47 $\pm$ 0.03***	0.53 $\pm$ 0.05*	0.52 $\pm$ 0.04***	0.46 $\pm$ 0.05	0.52 $\pm$ 0.03***
Protein%							
Scenario 1	0.45 $\pm$ 0.06	0.35 $\pm$ 0.05	0.44 $\pm$ 0.03	0.57 $\pm$ 0.05	0.51 $\pm$ 0.04	0.49 $\pm$ 0.05	0.52 $\pm$ 0.03
Scenario 2	0.51 $\pm$ 0.06*	0.43 $\pm$ 0.05**	0.49 $\pm$ 0.03**	0.59 $\pm$ 0.05	0.58 $\pm$ 0.04***	0.53 $\pm$ 0.05*	0.54 $\pm$ 0.03
Scenario 3	0.51 $\pm$ 0.06*	0.43 $\pm$ 0.05**	0.49 $\pm$ 0.03**	0.60 $\pm$ 0.05	0.58 $\pm$ 0.04***	0.53 $\pm$ 0.05*	0.54 $\pm$ 0.03
Scenario 4	0.51 $\pm$ 0.06*	0.41 $\pm$ 0.05*	0.52 $\pm$ 0.03***	0.60 $\pm$ 0.05*	0.58 $\pm$ 0.04***	0.52 $\pm$ 0.05	0.55 $\pm$ 0.03***
MKE							
Scenario 1	0.12 $\pm$ 0.07	0.09 $\pm$ 0.05	0.15 $\pm$ 0.03	0.12 $\pm$ 0.06	0.12 $\pm$ 0.05	0.17 $\pm$ 0.05	0.16 $\pm$ 0.04
Scenario 2	0.27 $\pm$ 0.07**	0.22 $\pm$ 0.05***	0.22 $\pm$ 0.03**	0.19 $\pm$ 0.06*	0.18 $\pm$ 0.04*	0.19 $\pm$ 0.05	0.13 $\pm$ 0.04
Scenario 3	0.30 $\pm$ 0.07**	0.26 $\pm$ 0.05**	0.23 $\pm$ 0.03**	0.26 $\pm$ 0.06*	0.18 $\pm$ 0.04*	0.18 $\pm$ 0.05	0.12 $\pm$ 0.04
Scenario 4	0.30 $\pm$ 0.07**	0.27 $\pm$ 0.05**	0.22 $\pm$ 0.03**	0.26 $\pm$ 0.06*	0.18 $\pm$ 0.04	0.18 $\pm$ 0.05	0.16 $\pm$ 0.04
DMI							
Scenario 1	0.13 $\pm$ 0.07	0.21 $\pm$ 0.05	0.21 $\pm$ 0.03	0.22 $\pm$ 0.06	0.24 $\pm$ 0.04	0.20 $\pm$ 0.05	0.19 $\pm$ 0.04
Scenario 2	0.07 $\pm$ 0.07	0.25 $\pm$ 0.05	0.32 $\pm$ 0.03***	0.26 $\pm$ 0.06	0.29 $\pm$ 0.04*	0.26 $\pm$ 0.05**	0.18 $\pm$ 0.04
Scenario 3	0.06 $\pm$ 0.07	0.25 $\pm$ 0.05	0.32 $\pm$ 0.03***	0.26 $\pm$ 0.06	0.30 $\pm$ 0.04	0.28 $\pm$ 0.05*	0.19 $\pm$ 0.04
Scenario 4	0.13 $\pm$ 0.07	0.26 $\pm$ 0.05	0.32 $\pm$ 0.03***	0.26 $\pm$ 0.06	0.31 $\pm$ 0.04*	0.30 $\pm$ 0.05**	0.22 $\pm$ 0.04
MBW							
Scenario 1	0.34 $\pm$ 0.06	0.29 $\pm$ 0.05	0.37 $\pm$ 0.03	0.45 $\pm$ 0.06	0.38 $\pm$ 0.04	0.26 $\pm$ 0.05	0.34 $\pm$ 0.03
Scenario 2	0.38 $\pm$ 0.06	0.43 $\pm$ 0.05***	0.45 $\pm$ 0.03***	0.51 $\pm$ 0.05**	0.44 $\pm$ 0.04**	0.35 $\pm$ 0.05***	0.43 $\pm$ 0.03***
Scenario 3	0.39 $\pm$ 0.06	0.44 $\pm$ 0.05***	0.45 $\pm$ 0.03***	0.51 $\pm$ 0.05**	0.44 $\pm$ 0.04**	0.36 $\pm$ 0.05***	0.43 $\pm$ 0.03***
Scenario 4	0.38 $\pm$ 0.06	0.42 $\pm$ 0.05***	0.45 $\pm$ 0.03***	0.47 $\pm$ 0.06	0.45 $\pm$ 0.04**	0.33 $\pm$ 0.05*	0.43 $\pm$ 0.03***

