Proc. Assoc. Advmt. Anim. Breed. Genet. 18:147-150

GENETIC MARKERS FOR LACTATION PERSISTENCY IN AUSTRALIAN DAIRY COWS

J.E. Pryce, M. Haile-Mariam, K. Verbyla, P.J. Bowman, M.E. Goddard and B.J. Hayes

Department of Primary Industries, 1 Park Drive, Bundoora, Victoria 3083

SUMMARY

A genome wide association study was used to detect quantitative trait loci (QTLs) for lactation persistency in dairy cattle. Persistency was defined as the slope after peak production calculated using test-day solutions for milk yield in Holsteins and Jerseys. As milk yield is correlated to persistency (r= 0.35), persistency was also adjusted for milk yield. Two strategies were used to search for QTLs: a SNP-by-SNP analysis where persistency solutions for sires (adjusted for fixed effects) were regressed on each single nucleotide polymorphism (SNP) in turn and a genomic selection method (BayesA) where all SNPs are fitted simultaneously. In each analysis, the discovery population comprised 743 Holstein bulls proven before 2005 and the validation datasets were 357 Holstein bulls proven after 2005 and 294 Jersey sires. A genomic region located between 21,408 kbp and 23,744 kbp on chromosome 6 had four SNPs that validated in both Holsteins and Jerseys and may indicate the presence of a QTL for persistency. The largest SNP effect from BayesA was in a similar genomic region to a SNP that validated in the single SNP analysis. False discovery rates were higher for persistency (>65%) than milk volume (24%). We hypothesise that there are many mutations that have a small effect on persistency. Genomic selection using a large number of markers appears to be a promising strategy to improve persistency. For sires not included in the prediction of SNP effects, pedigree information alone had a correlation of 0.16 with persistency EBVs, while combining genomic information with pedigree information increased the correlation with persistency EBV to 0.4.

INTRODUCTION

Persistency of lactation in dairy production is defined as the rate of decline in production after peak milk production has been reached (Cole and VanRaden 2006). Persistency may be a trait of economic importance because it can be used to select for extended lactations, which has a beneficial impact on food costs, health and fertility (Dekkers *et al.* 1998).

Dairy cattle are now routinely being genotyped for many thousands of single nucleotide polymorphisms (SNPs). One way in which SNPs are being used is through genomic selection (Meuwissen *et al.* 2001), where the effect on a trait of chromosome segments, defined by SNPs, are estimated simultaneously and used to predict breeding values. An alternative way of using SNPs is to apply a genome wide association study where individual associations between SNPs and a deregressed breeding value or daughter-yield-deviation of a trait of interest are sought. SNPs with strong relationships to traits of interest can be useful as part of a SNP panel to select animals or to improve our understanding of the biological control of a trait.

The aim of this study was to estimate individual SNP effects for persistency in dairy cattle in a reference dataset of Holstein bulls proven before 2005 and validate these SNPs in: 1) Holstein bulls proven after 2005 and 2) Jersey bulls. Secondly, we used a method of calculating a prediction equation for persistency using SNPs called BayesA (Meuwissen *et al.* 2001).

MATERIALS AND METHODS

Data. Genetic markers were obtained for Jersey and Holstein sires from Ilumina using the BovineSNP50 BeadChip (Illumina, San Diego, CA). The SNP data were edited to ensure that the

call rate was greater than 90% and the minor allele frequency was greater than 2%. SNPs that could not be mapped, or that were on the X chromosome were excluded. After editing, the number of SNPs available for analysis in each dataset was 39,048 out of the original 51,386 SNPs. Each SNP was biallelic (e.g. A and G alleles) and recoded to 1 or 2 according to the allele present at each locus. Missing genotypes were imputed using fastPHASE (Scheet and Stephens 2006).

First parity records on milk volume were extracted from the ADHIS database on cows that calved between 1999 and 2007. Only herds with daughters of bulls that had genotype information were kept. The data was edited to include only herd-test-days with a minimum of ten cows and sires with at least 30 daughters. The final datasets included records on 797,025 first lactation daughters of 3,459 sires in Holsteins and 68,230 first lactation daughters of 1,196 sires in Jerseys.

