Multi-Country Genomic Selection

C. Schrooten, and C. van der Linde
CRV, P.O. Box 454, 6800 AL, Arnhem, the Netherlands

Abstract

CRV is an international organization with breeding programs in multiple countries. Multiple reference populations for genomic selection are available within CRV. CRV implemented genomic selection in 2006 in its breeding programs in the Netherlands/Flanders, and subsequently in New Zealand. This paper describes how information from multiple reference populations can be combined. A comparison is made between using a US-based reference population with EBV based on local daughters only, vs. a reference population where also bulls with MACE EBV based on foreign daughters are included.

Introduction

CRV is an international cooperative cattle improvement organization with operations and breeding programs in Oceania, Latin America, Europe and the U.S. Across the various breeds and breeding programs, approximately 450 bulls are progeny tested each year. The largest testing program is for Holstein in the Netherlands and Flanders, with 200 progeny tested bulls. Other testing programs are for MRY (10) in the Netherlands and Flanders, Holstein (60) and Jersey (30) in New Zealand, Holstein (45) and Fleckvieh (Simmental) (40) in Czech Republic, and Fleckvieh (60) in Germany.

Figure 1. Overview of CRV locations worldwide.

In 2006, the use of genomic information was implemented in the breeding program in the Netherlands/Flanders (de Roos *et al.*, 2009), for pre-selection in the bulls to be progeny tested. Initially, information for genomic selection (GS) consisted of a reference population of approximately 1500 bulls with genotypes for 3072 SNP markers. From September 2007 onwards, genotypes were obtained using a custom 60K SNP Beadchip (CRV 60K SNP-chip). The reference population has gradually increased from 1500 in 2007 to more than 4000 in 2009. Early 2008, GS was also implemented in the breeding programs for Friesian and Jersey in New Zealand.

In September 2009, CRV USA started, with the operation of an American-based breeding program as one of the main activities. This program involves contracting bull dams and the genomic testing of a substantial number of young bulls each year. Currently about 1200 bulls with progeny-based estimated breeding values (EBV) in the United States have been genotyped by CRV, resulting in a reference population of 1200 bulls for the application of GS on the US base. This could be used as a "standalone" reference population, but on top of that, information from the reference population in the Netherlands/Flanders could add to the increase in reliability obtained from marker information. This study compares both alternatives.

Material and Methods

Data

Data comprised 5330 genotyped (CRV 60K SNP-chip) bulls. Most of the genotyped bulls (5258) had official EBVs in the United States for production traits, where 1156 bulls had EBVs based on US daughters, and 4102 bulls had MACE (Multiple Across Country Evaluation) EBVs based on foreign daughters. Official EBVs from the August 2009 evaluation were available for 9 traits listed in the 038-file of the USDA: production traits,

fertility, longevity, somatic cell count and Net Merit. Progeny-based EBVs on the US scale were included when trait reliabilities were at least 60% and the number of daughters at least 10. Number of daughters and reliabilities were available per trait. MACE EBVs were included for bulls with 60% or higher reliability. Only Black and White and Red and White Holstein (breed codes HO and WW) were included.

Evaluations

Two different evaluations were carried out:

1. EVAL_US: Evaluation including genotyped bulls with progeny-based EBVs in the US.
2. EVAL_ALL: Evaluation including genotyped bulls with US-progeny-based EBVs or MACE EBVs on the US scale.

Each evaluation consisted of two evaluations, differing in information included in the model:

A) pedigree information
B) pedigree and genomic information

The genomic breeding values (Direct estimated Genomic Values, DGVs) were estimated in a GS evaluation where the phenotypes of validation bulls were omitted from the data. In EVAL_US, this was done in 16 replicates, where in each replicate 1/16 of the phenotypes were omitted. As a result, the progeny-based EBV of each validation bull was included in 15 replicates and omitted in 1 replicate. In EVAL_ALL, DGVs were estimated in 3 replicates per trait with identical data in the analysis, i.e. phenotypes of all validation bulls were omitted in each replicate. The same validation bulls were included in all the three replicates.

