Proc. Assoc. Advmt. Anim. Breed. Genet. 20:138-141

ACROSS-BREED GENOMIC EVALUATION BASED ON BOVINE HIGH DENSITY GENOTYPES, AND PHENOTYPES OF BULLS AND COWS

C. Schrooten[1], G.C.B. Schopen[1], A. Parker[2], A. Medley[2] and P. Beatson[2]

[1]CRV BV, P.O. Box 454, 6800 AL Arnhem, the Netherlands
[2]CRV AmBreed, NZ PO Box 176, Hamilton 3240, New Zealand

SUMMARY

Most applications of genomic selection are based on a reference population of bulls only, genotyped with 50k SNP-chips. In some populations, the size of the reference population is limited, resulting in relatively low reliabilities of genomic breeding values. In this study we looked at the possibility of expanding the reference population by combining several breeds in one genomic evaluation, and making use of reference cows in addition to reference bulls. Because such an evaluation needs genotypes at higher density than 50k, high density (777k) SNP-chip genotypes were used. Presentation of results was limited to 7 traits. On average, reliabilities were 1-4% higher than reliabilities from a single breed evaluation using a bull reference population with 50k SNP-chip genotypes, and 0-2% higher than reliabilities from an across-breed evaluation based only on reference bulls and high density genotypes.

INTRODUCTION

Genomic evaluation at CRV Ambreed has been based on 50k SNP-chip genotypes and single-breed reference populations of bulls. The individual reference populations for Friesians and Jerseys consist of approximately 2,200 and 1,200 reference bulls, respectively. These reference populations are relatively small, compared to the reference populations in North America (VanRaden 2010) and Europe (Lund *et al*. 2011). Therefore, in these small populations, the use of genomic information is predicted to result in only a moderate increase in reliability of breeding values of animals without phenotype.

Reliabilities may increase further if reference populations are combined in a multi-breed genomic evaluation. To make use of genomic information across breeds, markers must be in Linkage Disequilibrium (LD) with the mutations affecting the trait of interest, and the linkage phase must be the same in the individual breeds. De Roos *et al.* (2008) looked at LD and phase persistency in Holstein Friesian, Jersey, and Angus populations in Australia, New Zealand and the Netherlands. They concluded that strong enough and persistent LD could be obtained when genotyping with at least 300k SNP. Therefore, to combine reference populations for CRV Ambreed, a higher density is needed than obtained with the currently used 50k SNP-chip.

The reference population can also be expanded by adding cows with phenotypic information to the reference population. Because reliability of phenotypic information is lower for cows than for bulls, the benefit of adding a certain number of cows to the reference population is lower than the benefit of adding the same number of bulls. Nevertheless, when no additional bulls are available and genotyping cost are sufficiently low, cows offer a good opportunity to expand the reference population.

The objective of this study was to estimate the effect on reliability of genomic breeding values, when single-breed reference populations are combined, and the reference population is augmented with high density genotypes and cow genotypes and phenotypes.

MATERIALS AND METHODS

Genotypes. Genotypes of 465 Friesians, 227 Jerseys and 57 crossbreds were obtained using the Illumina BovineHD BeadChip, containing 777k SNP-markers:

(http://www.illumina.com/Documents/products/datasheets/datasheet_bovineHD.pdf).

Genotypes of approximately 9,000 animals, obtained with 50k chips, were imputed to BovineHD with Beagle version 3.0 (Browning and Browning, 2007), using the 749 HD-genotyped animals as reference set for imputation. After data edits, 9,486 animals were available for evaluation. Ancestral haplotype scores were obtained for 622k loci on the 29 autosomes. To reduce computer requirements for genomic evaluation and because the full SNP-set contains redundant information due to complete or nearly complete LD between neighboring SNP, the number of HD loci based on hidden states (i.e. ancestral haplotypes) was reduced by considering only the first locus out of every 10 consecutive loci in genomic evaluation. After this reduction, 62,302 loci remained for evaluation.

