Genotype by environment interaction for age at slaughter in Irish dairy and beef crossbreds using a genomic reaction norm model

B. Gredler-Grandl[1], J. Vandenplas[1], A.J. Twomey[2] and M.P.L. Calus[1]

[1] Animal Breeding and Genomics, Wageningen UR, P.O. Box 338, 6700 AH, Wageningen, The Netherlands

[2] Teagasc, Moorepark, Animal Grassland Research and Innovation Centre, Fermoy, P61 P302, Co. Cork, Ireland

Abstract

In beef cattle production, the reduction of the number of days from birth until the target weight at slaughter is reached, represents a sustainable option to increase efficiency and reduce the environmental impact. In the presence of genotype by environment interaction (GxE), selection of resilient animals is important. We estimated GxE for age at slaughter in an Irish dairy and beef crossbred population using an analysis protocol modelling either homogeneous or heterogeneous SNP-(co)variances across the genome based on readily available BLUP software packages. We allowed for heterogeneous SNP (co)variances by using different weights across SNPs. In our approach, we divide the data set of interest in two subsets and follow a 2-step approach: (1) derive SNP (co)variances from SNP effects estimated in the first data set, and (2) weight the SNP genotypes using estimated SNP (co)variances from (1) to re-estimate SNP effects in the second data set. The data set consisted of 14,193 genotyped crossbred heifers, steers and bulls in 2,041 herds. Phenotypes used in the genomic analysis were yield deviations for age at slaughter. We estimated contemporary group (CG) effects for age at slaughter in a univariate BLUP analysis to be subsequently used as continuous environmental descriptor in the genomic reaction norm models. Results show large genetic variations for age at slaughter. The average heritability estimated across all CG was 0.24. The genetic parameters for age at slaughter estimated along the environmental gradient support the existence of GxE in extreme environments. Nevertheless, the genetic correlation between the majority of environments was greater than 0.89. Higher accuracy for young selection candidates were achieved when using genomic information instead of using pedigree information only. However, modeling homo- or heterogeneous SNP (co)variances across the genome resulted in similar accuracy of genomic breeding values for age at slaughter.

Key words: genotype by environment interaction, genomic reaction norm model, heterogeneous SNP variance, age at slaughter

Introduction

Age at slaughter is a novel trait with the potential to increase efficiency and reduce the environmental impact of beef production. Berry et al. (2017) have shown the potential of genetic selection for a younger age at slaughter exploiting large amounts of genetic variation for this trait. Selection of resilient animals is important, specifically when genotype by environment interaction (GxE) exists. Genotype by environment interaction is often modelled by a multi-trait approach, where the same trait measured in different environments is considered being a genetically different, but correlated trait (Falconer, 1952). Alternatively, reaction norm models are used, where the breeding values are modelled as a function of the environment defined as a continuous variable. Both, genomic multi-trait models or reaction norm models, implicitly assume the same (co)variance matrix for every SNP. Since certain regions in the genome may contain QTL, the assumption of equal (co)variances across the genome may be violated. To overcome this limitation, we have developed an

analysis protocol allowing for heterogeneous SNP (co)variances across the genome in genomic GxE models. The analysis protocol can be implemented using standard BLUP software packages. The analysis protocol has been tested in simulated data resulting in a slight increase in accuracy of genomic breeding values (Gredler-Grandl and Calus, 2021). The objective of this study was to quantify GxE for age at slaughter and to evaluate the accuracy of genomic reaction norm models allowing for heterogeneous SNP (co)variances in an Irish dairy and beef crossbred population.

