Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Integrate cow and bull data in a genomic evaluation for conformation traits and claw health

G.C.B. Schopen*, P.J.A.Vessies*, M.P.L. Calus†, C. Schrooten* and A.P.W. de Roos*
* CRV, PO Box 454, 6800 AL Arnhem, the Netherlands
†Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, Wageningen, the Netherlands

ABSTRACT: The two objectives of this study were to investigate and find methods to successfully integrate cow data in the bull reference population for genomic evaluation and to investigate the effect of adding reference cows on the DGV reliability for conformation traits and claw health. Information from about 25,000 bulls and about 15,000 cows was available. Bulls were genotyped with the Illumina 50K SNP chip and the cows with the Illumina 10K SNP chip. All animals were imputed to an equal 50K SNP set. After SNP edits 37,995 SNP remain for all animals. As phenotypes, yield deviations, deregressed proofs (DRPs) with adjustments for cows and DRPs calculated based on matrix deregression will be used. The three kinds of phenotypes will be validated to investigate the effect on the reliability of direct genomic value for conformation traits and claw health.
Keywords: cow data; genomic selection; conformation traits; claw health

Introduction

Direct genomic values (DGVs) are based on a large group of animals, usually progeny tested bulls, with known genotypes and phenotypes. To improve the reliability of DGVs for traditional traits and to develop DGVs for new traits, cow information can be added to the bull reference population (Buch et al., 2012).

CRV started a project called DataPlus (e.g. Stoop et al., 2014) to genotype cows to increase the reference population for genomics. Farmers from the Netherlands and Flanders can participate in this project. From the participating farms, all females, i.e. milking cows and young stock are genotyped. In addition, all new-born heifer calves will be genotyped after birth. The benefits for the farmers to genotype all their females are: more reliable estimation of the genetic potential of their animals, genomics integrated in official cow EBVs and used in the mating program.

In 2013, research has been performed in which the genotyped cows were added to the bull reference population to investigate the effect on the reliability of DGV. Unfortunately, results showed a decrease in reliability, also for the production traits, and were not reliable due to the phenotypes (deregressed proofs, DRPs) of the cows which were used. The deregression for the cows was done using the deregression method developed for only bulls. Therefore, this follow-up study has two objectives. First objective is to investigate and find methods to integrate cow data in the bull reference such that reliabilities of DGV increase. Second objective is to investigate the effect on the DGV reliability when adding cows, based on a genomic validation study for conformation traits and claw health.

Materials and Methods

Animals. The bull reference population consists of about 25,000 animals. The genotyping of cows is an ongoing process, meaning that the number of genotyped cows increases each week. At the end of 2013, about 15,000 cows and young stock have been genotyped within the DataPlus project. In total, the cow and bull reference population consists of about 40,000 animals.

Phenotypes. For bulls and cows, official national breeding values for all traits are available. Based on these breeding values, DRPs can be calculated and used as phenotypes. In a previous study, the use of DRPs, as calculated in the same way as for bulls, showed unreliable results. To use DRPs for cows and bulls, another method to calculate DRP is needed, that is able to accurately de-regress low reliability breeding values of cows.

Another type of phenotypes which can be used is yield deviations. Yield deviations are available for conformation traits and claw health. The conformation traits contain both high and low heritability traits. Claw health is a new trait with low amounts of data and therefore a trait with different characteristics than the conformation traits.

Genotypes. Bulls were genotyped for 50K SNP markers and the cows were genotyped for 10K SNP. The 50K genotypes were based on version 1 or 2 of a custom 50K SNP chip or on version 1 or 2 of the BovineSNP50 BeadChip (Matukumalli et al., 2009). The 10K genotypes were based on the Illumina 10K chip. Genotyping was done at the University of Liege or at GeneSeek. The genotypes of all bulls were imputed to a combined 50K SNP chip (containing SNP from the custom-made 50K chip and the Illumina 50K chip) to which also the 10K genotyped cows were imputed.

The following SNP edits were applied: call rate (>90%), minor allele frequency (>2.5%), and Hardy-Weinberg disequilibrium (maximum deviation of 0.15 between observed and predicted fraction of heterozygotes). Before imputation, a call rate of >90% also was applied per animal.