In the evaluation with only pedigree included, the pedigree EBV (PED) of the bull was calculated, and in the evaluation with phenotypes and genomic information included, the DGV was calculated. The PED and the DGV from the evaluation without the phenotype of the bull were used for the validation.

Bulls included in the validation were required to meet the following criteria:

- Bull is genotyped and progeny tested
- Sire of bull is genotyped and progeny tested
- Bull has no sons that are genotyped and progeny tested
- Bull is born between 1999 and 2004

The number of reference bulls differed per evaluation, but the number of validation bulls was 111 in all evaluations.

The DGVs and the PEDs were compared to the progeny-based EBVs by their squared correlation (R^2). The R^2-values were adjusted for reliabilities of phenotypes of the validation bulls being lower than 100% by dividing the R^2 by the average reliability of the phenotypes of the validation bulls. The DGVs and the PEDs were also compared to the progeny-based EBV by their linear regression coefficient (y = progeny-based EBV and x = DGV, $y = a + b*x$).

The difference in R^2-values ($R^2_{\text{DGVs, progeny-based EBVs}}$ minus $R^2_{\text{PEDs,progeny-based EBVs}}$) was considered to be the increase in reliability resulting from genomic information. Reliabilities obtained in this study were compared to reliabilities based on the Dutch/Flemish reference population. All reliabilities were adjusted for reliabilities of phenotypes of the validation bulls being lower than 100%.

Model

Genomic predictions were calculated with the Bayesian multiple QTL model of Meuwissen and Goddard (2004), but fitting SNP genotypes rather than haplotypes and identical-by-descent probabilities (Calus *et al.*, 2008). The following model was used:

$$y_i = \mu + u_i + \Sigma_{40653}\, z_{ij}\, q_j + e_i,$$

where

y_i progeny-based EBV of bull *i*
u_i random polygenic effect of bull *i*; var(u) = A σ^2_u where A is the relationship matrix; var(e) = σ^2_e
q_j vector of effects for SNP *j*; mixed distribution: most SNPs from $\sigma^2/100$, few SNPs from σ^2, with σ^2 = variance of putative SNP effects
z_{ij} incidence vector for bull *i* at SNP *j*: [2 0], [1 1], [0 2]
e_i residual for bull *i*

Bayesian techniques with Gibbs sampling were used to estimate all parameters. The Gibbs sampler was run for 10,000 iterations, and the first 2,000 iterations were considered as burn-in. DGVs were calculated as

$\mu + \text{u-hat}_i + \Sigma_{40653}\, z_{ij}\, \text{q-hat}_j$,

where u-hat_i and q-hat_j are the posterior means for u_i and q_j. PEDs were calculated from the same data, but using the following model:

$y_i = \mu + u_i + e_i$.

Results and Discussion

Table 1 shows reliabilities (R^2) of the PED and the DGV with the progeny-based EBV and the regression coefficients of the regression of progeny-based EBV (phenotype) on DGV for evaluation EVAL_US.

Table 1. Reliabilities and regression coefficients for GS evaluations of validation bulls (n=111) including only bulls with US progeny-based EBVs as reference bulls (EVAL_US).

Trait	Reliability		Difference	Regression coefficient		Number of reference bulls
	PED	DGV	(DGV–PED) / rel_{EBV}[1]	a	b	
Milk	0.519	0.625	0.111	14.97	1.02	1156
Fat	0.371	0.578	0.218	-0.61	1.11	1156
Protein	0.516	0.609	0.098	-0.88	1.08	1156
Fat percentage	0.328	0.660	0.349	0.26	0.98	1156
Protein percentage	0.332	0.513	0.190	0.19	0.91	1156
Somatic cell score	0.271	0.265	-0.007	56.71	0.80	1130
Daughter pregnancy rate	0.429	0.489	0.073	1.40	1.07	1028
Productive life	0.416	0.531	0.140	2.01	1.08	1108
Net Merit	0.425	0.522	0.108	27.68	1.01	1152

[1] Average reliability of proofs of validation bulls (Table 1).

