The Implementation of Genomic Evaluations in the UK

R. Mrode[1], T. Krzyzelewski[1], K. Moore[1], M. Winters[2] and M. Coffey[1]
[1]*SAC, Roslin Institute Building, Easter Bush, Midlothian, EH25 9RG UK*
[2]DairyCo *Stoneleigh Park, Kenilworth, Warwickshire, CV8 2TL, UK*

Abstract

Preliminary genomic evaluations for production and fitness traits in Holstein Friesian bulls in the United Kingdom (UK) were undertaken using 11480 bulls with 50k genotypes. A linear model with and without polygenic effects was implemented and results were compared. Linear equations were solved with Mix99. The accuracy of predictions in terms of correlations between deregressed proofs and direct genomic values in the validation set were about 0.68 for production traits and somatic cell count (SCC) but lower at 0.45 for longevity. Inclusion of polygenic effects increased the regression coefficients to about unity, thus improving the predictive ability of the model. The gain in reliability in genomic enhanced pedigree index (GPI) relative to traditional PI were about 23% for production traits and 15% for longevity.

Keywords: genomic evaluation, linear model, polygenic effect

Introduction

Since the introduction of the concept of genomic selection by Meuwissen *et al.* (2001), a number of countries have implemented genomic evaluations at the national level for dairy bulls and cows using the Illumina bovine chip SNP50k (VanRaden, 2008; Berry *et al.*, 2009). International trade in semen exits between a number of countries who now market genomically proven young bulls that do not have any daughters and which are not included in MACE evaluations. The objective of this study was to develop a genomic evaluation system for Holstein-Friesians in the UK for production and fitness traits.

Materials and Methods

Data from 11480 bulls with 50k genotypes were available for the analysis. However, 600 of these bulls were genotyped with the Illumina HD chip but only the corresponding SNPs on the 50k chip were extracted and used for these bulls. These genotypes are a combination of the North American Cooperative Dairy DNA Repository (CDDR), UK AI industry and SAC genotypes.

Minor allele frequency was set to 0.05, call rate for animals was 70% (across the various chips) and checks for Mendelian inconsistency were also carried out. Missing SNPs calls were replaced by the most frequent genotype. The total number of 41703 SNPs were selected for genomic evaluations after these various edits.

Deregressed sire proofs from the UK official January 2011 run and MACE proofs were used as input variables in the genomic evaluations. The software MiX99 (Lidauer *et al.*, 2011) was used for the deregression using a full animal pedigree with effective daughter contributions (EDCs) used as weights. However a high proportion of bull used for the genomic evaluations were foreign bulls with no daughters in the UK and so corresponding EDCs on the UK scale were not available. The EDCs on the UK scale for these bulls were computed using the EDCs in the foreign countries and the genetic covariance matrix as described by Liu (2009).

The statistical model used for the estimation of SNP effects is:

$$y_i = \mu + v_i + \sum_{j=1}^{m} z_{ij} u_j + e_i$$

where y_i is the deregressed proof of a bull, μ is the overall mean, v_i is the residual polygenic effect (10% of additive genetic variance) of *i*th bull, z is the genotype value coded as 0, and 2 for the homozygotes and 1 for the heterozygote, u_i is the random regression coefficient for *j*th SNPj and e_i is the residual effect. Analyses were also carried out with the polygenic effects excluded from the model and results were compared. The traits analyzed were milk, fat and protein yields, somatic cell count (SCC) and longevity.

The bulls genotyped were born from 1970 to 2006 and the age distribution is shown in Table 1. The bulls born before 2005 were used as the reference population and were used to estimate the SNP effects for production traits. Bulls born after 2004 were used for the purposes of validation. The bulls in both data sets were required to have at least 10 EDCs and a reliability of at least 69%. However for SCC and longevity, bulls born prior to 2004 were used as reference population and all other bulls were assigned to the validation set. There were 7469 and 1849 bulls in the reference and validation data set for production traits respectively. Corresponding figures for were 4448 and 1465 for SCC, and 6423 and 1309 for longevity.

A traditional evaluation of the genotyped bulls in the reference population was implemented using the additive genetic relationship matrix. This meant that pedigree index (PI) that includes equivalent information to genomic evaluations can be computed. The PI for bulls was computed using only the male pedigree, that is, 0.5(sire PTA + 0.25 maternal grandsire PTA + 0.125 grand maternal grandsire + …)

Reliabilities for direct genomic values (DGV) were computed by inversion of the MME for SNP effects (VanRaden, 2008).