The imputation was done using BEAGLE software (Browning and Browning, 2007), combined with PHASE-

BOOK software (Druet and Georges, 2010). After imputation, the same SNP edits as described previously were applied. After SNP editing, the final data set consisted of about 40,000 Holstein-Friesians with 37,995 SNP distributed across 29 autosomes.

Investigation other methods to integrate cow data. Different national genetic evaluation centers were contacted to assess if and how they integrate cow data in the genomic breeding value estimation. Questions posed included: which phenotypes do they use, i.e. EBVs, DRPs or yield deviations and which method do they exactly use to get these phenotypes. In addition a literature study was done about the different methods available to transform cow phenotypes into reliable input for genomic evaluation.

Based on the outcome of the methods used by other national genetic evaluation centers and in literature, different methods (i.e. phenotypes for cows as input for genomic selection) will be selected. The selected methods will be evaluated in a genomic validation study.

Genomic validation. The effect of adding cow information to the bull reference data on the reliability of DGVs will be assessed by validations using three different phenotypes: yield deviations, DRPs with adjustments for cows based on the mean and variance to make cow evaluations more like bull evaluations (Wiggans et al., 2011) and DRPs calculated with matrix deregression. The genomic validation procedure consists of four steps; (1) identification of validation bulls based on the following rules: bull and sire of bull were genotyped and progeny tested, bull had no sons that were genotyped and progeny tested, bull had no daughters that were genotyped and was a black & white Holstein-Friesian bull; (2) phenotypes of validation bulls will be omitted from the data and DGVs and pedigree EBVs (PEBV) will be estimated for these omitted bulls in a genomic evaluation; (3) comparison of DGV and PEBV of validation bulls based on their correlations with their phenotypes; (4) calculation of reliability of DGV and PEBV.

The model to estimate DGV's for one trait is known as Bayesian stochastic search variable selection (e.g. Verbyla et al 2009):

$$y_i = \mu + u_i + \sum_j^n X_{ij} a_j + e_i \qquad [1]$$

where y_i is the phenotype of bull *i*, μ is the overall mean, u_i is the random polygenic effect of bull *i*, n is the total number of loci, X_{ij} is the genotype at SNP *j*, a_j is the allele substitution effect at SNP *j*, and e_i is the residual for bull *i*. Each phenotype will be weighted. Using yield deviations, the weights will be based on formula's as described in Garrick et al., (2009). Using DRPs the weight will be based on effective daughter contributions (EDC, Fikse and Banos , 2001).

A Markov Chain Monte Carlo method using Gibbs sampling was used to obtain posterior estimates for all effects in the model. The conditional posterior density of a_j is:

$$N\left(\hat{a}_j; \frac{\omega_j \hat{\sigma}_e^2}{\mathbf{x}_j' \boldsymbol{R}^{-1} \mathbf{x}_j + \lambda_j}\right)$$

where $\hat{a}_j$ is the conditional mean of the allele substitution effect at locus *j*, $\lambda_j = \frac{\omega_j \hat{\sigma}_e^2}{\hat{\sigma}_a^2}$, where $\omega_j = 1$ (if $I_j = 1$) or $\omega_j = 100$ (if $I_j = 0$), and $\boldsymbol{R}^{-1}$ is a diagonal matrix containing the reciprocals of the weights on the diagonals. The conditional posterior density of σ_a^2 was: $\sigma_a^2 | a \sim \chi^{-2}(\nu_a + n, S_a^2 + \boldsymbol{\omega}' \hat{\mathbf{a}}^2)$, where $\hat{\mathbf{a}}^2$ is a vector with squares of the current estimates of the allele substitution effects of all loci, that is weighted by vector $\boldsymbol{\omega}$. The conditional posterior distribution of I_j was:

$$\Pr(I_j = 1) = \frac{f(r_j | I_j = 1)(1 - \pi)}{f(r_j | I_j = 0)\pi + f(r_j | I_j = 1)(1 - \pi)}$$

where $r_j = \mathbf{x}_j' \boldsymbol{R}^{-1} \mathbf{y}^* + \mathbf{x}_j' \boldsymbol{R}^{-1} \mathbf{x}_j \hat{a}_j$ where $\mathbf{y}^*$ are the conditional phenotypes, and $f(r_j | I_j = \delta)$ where δ is either 0 2. More details on the model and the implementation can be found in Calus et al. (2014). The Gibbs sampling will be run for four chains per trait.