Statistical methods. Sire solutions for 3rd degree Legendre polynomials fitted to first lactation test-day records of milk yield were estimated using a random regression BLUP model in ASReml within breed (Gilmour *et al.* 2006). The (co)variance components and model used in the BLUP analysis were obtained from the study of persistency reported by Haile-Mariam and Goddard (2008). A pedigree was not included in the model because this would increase the probability of SNPs being selected on the basis of relationships between animals, rather than SNPs that are related to persistency. The solutions obtained were equivalent to daughter-yield-deviations.

Persistency (PERS) was calculated as $S_{i,54}$–$S_{i,274}$ where $S_{i,d}$ is sire solution of sire *i* on day *d* of lactation and is the gradient from after peak lactation. A 300 day milk solution (VOL) was calculated as the sum of daily yields. Persistency adjusted for milk yield (PERSadj) was calculated by regressing PERS on VOL. EBVs for PERS and PERSadj were calculated by fitting pedigree.

The Holstein data was split into two according to age of sire, as prediction of genetic merit in younger animals using associations found in older animals is generally more useful for selection. The reference dataset included solutions for Holstein bulls that received their first proof prior to 2005 (n = 743). Holsteins that were first proven after 2005 (until 2007) formed the first validation dataset (n = 357). The second validation dataset comprised 294 Jersey sires.

PERS, PERSadj and VOL were regressed on individual recoded SNPs using ASReml (Gilmour et al., 2006) in the reference and validation datasets separately. The model included the SNP as a fixed effect and sire as a random effect, a pedigree was also included. A subset of SNPs were selected where the F-probability was P<0.005 in the reference dataset and P<0.05 in the validation dataset and the direction of the SNP estimate was the same in both datasets. The false discovery rate (FDR) was calculated as FDR=E(R*P/S, where R is the number of tests, P is the p-value used to in the F-test and S is the number of SNPs with significant F values; e.g. the expectation of the number of false discoveries by chance divided by the actual SNPs significant at this threshold.

The SNPs were also fitted simultaneously in a BayesA model, similar to the model described by Meuwissen et al. (2001), modified to include a polygenic effect. The SNP effects were estimated in the reference population and molecular breeding values (MBV) calculated for the Holstein validation sires by summing the SNP effects multiplied by the allele frequency at each SNP position. The correlation between the persistency EBVs estimated from the data plus pedigree and the corresponding MBV estimated using genomic information in the Holstein validation dataset was calculated. The BayesA SNP solutions were also used to verify SNPs detected using the single SNP approach.

RESULTS AND DISCUSSION

The number of significant SNPs (P<0.005) for persistency in the reference data set were 302 and 262 for PERS and PERSadj respectively, which is a false discovery rate (FDR) of 65% and 75% (Table 1). A low FDR indicates that more SNPs were statistically significant than expected by chance. Of the SNPs that were P<0.005 in the reference dataset, 20 and 22 SNPs were at

P<0.05 in the Holstein validation population and 12 and 9 were in the same direction as the reference population for PERS and PERSadj respectively. Seven and three SNPs validated in Jerseys. FDRs in the Holstein validation dataset were 66% to 69% for PERS and PERSadj and 101% to 164% in Jerseys for PERS and PERSadj.

Table 1. **Number of SNPs found to be significant, in the same direction in Holstein reference and Holstein and Jersey validation datasets and percentage of false discovery rates (FDRs), when SNPs were regressed individually on persistency (PERS), persistency adjusted for milk volume (PERSadj) and 300d milk volume (VOL)**

	PERS	PERSadj	VOL
Holstein reference P<0.005	302	262	738
Holstein validation P<0.05	22	20	152
Holstein same direction	12	9	144
Direction validated	55%	45%	95%
Jersey P<0.05	15	8	77
Jersey same direction	7	3	37
Direction validated	47%	38%	48%
FDR reference Holstein	65%	75%	26%
FDR validation Holstein	69%	66%	24%
FDR validation Jersey	101%	164%	48%

Table 2. SNPs validated at P<0.005 in the reference population and P<0.05 in the validation dataset in Holsteins (H) and Jerseys (J) persistency adjusted (PERSadj), only SNPs that validate in the same direction are shown for Holsteins