¹Scenario 1 = training population excludes herdmates; scenario 2 = training population includes herdmates; scenario 3 = training population includes herdmates, and the genomic prediction model includes both constant across-herd and herd-specific genetic marker effects; scenario 4 = 2-step genotype-by-herd interaction model. Step 1 = regress phenotypes within each herd on EBV generated under scenario 1; step 2 = examine step 1 residuals within each herd for remaining herd-specific genetic variance, estimate herd-specific marker effects, and adjust breeding values from scenario 1 (adjusted EBV = step 1 fitted value + step 2 EBV). MY = milk yield; MKE = milk energy content; MBW = metabolic BW.

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ (Hotelling's overlapping correlated correlation coefficient comparison test for differences from scenario 1 correlation estimates).

Table 4. Regression coefficients ($\pm$ SE) for milk production and feed efficiency phenotypes regressed on genomic predicted breeding values determined under 4 scenarios

Item ¹	Herd						
	Canada	Florida	Iowa	Michigan	Maryland	Wisconsin 1	Wisconsin 2
MY							
Scenario 1	0.52 $\pm$ 0.32	0.12 $\pm$ 0.25	0.71 $\pm$ 0.17	0.67 $\pm$ 0.31	0.88 $\pm$ 0.23	1.38 $\pm$ 0.28	1.13 $\pm$ 0.19
Scenario 2	1.04 $\pm$ 0.23	0.94 $\pm$ 0.24	0.82 $\pm$ 0.15	1.05 $\pm$ 0.28	1.06 $\pm$ 0.20	1.34 $\pm$ 0.24	0.98 $\pm$ 0.20
Scenario 3	0.98 $\pm$ 0.19	0.90 $\pm$ 0.17	0.99 $\pm$ 0.18	0.99 $\pm$ 0.18	1.02 $\pm$ 0.19	1.23 $\pm$ 0.23	1.08 $\pm$ 0.22
Scenario 4	1.06 $\pm$ 0.20	1.04 $\pm$ 0.20	1.00 $\pm$ 0.17	1.03 $\pm$ 0.19	0.94 $\pm$ 0.18	1.00 $\pm$ 0.18	1.00 $\pm$ 0.17
Fat %							
Scenario 1	0.19 $\pm$ 0.09	0.63 $\pm$ 0.10	0.85 $\pm$ 0.08	1.10 $\pm$ 0.14	0.96 $\pm$ 0.09	1.14 $\pm$ 0.13	1.40 $\pm$ 0.10
Scenario 2	0.52 $\pm$ 0.11	0.60 $\pm$ 0.10	1.03 $\pm$ 0.07	1.17 $\pm$ 0.12	0.95 $\pm$ 0.08	1.08 $\pm$ 0.11	1.13 $\pm$ 0.07
Scenario 3	0.56 $\pm$ 0.11	0.64 $\pm$ 0.11	1.00 $\pm$ 0.06	1.06 $\pm$ 0.11	0.96 $\pm$ 0.08	1.10 $\pm$ 0.12	1.10 $\pm$ 0.07