Materials and Methods

Phenotypes and genotypes

14,193 genotyped and phenotyped bulls, steers and heifers were available. The data set included purebred Holstein, Limousin, Charolais, Aberdeen Angus, Belgian Blue, Hereford, Simmental, Saler, Aubrac, Blonde d'Aquitaine, Parthenaise animals as well as crosses thereof. The genotypes were imputed to a high density SNP chip level and consisted of 662,011 SNP. Yield deviations (YD) for age at slaughter were used as phenotypes for genomic GxE analyses. The effects accounted for were contemporary group (CG), interactions between gender (bull, steer, heifer) and carcass weight and as well as gender and carcass fat, parity of dam, herd source (dairy or beef herd), heterosis class and recombination loss. Contemporary group effects, reflecting management and environmental conditions and estimated in the model described above, were used as continuous environmental descriptor. Animals of the same gender, and similar birthday or day purchased into the same herd in close period of time were assigned to the same CG.

Analysis protocol

A protocol consisting of several steps has been developed to allow for heterogeneous SNP (co)variances across the genome in genomic GxE models. Firstly, the data set is split in two subsets: the subset 1 is used to estimate SNP effects $\hat{\alpha}$ using a model that assumes equal (co)variances for all SNP; the subset 2 is analysed with a model with heterogenous SNP covariances computed as described below.

A K-means clustering approach (Saatchi et al., 2011) applied to the genomic relationship matrix of the herds has been used to assign the animals to subset 1 and 2. Average genotypes per herd were calculated and the genomic relationship matrix computed as $\mathbf{G_{herds}} = \frac{\mathbf{ZZ'}}{2\sum p_k(1-p_k)}$, where $\mathbf{G_{herds}}$ is the genomic herd relationship matrix, $\mathbf{Z}$ is the incidence matrix containing average genotypes for all herds for all SNP and p_k is the allele frequency of SNP *k* in the genotyped animals. A dissimilarity matrix between all herds was calculated based on the elements of the genomic relationship matrix. The number of clusters were set to 12 to ensure that all main breeds are represented at least by three clusters. Four clusters mainly representing the main breeds Holstein, Limousin, Charolais and Angus with the highest importance for crossbreeding have been chosen for subset 1. The remaining eight clusters were used in subset 2.

Analysis subset 1

A univariate linear genomic reaction norm model has been applied to subset 1 using the software package mtg2 (Lee et al., 2016):

$$\mathbf{y} = \mathbf{1}\mu + \boldsymbol{\beta_0} + \mathbf{Q}\boldsymbol{\beta_1} + \boldsymbol{\tau_0} + \boldsymbol{\tau_1}$$

where $\mathbf{y}$ is a vector of YD, μ is an overall mean, $\boldsymbol{\beta_0}$ and $\boldsymbol{\beta_1}$ are the vectors of intercept and first order of regression coefficients for the random genetic effects, $\mathbf{1}$ is a vector of ones, $\mathbf{Q}$ is a (diagonal) incidence matrix storing the environmental values (contemporary group effects) for each individual, and $\boldsymbol{\tau_0}$ and $\boldsymbol{\tau_1}$ are intercept and second order of regression coefficients for the random residual effects to account for heterogeneous residual variances across environments. It is assumed that $\begin{bmatrix}\boldsymbol{\beta_0}\\ \boldsymbol{\beta_1}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \mathbf{G_{VR}}\begin{bmatrix}\sigma^2_{\beta 0} & \sigma_{\beta 0 \beta 1}\\ \sigma_{\beta 0 \beta 1} & \sigma^2_{\beta 1}\end{bmatrix}\right)$, where $\mathbf{G_{VR}}$ is a genomic relationship matrix of the animals in subset 1 using the first method of VanRaden (2008). It is assumed that

$$\begin{bmatrix}\boldsymbol{\tau_0}\\ \boldsymbol{\tau_1}\end{bmatrix} \sim N\left(\begin{bmatrix}\mathbf{0}\\ \mathbf{0}\end{bmatrix}, \mathbf{I}\begin{bmatrix}\sigma^2_{\tau 0} & \sigma_{\tau 0 \tau 1}\\ \sigma_{\tau 0 \tau 1} & \sigma^2_{\tau 1}\end{bmatrix}\right).$$

