Implementation of Genomics in Australia

G.J. Nieuwhof[1,2], K.T. Beard[2], K.V. Konstantinov[1,2], P.J. Bowman[1], B.J. Hayes[1]
[1] Computational Biology Group, Department of Primary Industries Victoria, Bundoora, Australia, [2] Australian Dairy Herd Improvement Scheme, 22 William Street, Melbourne, Australia
Gert.Nieuwhof@dpi.vic.gov.au

Abstract

Australia is in the process of implementing genomic estimated breeding values for Holstein. Direct genetic values (DGVs) are estimated with BLUP, but an animal specific reliability is calculated separately. Blending is used to combine pedigree, herd recording and genomic information (following Harris and Johnson, 2010). The reference set used to derive the prediction equation consists of 2193 bulls. The addition of genomic information to the parent average leads to an increase in reliability of 20 to 31% for the traits investigated. Once a bull has a first set of daughters, inclusion of genomic information does not greatly alter its BV or reliability. The first release of a limited set of traits is planned for September 2010 followed by an official release for all traits and both sexes in April 2011.

Introduction

The Australian Dairy Herd Improvement Scheme (ADHIS) has published breeding values for Australian dairy cattle since the 1980s. Over time the system has evolved to include a range of more than 40 traits relevant to Australian dairy farmers.

With the demonstration that genomic information can improve the reliability of estimated breeding values for young bulls with no progeny test information (e.g. Van Raden *et al.*, 2009, Hayes *et al.*, 2009, Luan *et al.*, 2009), ADHIS has developed procedures to include genomic information in the existing evaluations. Following consultation with industry, only Australian data has been used to estimate marker effects, and individual animal reliabilities are calculated.

The aim of this paper is to describe the inclusion of genomic information in ADHIS evaluations, and to evaluate the reliabilities of genomic breeding values achieved.

Methods

The flow of data and program for evaluation consists of four major steps; genotype QA, imputation, estimation of DGVs and blending.

Quality assurance of genotypes include evaluating call rate and genetrain scores for each marker in a batch, check for lack of variation in the X-chromosome for males, check for duplicates in a batch indicating sampling issues and duplicate genotypes for different animals across batches indicating monozygotic twins or clones (which may cause dependencies in the analysis), Hardy Weinberg equilibrium and genotype inconsistencies given the pedigree.

Imputation of missing genotypes or genotypes failing to meet the minimum genetrain score is either full including all genotyped animals using fastPhase or a quick imputation for animals genotyped after the latest full imputation only. With the current data FastPhase takes about 6 days to run, and therefore cannot be used as part of routine runs.

DGVs are estimated using BLUP as described as RR-BLUP by Moser *et al.* (2009), in this paper it will be referred to as SNP BLUP. It is assumed that SNP effects are random and the DGV for bull *i* called g_i is defined as:

$$g_i = \sum_{k=1}^{p} x_{ik}\beta_k$$

With x_{ik} a vector describing the genotype of bull *i* for *p* SNPs and β_k is a vector with k SNP effects. The SNP effects are found by solving:

$$\beta=(X'X+I\lambda)^{-1}X'y$$

with y the phenotype and X a matrix with vectors x_{ik} for all bulls. In this equation λ is defined as σ_e^2/σ_g^2, with σ_g^2 the genetic variation captured by the SNPs and σ_e^2 the error. In our experience σ_g^2 is difficult to estimate and we decide to estimate λ by cross-validation.

Blending follows Harris and Johnson (2010) with genomic breeding value GEBV=

$$\frac{(1-R_A)(1-R_G)\underline{\hat{a}_N}-(1-R_N)(1-R_G)\underline{\hat{a}_A}+(1-R_N)(1-R_A)\underline{\hat{a}_G}}{1-2R_A-R_NR_G+R_AR_G+R_NR_A}$$

With R_A, R_G and R_N the reliabilities of the standard EBV ($\underline{\hat{a}_N}$), the BV based on genotyped animals only ($\underline{\hat{a}_A}$) and the DGV ($\underline{\hat{a}_G}$) respectively (the last reliability is derived from the standard blup equations with the relationship matrix replaced by the genomic relationship matrix).

