Accuracy of pedigree and genomic predictions of meat quality in multi-breed cattle using single and two step model

R.D. Evans, D.P. Berry, S., Naderi

Objectives:

- Compare pedigree, 2-step and single step genomic methods
- Assess impact of half sibs in training population

Methods

- Multi-trait (x3) across breed animal model
- 2 step SNPBLUP and blending *(Van Raden, 2009)*
- Single step GBLUP *(Christensen & Lund, 2010)*
- Mix99 (LUKE, Finland)

Dataset

- 5912 trained sensory panel phenotypes
- 3 traits: Tenderness, Flavor, Juiciness
- All commercial multi-breed cattle
- 90% of animals also genotyped (50k IDB)

Validation

- Sire level: 49 AI sires with a min of 25 progeny
  - *r(DYD,GEBV), regression slopes*
- Animal level: random 12 progeny from validation sires
  - *r(YD,GEBV), regression slopes*

Conclusions

- Single step yielded best prediction accuracy for both correlation and slopes
- Increasing half sib representation in training increased prediction accuracy
- Single step GEBVs for meat quality are now routinely published in Ireland

revans@icbf.com

Accuracy of pedigree and genomic predictions of meat quality in a multi-breed cattle using single and two step model

S. Naderi[1], R. Evans[1*] and D. P. Berry[2]

[1] The Irish Cattle Breeding Federation, Highfield House, Bandon, Co. Cork,Ireland
[2] Department of Animal Bioscience, Animal and Grassland Research and Innovation Centre, Teagasc, Moorepark, Fermoy, Co. Cork, Ireland; revans@icbf.com

Abstract
The objective of this study was to develop single step genomic breeding values (ssGEBV) for meat tenderness, juiciness and flavour and compare with two-step genomic breeding values as well as non-genomic (i.e., traditional) breeding values (EBV); the comparison also involved testing the impact on the accuracy of the (G)EBVs by altering the number of half sibs in the training population. A residual polygenic single step GBLUP and a two-step genomic approach was applied to 5,262 genotyped and phenotyped animals. Including genomic information improved the prediction accuracy and this accuracy also improved as the number of half sib-progeny in the training population increased. Moreover, applying single step genomic approaches were associated with further improvements in prediction performance across all traits.

Introduction
Meat eating quality (MEQ) impacts repeat purchase of meat by consumers; this is particularly important for beef meat which is often more variable in quality than many other meat types. MEQ is often depicted by meat tenderness, juiciness, and flavour. Whereas heterogeneity is often viewed as an inconvenience, animal breeders require variability to achieve genetic gain. Nonetheless, genetic gain eventually translates into phenotypic changes which is a function of how well estimates of genetic merit translate into expected phenotypic differences but also the acceptance of stakeholders that (products from) individuals divergent in genetic merit will truly be reflective in phenotypic differences. Validation of genetic evaluations is therefore key to stakeholder acceptance. The accuracy of estimates of genetic merit can be improved using widely available genomic information. Integrating phenotypic, pedigree and genomic data to achieve high prediction performance is a modelling challenge. Therefore, the objective of the present study was to evaluate alternative strategies of combining these data to predict genetic merit and, by extension, phenotypic performance.

Materials & Methods

Data. Tenderness, juiciness and flavour sensory data were available on 5,912 crossbred cattle consisting of 2,137 steers, 2,093 heifers and 1,682 young bulls. The main breed proportion of the animals was Angus (27%), Belgian Blue (5%), Charolais (12%), Holstein-Friesian (4%), Hereford (14%), Limousin (21%) and Simmental (8%). The phenotyped population was constructed to be as diverse as possible with the aim of a maximum of 20 progeny per sire. The generation of the meat sensory data including the standard operating procedures are described in detail by Judge et al. (2021) for a subset of the data set used. Each of the three traits were scored on a 1 to 10 scale: 1 = very tough and 10 = very tender; juiciness: 1 = not juicy and 10 = very juicy; beef flavour: 1 = no beef flavour and 10 = very strong beef flavour. Of the 5,912 animals with meat quality phenotypes, 5,262 were genotyped with the International Dairy and

Beef genotyping platform version3 (IDBV3; 50,855 SNPs). The number of animals in the pedigree was 95,674.

