Genomic Selection at CRV

A.P.W. de Roos, C. Schrooten, E. Mullaart, S. van der Beek, G. de Jong, and W. Voskamp
CRV, P.O. Box 454, 6800 AL, Arnhem, the Netherlands

Abstract

CRV implemented genomic selection in 2006 and currently uses it in its breeding programs in the Netherlands/Flanders and New Zealand. Genomic predictions are combined with national breeding values and subsequently published.

Introduction

The aim of this paper is to present the research, development and implementation of genomic selection at CRV and the implications for the national genetic evaluations in the Netherlands and Flanders, and in New Zealand.

First adopter

CRV and formerly Holland Genetics and CR Delta/VRV have invested substantially in QTL (fine) mapping and marker assisted selection projects through own research and supporting research at universities and research institutes, predominantly in Wageningen and Lelystad, the Netherlands (Prof. Johan van Arendonk and Dr. Roel Veerkamp), in Liege, Belgium (Prof. Michel Georges) and Ås, Norway (Prof. Theo Meuwissen). The research aimed at whole genome QTL detection (e.g. Schrooten *et al.*, 2000), QTL fine mapping (e.g. Blott *et al.*, 2003), marker assisted breeding value estimation (e.g. Meuwissen and Goddard, 2004), and the design of marker assisted breeding programs (e.g. Schrooten *et al.*, 2005). One of the most important findings was the discovery of the K232A mutation in the DGAT1 gene which has a large effect on milk production and fat and protein percentage (Grisart *et al.*, 2002). Although the QTL (fine) mapping projects were successful the overall impact on the breeding program was limited. When Meuwissen *et al.* (2001) introduced the idea of genomic selection, the number of markers available in cattle was too small and the costs for genotyping was too high for application. Meuwissen and Goddard (2004) developed a Bayesian multiple QTL model, which is based on haplotypes and identical-by-descent probabilities based on linkage and linkage disequilibrium. This facilitated breeding value estimation using densely genotyped parts of the genome (De Roos *et al.*, 2007). In 2005, the G-Lection project was initiated as a collaboration between the Universities of Wageningen and Liege, the Animal Sciences Group in Lelystad, Prof. Theo Meuwissen, and the breeding companies CRV (cattle), IPG (pigs) and Nutreco/Hendrix Genetics (poultry, pigs). In 2006, CRV was the first animal breeding organisation in the world to use genomic selection in its breeding program. The reference population comprised ~1500 progeny tested bulls, and over a period of one year ~1000 selection candidates were genotyped. The animals were genotyped for 3072 public SNPs, using a custom Illumina GoldenGate assay. This application gave ~5-10% higher reliability compared to parent average and a lot of experience with using genomic selection in the breeding program.

Illumina 60K SNP assay

Early 2007, CRV and the University of Liege developed a custom 60K SNP Beadchip for the Illumina iSelect platform, using publicly available SNPs. The aim was to have at least the same level of linkage disequilibrium between markers and QTL within the Holstein Friesian population as in the simulation by Meuwissen *et al.* (2001). The initial reference population comprised ~1500 progeny tested bulls, which was extended to ~3600 bulls in 2008. Around 48,000 SNPs were polymorphic in our population and used in genomic predictions since October 2007. The total number of animals genotyped to date exceeds 10,000.

Genomic predictions

Genomic predictions are calculated with the Bayesian multiple QTL model of Meuwissen and Goddard (2004), but fits SNP genotypes rather than haplotypes and identical-by-descent probabilities, because of the higher marker density (Calus *et al.*, 2008). The model includes a random polygenic effect and a random effect for each SNP. Subsequently, the same data is also analysed with a model that includes only a random polygenic effect. The difference in the posterior means of the total breeding values between the genomic prediction model ($\hat{u}_{genomic}$) and the polygenic model ($\hat{u}_{polygenic}$) is used as the marker effect of an animal ($\hat{u}_{mark} = \hat{u}_{genomic} - \hat{u}_{polygenic}$).