Genomic evaluation. Genomic evaluation was performed for 26 traits, but presentation of results will be restricted to a subset of 7 traits with moderate to high heritability that are part of the New Zealand Merit Index (NZMI, http://www.crv4all.co.nz/Library/NZMI.html). Depending on the trait, de-regressed proofs (DRP) of 3,200-3,700 bulls and 1,300 – 2,600 cows were available. Effective daughter contributions (EDCs) were used as weight for the phenotype. Phenotypes of the youngest cohorts of bulls (born from 1-1-2005 onwards), and their daughters were not used to estimate effects. About 150-200 cow phenotypes for each breed were not included in the genomic evaluation because these cows were sired by a validation bull. The bulls from these cohorts were considered a validation bull if they had a genotyped sire with a phenotype, and no genotyped sons with phenotype. Furthermore, animals (cows) with phenotype reliability below a trait-dependent threshold (ranging from 0.25 to 0.50) were not used to estimate effects, because initial analyses indicated a negative effect on reliability when including low reliability phenotypes. The minimum reliability per trait was chosen on the one hand to discard records with a very low reliability and on the other hand to keep enough records to be able to estimate the impact of cow data on the reliability of genomic proofs.

The following model was used for genomic evaluation to obtain genomic breeding values (GBV):

$$y_i = \mu + u_i + \sum_j^n (q_{ij1} + q_{ij2})v_j + e_i$$

where y_i is the deregressed proof 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, v_j is the direction vector of the effects of the haplotypes at locus j, $q_{ij1\,(2)}$ is the size of the effect for the first (second) haplotype ID of animal i at locus j and e_i is the residual for bull i. A Markov Chain Monte Carlo method using Gibbs sampling was used to obtain posterior estimates for all effects in the model. The Gibbs sampler was run for 10,000 iterations with a 2,000 burn-in. Four replicates per trait were run. More details on the method can be found in Meuwissen and Goddard (2004) and Calus et al. (2008).

Pedigree based breeding values (PBV) were estimated with the same data using a comparable model without genomic information:

$$y_i = \mu + u_i + e_i$$

Validation. The genomic prediction of the validation bulls was compared to their daughter-based phenotype, as an approximation of increased reliabilities due to genomic information. Squared correlations (R^2) between DRP and both GBV and PBV were computed and compared to each other to obtain ΔR^2 using the following formula:

$$\Delta R^2 = \frac{R^2_{GBV,DRP} - R^2_{PBV,DRP}}{REL_{DRP}}$$

where $R^2_{GBV,DRP}$ is the squared correlation between GBV and DRP, $R^2_{PBV,DRP}$ is the squared correlation between PBV and DRP, and REL_{DRP} is the reliability of the DRP.

The increase in GBV reliability (measured as ΔR^2) in the reduced HD loci subset was compared to the increase in GBV reliability when using only HD genotyped bulls, and compared to the increase in GBV reliability of the routine genomic evaluation (50k bull genotypes) obtained in earlier validations.

RESULTS AND DISCUSSION

In this paper, presentation of results was limited to a subset of 11 traits out of 26 traits, all with heritability >= 0.15. Nine of the 26 analyzed traits had a heritability below 0.15. Only one (Jerseys) or three traits (Friesians) out of these nine traits benefited from genomics when a single-breed reference population consisting of bulls was used. Therefore, the comparison with results from a multi-breed evaluation with cows included was not made for traits with heritability below 0.15.

The number of reference bulls ranged from 3,276 (udder, Table 1) to 3,548 (protein and milk). The number of reference cows showed more variation, from 1,584 (rump angle) to 2,640 (milk), mainly due to the applied threshold for minimum reliability of the phenotype. The number of bulls used for validation was 345 or 346 for Friesians, 163 for Jerseys and 56-67 for crossbreds. Average increase in R^2 due to genomic information was 9.4% and 10.9% for Friesians and Jerseys, respectively, when an across-breed evaluation was performed using high density genotype data and reference bulls as well as reference cows. With only bulls genotyped with 50k as reference animals in a single breed analysis, increase in R^2 was 8.6% (Friesians) and 7.0% (Jerseys). This indicates that Jerseys, which initially had the smallest reference population, gained most from expanding the reference population. Without cows in the reference population (results not shown), average increase in R^2 due to genomic information was 9.3% (Friesians), 8.5% (Jerseys), and 7.6% (crossbreds). For Jerseys, both reference cows and animals from the other breed seem to contribute to higher reliabilities of genomic breeding values, although not all traits show this result. For crossbreds, increase of R^2 was 1.2% higher when cows were added to the reference population.