The accuracy of genomic evaluation was computed as correlations between direct genomic values (DGV) and de-regressed evaluations for bulls in the validation set.

Final genomic evaluations were computed for bulls in the reference population using the selection index approach of VanRaden *et al.* (2009) to combine the DGV with PTAs from the subset evaluations and official PTAs. Correspondingly, genomic evaluations (GPI) for the validation set were computed from a combination the DGV, pedigree index (PI), based on only male pedigree from the subset evaluations and PI from official evaluations. The gain in reliability from the traditional PI and GPI were then computed and are summarized for the various traits.

Results and Discussion

The estimation of SNP effects for the production traits took about 16 minutes per trait with MiX99, however run times increased to 85 minutes with the inclusion of polygenic effects. However run times for SCC and longevity were much less at 26 minutes with polygenic effects. The accuracy of evaluations in terms of relationships between DGVs and deregressed proofs in the validation are given in Table 2. In general correlations were about 0.68 for production traits and SCC but lower at 0.45 for longevity. These are similar to those reported by Liu *et al.* (2011) for German Holsteins. The inclusion of polygenic effects in the model resulted in a slight decrease in correlations for production traits but increased the correlation for longevity and had no effect for SCC. However for all traits the inclusion of the polygenic effect increased the regression coefficients to about unity. Thus the inclusion of polygenic effects resulted in the increased predictive ability of the model (Liu *et al.*, 2011).

In Table 3, the gain in reliability from comparing tradition PI with PI enhanced with genomic predictions (GPI) are shown. The gain in reliability was about 22 to 23 % for production traits and SCC but lower at 15% for longevity. These are consistent with the estimates reported by VanRaden *et al.* (2009) for the Holstein population in the USA.

Developments will continue on the reported traits and will be extended to include fertility and calving traits prior to the official publication of results in the UK

References

Berry, P.D., Kearney, F. & Harris, B. 2009. Genomic selection in Ireland. *Interbull Bulletin 39,* 29-34.

Lidauer, M., Matilainen, K., Mantysaari, E. & Straden, I. 2011. *MiX99 Manual.* MTT , Jokioinen, Finland.

Liu, Z. 2009. Deregressing MACE Proofs for genomic evaluations. *Paper presented at PROTEJE meeting in Brussels,* 2009.

Liu, Z., Seefried, F.R., Reinhardt, F., Rensing, S., Thaller, G. & Reents, R. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. *Genetics Selection Evolution 43,* 19.

Meuwissen, T.H, Hayes, B. & Goddard, M.E. 2001. Prediction of total genetic value using genowide-wide dense marker maps. *Genetics 157,* 1819-1829.

VanRaden, P.M. 2008. Efficient method to compute genomic predictions. *J. Dairy Sci. 91,* 4414-4423.

VanRaden, P.M., Tassel, C.P., Wiggans, G.R, Sionstegard, R.D, Taylor, J.F. & Schenkel, F.S. 2009. Reliability of genomic predictions for north American Holstein bulls. *J. Dairy Sci. 92,* 16-24.

Table 1. Distribution of genotyped with data by year of birth.

Year of birth	Bulls with UK daughters	Bulls with only MACE proofs
<1994	659	657
1995	101	485
1996	90	513
1997	152	697
1998	106	391
1999	139	421
2000	112	411
2001	158	501
2002	97	637
2003	106	1245
2004	102	1435
>2005	138	1259

Table 2. Correlations (Corr), regressions (Reg) and mean differences between de-regressed PTAs and DGVs (MD).

Trait	No polygenic effect			10% polygenic effect		
	Corr	Reg	MD	Corr	Reg	MD
Milk yield	0.68	0.83	-11	0.66	0.99	25
Fat yield	0.68	0.87	-0.22	0.67	1.03	1.1
Protein yield	0.65	0.82	0.21	0.64	0.98	1.1
SCC	0.69	0.91	0.56	0.69	1.10	-1.0
Longevity	0.45	0.63	0.02	0.49	1.14	0.05

Table 3. Reliabilities for bulls in the validation data set.

Trait	Pedigree Index	Genomic prediction	Gain
Milk yield	31	53	22
Fat yield	31	54	23
Protein yield	31	53	22
SCC	31	51	20
Longevity	30	45	15