Application of genomic information in a dairy cattle breeding scheme

R.J. SPELMAN*, J. ARIAS, M. KEEHAN, V. OBOLONKIN, A. WINKELMAN,
D. JOHNSON and B.L. HARRIS.

Livestock Improvement Corporation, Private Bag 3016, Hamilton 3240, New Zealand
*Corresponding author: rspelman@lic.co.nz

ABSTRACT

Marker assisted selection has been applied in dairy cattle breeding schemes over the last decade with minor to moderate improvements in genetic gain. In most cases, the cost effectiveness of marker assisted selection has been neutral, at best. The completion of the sequencing of the bovine genome in 2006 has generated a large number of single nucleotide polymorphisms that has allowed the application of genomic selection. In addition, the cost of genotyping single nucleotide polymorphisms is considerably lower than microsatellite genotyping. Livestock Improvement Corporation has undertaken extensive genotyping of the sires that have been progeny tested in the last 20 to 30 years. Analysis of these data by Livestock Improvement Corporation researchers has identified combinations of markers that were used to obtain genetic evaluation of bulls from their DNA. By utilising this genomic information, semen from two-year-old sires has been sold to New Zealand dairy farmers. Over the past two years, the genomically selected bulls have accounted for approximately 15 to 20% of the inseminations. It is expected that selection based on genomic information, termed genomic selection, will increase the rate of genetic gain in New Zealand by 30 to 50%.

Keywords: genomics; dairy cattle; breeding scheme.

INTRODUCTION

Livestock Improvement Corporation (LIC) has been investing in DNA technology since the early 1990s. The first application of DNA information in the LIC breeding scheme was for parentage testing in the mid 1990s. At the same time, the detection of quantitative trait loci (QTL) for dairy cattle had just commenced (Georges *et al*., 1995). Numerous QTL have been identified for dairy cattle traits but there are only three instances where the underlying causative mutation for the QTL has been identified; chromosome 14 – DGAT1 (Grisart *et al*., 2002), chromosome 20 – GHR (Blott *et al*., 2003) and chromosome 6 (Ron *et al*., 2006). LIC researchers were involved in the chromosome 14 and 20 work and these genes have been utilised within the LIC breeding scheme.

Utilisation of QTL in dairy cattle breeding schemes via marker-assisted selection started at LIC in 1998 using six QTL that had been identified for protein, fat and milk volume. The MAS breeding scheme relied on the generation of multiple full-sib sons from given sire and dam combinations. The resulting male offspring were genotyped and the sons that received the desirable alleles for the QTL were selected to enter progeny testing. The reproductive performance of the donor cows was poor resulting in very few of the families having enough sons to allow within-family marker-assisted selection. After two years of poor reproductive performance the within-family MAS was abandoned. LIC recommended MAS in 2003 after the identification of both DGAT1 and GHR genes.

All of the bull dams and bulls entering the progeny testing scheme were genotyped for these two genes for the next four years.

Meuwissen *et al*. (2001) first proposed the use of dense marker maps in a genomic selection setting. Through simulation they found that the reliabilities of the estimated breeding values from genomic information were 0.75 to 0.85. At the time the paper was published, the bovine genomics community was restricted to using approximately 4,000 microsatellite markers. Genotyping 2,000 sires with the 4,000 microsatellites would have cost approximately NZ$20 million.

In 2006, the sequencing of the bovine genome by the international consortium was completed (Kappes *et al*., 2006). The sequencing generated in excess of a million single nucleotide polymorphisms (SNPs) compared to the thousands of the previous type of microsatellites. In addition to the greater number of SNPs, the cost of genotyping decreased from NZ$2.50 per genotype for a microsatellite to less than NZ$0.01 per marker when tens of thousands of SNP are typed in parallel. The technology shift to large-scale SNP genotyping has had, and will continue to have, a major effect on the utilisation of markers in dairy cattle breeding schemes. This paper describes the application of genomic selection in the LIC breeding scheme.

