Holstein-Friesian Relationships and the Impact on the Accuracy of an Across-Breed Evaluation.

A. Brown,[*] G. Banos,[*†] M.P Coffey,[*] J.A Woolliams[†] and R.A Mrode.[*]
[*]Scotland's Rural College, Edinburgh, UK[†] The Roslin Institute and R(D)SVS, University of Edinburgh, Midlothian, UK

ABSTRACT: Genomic evaluations are limited in their success for some cattle breeds due to the small size of reference populations. The objectives of this study are to investigate the level of relationships between animals in two breeds and the impact of the relationship on the accuracy of an across-breed genomic evaluation. The dataset consisted of 4,726 genotyped Holstein and Friesian individuals used in the UK, split into reference and validation populations based on year of birth. Principal components analysis based on the numerator relationship matrix (**A**) and genomic relationship matrix (**G**) was used to visualise relationships between breeds, and genomic predictions for yield traits were calculated. Results suggested that accuracies increase as the relationship between breeds increases, and that British Friesians seem to be more closely related to Holstein cattle than to Irish Friesians.
Keywords: Genomic evaluation; Multi-breed; Principal component analysis

Introduction

Genomic evaluations are now routinely published for a number of dairy cattle breeds worldwide. The UK has so far published commercial genomic breeding values (GEBVs) for Holstein cattle (Mrode et al. (2011)), and plans to release a Friesian evaluation in the near future. However, there are very few Friesian animals with genotype data available for inclusion in a within-breed genomic evaluation. There has been a reasonable degree of crossbreeding between Holstein and Friesian cattle in the UK, to the point where the breeds are collectively known as 'Holstein Friesian' cattle, and so it may be possible to use Holstein cattle in a Friesian genomic evaluation. However, data from Ireland suggests that the breeds are sufficiently different that a Holstein reference cannot successfully be used to predict GEBVs in Friesian cattle (Berry, personal communication). It is possible that this could be influenced by the degree of relatedness between the two breeds. This study aims to (i) investigate the relationship between Holstein and Friesian cattle raised in the UK and (ii) explore the impact of the relationship on the accuracy of across-breed genomic evaluation.

Materials and Methods

Data: The initial dataset comprised 7,886 Holstein and Friesian bulls used in the UK that were born between 1960 and 2008 and had both genotype data and Interbull de-regressed proofs for milk, fat and protein yield. This population was split into 3 categories based on percentage of each breed (PEB), where animals calculated as ≥87.5% Holstein were classified as purebred Holstein, animals calculated as ≥87.5% Friesian were classified as purebred Friesian, and animals calculated as 25-75% Holstein were classified as crosses. Animals that did not fit into these categories were discarded from further analysis. Friesians were of either British or Irish origin. A cut off year of 1997 was used to divide the population into reference and validation sets. The reference population consisted of 2,075 Holsteins, 21 Friesians of British origin, 29 Friesians of Irish origin and 29 crosses (total = 2,154). A random sample of 2,075 Holsteins born after 1997 were included in the validation set, along with 19 Friesians of British origin, 21 Friesians of Irish origin and 7 crosses (total = 2,122). The final dataset comprised of a total of 4,276 animals - 4,150 Holsteins, 40 Friesians of British origin, 50 Friesians of Irish origin, and 36 crosses. The pedigree for these animals consisted of 35,599 animals across 10 generations.

Genotype Data: Animals were genotyped using either the Illumina 50k bovine or the Illumina BovineHD SNP chips, but only SNPs equivalent to those on the 50k chip were used. SNPs with a minor allele frequency below 0.05, a call rate of <0.95 or GC content below 0.6 were filtered out, along with those that were detected as not being in Hardy-Weinberg equilibrium, leaving a dataset containing information for 43,121 markers.

