Proc. Assoc. Advmt. Anim. Breed. Genet. ***17****: 231-234*

GENOME WIDE SELECTION IN DAIRY CATTLE BASED ON HIGH-DENSITY GENOME-WIDE SNP ANALYSIS: FROM DISCOVERY TO APPLICATION

H.W. Raadsma[1,2], K.R. Zenger[1,2], M.S. Khatkar[1,2], R. Crump[1,4], G. Moser[1], J. Solkner [1,2], J.A.L. Cavanagh[1,2], R.J. Hawken[1,3], M. Hobbs[1,2], W. Barris[1,3], F.W.Nicholas[1,2,], B.Tier[1,4]

[1]Co-operative Research Centre for Innovative Dairy Products-CRC IDP, [2]ReproGen – Centre for Advanced Technologies in Animal Genetics and Reproduction, Faculty of Veterinary Science, The University of Sydney, Camden, Australia, [3] CSIRO Livestock Industries, Brisbane, Australia, [4] AGBU, University of New England, Armidale, Australia

SUMMARY

A genome wide selection (GWS) platform was developed for prediction of genetic merit in dairy cattle. The critical components of the GWS platform included a genome wide SNP analysis assay representing 15,036 SNPs, 1546 progeny tested Holstein Friesian sires with EBV (ABV) for 42 lactation performance traits, and a series of complexity reduction methods with internal and external cross validation. Derived Molecular Breeding Values (MBV) using a fraction of the available SNP information, were shown to have high predictive value for genetic merit (r=0.65-0.87 with ABV) in bulls not used in the training data from which the SNP effects were derived. GWS can be used in the absence of SNP location and pedigree to make potentially highly accurate predictions of genetic merit at an early age from DNA analyses.

INTRODUCTION

Development of high-density large-scale single nucleotide polymorphism (SNP) genotyping platforms (Hardenbol *et al.* 2005) has opened the possibility of Genome Wide Selection (GWS) in cattle. In this paper we discuss the development of a GWS platform for dairy cattle. Tier *et. al.* (2007) discuss the ramifications and technical issues of GWS for applications in animal breeding. GWS is performed by selecting subsets of marker combinations from a panel of markers located across the genome, to maximize the relationship between genetic merit derived from molecular data and true breeding value (TBV) or a proxy based on Estimated Breeding Value (EBV). Other than the requirements that markers are located across the genome, no additional information such as marker location or pedigree is required to assemble a molecular breeding value (MBV). GWS thus differs substantially from QTL, candidate gene and high density marker association approaches, which all focus on estimation of single locus effects and then assembling these in a multi-locus or multi-marker model to predict genetic merit.

DEVELOPMENT OF GWS PLATFORM FOR DAIRY CATTLE

1546 Australian progeny-tested Holstein Friesian dairy bulls were screened for 15,036 SNP markers. All bulls were sourced from Genetics Australia, and represented a cohort of sires used for ongoing commercial use (proven) and rejected (non selected) following progeny testing. The sires were born between 1955 and 2001, with >96% of sires born after 1980. All sires had ABV data calculated by ADHIS for 42 traits associated with lactation performance, conformation, reproductive fitness and disease resistance. All sires had a mean reliability of 0.79 for APR. Over 85% of sires had a reliability greater than 0.70.

SNP discovery. The GWS platform is built on a commercial SNP genotyping platform (Parallele-Affymetrix) incorporating 10,410 public domain SNP markers and 4,626 proprietary SNP markers. The proprietary markers were selected to cover regions in the genome predicted to be marker-sparse, known QTL regions, and candidate genes from the CRC-IDP candidate gene data base, using both *in-silico* discovery and re-sequencing strategies.
SNP performance. The 22.5 million data points resulted in the following summary performance statistics: 99.4 % conversion rate to genotype assays; 88.1 % informative SNP markers; 91.1 % placed with predicted position based on Btau3; 97.1% on an integrated bovine map: 74.6 % with minor allele frequency >0.05, and a reproducibility of 99.2 % for repeat informative assayable SNPs. After editing and correction for discordant SNPs, 10,715 high utility SNPs were used in GWS as shown in Table 1.

