J. Dairy Sci. 100:1–10
https://doi.org/10.3168/jds.2016-11606

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

C. Yao,*[1] G. de los Campos,† M. J. VandeHaar,† D. M. Spurlock,‡ L. E. Armentano,* M. Coffey,§ Y. de Haas,# R. F. Veerkamp,# C. R. Staples,‖ E. E. Connor,¶ Z. Wang, M. D. Hanigan,†† R. J. Tempelman,† and K. A. Weigel***

*Department of Dairy Science, University of Wisconsin, Madison 53706
†Department of Animal Science, Michigan State University, East Lansing 48824
‡Department of Animal Science, Iowa State University, Ames 50011
§Scottish Agricultural College, Easter Bush, Midlothian, EH25 9RG, UK
#Wageningen UR Livestock Research, Wageningen, 6700 AH, the Netherlands
‖Department of Animal Sciences, University of Florida, Gainesville 32611
¶Animal Genomics and Improvement Laboratory, Agricultural Research Service, USDA, Beltsville, MD 20705
**Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB, T6G 2P5, Canada
††Department of Dairy Science, Virginia Tech, Blacksburg 24061

ABSTRACT

Feed efficiency in dairy cattle has gained much attention recently. Due to the cost-prohibitive measurement of individual feed intakes, combining data from multiple countries is often necessary to ensure an adequate reference population. It may then be essential to model genetic heterogeneity when making inferences about feed efficiency or selecting efficient cattle using genomic information. In this study, we constructed a marker × environment interaction model that decomposed marker effects into main effects and interaction components that were specific to each environment. We compared environment-specific variance component estimates and prediction accuracies from the interaction model analyses, an across-environment analyses ignoring population stratification, and a within-environment analyses using an international feed efficiency data set. Phenotypes included residual feed intake, dry matter intake, net energy in milk, and metabolic body weight from 3,656 cows measured in 3 broadly defined environments: North America (NAM), the Netherlands (NLD), and Scotland (SAC). Genotypic data included 57,574 single nucleotide polymorphisms per animal. The interaction model gave the highest prediction accuracy for metabolic body weight, which had the largest estimated heritabilities ranging from 0.37 to 0.55. The within-environment model performed the best when predicting residual feed intake, which had the lowest estimated heritabilities ranging from 0.13 to 0.41. For traits (dry matter intake and net energy in milk) with intermediate estimated heritabilities (0.21 to 0.50 and 0.17 to 0.53, respectively), performance of the 3 models was comparable. Genomic correlations between environments also were computed using variance component estimates from the interaction model. Averaged across all traits, genomic correlations were highest between NAM and NLD, and lowest between NAM and SAC. In conclusion, the interaction model provided a novel way to evaluate traits measured in multiple environments in which genetic heterogeneity may exist. This model allowed estimation of environment-specific parameters and provided genomic predictions that approached or exceeded the accuracy of competing within- or across-environment models.

Key words: genomic selection, interaction model, feed efficiency

INTRODUCTION

Feed efficiency in dairy cattle has gained the attention of researchers, dairy farmers, and breeding companies internationally. Early studies aimed to estimate variance components for residual feed intake (**RFI**; Van Arendonk et al., 1991; Veerkamp et al., 1995), identify additive and epistatic SNP associated with RFI (Yao et al., 2013), and combine health history data with SNP genotypes to predict future phenotypes (Yao et al., 2015). These studies were based on a few hundred animals from a single research station due to the exorbitant cost associated with measuring individual feed intakes. Next, scientists combined data collected under similar

Received June 13, 2016.
Accepted November 22, 2016.
[1]Corresponding author: Chen.Yao225@gmail.com

climate and management conditions to estimate genetic parameters for RFI (Williams et al., 2011; Hardie et al., 2015) and to predict estimated breeding values for RFI (Pryce et al., 2012; Davis et al., 2014). To further improve the reliability of prediction, researchers pooled data globally. Berry et al. (2014) estimated genetic parameters for DMI, and Tempelman et al. (2015) estimated overall and country-specific heritabilities of RFI and explored heterogeneous relationships between RFI and traits relating to energy utilization. Both studies concluded that genetic evaluation using data collated from international populations appeared promising. The challenge lies in considering the heterogeneous genetic variance components and potential genotype × environment interactions when combining data from multiple sources.