Scenario 4	0.99 $\pm$ 0.22	1.01 $\pm$ 0.15	1.00 $\pm$ 0.06	0.97 $\pm$ 0.10	0.98 $\pm$ 0.07	0.95 $\pm$ 0.10	1.01 $\pm$ 0.06
Protein %							
Scenario 1	0.81 $\pm$ 0.11	0.71 $\pm$ 0.10	0.89 $\pm$ 0.06	1.38 $\pm$ 0.12	0.99 $\pm$ 0.08	1.00 $\pm$ 0.10	1.21 $\pm$ 0.07
Scenario 2	0.83 $\pm$ 0.10	0.89 $\pm$ 0.10	0.96 $\pm$ 0.06	1.26 $\pm$ 0.11	1.01 $\pm$ 0.06	1.06 $\pm$ 0.09	1.08 $\pm$ 0.06
Scenario 3	0.84 $\pm$ 0.10	0.89 $\pm$ 0.10	0.98 $\pm$ 0.06	1.16 $\pm$ 0.10	0.99 $\pm$ 0.06	1.07 $\pm$ 0.09	1.09 $\pm$ 0.06
Scenario 4	0.99 $\pm$ 0.12	1.08 $\pm$ 0.13	1.02 $\pm$ 0.06	0.98 $\pm$ 0.08	0.98 $\pm$ 0.06	0.99 $\pm$ 0.09	1.00 $\pm$ 0.05
MKE							
Scenario 1	0.60 $\pm$ 0.33	0.55 $\pm$ 0.33	1.05 $\pm$ 0.23	0.65 $\pm$ 0.34	0.64 $\pm$ 0.24	1.17 $\pm$ 0.36	0.83 $\pm$ 0.18
Scenario 2	1.02 $\pm$ 0.25	1.12 $\pm$ 0.27	1.01 $\pm$ 0.15	0.96 $\pm$ 0.32	0.93 $\pm$ 0.24	1.23 $\pm$ 0.35	0.69 $\pm$ 0.20
Scenario 3	0.98 $\pm$ 0.22	0.95 $\pm$ 0.19	1.00 $\pm$ 0.14	0.94 $\pm$ 0.22	1.01 $\pm$ 0.25	1.49 $\pm$ 0.45	0.88 $\pm$ 0.26
Scenario 4	1.04 $\pm$ 0.23	1.01 $\pm$ 0.20	0.95 $\pm$ 0.14	0.99 $\pm$ 0.23	1.00 $\pm$ 0.26	1.03 $\pm$ 0.30	0.83 $\pm$ 0.18
DMI							
Scenario 1	0.68 $\pm$ 0.35	1.04 $\pm$ 0.27	1.09 $\pm$ 0.17	0.77 $\pm$ 0.22	0.97 $\pm$ 0.18	0.92 $\pm$ 0.25	0.77 $\pm$ 0.14
Scenario 2	0.34 $\pm$ 0.34	1.16 $\pm$ 0.24	1.11 $\pm$ 0.11	0.84 $\pm$ 0.20	0.98 $\pm$ 0.15	1.06 $\pm$ 0.21	0.78 $\pm$ 0.15
Scenario 3	0.35 $\pm$ 0.38	1.17 $\pm$ 0.24	1.06 $\pm$ 0.10	0.91 $\pm$ 0.21	1.02 $\pm$ 0.15	1.03 $\pm$ 0.19	0.91 $\pm$ 0.17
Scenario 4	0.68 $\pm$ 0.35	1.10 $\pm$ 0.22	1.03 $\pm$ 0.10	1.03 $\pm$ 0.24	1.05 $\pm$ 0.15	1.06 $\pm$ 0.19	1.00 $\pm$ 0.16
MBW							
Scenario 1	0.86 $\pm$ 0.16	0.81 $\pm$ 0.15	1.00 $\pm$ 0.08	1.24 $\pm$ 0.16	0.97 $\pm$ 0.11	0.72 $\pm$ 0.15	0.96 $\pm$ 0.09
Scenario 2	0.87 $\pm$ 0.14	1.12 $\pm$ 0.13	0.95 $\pm$ 0.06	1.18 $\pm$ 0.13	0.93 $\pm$ 0.09	0.91 $\pm$ 0.13	1.04 $\pm$ 0.08
Scenario 3	0.89 $\pm$ 0.15	1.05 $\pm$ 0.12	0.98 $\pm$ 0.06	1.19 $\pm$ 0.13	0.95 $\pm$ 0.09	0.91 $\pm$ 0.13	1.06 $\pm$ 0.08
Scenario 4	1.03 $\pm$ 0.17	1.09 $\pm$ 0.13	0.98 $\pm$ 0.06	1.01 $\pm$ 0.12	0.95 $\pm$ 0.09	1.06 $\pm$ 0.17	1.00 $\pm$ 0.07