Integration with national EBVs

The marker effect ($\hat{u}_{mark}$) is subsequently combined with the official national breeding value or parent average of the animal ($\hat{u}_{nat}$). The national breeding value is therefore divided into two parts, a sire pedigree index (PI) and the rest, which includes the Mendelian sampling effects of the animal and its maternal pedigree (MS):

$$\hat{u}_{nat,PI} = \tfrac{1}{2}\hat{u}_{nat,sire} + \tfrac{1}{4}\hat{u}_{nat,mgs} + \ldots$$

$$\hat{u}_{nat,MS} = \hat{u}_{nat} - \hat{u}_{nat,PI}$$

The selection index is:

$$\hat{u}_{tot} = \hat{u}_{nat,PI} + b_{mark}\hat{u}_{mark} + b_{nat,MS}\hat{u}_{nat,MS}$$

$$\mathbf{b} = \mathbf{P}^{-1}\mathbf{G}$$

$$\begin{pmatrix} b_{mark} \\ b_{nat} \end{pmatrix} = \begin{pmatrix} r^2_{mark} & \dfrac{r^2_{mark} r^2_{nat,MS}}{1 - r^2_{nat,PI}} \\ symm & r^2_{nat,MS} \end{pmatrix}^{-1} \begin{pmatrix} r^2_{mark} \\ r^2_{nat,MS} \end{pmatrix} \text{ and}$$

the associated reliability is:

$$r^2_{tot} = \mathbf{b}'\mathbf{P}\mathbf{b}$$

$$= r^2_{nat,PI} + b_{mark} r^2_{mark} + b_{nat} r^2_{nat,MS}$$

where r^2_{mark} is the reliability of the marker effect, which is obtained from a validation study (see next section), $r^2_{nat,PI}$ is the reliability of the national male pedigree index and $r^2_{nat,MS} = r^2_{nat} - r^2_{nat,PI}$, where r^2_{nat} is the reliability of the national breeding value. VanRaden *et al.* (2009) also proposed a selection index for combining genomic predictions with national breeding values, but they used a theoretical reliability of the genomic prediction in the selection index, which was substantially higher than the observed reliability in a validation study.

Validation

To assess the reliability of the marker effects (r^2_{mark}), the phenotypes of all progeny tested bulls born in or after 2001 were omitted from the evaluation and the remaining 3160 reference bulls were used to predict their breeding value, using the genomic prediction model and the polygenic model. Squared correlations between predicted and national breeding values (R^2) were computed for a subset of 260 bulls, including only Black-and-white Holstein Friesian bulls born in 2001 or 2002, with a sire but no sons in the reference population. The reliability of the marker effects was computed as:

$$r^2_{mark} = \frac{R^2_{genomic} - R^2_{polygenic}}{\bar{r}^2_{nat}}$$

where $\bar{r}^2_{nat}$ is the average reliability of the national breeding values of the 260 validation bulls. VanRaden *et al.* (2009) calculated reliabilities as $r^2_{tot} = r^2_{nat} + \dfrac{R^2_{genomic} - R^2_{polygenic}}{\bar{r}^2_{nat}}$ rather than deriving it directly from their selection index. This implicitly assumes that $\hat{u}_{mark}$ and $\hat{u}_{nat,MS}$ are independent, while they both partially explain the Mendelian sampling effect of the maternal pedigree.

Reliabilities

The reliabilities of the combined breeding values (r^2_{tot}) for animals without own or progeny performance varied from 0.27 (maternal calving ease) to 0.78 (fat percentage) and were on average 0.66 for production traits (n = 6), 0.52 for type traits (n = 23) and 0.44 for functional traits (n = 12). Results for a subset of these traits are listed in Table 1. The average increase in reliability was 0.31 for production traits, 0.22 for type traits and 0.17 for functional traits. The reliabilities were on average 0.04 higher than those reported by VanRaden *et al.* (2009) for a subset of 27 traits analysed in both studies, while the reliabilities across traits followed a very similar ranking (correlation = 0.77).

Table 1. Reliabilities of national (Nat.) and combined (Tot.) breeding values for animals without own or progeny performance.

Trait	Nat.	Tot.	Diff.
kg milk	0.35	0.61	0.26
kg fat	0.35	0.68	0.33
kg protein	0.35	0.64	0.29
rump angle	0.33	0.53	0.20
body condition	0.29	0.53	0.24
udder depth	0.34	0.57	0.23
locomotion	0.27	0.54	0.27
somatic cell score	0.32	0.55	0.23
fertility index	0.30	0.41	0.11
longevity	0.22	0.37	0.15
direct calving ease	0.29	0.49	0.20
mat. calving ease	0.23	0.27	0.04

From validations with a reduced number of reference bulls, it was concluded that the reliabilities for most traits increase almost linearly with the number of reference bulls, which was also observed by VanRaden *et al.* (2009). Based on these conclusions, CRV considers to extend the reference population to ≥5500.