Table 1. SNP usage and characteristics from a standard 10k Affymetrix bovine panel and 5k CRC custom panel

Panel	10K	Custom	Total	%Total
total SN	10410	4626	15036	100.0
typed successfully	10341	4602	14943	99.4
polymorphic	9928	3311	13239	88.1
MAF> 0.05	8713	2511	11224	74.6
departing from HWE	751	186	937	6.2
typed in <20% animals	172	74	246	1.6
autosomal	9242	4218	13460	89.5
used in GWS	928372	2343	10715	71.3

MAF, minor allele frequency; HWE, Hardy Weinberg Equilibrium; GWS, genome wide selection

Technical challenges with whole genome SNP analyses included preparation of DNA to suitable quality for SNP-chip analysis, allelic drop out, mistyping and non-Mendelian segregation, and in one case plate transposition. The use of appropriate controls and sample structures via related animals seems warranted.
SNP complexity reduction. The challenge of dealing with over parameterized data sets where the number of SNP variables greatly exceed the number of observations is dealt with at this meeting by Crump *et al.* Moser *et al.*(2007). Powerful approaches for analyzing high-dimensional whole-genome SNP data such as supervised dimension reduction through partial least squares (PLS), Principal Component Analyses, and use of optimal search algorithms for exploring the parameter space were used for prediction of genetic merit based on Molecular Breeding values (MBV). Additional non statistical SNP reduction methods will exploit use of tag SNPs in defined haplotypes.

Prediction and validation of MBV. A remarkable feature of model selection and cross validation methods has been the accurate prediction of true breeding value (TBV) via EBV (Fig. 1) Accuracies of prediction within the range of 0.7-0.85 in the absence of pedigree, and QTL/gene information have been obtained. Typically only a fraction of the available SNP (<1 %) are used to predict MBV for all major traits used in dairy cattle selection. Furthermore no loss of efficiency was observed when 6000 of the available SNPs were used in GWS development. Application of GWS may make QTL and

gene discovery programmes for Marker Assisted Selection (MAS) redundant and offer significant new promise.

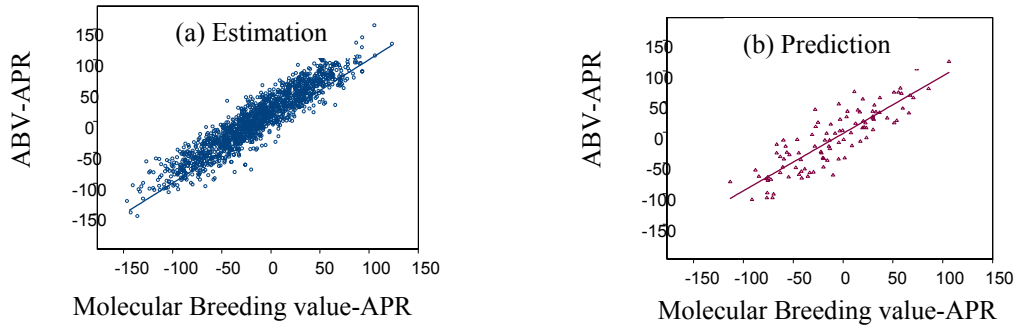

Fig. 1 Relationship between Molecular Breeding Value (MBV) and realized Estimated Breeding Value (ABV) in Australian profit rank (APR) for (a) cohort of 1346 bulls used in the estimation data set, and (b) an untested set of 200 bulls in the prediction data set. Line depicts MBV=ABV.

Utility of MBV prediction of EBV over use of pedigree alone at time of selection is shown in Fig 2. Relationship between Pedigree EBV-ABV r=0.69, MBV-ABV r=0.87, and pedigree EBV-MBV r=0.64.