¹Scenario 1 = training population excludes herdmates; scenario 2 = training population includes herdmates; scenario 3 = training population includes herdmates, and the genomic prediction model includes both constant across-herd and herd-specific genetic marker effects; scenario 4 = 2-step genotype-by-herd interaction model. Step 1 = regress phenotypes within each herd on EBV generated under scenario 1; step 2 = examine step 1 residuals within each herd for remaining herd-specific genetic variance, estimate herd-specific marker effects, and adjust breeding values from scenario 1 (adjusted EBV = step 1 fitted value + step 2 EBV). MY = milk yield; MKE = milk energy content; MBW = metabolic BW.

mation. The simple inclusion of herdmates in the training data set while assuming a constant marker effect across herds revealed large improvements in prediction accuracy, whereas explicitly modeling marker-by-herd interactions often resulted in only modest further improvement.

Results from this study reveal herd-specific genetic variance contributes to phenotypic variance. Interestingly, the herds included in this study were previously analyzed as part of a larger data set to identify genotype-by-region interactions for North America, Scotland, and the Netherlands (Yao et al., 2017). The authors observed that prediction accuracy was not a reflection of the size of the reference population. For example, the North American population was largest but accuracies were not the best for any trait, whereas the Scotland reference population was the smallest, composed of only 2 herds, but had the highest accuracies for all traits examined except residual feed intake. The authors concluded greater genetic and environmental variation

in the North American data may have contributed to the reduced accuracy of that region. The herd-specific genetic effects observed among the North American herds in our current study support their conclusion and demonstrate opportunities for improvement of genomic evaluations. For example, the within-herd heritability estimate for milk production in the Florida herd was relatively high, given one might expect the increased environmental stress in that geographic region would result in lower heritability. Results from this study indicate that within-herd heritability for the Florida herd would be much lower if herd-specific genetic effects were not modeled. A large portion of the genomic variance within the Florida herd is herd-specific, and therefore, within-herd ranking of animals will benefit more from EBV adjustment for herd-specific genetic effects. The herdmate information adjusted and unadjusted EBV were very poorly correlated. The climate of the Florida herd is very different compared with that of the remaining herds in the study and may be a source of

herd-specific genetic effects on milk yield. Adjustment of EBV within the Florida herd for the herd-specific genetic effects results in a substantial increase in ranking accuracy.

Practical application of the study findings for genotyped herds may entail an examination of nationally EBV correlation with observed phenotypes within a herd. The national EBV has the benefits of a very large reference population and ability to most accurately estimate small constant marker effects across herds. Poor correlation may indicate unaccounted for herd-specific genetic effects. Available phenotypes within the herd (ideally from a single snapshot of time to reduce random noise) should be regressed on the national EBV. The regression coefficient from this model represents a herd-specific expansion or shrinkage adjustment of the EBV. The EBV is the predicted deviation from the mean phenotypic performance and assumes that a one-unit difference in EBV equates to a one-unit phenotypic difference for that trait. Fitted values from this model adjust for the expansion or shrinkage of EBV within a particular herd but do not result in the reranking of animals. Although, expansion or shrinkage of EBV for component traits of net merit indices could result in reranking of animals within a herd based on net merit.

In an effort to more accurately rank animals within a herd, phenotypes preadjusted for national EBV effects can be examined for additional genetic variance not captured by the nationally estimated constant marker effects. If the genetic variance is greater than zero, cross validation can be used to determine whether within-herd prediction accuracy improves with modeling of herd-specific genetic effects. Results from this study indicate this 2-step approach is an effective method of modeling herd-specific genetic effects and improving within-herd prediction accuracy. Even small changes in the ranking of animals due to herd-specific genetic effects may be effective when dairy producers are making selection and culling decisions every day.

The authors acknowledge the sample size in this study is small for accurate variance component estimation. Replacement of the scenario 1 EBV in this study with a national EBV determined from a large reference population would more accurately estimate the common genomic variance within herds, which might in turn reduce the gains in prediction accuracy observed in this study. Further research in this area is warranted.