For Friesians and Jerseys, the across-breed evaluation with cows added to the reference population gave higher reliabilities for 10 out of 14 analyzed trait-breed combinations, compared to reliabilities when only 50k-genotyped reference bulls were used. Exceptions were protein and milk in Friesians, and milk and udder overall in Jerseys.

The size of the bull reference population ranged from 3,276 to 3,548, whereas the number of cows added to the reference population was lower, ranging from 1,584 to 2,640. Because reliability of cow phenotypes is lower, the amount of information added to the genomic evaluation is relatively low when converted to bull equivalents. Nevertheless, there was benefit from adding a relatively low number of cows to the reference population for most trait-breed combinations presented here. This offers opportunities to further increase reliabilities of genomic breeding values by adding more cows to the reference population.

Table 1. For each trait: number of reference bulls (Nrefb) and cows (Nrefc), and for each breed: increase in R^2 due to genomics for the across-breed evaluation based on BovineHD genotypes and a reference population of bulls and cows (ΔR^2), increase in R^2 in a single-breed evaluation based on 50k genotypes and a reference population of bulls only (ΔR^2s). For crossbreds, number of validation bulls (Nval) is indicated

			Friesian[1)]		Jersey[1)]		Crossbred[2)]	
Trait	Nrefb	Nrefc	ΔR^2 (%)	ΔR^2s (%)	ΔR^2 (%)	ΔR^2s (%)	Nval[1)]	ΔR^2 (%)
Protein	3548	1985	5.4	11.7	7.0	4.7	67	12.2
Milk	3548	2640	11.6	15.4	7.0	20.4	67	9.8
Liveweight	3343	2313	7.8	3.5	16.4	4.9	59	7.7
Somatic Cells	3493	2275	8.8	6.9	18.5	5.1	66	12.5
Capacity	3470	2072	10.3	9.5	8.7	4.0	59	-0.2
Rump angle	3281	1584	12.9	11.2	11.2	1.3	57	11.1
Udder overall	3276	2481	9.3	2.1	7.2	8.4	56	8.5
Average			9.4	8.6	10.9	7.0		8.8

1) Number of validation bulls was 345-346 for Friesians, and 163 for Jerseys
2) No results from single breed evaluation based on 50k genotypes available for Crossbreds

CONCLUSIONS

Averaged across 7 analyzed traits, across-breed genomic evaluation resulted in 0.7% (Friesian) and 1.0% (Jersey) higher reliabilities than the single-breed genomic evaluation based on 50k genotypes. Adding cows to the reference population was beneficial in Jerseys and crossbreds, where reliabilities increased with 2.2% (Jersey) and 1.2% (crossbreds). Considering that a relatively low number of cows was added, higher reliabilities can be obtained by adding more cows.

REFERENCES

Browning S.R. and Browning B.L. (2007) *Am. J. Hum. Genet.* **81**: 1084.
Calus M.P.L., Meuwissen T.H.E., de Roos A.P.W., and Veerkamp R.F. (2008) *Genetics* **178**: 553.
De Roos A.P.W, Hayes B.J, Spelman R.J. and Goddard M.E. (2008) *Genetics* **179**: 1503.
Lund M.S., de Roos A.P.W., de Vries A.G, Druet T., Ducrocq V., Fritz S., Guillaume F., Guldbrandtsen B., Liu Z., Reents R., Schrooten C., Seefried F. and Su G. (2011) *Genet. Sel. Evol.* **43**: 43.
Meuwissen T.H.E. and Goddard M.E. (2004) *Genet. Sel. Evol.* **36**: 261.
VanRaden P. (2010). *Proc. 9th WCGALP* **27**: 8.