Principal Components Analysis: Relationship matrices based on pedigree data (**A**) and SNP data (**G**) were calculated for (a) the full genotyped population (n=4,276), (b) a subset of this data, comprising Holstein reference animals along with all Friesian and Holstein Friesian cross animals (n=2,201), and (c) a subset comprising Holstein validation animals along with all Friesian and Holstein Friesian cross animals (n=2,201). The **A** matrix in each case was calculated with RelaX2 (Stranden and Vuori, (2006)), and the **G** matrices were calculated using the method of VanRaden (2008). Principal component analysis was performed using the R function "princomp" (R core team, (2013)).

Estimation of genomic breeding values: The single-step method – HBLUP (Legarra et al., (2009)); Misztal et al., (2009)); Christensen and Lund., (2010)) was used to estimate GEBVs using de-regressed bull proofs (DRPs) as phenotypes. The de-regression was carried out on the official UK proofs for December 2013 using national parameters. The BLUPF90 software package (Misztal et al., (2002)) was used to fit the following mixed model:

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

where **y** is a vector of DRPs, **b** is the vector of fixed effects consisting of the mean, **a** is a vector of animal effects, and **e** is the vector of residual effects; **X** and **Z** are respective incidence matrices. Random effects were assumed to be normally distributed with a mean of zero, and the following covariance structure:

$$Var\begin{bmatrix}\boldsymbol{a_1}\\ \boldsymbol{a_2}\\ \boldsymbol{a_3}\\ e_1\\ e_2\\ e_3\end{bmatrix} = \begin{bmatrix} \boldsymbol{H}\sigma^2_{a_1} & \boldsymbol{H}\sigma_{a_{12}} & \boldsymbol{H}\sigma_{a_{13}} & 0 & 0 & 0 \\ & \boldsymbol{H}\sigma^2_{a_2} & \boldsymbol{H}\sigma_{a_{23}} & 0 & 0 & 0 \\ & & \boldsymbol{H}\sigma^2_{a_3} & 0 & 0 & 0 \\ & & & \boldsymbol{I}\sigma^2_{e_1} & \boldsymbol{I}\sigma_{e_{12}} & \boldsymbol{I}\sigma_{e_{13}} \\ & & & & \boldsymbol{I}\sigma^2_{e_2} & \boldsymbol{I}\sigma_{e_{23}} \\ & & symm & & & \boldsymbol{I}\sigma^2_{e_3} \end{bmatrix}$$

Where indices 1-3 indicate milk, fat and protein, **I** is an identity matrix, and **H** is the unified relationship calculated using a blend of Van Raden's (2008) **G** matrix and the **A** matrix. This **G** was computed as:

$$\mathbf{G} = 0.95\,\frac{\mathbf{SS'}}{2\sum_{i=1}^{n} p_i(1-p_i)} + 0.05\mathbf{A}$$

Where **S** is a centred incidence matrix of SNP genotypes, *n* is the number of SNPs, and p_i is the allele frequency of marker *i*. **H** is equivalent to that described in Appendix 1 of Aguilar et al. (2010). GBLUP was also carried out using the same program, by replacing the **H** matrix with the **G** matrix (i.e. removing pedigree information).

Study design: GEBVs were calculated using 3 reference populations, (i) Holstein only reference population, (ii) reference population containing both purebred Holstein and purebred Friesian animals, and (iii) the full reference population made up of both purebreds and crosses. Each of theses are used to predict accuracies in a validation population containing Holsteins, Friesians and crosses, and also for Holstein and Friesian subsets of the full validation population. Finally, conventional evaluations (EBVs) were calculated using the same model and software as HBLUP, but the **H** matrix was replaced with the **A** matrix calculated from the pedigree (PBLUP).

Accuracy of evaluation: In all cases the accuracy of evaluation was calculated as the correlation between DRPs and GEBVs. Accuracies were calculated separately for the full validation population, the Holstein validation population and the Friesian validation population. The error of prediction was calculated as the mean of differences between DRP and GEBV.