MATERIALS AND METHODS

Genotyping

Approximately eighteen hundred Holstein-Friesian sires, 1,200 Jerseys sires and 300

TABLE 1: Estimating DNA breeding values (BV) for two animals based on the additive effects for three SNP loci.

(a)

Parameter	Marker 1 effect[1]	Marker 2 effect[1]	Marker 3 effect[1]
Additive effect	-1	0.5	0.1

[1]The estimated effects are half the difference between the 22 genotypic class and the 11 genotypic class, with the 12 genotypic class centered on zero.

(b)

Animal	Marker 1 genotype	Marker 2 genotype	Marker 3 genotype	DNA BV
A	11	12	22	1.1
B	22	11	22	-1.4

KiwiCross[TM] sires were genotyped over the Illumina 50K SNP panel (Illumina Inc. San Deigo, USA). The progeny tested sires ranged in birth year from the early 1980s to the mid 2000s. KiwiCross[TM] is the name given to crossbred animals that are less than 87.5% of one of the main dairy breeds.

Statistical analysis

The dataset was split into two, a research dataset and a validation dataset. The Research dataset contained all bulls that were progeny tested prior to 2002 and totaled 2,450 sires. The Validation dataset consisted of sires born after 2002 that were progeny tested and consisted of 850 sires. The Research dataset was split into three components, all 2,450 animals, 1,400 Holstein-Friesian animals and 1,000 Jersey animals. These datasets were analysed to identify which SNPs affected the phenotype of interest. Estimated breeding values were used as the

TABLE 2: Correlations between DNA-based and progeny test BVs in the validation population for the Holstein-Friesian, Jersey and KiwiCross[TM] breeds.

Trait	Holstein-Friesian[1]	Jersey[2]	KiwiCross[3]
Protein yield	0.65	0.50	0.72
Milk fat yield	0.53	0.54	0.59
Milk volume	0.67	0.62	0.76
Live weight	0.60	0.57	0.75
Fertility	0.68	0.49	0.52
Somatic cell score	0.57	0.58	0.57
Total longevity	0.65	0.49	0.48
Shed temperament	0.50	0.46	0.48
Farmer opinion	0.46	0.38	0.37
Udder overall	0.47	0.50	0.49

[1]SNP effects estimated from the Holstein-Friesian Research data set.
[2]SNP effects estimated from the Jersey Research data set.
[3]SNP effects estimated from all of the animals in the Research data set.

phenotype for 25 different traits. Results presented in this paper are for a subset of these phenotypes. The three components of the Research dataset were analysed using the methodology termed BLUP by Meuwissen *et al.* (2001). In brief, each of the marker loci were simultaneously fitted as random effects with equal genetic variance. This means it was assumed that each locus contributed equally to the genetic component of the phenotype.

Once the genetic effects for each of the SNPs had been estimated from the research dataset, the accuracy of the estimates was tested in the validation dataset. DNA breeding values were estimated for each animal, given its DNA profile, in the validation dataset by summing the appropriate estimated SNP effects from the research dataset. Table 1 demonstrates this for an example where two validation animals have DNA breeding values (BV) calculated from three SNP markers that have had effects estimated from the Research dataset. In this example, each of the SNPs have three genotypic classes (11,12, 22) and the estimated effect is half the difference between the 22 genotypic class and the 11 genotypic class. The 12 genotypic class is centered on zero for each of the loci. Animal A has a marker profile of 11 for Marker one and therefore has a marker effect of one, whereas Animal B has the marker profile of 22 and therefore has a marker effect of –1. To calculate the DNA BV for Animal A the sum of the appropriate effects is taken as (1 + 0 + 0.1 = 1.1).

The validation of the SNP effect estimates was undertaken by comparing the DNA BVs with the progeny test BVs estimated for the relevant sires. The degree of accuracy of the DNA-based BVs was measured by their correlation with the progeny test BVs. Three different DNA BVs were estimated for each validation animal utilising the SNP effects from the three individual research datasets.