Breed	Direction*	SNP name	Chromosome	Position (bp)	F-Prob*
H	-	BTB-01600593	2	16,000,786	0.017
H	-	ARS-BFGL-NGS-112143	4	10,139,426	0.038
H	-	ARS-BFGL-NGS-27962	6	21,593,191	0.017
H	-	UA-IFASA-1756	6	21,620,640	0.023
H	-	BTA-82896-no-rs	8	11,649,044	0.041
H	-	Hapmap59058-rs29016195	12	8,844,600	0.026
H	-	BTA-42074-no-rs	17	11,271,481	0.004
H	+	ARS-BFGL-NGS-38620	18	64,466,895	0.045
H	+	BTB-00920286	26	3,410,026	0.013
J	-	BTB-00780124	1	144,105,011	0.042
J	+	BTA-47105-no-rs	5	113,682,010	0.043
J	-	ARS-BFGL-NGS-60840	6	13,520,548	0.025
J	+	BTB-00245990	6	21,408,490	0.003
J	-	BTB-00245990	6	23,744,743	0.012
J	-	ARS-BFGL-BAC-35623	6	70,432,390	0.022
J	+	ARS-BFGL-NGS-60840	13	37,444,034	0.041
J	+	BTB-00245990	19	33,644,562	0.007

*In validation population

The correlation between PERS and VOL was 0.35 in the Holstein combined reference and validation datasets. Therefore, PERSadj may be better measures of persistency as it is independent of yield. However, the correlation between PERS and PERSadj was 0.94, so they are similar traits.

Although no single SNP validated for PERSadj in both Holsteins and Jerseys (Table 2) there was a genomic region between 21,408 kbp and 23,744 kbp on chromosome 6 that had SNPs validating in both breeds. This is not in the same region as mutations previously found to affect milk production, for example ABCG2 which is located at approximately 37 Mbp (Cohen Zinder *et al.* 2005). The largest effect from the Bayes A analysis (located on chromosome 2 at 15,658 kbp) was close to a SNP that validated in Holsteins (Table 2). There were two SNPs that validated in both Jerseys and Holsteins for milk volume (VOL), both on chromosome 14 and close to the mutation known to exist on chromosome 14 that affects milk yield (DGAT1; Grisart *et al.* 2002).

The correlation between the equivalent PERS EBV and MBV in the Holstein validation dataset (i.e. sires not included in the prediction of SNP effects) using BayesA was 0.40 for PERS and 0.28 for PERSadj, the equivalent correlations between the EBV and parent average EBV was 0.16 and 0.25 for PERS and PERSadj respectively. Thus, genomic selection may be a suitable way to improve persistency.

In the BayesA method all SNPs were fitted simultaneously, while each SNP was fitted in turn in the SNP by SNP analysis. The SNP by SNP analysis is therefore much more likely to select close neighbours, while the BayesA analysis will lead to the selection of the SNP with the largest effect in a groups of neighbours and is therefore more likely to be close to the causative mutation in cases where a number of candidate SNPs have been discovered in the SNP by SNP analysis.

CONCLUSION

Having two validation populations, one being the same breed as the discovery dataset and the other an alternative breed, provides a powerful way to validate SNPs discovered using GWAS and minimise the risk of pursuing incorrect genomic regions when false discovery rates are high. This is especially important for low heritability traits such as persistency. However our results suggest that there are many mutations of small effect on persistency. Genomic selection using a large number of markers appears to be a promising alternative.

ACKNOWLEDGEMENT

We would like to thank the Gardiner Foundation for funding this research.

REFERENCES

Cohen-Zinder, M., Seroussi, E., Larkin, D.M., Loor, J.J. and Everts-van der Wind, A. (2005) *Genome Res*. **15**:936.

Cole, J.B. and VanRaden, P.M. (2006) J. Dairy Sci. **89**:2722.

Dekkers, J.C.M., Ten Hag, J.H. and Weersink, A. (1998) *Livest. Prod. Sci.* **53**:237.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. and Thompson, R. (2006) "ASReml User Guide Release 2.0" VSN International Ltd, Hemel Hempstead, UK.

Grisart, B., Coppieters, W., Farnir, F., Karim, L., Ford, C. Berzi, P., Cambisano, N., Mni, M., Reid, S., Simon, P., Spelman, R., Georges. M. and Snell, R. (2002) *Genome Res*. **12**:222.

Haile-Mariam, M. and Goddard, M.E, (2008) *Animal*, **2**:325.

Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. (2001) *Genetics* **157**:1819.

Scheet P. and Stephens, M.A. (2006) *Am J Hum Genet* **78**:629.