In Australia, to date 2332 Holstein bulls, 512 Jersey bulls and 529 Holstein cows have been genotyped for the Illumina 50K SNP chip. Based on the availability of daughter records a reference population for Holstein was created consisting of 2193 bulls (Table 1).

Evaluation of the approach

Data

For the purpose of QA, the Holstein reference population was split in a smaller training population of 1873 with highly reliable BVs for most traits and 320 animals that were progeny tested between 2006 and 2008, and had a parent average but no daughters in 2005. Validation animals were excluded from the analysis for a specific trait if they had less than 50 daughters with fertility or survival records or less than 20 with overall type.

In April 2010 a test genomics evaluation was run based on the March/April 2010 evaluations with marker effects estimated from Australian daughters only, and the EBV from Interbull where available (and Australian otherwise). The aim of the evaluation presented below was to compare the reliabilities of the 2005 Parent Average and the reliability of the GEBV, with the 2010 EBVs based on daughter records. A secondary aim was to look at the effect of adding genomic information to the already highly reliable 2010 EBVs.

This work focused on 6 traits; protein, fat, milk, overall type, survival and fertility. The first 5 traits are analysed based on their Daughter Trait Deviations (DTDs) which are equivalent to DYDs for the yield traits. The survival EBV as routinely published is an index of direct survival, overall type, pin set, likability and udder depth. As there is no DTD for this index, and since not all of its components were included in the current analysis, analysis for survival is for the moment based on the deregressed survival index. Three additional indices (protein%, fat%, ASI) were included in the analysis based on their component traits (protein, fat and milk). The Australian Selection Index (ASI) represents the net-value in Aus$ of milk, fat and protein. Yield traits are assumed to have the same heritability and reliability. Non-production traits in Australia are expressed relative to the phenotypic mean of 100, so that each unit deviation from 100 equals a 1% of deviation from the phenotypic mean.

Results

Table 2 shows the mean parent average without and with genomics for the validation bulls for the 6 traits and 3 indices, as well as the difference between the two analyses. It is clear that the reliability increases dramatically. For production traits it doubles from 28 to 56%, and other traits also show increases. Addition of genomics to the PA leads to a decrease in average BV for all traits, except survival which is considerably higher when genomics are included.

The same 320 bulls are represented in Table 3, but now their daughters are included. It is clear that there is limited benefit in terms of reliability from adding genotypes for yield

traits after daughters are included. For the fertility and survival, for which the reliability based on daughters was lower, there is still a considerable increase in reliability. Also note that for survival the standard deviation is larger for the GEBV than the EBV.

The analysis in Table 4 shows that for yield traits and in terms of average breeding value the combination of PA plus genomics has a smaller overestimation than the PA on its own. For the non-yield traits the picture is less clear. Since the reliability of non-yield traits is still relatively low, it is not evident whether the apparent bias in prediction of the EBV with daughters based on PA plus genomics is due to the genomics or indeed to the lack of precision as indicated by the reliability of the EBV with daughters.

A similar picture emerges when looking at the correlation of the EBV based on daughters with PA on one hand and PA plus genomics on the other. For yield traits this correlation is 55 to 58% (and much higher for fat and protein%), but much lower for the non-production traits, especially fertility and survival.

There is very little difference in EBV following progeny testing with or without genomics. The general trend is for the PA plus genomics to be closer to the BV with daughters, but there are exceptions.

Discussion

The genomic analysis of Australian Holstein gives an increase of 20 to 31% in reliability compared to the parent average for the six traits investigated. This is similar to results found in other countries (de Roos *et al*., 2009; Reinhardt *et al.*, 2009; Van Doormaal *et al.*, 2009; AIPL, 2010). These genomic breeding values also remove some of the bias in the parent average. Once daughter records are available, the addition of genomics does have a very small effect on reliability and average BV for the highly reliable yield traits, while there is still a useful increase for fertility and survival.