Results and Discussion

Results of the principal components analysis based on **A** differed significantly from those based on **G** (Figure 1). Populations (b) and (c) based on A show Holsteins and Friesians separating along PC1, whereas the equivalent populations show a long cluster of Holsteins along PC1, with the Friesians and crosses separating into two distinct clusters, the first close to the Holstein animals, and the other separated from the other clusters along PC2. This separation can be attributed to animals of differing origin – Friesians and crosses of UK origin cluster close to the Holstein animals, and Friesians and crosses of Irish origin cluster distinctly. This may suggest that Irish Friesians are genetically more distinct from Holsteins than are British Friesians. However, the graph from population (a) based on the **G** matrix shows both British and Irish Friesians spread evenly across one large cluster, and so the results in Figure 1 may just be a consequence of small sample size. This should therefore be investigated further with a larger dataset containing more Friesian animals before any firm conclusions can be made.

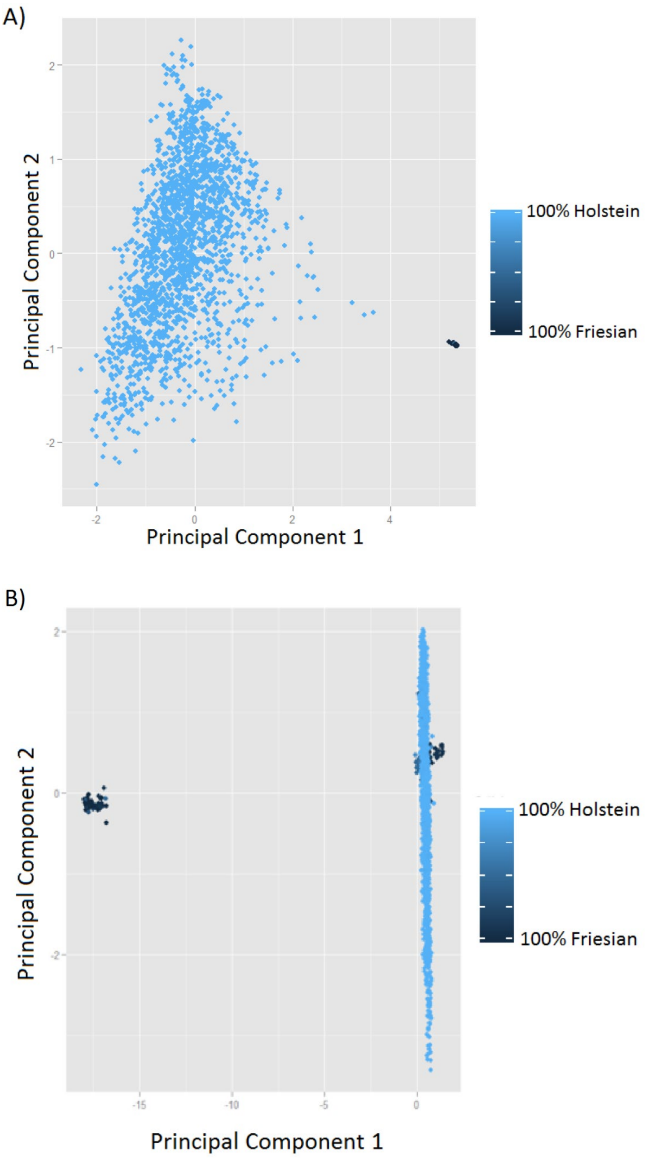

Figure 1. A) Scatterplot of principal components 1 and 2 calculated using the A matrix and coloured according to PEB, and B) Scatterplot of principal components 1 and 2 calculated using the G matrix and coloured according to PEB.