The 2-step approach in this study focused on evaluating the presence of herd-specific genetic effects and their effect on prediction accuracy. A logical direction

Table 5. Within-herd correlation of genomic predicted breeding values determined under scenarios with exclusion or inclusion of herdmate information and differences in explained phenotype variation (R^2) and ranking accuracy for milk-production and feed-efficiency traits

Item ¹	Herd						
	Canada	Florida	Iowa	Michigan	Maryland	Wisconsin 1	Wisconsin 2
MY							
EBV correlation (Sc1, Sc4)	0.34	-0.03	0.72	0.38	0.74	0.90	1.00
ΔR^2 (Sc4 - Sc1)	0.10	0.07	0.02	0.09	0.03	0.02	0.00
Δ Ranking accuracy (Sc4 - Sc1)	0.22	0.19	0.06	0.17	0.04	0.03	0.00
Fat %							
EBV correlation (Sc1, Sc4)	0.44	0.99	0.66	0.82	0.85	0.94	0.87
ΔR^2 (Sc4 - Sc1)	0.07	0.00	0.12	0.07	0.07	0.02	0.06
Δ Ranking accuracy (Sc4 - Sc1)	0.14	0.01	0.15	0.08	0.07	0.02	0.06
Protein %							
EBV correlation (Sc1, Sc4)	0.88	0.85	0.84	0.96	0.88	0.94	0.94
ΔR^2 (Sc4 - Sc1)	0.06	0.04	0.08	0.03	0.07	0.02	0.04
Δ Ranking accuracy (Sc4 - Sc1)	0.04	0.06	0.08	0.01	0.06	0.01	0.03
MKE							
EBV correlation (Sc1, Sc4)	0.43	0.35	0.69	0.38	0.66	0.95	1.00
ΔR^2 (Sc4 - Sc1)	0.07	0.06	0.03	0.05	0.02	0.00	0.00
Δ Ranking accuracy (Sc4 - Sc1)	0.15	0.13	0.09	0.16	0.06	0.02	0.00
DMI							
EBV correlation (Sc1, Sc4)	1.00	0.75	0.65	0.84	0.78	0.69	0.89
ΔR^2 (Sc4 - Sc1)	0.00	0.03	0.06	0.02	0.04	0.05	0.01
Δ Ranking accuracy (Sc4 - Sc1)	0.00	0.06	0.12	0.03	0.08	0.09	0.02
MBW							
EBV correlation (Sc1, Sc4)	0.90	0.69	0.81	0.94	0.87	0.80	0.79
ΔR^2 (Sc4 - Sc1)	0.03	0.09	0.07	0.03	0.05	0.04	0.07
Δ Ranking accuracy (Sc4 - Sc1)	0.07	0.13	0.08	0.01	0.07	0.05	0.10

¹Scenario 1 (Sc1) = training population excludes herdmates; scenario 4 (Sc4) = 2-step genotype-by-herd interaction model. Step 1 = regress phenotypes within each herd on EBV generated under scenario 1; step 2 = examine step 1 residuals within each herd for remaining herd-specific genetic variance, estimate herd-specific marker effects, and adjust breeding values from scenario 1 (adjusted EBV = step 1 fitted value + step 2 EBV). ΔR^2 = difference in R^2 values determined from phenotype regression on EBV; Δ Ranking accuracy = difference in Spearman's rank correlation coefficients for phenotype and EBV; MY = milk yield; MKE = milk energy content; MBW = metabolic BW.

for future research would be to identify regions of the genome associated with large herd-specific genetic effects. A similar 2-step approach could be performed to identify markers within a herd associated with phenotypes preadjusted for national EBV. Replication of associated markers in multiple herds would serve as validation and facilitate mapping of candidate genes that may serve as targets for management strategies to optimize a herd's genetic potential.