Unlike many other countries, Australia has chosen to use Daughter Trait Deviations (DTDs) rather than deregressed breeding values to estimate SNP effects. The reason for this is that they can be considered to be a more independent and accurate measure of phenotypic performance of a bull's daughters. DTDs include both genetic and residual effects and the latter may include major gene effects, which would not be fully included in a deregressed BV. It is anticipated that inclusion of cows in the reference population will be through their Trait Deviation. In practical terms, deregressed values of lowly reliable EBVs of cows would contain large prediction errors. A downside of the use of DTDs is that it can only be applied to Australian daughters, and, if that were desired, not use additional foreign daughters. The use of DTDs to estimate SNP effects and hence DGVs from Australian daughters, but including the Interbull MACE EBV in the GEBV reduces the correlation between DGV and GEBV for bulls in the reference population.

The current estimation procedure includes two programs to estimate the DGV, first SNPBLUP, then as part of the blending procedure (GBLUP). The GEBV is based on the DGV estimate from the first and the reliability from the latter. An important feature of GBLUP is its ability to estimate a DGV reliability specific for each genotyped animal. The reason for still using SNPBLUP is that it will give a great deal of future flexibility, for example if the number of SNPs becomes extremely large, or if a cow chip is produced with fewer, trait specific, SNPs, and DGV must be calculated from this chip. In both these cases SNPBLUP could be used with little or no modification to calculate DGV.

The blending of high reliable bulls, especially those in the reference population proved difficult. These animals would have EBVs and DGVs with reliabilities of over 90%, with the estimates for EBV and DGV at times quite different. Using a standard procedure, the GEBV deviated more from the EBV than would seem acceptable to industry. This may be because of a bias in the EBV, or a problem with the blending. In order to account for this, the reliability of the DGV and uhat were regressed back depending on the

reliability of the EBV. This resulted in little effect of genomics on EBVs with a reliability around 85%. It is proposed, at least in the short term, to publish EBVs rather than GEBVs for the highly reliable bulls.

ADHIS plans its first unofficial public release of genomic breeding values for September 2010. This will only include Holstein progeny test bulls. Further development work should enable the release of genomic breeding values for all traits and for Holstein cows as well as bulls in December 2010 with the first full official release in April 2011.

Further work focuses on the inclusion of cows in the reference population and expansion of the Jersey reference population to a sufficient size.

ADHIS is actively seeking the exchange of genotypes with other countries, with focus on those bulls that have significant numbers of well-recorded daughters in Australia.

Conclusions

In Australia Holstein bulls, the addition of genomics to the parent average leads to an increase in reliability of 20 to 31% for the traits investigated. Once a bull has a first set of daughters, inclusion of genomics does not greatly alter its BV or reliability.

The Australian genomic evaluation system is open for all players in the Australian dairy industry, including breeding companies and individual farmers. The first unofficial GEBVs will be released in September 2010, followed by an official release including all traits and both sexes in April 2011 for Holstein.

Acknowledgments

This project was jointly funded by the Australian Dairy Herd Improvement Scheme (ADHIS) and DPI Victoria.

References

AIPL 2010. Comparison of April 2010 genomic and traditional evaluations. http://aipl.arsusda.gov/eval/summary/comparexml_menu.cfm?R_menu=v_1004.v_Young_Bulls.v_Holstein_wddx#StartBody.

de Roos, A.P.W., Schrooten, C., Mullaart, E., van der Beek, S., de Jong, G. & Voskamp, W. 2009. Genomic selection at CRV. *Interbulletin 39*, 47-50.

Harris, B.L. & Johnson, D.L. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. *J. Dairy Sci. 93*, 1243–1252.

Hayes, B.J., Bowman, P.J. Chamberlain, A.J. & Goddard, M.E. 2009. Invited Review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci. 92*, 433-443.

Hayes, B.J., Raadsma, H.W., Tier, B., Khatkar, M.S., Crump, R.E., Moser, G., Bowman, P.L., Chamberlain, A.J., Pryce, J. & Haile-Mariam, M. 2010. *Genomic Selection for Core ADHIS traits. CRC 13039 Final Report.*

Luan, T., Woolliams, J.A., Lien, S., Kent, M., Svendsen, M. & Meuwissen, T.H.E. 2009. The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation. *Genetics 183*, 1119–1126.

Moser, G., Tier, B., Crump, R.E., Khatkar, M.S. & Raadsma, H.W. 2009. A comparison of five methods to predict genomic breeding values of dairy bulls from genome-wide SNP markers. *Gen. Sel. Evol. 41*, 56-71.