Accuracies of prediction for milk yield based on PBLUP and GBLUP can be seen in Table 1. Results of predictions for fat and protein content are not shown here for the sake of brevity, but show similar patterns to those seen in the results for milk yield. HBLUP results are not reported as they were extremely similar to GBLUP accuracies, and the results from PBLUP and GBLUP can be more directly compared with the results of the principal components analysis. Accuracy of predictions for the full validation population ranged from 0.67 – 0.69 for PBLUP, and from 0.74 – 0.80 for GBLUP, and GBLUP accuracies increased as Friesians and crosses were added to the reference population. Prediction accuracies of the Holstein only

validation population were similar to those obtained from full validation HBLUP, but they did not benefit from the addition of Friesians and crosses. It was not possible to predict GEBVs for Friesians with reasonable accuracy based on a Holstein only reference population (approximate standard error of correlation = 0.15), supporting evidence from other studies (Hayes et al., (2009)). Surprisingly, the accuracy of predictions for Friesians increased to 0.87 when Friesians were included in the reference population; however, the prediction error was almost twice that of the Holstein prediction, which is probably due to small sample size of the Friesians. Also, the additive genetic relationship between Friesians in the reference and validation was greater than 0.2 for close to one third of the bulls, which could contribute to the high accuracy of predictions in Friesians. The accuracy of evaluations for British and Irish subsets of Friesians is not reported as sample size was too small to draw conclusions, but prediction accuracy was generally higher in British Friesians than Irish Friesians.

Table 1. Correlations (Corr) between milk yield deregressed proofs and EBV from PBLUP, and GEBV from GBLUP, for animals in the validation population, and the related prediction error (Error). **H**- Holstein only reference population; **HF** – combined Holstein and Friesian reference population; **HFX** - reference population containing Holsteins, Friesians and crosses; HFX - the full validation population; H - Holstein only subset of the validation population; F – Friesian only subset of the validation population.

Ref.	Val.	Corr (PBLUP)	Error (PBLUP)	Corr (GBLUP)	Error (GBLUP)
H	HFX	0.68	-89.97	0.74	144.53
HF	HFX	0.67	-255.87	0.78	131.58
HFX	HFX	0.69	-290.69	0.79	124.20
H	H	0.66	-83.05	0.75	148.98
HF	H	0.59	-252.79	0.74	133.40
HFX	H	0.58	-289.75	0.73	124.50
H	F	0.05	-576.96	-0.17	-242.79
HF	F	0.80	-476.60	0.87	-48.71
HFX	F	0.77	-376.10	0.85	90.34

Conclusion

The results of this pilot study suggest that i) inclusion of Friesian animals in the reference population improves the accuracy of predictions, and ii) that there is a tendency for British Friesians - which from the principal components analysis appear to be more closely related to Holsteins relative to Irish Friesians - to be predicted with higher accuracy than Irish Friesians. Follow up studies with larger sample sizes, and also HD SNP data should be carried out to test this emerging hypothesis, and whether this should be taken into account when implementing multi-breed genomic selection.

Acknowledgements

This paper is part of a 4 year PhD project sponsored by Biosciences Knowledge Transfer Network (KTN), Biotechnology and Biological Sciences Research Council (BBSRC) and DairyCo. The authors gratefully acknowledge the Irish Cattle Breeding Federation, Teagasc, the North American Cooperative Dairy DNA Repository (CDDR) and UK industry for providing genotype data. A. Brown would also like to acknowledge the British Society of Animal Science (BSAS) for providing financial support.

Literature Cited

Christensen, O.F., & M.S. Lund. (2010). Genet. Sel. Evol. 42:2.

Hayes, B.J., P.J. Bowman, A.C. Chamberlain, K. Verbyla, & M.E. Goddard. (2009). Genet. Sel. Evol. 41:51.

Legarra, A., I. Aguilar, & I. Misztal. (2009). J. Dairy Sci. 92:4656-4663.

Misztal, I., S. Tsuruta, T. Strabel, B. Auvray, T. Druet, & D.H. Lee. (2002). Proc 7th WCGALP. 28:07

Misztal, I., A. Legarra, & I. Aguilar. (2009). J. Dairy Sci. 92:4648–4655.

Mrode, R., Krzyzelewski, T., Moore, K. et al. (2011) Interbull Bulletin.44:173-174

R core team (2013). http://www.R-project.org/

Strandén, I. & K. Vuori. (2006). Proc 8th WCGALP. 27:27-30

VanRaden, P.M. (2008). J. Dairy Sci. 91:4414-4423