RESULTS

Accuracy of genomic estimates

Correlations between the DNA BV and the progeny test BV varied from 0.37 to 0.76 for the production and non-production traits in the validation populations (Table 2). In general, the correlations for the production traits were higher than those for the non-production traits. The correlations for the KiwiCross[TM] population were on average higher than those for the other two breeds for the traits used in the calculation of Breeding Worth where the selection index included the traits of milk fat, milk protein, milk volume, live weight, somatic cell count, fertility and longevity. One exception to this was fertility in

the Holstein-Friesian breed where it appears that the SNP markers identified strain differences between New Zealand and North American genetotypes. The correlations dropped dramatically when the SNP effects were estimated from the other breed in the research dataset to the breed that they were validated in. For example, when the SNP effects were estimated in the Jersey research dataset the correlations between the DNA BVs and the progeny test BVs for the Holstein-Friesian bulls in the validation dataset were between –0.05 and 0.16. This was also observed for the situation where Holstein-Friesian SNP effects were used in the Jersey validation dataset. Therefore, the SNP effects estimated are breed specific. However, when the combined breed research dataset was used and then validated in the individual populations the correlations were nearly identical to those presented in Table 2 for the Holstein-Friesian and Jersey breeds.

Breeding value estimation

Up until mid 2009, three sources of information were used in the estimation of breeding values for New Zealand dairy sires; parental information, own performance and progeny performance. DNA is now the fourth source of information. A dairy bull of one year of age has only information from his parents for dairy-related traits. For a sire with DNA information this is combined with parental information to estimate a genomic breeding value for the bull. Approximately 50 to 60% of the genomic breeding value is parental information with the balance being DNA information. The reliabilities for the genomic breeding values for the bulls with parental and DNA information are approximately 50 to 55%. This is in contrast to a reliability of 75% to 85% for a progeny tested bull and 30 to 35% for a bull with only parental information.

DISCUSSION

The correlations presented in this study are less than those generated through stochastic simulation by Meuwissen *et al*. (2001). The initial parameters assumed by Meuwissen *et al*. (2001) regarding the underlying distribution of genetics effects assumed that there were more loci with larger effects rather than the large number of loci with small effects that have been found in real data. The correlations in our study are in concordance with the theoretical expectations that have been modeled recently by Goddard and Hayes (2009). The Goddard and Hayes (2009) paper clearly demonstrates that in order to increase the correlations to 0.75 and above, between 10,000 and 20,000 animals need to be genotyped depending on the heritability of the trait.

Genomically evaluated bulls were first commercially released in New Zealand in 2008 with the creation of genomically selected bull teams. The genomically selected teams are marketed as "DNA Proven" whereas the teams with sires that are progeny tested are marketed as "Daughter Proven". The DNA proven teams included bulls that were two or three years of age and thus did not have lactating daughters. In the 2008 mating season, the genetic superiority of the DNA proven teams over the daughter proven teams were approximately 10, 20 and 30 BW points for the Holstein-Friesian, Jersey and KiwiCross teams respectively. Ten BW is approximately equivalent to the genetic gain achieved per annum in the dairy industry. During the 2008 mating season, the Holstein-Friesian DNA Proven team was withdrawn from the market. This was driven by further information from the 2005 sire proving scheme where the DNA proof for the Holstein-Friesian team did not predict the daughter proof as well as had been seen in the two previous sire proving scheme years. Given that the Holstein-Friesian DNA Proven team had an advantage of only approximately 10 BW it appeared that there was a significant probability that the DNA Proven team could have a lower BW than the Daughter Proven team. Over the 2008 and 2009 mating seasons the three DNA Proven Holstein Friesian, Jersey, KiwiCross[TM] teams, undertook approximately 15 to 20% of the total inseminations.