The number of reference bulls per trait ranged from 1028 for daughter pregnancy rate to 1156 for production traits. Differences in reliability between PED and DGV, adjusted for average reliability of phenotypes of validation bulls, ranged from –0.01 (somatic cell score) to +0.35 (fat percentage). The average difference in reliability of the 9 traits was 0.142. Regression coefficients of DGVs on phenotypes were between 0.80 and 1.11.

The increase in reliability was 0.06 lower compared to results found in the Dutch/Flemish reference population of size 1400. The increase in reliability in that situation ranged from .08 (longevity) to .38 (fat percentage) (unpublished results).

Table 2 shows reliabilities (R^2) of the PED and the DGV with the progeny-based EBV and the regression coefficients of the regression of progeny-based EBV on DGV for evaluation EVAL_ALL. Figure 2 shows the increase in reliability due to genomic information in evaluations EVAL_US and EVAL_ALL.

The number of reference bulls per trait ranged from 4325 for productive life to 5093 for production traits. Differences in reliability between PED and DGV, adjusted for average reliability of phenotypes of validation bulls, ranged from 0.13 (daughter pregnancy rate) to 0.41 (fat percentage). The average difference in reliability of the 9 traits was 0.229. The average difference in reliability between PED

and DGV was 0.09 higher in evaluation EVAL_ALL compared to evaluation EVAL_US. Regression coefficients of DGVs on phenotypes were between 0.93 and 1.31.

Table 2. Reliability and regression coefficients for GS evaluations of validation bulls (n=111) including US progeny-based EBVs and MACE EBVs as phenotype (EVAL_ALL).

Trait	Reliability		Difference	Regression coefficient		Number of reference bulls
	PED	DGV	(DGV–PED) / rel_{EBV}[1]	a	b	
Milk	0.519	0.670	0.159	3.81	1.19	5093
Fat	0.360	0.604	0.257	-1.21	1.26	5093
Protein	0.487	0.675	0.198	-3.05	1.31	5093
Fat percentage	0.341	0.732	0.412	0.11	0.99	5093
Protein percentage	0.354	0.676	0.339	-0.02	0.97	5093
Somatic cell score	0.218	0.451	0.259	17.90	0.93	4995
Daughter pregnancy rate	0.410	0.512	0.127	0.77	1.15	4815
Productive life	0.426	0.537	0.137	2.06	1.14	4325
Net Merit	0.412	0.566	0.171	9.01	1.19	4959

[1] Average reliability of proofs of validation bulls (Table 1).