Calculation of SNP specific weights

For the model with heterogeneous SNP (co)variances (HET), SNP specific weights for each SNP *k* for each coefficient *i* of the reaction norm model (i.e. intercept β_0 and the linear regression coefficient β_1) were calculated as

$$D_{k_i} = \sqrt{2p_k(1-p_k)}\hat{\alpha}_{k_i}$$

where D_{k_i} is diagonal element *i* of diagonal matrix $\mathbf{D_k}$ that stores the weights for SNP *k*, p_k is the allele frequency of SNP *k*, and $\hat{\alpha}_k$ is the estimated effect of SNP *k* for coefficient *i*. The SNP effects $\hat{\alpha}_k$ for intercept and linear regression coefficient were obtained by backsolving GEBV for β_0 and β_1 obtained from the genomic reaction norm model and the subset 1. SNP effects were calculated following the approach described in Bouwman et al. (2017) and implemented in the companion program compute_SNP_effects of calc_grm (Calus and Vandenplas, 2016).

Analysis subset 2

In subset 2, a seven-fold cross validation has been applied, where each cluster has been used once as validation set for genomic prediction. The number of animals in each cross validation is shown in Table 1.

The following SNP-BLUP model was applied to subset 2 using the MiXBLUP software (ten Napel et al. 2020):

$$\mathbf{y} = \mathbf{1}\boldsymbol{\mu} + \mathbf{Z}\boldsymbol{\gamma_0} + \mathbf{QZ}\boldsymbol{\gamma_1} + \mathbf{e}$$

where $\mathbf{y}$ is the vector of YD of animals in the training set of each cross validation run, μ is an overall mean, $\mathbf{Z}$ is a matrix including the centered genotypes for each SNP, $\mathbf{Q}$ is a diagonal matrix storing the environmental values for each individual, $\boldsymbol{\gamma_0}$ and $\boldsymbol{\gamma_1}$ are vectors of estimated SNP effects for random intercept and linear regression coefficient, respectively, and $\mathbf{e}$ is a random residual term. For HET the following (co)variance matrix is used for SNP *k*:

$$Var([\boldsymbol{\gamma_0}, \boldsymbol{\gamma_1}]') = \mathbf{D_k} * \mathbf{G} * \boldsymbol{D'_k}$$

where $\mathbf{G}$ is the estimated genetic (co)variance matrix between intercept and quadratic regression coefficient obtained from the reaction norm model in the analysis of subset 1. For HOM, homogeneous SNP variances for intercept and linear regression coefficient are provided by $\sigma^2_g/2\sum p_k(1-p)$, where σ^2_g is the genetic variance for either intercept or linear regression coefficient estimated in subset 1. The GEBV for validation animals were calculated as $\mathbf{GEBV} = \mathbf{1}\hat{\mu} + \mathbf{Z}\hat{\boldsymbol{\gamma}}_\mathbf{0} + \mathbf{ZQ}\hat{\boldsymbol{\gamma}}_\mathbf{1}$. The accuracies of GEBV for individuals in the validation set were obtained as the correlation coefficient between the observed YD and predicted GEBV divided by the square root of heritability.

Results & Discussion

The estimated genetic variances for intercept and the quadratic regression coefficients were 670.87 and 19.99, respectively. The estimated heritability across all environments is shown in Figure 1. Highest heritabilities were observed for very extreme environments (CG effects). The average heritability across all environments was 0.242 (SD 0.017). Estimated heritabilities in this study are similar as reported by Berry et al. 2017.

Genetic correlations between environments were in the range between 0.57 and 1. The lowest genetic correlation was observed between extreme and intermediate environments, indicating that breeding values may change in different environments. However, for the majority of environments (CG effects greater than -1.8 or below 1.8) all genetic correlations were higher than 0.89 so that one breeding program seems justified across all environments.