The BW reliability of the DNA Proven sires is approximately 55%. For a sire that has an estimated BW of 250 with a reliability of 55%, the 95% confidence interval is between 160 and 340. For a Daughter Proven sire with a 85% reliability the 95% confidence interval is between 200 and 300. Utilising the bulls in teams of 15 to 18 reduces the effect of the lower reliability of the DNA Proven sires. Team reliabilities of 97% are achieved. This is reflected in that the 95% confidence interval is ± 23 BW points of that estimated from the DNA proofs. In contrast for the Daughter Proven team the 95% confidence interval of the mean of the team will be ± 13 BW points from that estimated from progeny testing. Therefore farmers using teams of DNA Proven bulls can expect to see greater volatility in the proofs of the DNA team compared to the Daughter Proven team.

The LIC breeding scheme, which has been based on progeny testing 300 bulls per annum, has been altered to utilise the incorporation of DNA information. In 2009, LIC screened over 1,000 bulls based on their DNA profile and then selected the best 160. These bulls enter progeny test as one-year-olds and then their semen will be sold commercially as a two-year-old. The bulls will be retained until they receive their progeny test evaluation and semen will be sold to farmers who prefer to purchase genetics based on progeny test information. With the shorter generation interval due to bulls being

used as two-year-olds instead of five-year-olds, it is estimated that the rate of genetic gain in the New Zealand population could increase by 30 to 50% (A. Winkelman, Unpublished data).

The current DNA BVs were estimated using the data from 50,000 SNP markers. High density marker panels with 500,000 to 800,000 markers will be available mid 2010. We expect that the higher density of markers will enable the SNP effects to be effective across breeds rather than the current breed specific effects that are estimated from the 50K marker panel. This is due to the significant markers being closer to the functional mutations compared to those from the lower density panels. Therefore the linkage disequilibrium between marker and functional mutation should hold across breeds. In addition to the technology change with genotyping, a major shift is occurring with sequencing. When the bovine genome was sequenced in 2006-08, it cost approximately $US 50 million. If repeated this year, it is expected that the sequencing could be completed for less than $US 100,000 and by 2012 there are indications that a genome could be sequenced for $US 1,000. Given the current trends, the expectation is that sequencing will become part of the selection process for dairy breeding schemes in the next 10 years.

The utility of DNA in dairy cattle breeding schemes has now reached the level of accuracy that enables dramatic changes and improvements to breeding schemes. With denser marker maps becoming available in the coming years, more sophisticated statistical tools, and in the longer term, with sequencing, it is expected that the level of accuracy from genomic information will continue to improve.

REFERENCES

Blott S.; Kim, J.J.; Moisio, S.; Schmidt-Kuntzel, A.; Cornet, A.; Berzi, P.; Cambisano, N.; Ford, C.; Grisart, B.; Johnson, D.; Karim, L.; Simon, P.; Snell, R.; Spelman, R.; Wong, J.; Vilkki, J.; Georges, M.; Farnir F.; Coppieters, W. 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.

Georges, M.; Nielsen, D.; Mackinnon, M.; Mishra, A.; Okimoto, R.; Pasquino, A.T.; Sargeant, L.S.; Sorenson, A.; Steele, M.R.; Zhao, X.; Womack, J.E.; Hoeschele, I. 1995: Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* **139**: 907-920.

Goddard, M.E.; Hayes, B.J. 2009: Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**: 381-391.

Grisart, B.; Coppieters, W.; Farnir, F.; Karim, L.; Ford, C.; Cambisano, N.; Mni, M.; Reid, S.; Spelman, R.; Georges, M.; Snell, R. 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 Research* **12(2)**: 222-231.

Kappes, S.M.; Green, R.D.; Van Tassell, C.P. 2006. Sequencing the bovine genome and developing a haplotype map: approaches and opportunities. *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production. Belo Horizonte, Minas Gerais, Brazil.* **22**: 937.

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.

Ron, M.; Cohen-Zinder, M.; Peter, C.; Weller, J.I.; Erhardt, G. 2006. Short communication: a polymorphism in ABCG2 in *Bos indicus* and *Bos taurus* cattle breeds. *Journal of Dairy Science* **89(12)**: 4921-4923.