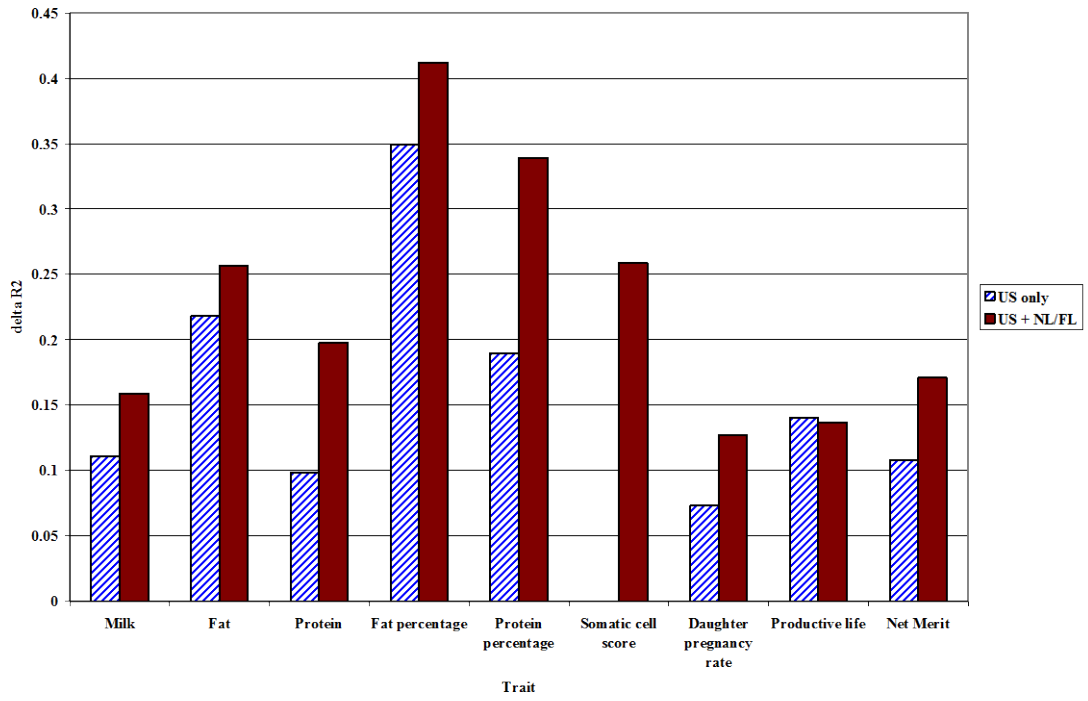

Figure 2. Increase in R^2 (delta R^2) between DGV and progeny-based EBV, relative to R^2 between PED and progeny-based EBV, for two reference populations: only genotyped bulls with EBV based on US progeny (EVAL_US), and all genotyped bulls with official EBV based on the US scale (EVAL_ALL)

The increase in reliability was 0.03 lower compared to results found in the Dutch/Flemish reference population of 3200 bulls. There, the increase in reliability ranged from 0.11 (daughter pregnancy rate) to 0.42 (fat percentage) (unpublished results). VanRaden *et al.* (2009) studied increase in reliability for a reference population of 3576 bulls, comparing genomic reliability to reliability of traditional parent average. For the 9 traits listed in Table 2, they found on average 0.03 higher increase in reliability compared to the current study.

For all traits, except productive life, there was a further increase in reliability when foreign reference bulls were added with their MACE proofs. Benefit was highest for somatic cell score, but the increase in reliability using the US-based reference population was unexpectedly low. It has not been clarified what caused this result.

Regression coefficients increased when a large number of bulls with their MACE EBV was added to the reference population. Due to genetic correlations between countries being lower than 1.0, MACE EBV show less variance, which may also cause an underestimation of the variance of genomic effects. This results in a slight bias of DGV, and a higher regression coefficient compared to the evaluation based on a US-only reference population (EVAL_US). An analysis based on weighted Deregressed Proofs (DRP) or Daughter Yield Deviations (DYD) instead of progeny-based EBVs is preferred, as also indicated by Garrick *et al.* (2009), but DRP and DYD were not available for this study.
Another way to deal with a genetic correlation between countries lower than one is to analyze phenotypes on each country's scale as two correlated traits in a multi-trait analysis. This is currently being explored.

Summary and Conclusions

- Reliability of DGV based on a reference population of 1000-1200 bulls with US-progeny-based EBV was 10-20% higher than reliability of PED.
- A further 5-15% increase in reliability of DGV was obtained when almost 4000 reference bulls with MACE EBV based on foreign daughters were added to the reference population.
- Reliabilities of genomic selection on the Dutch/Flemish population and North American populations were comparable taking into account the size of the reference population.
- Results might be even better when alternative phenotypes are used, like DRP or DYD combined with appropriate weighting factors, or when analyzing the same trait in different countries as two correlated traits.

References

Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W. & Veerkamp, R.F. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics 178*, 553-561.

de Roos, A.P.W., Schrooten, C., Mullaart, E., van der Beek, S., de Jong, G. & Voskamp, W. 2009. Genomic Selection at CRV. INTERBULL international workshop, Uppsala, Sweden, January 26-29. *Interbull Bulletin 39*, 47-50.

Garrick, D.J., Taylor, J.F. & Fernando, R.L. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet. Sel. Evol. 41*, 55-62.

Meuwissen, T.H.E. & Goddard, M.E. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multitrait data. *Genet. Sel. Evol. 36*, 261-279.

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

VanRaden, P.M., VanTassel, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S. Invited review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci. 92*, 16-24.