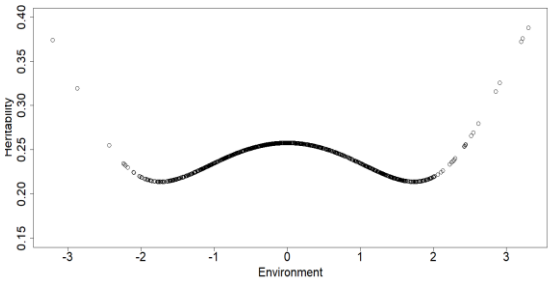

Figure 1. Heritability of age at slaughter across environments (i.e. CG effects standardized to mean of 0 and SD of 1).

Accuracies of genomic breeding values modelling homogeneous (HOM) and heterogeneous (HET) SNP (co)variances are shown in Table 1. Accuracy for HOM were between 0.193 and 0.607 across all CV. For HET, accuracies were slightly lower compared to HOM across all CV except for CV2, where Holstein and Aberdeen Angus are the main breeds contributing gene proportion. Overall, modelling heterogeneous SNP (co)variances did not result in higher accuracies of genomic breeding values for age at slaughter. Reasons may be that the genetic architecture of age at slaughter is highly polygenic or that it is difficult to take advantage of the proposed method given the very heterogeneous data set including different breeds and crossbred animals.

Table 1. Number of animals in the training (TRAIN) and validation (VAL) set for each cross validation (CV) and accuracy of genomic prediction for HOM and HET

CV	TRAIN	VAL	HOM	HET
1	7,333	1,664	0.298	0.279
2	8,495	502	0.607	0.613
3	7,476	1,521	0.193	0.184
4	7,248	1,749	0.319	0.295
5	8,454	543	0.237	0.191
6	7,626	1,371	0.329	0.314
7	7,350	1,647	0.334	0.315

Conclusions

The results show large genetic variations for age at slaughter in an Irish dairy and beef crossbred population. The estimated genetic correlations for age at slaughter between different environments suggest the existence of GxE to some extent. Modelling heterogeneous SNP (co) variances did not increase the accuracy of genomic breeding values.

Acknowledgments

We thank the Irish Cattle Breeding Federation for the interest in the study and the permission to use the data set. Donagh Berry (Teagasc, Ireland) is acknowledged for helpful discussions.

This work has been carried out within the project GenTORE – Genomic management Tools to Optimize Resilience and Efficiency. This project has received funding from the European Union's Horizon 2020 research and innovation program under Grant Agreement No 727213.

References

Berry DP, Cromie AR, Judge MM. Rapid Communication: Large exploitable genetic variability exists to shorten age at slaughter in cattle. J Anim Sci. 2017;95(10):4526-32.

Bouwman, A.C., Hayes, B.J., and Calus, M.P.L. 2017. Estimated allele substitution effects underlying genomic evaluation models depend on the scaling of allele counts. Genet. Sel. Evol. 49, 79.

Calus M., and Vandenplas, J. 2016. Calc_grm – a program to compute pedigree, genomic, and combined relationship matrices. ABGC, Wageningen UR Livestock Research.

Falconer, D. S. 1952. The problem of environment and selection. Am. Nat. 86:293-298.

Gredler-Grandl, B. and Calus, M.P.L. 2021. Genomic GxE approaches modelling heterogeneous SNP variances: applied to simulated data. Interbull Bulletin 56, 192-193.

Lee, S.H. and van der Werf, J.H. 2016. MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. Bioinformatics 32, 1420-1422.

Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, et al. Accuracies of genomic breeding values in American

Angus beef cattle using K-means clustering for cross-validation. Genet Sel Evol. 2011;43(1):40.

ten Napel, J., J. Vandenplas, M. Lidauer, I. Stranden, M. Taskinen, E. Mäntysaari, M. P. L. Calus, and R. F. Veerkamp. 2020. MiXBLUP, the Mixed-model Best Linear Unbiased Prediction software for PCs for large genetic evaluation systems – Manual V2.2 - 2020 - 05, Wageningen, the Netherlands www.mixblup.eu

vanRaden, P. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91(11):4414-23.