Estimation of breeding values. Traditional breeding values (EBVs) as well as genomic breeding values (GEBVs) were estimated for three traits using the following linear mixed model:

$\mathbf{y = Xb + Wu + e}$

where **y** is a vector of one of the sensory phenotypes for all animals in training set, **X** is the incidence matrix for fixed effects, **b** is the vector of fixed effects consisting of herd-by-date of slaughter, date-by-location of sensory analysis, gender of the animal, the age in months since slaughter of the meat sample, heterosis proportion class, month of analysis and breed proportion, **W** is the incidence matrix of random (genomic) effects, **u** is a vector of EBVs or GEBVs for all individuals, and **e** is a vector of random residuals. For the traditional genetic evaluation, it was assumed that $u \sim N\ (0,\ \boldsymbol{A}\sigma_u^2)$ where σ_u^2 is the additive genetic variance and **A** is pedigree relationship matrix. For the estimation of GEBV using the GBLUP model, the **A** matrix was replaced with a genomic relationship matrix. In the single-step GBLUP model, the **A** matrix was replaced with a **H** matrix where **H** was the combined relationship matrix (Christensen et al. 2012). The **H** matrix is constructed by blending the pedigree relationship matrix with the genomic relationship matrix) considering information from non-genotyped and genotyped animals simultaneously. The inverse of **H**-matrix was calculated according to Christensen & Lund (2010) and Aguilar et al. (2010):

$$\mathbf{H^{-1}} = \mathbf{A^{-1}} + \begin{bmatrix} 0 & 0 \\ 0 & \boldsymbol{G_w^{-1}} - \mathbf{A_{22}^{-1}} \end{bmatrix}$$

where $\mathbf{A^{-1}}$ was the inverse of the pedigree-based relationship matrix, $\mathbf{A_{22}^{-1}}$ was the inverse of the sub-matrix of the pedigree-based relationship matrix for genotyped animals and $\mathbf{G}_w$ was calculated as follow:

$\mathbf{G}_w = (\alpha \times \mathbf{G} + \beta \times \mathbf{A}_{22})$

where $\mathbf{G}_w$ was the genomic relationship matrix, $\beta = 1 - \alpha$ where $\alpha = 0.70$. For all the computations, the MiX99 software (MiX99 development team, 2020) was used.

The GEBVs from the two-step genomic evaluation are the result of blending the Direct Genomic Value (DGV) which was calculated by multiplying the SNP effects with the animals' genotypes, traditional Parent Average (PA) and the EBV_{ref} from a traditional evaluation including only relationships among genotyped animals. The method for blending three sources of information (i.e., DGV, PA and EBVref.) to calculate the GEBV are described in detail by VanRaden et al. (2009).

Training and testing population design. Three different scenarios were explored on a group of validation animals to assess prediction accuracy. First a group of 49 "well" proven sires with at least 25 progeny was identified. To build a validation set, a random 12 progeny from each of the 49 sires was sampled (i.e., 588 animals) with the remaining progeny of these sires, who did not appear in the validation set, being hereafter referred to as the *remaining set*. In the first scenario (S_I), none of the progeny from the well-proven sires were included in the training population. In the second scenario (S_II), a random seven progeny of each well proven sire from the *remaining set* was added to the training set; in the third scenario (S_III), all the *remaining set* progeny were added to the training set; only the phenotypic records of validation set were masked in this scenario.

Prediction accuracy. The correlation between (G)EBV and the phenotypes (following adjustment for all non-genetic effects) of the validation animals was considered to represent the prediction accuracy at the animal level for all three scenarios. In addition, to assess prediction performance at sire level, the correlation between validation sires' (G)EBVs from S_I in which

none of their progeny were in training set with their daughter yield deviation (DYD) was calculated. DYD were calculated based on the average yield deviation from the progeny of each sire including all phenotypic records (full data set).

Results and Discussion

Validation results from (single)two-step genomic evaluation as well as traditional EBVs for three scenarios at the animal level are presented in Table 1. Including genomic information and applying a single or two step genomic approach improved the prediction accuracy for all three traits compared with the EBVs from the pedigree-based relationship matrix. For example, the prediction accuracy for flavour increased from 0.09 to 0.17 following the incorporation of genomic information into the model and using two-step approach (i.e., S_III) with a similar pattern detected for the two other traits. Further increases in prediction accuracy were observed by implementing the single step genomic approach compared with the two-step method for all three traits and scenarios. Aguilar et al. (2010) demonstrated that ssGBLUP is a quick alternative to estimate GEBV when phenotypes, genotypes and pedigree are jointly available arguing that the automatic and proper definition of weighting factors to blend different sources of information were the main advantage of implementing this approach. In addition, Legarra et al. (2014) indicated that the appropriate weight to integrate various source of information avoids double counting of contributions emerging from genetic relationships and phenotypic records.