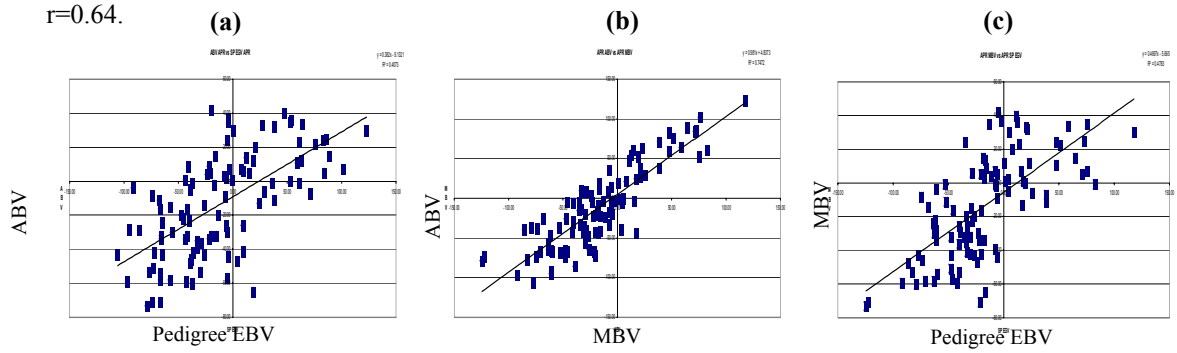

Fig. 2 Relationship for predicted progeny performance for APR in 200 bulls not used in GWS training set (a) pedigree EBV and ABV, (b) MBV and ABV, (c) pedigree EBV and MBV.

Utility and application of GWS. Deriving MBV from a population in which future predictions have to be made offers immediate use in young sire and elite dam selection (Tier *et al.* these proceedings) Features of GWS can readily be incorporated with advanced reproductive technologies, leading to greatly increased rates of genetic gain and potential significant cost reduction as breeding programmes move from progeny testing in sire selection to progeny validation. Use of MBV allows for screening of suitable germplasm from global sources, and may possibly extend to incorporate GxE and GxG and a numerator relationship matrix(NRM) based on shared genome content in genetic evaluation. Molecular Keys for GWS can be readily updated as new sires enter the industry.

To what extend the same molecular key can predict MBV and therefore TBV in other breeds is unknown, but warrants exploration. Similarly the frequency by which the molecular relationships with MBV need updating is unknown, but it would be reasonable for dairy cattle at least, that all new sires with significant progeny records will have a high density genome scan performed.
Additional applications. In addition to GWS, the SNP information is being used in the assessment of genome wide and population diversity, mate selection, management of inbreeding, study of inherited disorders, pedigree validation, Linkage Disequilibrium (LD) studies and assembly of the bovine Hapmap (Khatkar *et al.* these proceedings) and high-density integrated maps.

CONCLUSIONS

MBV appears to be potentially accurate predictors for EBV/ABV with requirements for relatively little additional information such as pedigree and SNP location. GWS derived SNP account for a high proportion of the additive genetic variance with a modest number of SNP makers. Never the less single SNP marker effects are relatively small typically accounting for <1% of additive genetic variation each. A potential new paradigm in genetic prediction is envisaged which is QTL free and may be relatively easy to implement. GWS could be expanded for multi-trait evaluation, potential GxE and GxG applications. The ultimate use of GWS applies to the forward prediction of MBV in young cohorts of bulls, and this has yet to be shown.

ACKNOWLEDGMENTS

The programme was supported by the CRC for Innovative Dairy Products, and platform establishment grant from Dairy Australia.

REFERENCES

Crump, R., *et al.* (2007). *Proc. Assoc. Advancement. Anim. Breed. Genet.* **17:** 304.
Hardenbol, P., Yu, F., Belmont, J., *et al.* (2005). *Genome Res* **15**: 269.
Khatkar, M. *et al.* (2007). *Proc. Assoc. Advancement. Anim. Breed. Genet*. **17:** 215.
Moser, G. *et al* (2007). *Proc. Assoc. Advancement. Anim. Breed. Genet*. **17:** 227.
Tier, B. *et al.* (2007). *Proc. Assoc. Advancement. Anim. Breed. Genet*. **17:** 308.
Zenger, K.R. *et al.* (2007). *Proc. Assoc. Advancement. Anim. Breed. Genet*. **17:** 123.