InSire

CRV's Holstein Friesian breeding program in the Netherlands and Flanders has been adjusted to the use of genomic selection. A group of 1000 bull calves are genotyped each year, out of which the highest 200 are progeny tested. These genomically selected bulls are named "InSire" bulls. Five hundred bull calves come from commercial breeders and the other 500 from CRV's nucleus program. Within the nucleus program, 500 heifers are genotyped per year, out of which the highest 100 are used as dams. The highest ranking InSire bulls are used as sires. The highest ranking 3- and 4-year old InSire bulls for a certain market segment (or breeding objective) are commercially available in "six-packs", i.e. packages of six InSire bulls with 5 or 10 straws of semen per bull.

National genetic evaluation

The official genetic evaluation for bulls in the Netherlands and Flanders is carried out by CRV, under the responsibility of NVO (Dutch Cattle Improvement Organisations). Genomic information is not yet used in the official genetic evaluation. Meanwhile, CRV reports the combined breeding values including genomic information for young InSire bulls available for progeny testing in its monthly CRV magazine. Methods for propagation of genomic information to non-genotyped relatives of genotyped animals (e.g. Gengler and VanRaden, 2008) need to be studied and developed. Furthermore, the overestimation of some high ranking cows in the national genetic evaluation needs to be studied to prevent biases in combined breeding values.

CRV Ambreed, New Zealand

CRV Ambreed is the second largest dairy breeding company in New Zealand with a market share of 20-25%. CRV has established a reference population which currently comprises 1040 New Zealand Friesian bulls and 440 New Zealand Jersey bulls. Breeding values are computed using the same methods as in the Netherlands/Flanders. Because of the limited size of the New Zealand reference population, multiple trait genomic predictions (Meuwissen and Goddard, 2004) are currently being developed to account for genotype by environment interaction between countries. This model would allow simultaneous analysis of New Zealand and Dutch/Flemish phenotypes, and is therefore expected to result in higher reliabilities and direct conversions to each scale.

The national genetic evaluation in New Zealand is carried out by New Zealand Animal Evaluation Limited (NZAEL). In New Zealand, both Livestock Improvement Corporation (LIC) and CRV have established reference populations and genomic predictions. Researchers at NZAEL have proposed a method to combine genomic information from multiple companies in the national genetic evaluation, while respecting each company's intellectual property, using a genomic relationship matrix. Other alternatives to integrate both company's genomic information would be to merge the reference populations or to integrate genomic predictions.

References

Blott, S., Kim, J.-J., Moisio, S., Schmidt-Küntzel, A., Cornet, A. *et al.* 2003. Molecular dissection of a quantitative trait locus: A phenylalanine-to-tyrosine substitution in the transmembrane domain of the bovine growth hormone receptor is associated with a major effect on milk yield and composition. *Genetics 163*, 253-266.

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., Calus, M.P.L. & Veerkamp, R.F. 2007. Breeding value estimation for fat percentage using dense markers on *Bos taurus* autosome 14. *J. Dairy Sci. 90*, 4821-4829.

Gengler, N. & VanRaden, P.M. 2008. Strategies to incorporate genomic prediction into population-wide genetic evaluations. *J. Dairy Sci. 91, E-Suppl.*, 506.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C. *et al.* 2002. Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine DGAT1 gene with major effect on milk yield and composition. *Genome Res. 12*, 222-231.

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.

Schrooten, C., Bovenhuis, H., Coppieters, W. & van Arendonk, J.A.M. 2000. Whole genome scan to detect quantitative trait loci for conformation and functional traits in dairy cattle. *J. Dairy Sci. 83*, 4795-4806.

Schrooten, C., Bovenhuis, H., van Arendonk, J.A.M. & Bijma, P. 2005. Genetic progress in multistage dairy cattle breeding schemes using genetic markers. *J. Dairy Sci. 88*, 1569-1581.

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.