Table 1: Correlations between the (genomic) estimated breeding value and adjusted phenotype for the three different scenarios at the animal level from three different approaches.

Traits	EBV			Two-step GEBV			Single step GEBV		
	S_I[1]	S_II[2]	S_III[3]	S_I	S_II	S_III	S_I	S_II	S_III
Tenderness	0.03	0.09	0.13	0.09	0.11	0.15	0.11	0.14	0.17
Flavor	0.09	0.14	0.20	0.17	0.18	0.22	0.21	0.21	0.26
Juiciness	0.008	0.06	0.12	0.08	0.10	0.14	0.09	0.10	0.15

1 S_I: none of the progeny from the well-proven sires were in the training population

2 S_II: a random seven progeny of each well proven sire from the *remaining set* was included in the training set

3 S_III: all the *remaining set* progeny were included in the training set

By increasing number of half sibs in the training population (i.e., S_I to S_III), the accuracy increased for all three prediction approaches (i. e., traditional, (single)two-step GEBV). This improvement could be explained by increasing the genetic relationships between training and validation sets but also an enlargement of the training population size. De Los Campos et al. (2013) indicated that the genetic relationships between training and validation set as well as training population size were the most important factors affecting prediction performance for GBLUP.

Validation results at the sire level for the three different prediction approaches for S_I in which none of the identified sires' progeny were included in the training population are shown in Table 2. The results demonstrate that incorporating genomic information and applying the (single)two-step genomic approach improved the prediction accuracy for all three traits. For instance, the prediction accuracy for tenderness increased from 0.17 to 0.38 following the inclusion of genomic information in the model and applying a two-step genomic model.

Using single step genomic approaches generated further improvements in prediction accuracies for all three traits. Across all traits, the slope of DYD on (G)EBV of the validation sires where the expectation of the slope was 0.5, improved with the inclusion of genomic information in the

model and improvement further when a single step genomic approach was applied compared to a two-step approach indicating ssGBLUP could reduce potential biases.

Table 2: Correlations between the (genomic) estimated breeding value and daughter yield deviations (DYD) and the slope of the DYD on the (G)EBV from three different approaches.

Traits	Statistic	EBV	Two-step GEBV	Single step GEBV
Tenderness	Correlation	0.17	0.38	0.42
	Slope	0.12	0.31	0.39
Flavor	Correlation	0.13	0.36	0.40
	Slope	0.10	0.34	0.36
Juiciness	Correlation	0.16	0.34	0.41
	Slope	0.09	0.27	0.32

References

Aguilar I., Misztal I., Johnson D. L., Legarra A., Tsuruta S. and et al. (2010) J. Dairy Sci. 93 (2): 743–752. https://doi.org/10.3168/jds.2009-2730.

Christensen O. F. and Lund M. S. (2010). Genet. Sel. Evol. 42 (2). https://doi.org/10.1186/1297-9686-42-2.

Christensen O. F., Madsen P., Nielsen B., Ostersen T. and Su G. (2012). Animal 6 (10): 1565–1571. https://DOI.10.1017/S1751731112000742.

Judge M. M., Conroy S., Hegarty P. J., Cromie A. R., Fanning R and et al. (2021). Meat Science, 172: 108371. https://doi.org/10.1016/j.meatsci.2020.108371.

Legarra A., Chistensen O.F., Aguilar I. and Misztal I. (2014). Livest. Prod. Sci. 166: 54–65. https://doi.org/10.1016/j.livsci.2014.04.029

MiX99 Development Team, (2020) MiX99: Natural Resources Institute Finland (Luke). Jokioinen, Finland. URL: http://www.luke.fi/mix99.

VanRaden P., Van Tassell C., Wiggans G., Sonstegard T., Schnabel R. and et al. (2009) J. Dairy Sci. 92 (1): 16-24. https://doi.org/10.3168/jds.2008-1514

de Los Campos G., Hickey J. M., Pong-Wong R., Daetwyler H. D. and Calus M. P. L. (2013) Genetics. 193 (2): 327–345. https://doi.org/10.1534/genetics.112.143313.