Multiple statistical models have been explored to address the genetic heterogeneity of data consisting of several subpopulations (e.g., multienvironment or multibreed data). One simple option is to analyze subsets of data within each subpopulation separately (i.e., within-environment or within-breed analyses). A major drawback is that this method reduces sample size, and therefore hampers the accuracy of genomic selection. Additionally, this approach does not allow sharing of information between subpopulations, which is important if marker effects are correlated across subpopulations. A second option is to combine data across subpopulations and ignore potential heterogeneity (i.e., across-environment or across-breed analyses). This method increases sample size and ensures sharing of information across subpopulations, but assumes that marker effects are constant across subpopulations, which may limit the potential benefits of a combined analyses. Hayes et al. (2009) compared the prediction accuracy of a multibreed analyses with that of a within-breed analyses using data from Holsteins and Jerseys. Those authors found no benefits of a combined analysis on prediction accuracy for the breed with a larger reference panel (Holstein), and observed that gains were minimal in prediction accuracy for the group with a smaller reference panel (Jersey).

Recently, de los Campos et al. (2015) suggested an intermediate multivariate marker × environment approach between the within-environment and across-environment models (i.e., the interaction model) to tackle potential genetic heterogeneity. This approach explicitly model interactions between whole-genome markers and environments, which allows marker effects to be constant and group-specific across environments. It also provides estimates of genomic correlations between subpopulations for each trait, which can be used to assess genetic similarity between subpopulations. The interaction model is also more straightforward to interpret and easier to implement than the traditional multivariate models (Karoui et al., 2012; Olson et al., 2012). Many studies in plant breeding have demonstrated the superiority of the interaction model relative to within-environment or across-environment analyses in terms of minimizing residual variance (Crossa et al., 2015) and increasing prediction accuracies of genomic selection (de los Campos et al., 2015; Lopez-Cruz et al., 2015); however, the interaction model failed to improve prediction accuracy in another recent study (Lehermeier et al., 2015). In the current study, the aforementioned interaction model was assessed and compared with within-environment and across-environment models using data from multiple environments to estimate genomic variances and assess the accuracy of genomic predictions for RFI and its component traits.

MATERIALS AND METHODS

Phenotypic Data

The data used in this study consisted of 3,656 cows from 3 broadly defined environments: 2,329 cows from North America (**NAM**), representing the United States and Canada; 875 cows from the Netherlands (**NLD**); and 452 cows from Scotland (**SAC**). More detailed information about the data sources can be found in Tempelman et al. (2015).

Phenotypic data included 4 traits, RFI, DMI, net energy in milk (**MilkE**), and metabolic body weight (**MBW**), per cow per day during the period from 50 to 200 d postpartum. Individual feed intakes were recorded daily using electronic measurement systems or manual weigh-backs. Dry matter percentage dried at 60°C was analyzed periodically (typically once a week), and individual DMI of cows were calculated weekly using the DM percentage of feed offered. According to NRC (2001), daily MilkE was calculated based on the gross energy per kilogram in fat, protein, and lactose as

$$\begin{aligned}\text{MilkE (Mcal)} = (&0.0929 \times \text{fat\%} + 0.0563 \times \text{protein\%} \\ &+ 0.0395 \times \text{lactose\%}) \times \text{daily milk yield (kg)}. \quad [1]\end{aligned}$$

Body weight was typically measured weekly, and MBW 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. Residual feed intake was calculated as the deviation of the feed intake from the expected feed intake as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}, \quad [2]$$