PEBVs were estimated from the same data using the following model:

$$y_i = \mu + u_i + e_i \qquad [2]$$

where y_i is the phenotype of bull *i*, μ is overall mean, u_i is the random polygenic effect of bull *i* and e_i is the residual for bull *i*. Each phenotype will be weighted as described above.

Squared correlations (R^2) between DRP and both DGV and PEBV were computed and compared to each other to obtain ΔR^2 using the following formula:

$$\Delta R^2 = \frac{R^2_{DGV,DRP} - R^2_{PEBV,DRP}}{REL_{DRP}},$$

where $R^2_{DGV,DRP}$ is the squared correlation between DGV and DRP, $R^2_{PEBV,DRP}$ is the squared correlation between PEBV and DRP, and REL_{DRP} is the average reliability of DRPs of all validation bulls.

Values for ΔR^2 using the different phenotypes will be compared to ΔR^2 obtained in a validation using only the bull reference population using the same validation animals as in the other studied alternatives.

Results

The study described in this paper is still ongoing and will finish at the beginning of June. Therefore, no information on the effect on the reliability of DGV when adding cows to the bull reference is available yet. However, the first phase of the study, to contact other national genetic evaluation centers and doing a literature study, was carried out. The results of this first phase of the study are described below.

Information for adding cows to the bull reference population was available for about 10 national genetic evaluation centers and research institutes. Over the different national genetic evaluation centers, two different types of phenotypes when adding cows were used. The first type of phenotypes was yield deviations calculated by correction for fixed effects, heterosis and permanent environment. The second type of phenotypes when adding cows to the bull reference was DRPs. For the DRPs, the difference among the national genetic evaluation centers was in the method which was used to calculate the DRP. One method to calculate DRPs was based on the method as described by VanRaden et al. (2009) and Wiggans et al. (2011 and 2012) which are both based on the difference between EBV or PTA and parent average (PA). The major difference between the method of VanRaden et al. (2009) and Wiggans et al. (2011 and 2012) is that Wiggans et al. (2011 and 2012) made adjustments for the cows based on the mean and variance. The adjustments for the cows were done to make cow evaluations more like bull evaluations resulting in an improved estimation of the SNP effects. Therefore, Wiggans et al (2011) showed a different deregression for cows and bulls whereas VanRaden et al. (2009) makes no difference between the deregression of the cows and bulls. Another method to calculate DRPs was based on matrix deregression in which also non-genotyped ancestors were included (Harris and Johnson, 2010 and Berry et al., 2013).

Conclusions

Based on this first phase of the project, yield deviations, DRPs with adjustments for the cows and DRPs calculated using matrix deregression will be validated to see the effect on the reliability of DGV when adding cows to the bull reference.

References

Berry, D. P., Kearney, F., and Harris, B. L. (2013). Interbull Bulletin no. 47. Nantes, France, August 23-25

Browning, S. R., and Browning, B. L. (2007). Am. J. Hum. Genet. 81:1084–1097

Buch, L. H., Kargo, M., Berg, P., et al. (2012). Animal. 6:880-886.

Calus, M. P. L. (2014). Genet. Sel. Evol. In press

Druet, T., and Georges, M. (2010). Genetics 184: 789 - 798

Fikse, W. F., and Banos, G. (2001). J. Dairy Sci. 84:1759-1767

Garrick, D. J., Taylor, J. F., and Fernando, R. L. (2009). Genet. Sel. Evol. 41:55

Harris, B. L., and Johnson, D. L. (2013). J. Dairy Sci. 93: 1243-1252

Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., et al. (2009). PLoS ONE 4:e5350.

VanRaden, P. M., C. P. Van Tassell, C. P., Wiggans, G.R., et al. (2009). J. Dairy Sci. 92:16–24

Verbyla, K. L., Hayes, B. J., Bowman, P. J., et al. (2009). Genet. Res. 91: 307-311

Stoop, W. M., Mullaart, E., and de Jong, G. (2014). Entered for "Proceedings of the 10th World Congress on Genetics Applied to Livestock Production, Vancouver, Canada, August 17-22"

WiWiggans, G.R., Cooper, T. A., VanRaden, P. M., et al. (2011). J. Dairy Sci. 94: 6188-6193

WiWiggans, G.R., VanRaden, P. M., and Cooper, T. A. (2012). J. Dairy Sci. 95: 3444-3447