Reinhardt, F., Liu, Z., Seefried, F. & Thaller, G. 2009. Implementation of genomic evaluation in German Holsteins. *Interbulletin 40*, 219-226.

Van Doormaal, B.J., Kistemaker, G.J., Sullivan, P.G., Sargolzaei, M. & Schenkel, F.S. 2009. Canadian implementation of genomic evaluations. *Interbulletin 40*, 214-218.

Van Raden, P.M., Van Tassel, C.P., Wiggan, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. & Schenkel, F.S. 2009. Invited Review: Reliability of genomic predictions for North American Holstein bulls. *J. Dairy Sci. 92*, 16-24.

Table 1. Average number of daughters with records in Australia and average reliability for Holstein reference population (number of bulls), training set (number of bulls) and validation set, as well as number of bulls in validation set, depending on trait.

	Reference (2193)		Training (1873)		Validation (2010 values)		
	Daughters	Reliability	Daughters	Reliability	Bulls	Daughters	Reliability
Production	612	90%	703	91%	320	74	85%
Overall Type	95	60%	109	62%	83	39	74%
Fertility	399	59%	458	60%	194	66	54%
Survival	535	67%	616	70%	185	74	54%

Table 2. Predicted average BV and reliability (standard deviation) for validation bulls based on 2005 parent average or 2005 parent average plus genomics.

Trait	PA	Rel	GEBV	Rel	GEBV-PA	Rel-rel
Milk	494(307)	28(2)%	443(281)	56(6)%	-51(240)	28(6)%
Fat	19.3(11.9)	28(2)%	14.0(9.6)	56(6)%	-5.7(8.7)	28(6)%
Prot	16.8(7.7)	28(2)%	13.4(6.5)	56(6)%	-3.4(6.2)	28(6)%
Fat %	-0.02		-0.07		-0.05	
Prot %	0.07		0.02		-0.05	
ASI	106		81		-25	
OType	102.8(2.9)	21(3)%	100.3(3.3)	42(1)%	-2.5(2.6)	21(8)%
Fert	99.8(2.1)	18(6)%	99.0(2.2)	38(10)%	-0.7(2.3)	20(8)%
Surv	97.7(1.3)	20(4)%	104.8(7.7)	50(7)%	7.1(7.9)	31(7)%

Table 3. Predicted average BV and reliability (standard deviation) for validation bulls based on 2010 breeding value (including daughters) without and with genomics.

Trait	EBV	Rel	GEBV	Rel	GEBV(g)-EBV	Rel-rel
Milk	277(406)	85(4)%	276(396)	85(4)%	-1(35)	+0.3(0.4)%
Fat	8.0(16.3)	85(4)%	8.5(15.4)	85(4)%	0.5(1.5)	+0.3(0.4)%
Prot	10.4(9.1)	85(4)%	10.2(8.8)	85(4)%	-0.2(0.9)	+0.3(0.4)%
Fat %	-0.06		-0.05		0.01	
Prot %	0.06		0.05		-0.01	
ASI	62		61		-1	
OType	101.8(6.2)	74(5)%	101.9(4.6)	75(5)%	0.1(2.2)	+0.5(0.5)%
Fert	100.3(2.8)	54(5)%	101.7(1.9)	58(4)%	1.4(1.7)	+4.2(1.7)%
Surv	101.2(2.1)	54(7)%	103.3(3.9)	61(4)%	2.0(3.7)	+6.3(3.9)%

Table 4. Difference between mean (standard deviation) PA with genomics and EBV based on daughters.

Trait	EBV - PA	Rel	EBV - GEBV	Rel
Milk	-217(414)	57(5)%	-166(335)	29(6)%
Fat	-11.7(15.7)	57(5)%	-6.0(13.5)	29(6)%
Prot	-6.3(9.9)	57(5)%	-3.0(7.8)	29(6)%
ASI	-44		-19	
Fat %	-0.04		0.01	
Prot %	-0.01		0.04	
OType	-1.0(5.5)	53(6)%	1.5(5.5)	32(9)%
Fert	0.5(2.8)	36(6)%	1.3(3.6)	16(9)%
Surv	3.5(1.9)	34(7)%	-3.6(7.8)	4(9)%