CONCLUSIONS

Herdmates tend to be more genetically similar to each other than animals belonging to separate herds, in addition to sharing a common environment and production management system. Thus, herdmates provide an excellent example of how the test animal's genetics will perform in the current herd environment. Significant gains in within-herd prediction accuracy are attainable by well-representing herdmates of the test animals in the training population. We constructed a novel, but practical, approach to incorporate herd-specific marker effects in genomic predicted breeding values. The approach can be applied on an individual-herd basis by performing a post hoc adjustment of national EBV for herd-specific genetic effects. Herd-customized approaches will aid dairy producers in optimally using their genomic data investment and improve selection decisions involving animals in their herds.

ACKNOWLEDGMENTS

This project was supported by Agriculture and Food Research Initiative Competitive Grants no. 2017-67012-26108 and 2011-68004-30340 from the USDA National Institute of Food and Agriculture (Washington, DC). This study uses data generated by the USDA–Agriculture and Food Research Initiative project supported by grant no. 2011-68004-30340 (VandeHaar et al., 2011), and data have been made publicly available in the form of data sharing with scientists from the USDA–Agricultural Research Service and Council on Dairy Cattle Breeding (Bowie, MD).

REFERENCES

- Cheng, H., D. J. Garrick, and R. L. Fernando. 2017. Efficient strategies for leave-one-out cross validation for genomic best linear unbiased prediction. *J. Anim. Sci. Biotechnol.* 8:38.
- Covarrubias-Pazaran, G. 2016. Genome-assisted prediction of quantitative traits using the R package sommer. *PLoS One* 11:e0156744.
- Edwards, S. M., J. B. Buntjer, R. Jackson, A. R. Bentley, J. Lage, E. Byrne, C. Burt, P. Jack, S. Berry, and E. Flatman. 2018. The effects of training population design on genomic prediction accuracy in wheat. *bioRxiv*. <https://doi.org/10.1101/443267>.
- Habier, D., R. L. Fernando, and J. C. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397.
- Meuwissen, T. H., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.
- Schultz, N. E., and K. A. Weigel. 2018. An improved genomic prediction model in populations featuring shared environments and familial relatedness. Page 520 in *Proc. World Congr. Genet. Appl. Livest. Prod.*, Auckland, New Zealand.
- Strandberg, E., S. Brotherstone, E. Wall, and M. P. Coffey. 2009. Genotype by environment interaction for first-lactation female fertility traits in UK dairy cattle. *J. Dairy Sci.* 92:3437–3446.
- Tempelman, R. J., D. M. Spurlock, M. Coffey, R. F. Veerkamp, L. E. Armentano, K. A. Weigel, Y. de Haas, C. R. Staples, E. E. Connor, Y. Lu, and M. J. VandeHaar. 2015. Heterogeneity in genetic and nongenetic variation and energy sink relationships for residual feed intake across research stations and countries. *J. Dairy Sci.* 98:2013–2026.
- Tiezzi, F., G. de Los Campos, K. L. Parker Gaddis, and C. Maltecca. 2017. Genotype by environment (climate) interaction improves genomic prediction for production traits in US Holstein cattle. *J. Dairy Sci.* 100:2042–2056.
- Weir, B. S., and W. G. Hill. 2002. Estimating F-statistics. *Annu. Rev. Genet.* 36:721–750.
- Wientjes, Y. C. J., P. Bijma, J. Vandenplas, and M. P. L. Calus. 2017. Multi-population genomic relationships for estimating current genetic variances within and genetic correlations between populations. *Genetics* 207:503–515.
- Wiggans, G. R., J. B. Cole, S. M. Hubbard, and T. S. Sonstegard. 2017. Genomic selection in dairy cattle: The USDA experience. *Annu. Rev. Anim. Biosci.* 5:309–327.
- Wiggans, G. R., T. A. Cooper, P. M. VanRaden, C. P. Van Tassell, D. M. Bickhart, and T. S. Sonstegard. 2016. Increasing the number of single nucleotide polymorphisms used in genomic evaluation of dairy cattle. *J. Dairy Sci.* 99:4504–4511.
- Yao, C., G. De Los Campos, M. VandeHaar, D. Spurlock, L. Armentano, M. Coffey, Y. De Haas, R. Veerkamp, C. Staples, and E. Connor. 2017. Use of genotype × environment interaction model to accommodate genetic heterogeneity for residual feed intake, dry matter intake, net energy in milk, and metabolic body weight in dairy cattle. *J. Dairy Sci.* 100:2007–2016.