Proc. Assoc. Advmt. Breed. Genet. 21: 213-216

A PUBLIC:PRIVATE COLLABORATION TO EVALUATE THE POTENTIAL VALUE OF GENOMIC INFORMATION TO A VERTICALLY-INTEGRATED COMMERCIAL BEEF CATTLE ENTERPRISE

B.T Welly[1], M.D. MacNeil[2], A.R. McClain[3], R.C. Raymond[3], E.B. Simpson[4], and A.L. Van Eenennaam [1]

[1] Department of Animal Science, University of California, Davis, CA, USA
[2] Delta G, Miles City, MT, USA
[3] J. R. Simplot Land and Livestock, Grand View, ID, USA
[4] GeneSeek, a Neogen company, Lincoln, NE, USA

SUMMARY

The objective of this collaborative research project was to use phenotypes collected from Charolais-sired crossbred calves in a commercial feedlot and processing plant to develop genomically-enhanced EBVs. Phenotypes and genotypes were collected from 4,195 crossbred feedlot calves and genomic breeding values (GBV) were calculated for post-weaning average daily gain, hot-carcass weight, marbling (MRB), ribeye area, and external fat thickness (FAT). Estimated breeding values (EBV) for Charolais sires with 10 or more progeny were calculated using an animal model with MTDFREML. Correlations of GBV with EBV ranged from 0.84 to 0.93 when all calves were included in the data, but dropped to between 0.13 and 0.31 when sire's own progeny were removed from the data set using a 5-fold cross-validation approach. Correlations increased when narrowing the evaluation to only those sires with 15 or more progeny, resulting in trait GBV accuracies ranging from 0.18 to 0.45 for FAT and MRB, respectively. The inclusion of additional progeny in subsequent years of this project is expected to improve the accuracies of genomic predictions, and data will be used to evaluate the potential uses, costs and predicted benefits of using genomic information to optimize breeding program design and management on this vertically-integrated beef operation.

INTRODUCTION

To explore the potential economic value of genomic information to a large, vertically-integrated beef cattle enterprise, a collaborative research agreement was formed between J.R. Simplot Land and Livestock and their genetic consultant Dr. Michael MacNeil, Delta G; the University of California, Davis; and Neogen/GeneSeek. Objectives of the project are i) to develop genomically-enhanced EBVs using data collected from commercial calves in the feedlot for the selection of terminal sire seedstock, ii) determine the cost:benefit of incorporating genomics into seedstock selection for an enterprise that derives value improvement in feedlot and processor economically-relevant traits, and iii) examine other opportunities for deriving additional value from the genomic information such as marker-assisted management of the feedlot calves or an optimized breeding program design for this enterprise. Here we provide a preliminary report of results from the first year of data collection.

MATERIALS AND METHODS

Phenotypes were collected from 4195 crossbred feedlot calves sired by Charolais terminal sires. Crossbred calves were genotyped with the "Low Density GeneSeek Genomic Profiler" (**GGP_LD)** bead chip that includes 19,725 SNPs. The phenotypic data collected for this project includes sex and polled status and encompasses traits involved in feedlot performance and carcass merit. Current traits analyzed for this project include post-weaning average daily gain (**ADG**), hot-carcass weight (**CWT**), marbling (**MRB**), ribeye area (**REA**), and backfat thickness (**FAT**). A

total of 629 Charolais sires were genotyped, 415 with the "High Density GeneSeek Genomic Profiler" (**GGP_HD**) that includes 76,883 SNPs and 214 with the GGP_LD. Sire assignment was performed on calves with all sires as potential candidates for each run using an exclusion-based method implemented by SireMatch (J. Pollak, Cornell University). Two sets of 500 SNPs, selected based on high minor allele frequency (**MAF**) and high call rate, were utilized for the assignment of sires.

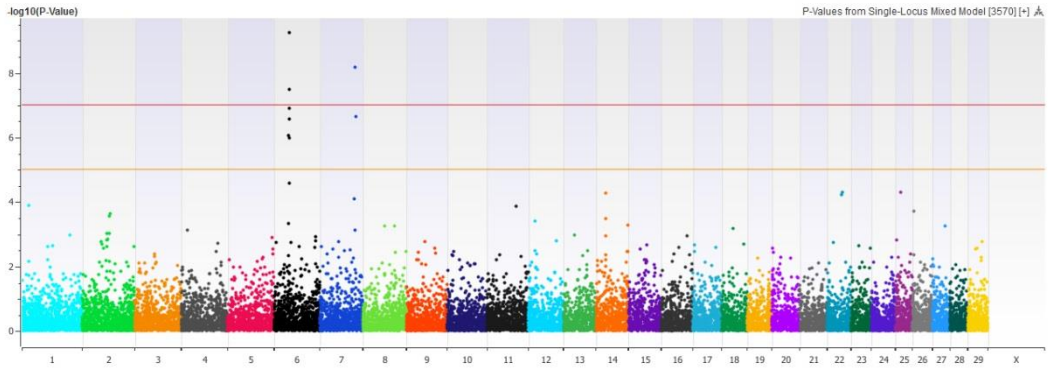
A genome wide association study was performed on all traits using the Efficient Mixed-Model Association eXpedited (EMMAX) model implemented in Golden Helix. The mixed model equations consisted of contemporary groups based on sex, ranch origin (10 ranches), and harvest date. The GWAS was carried out utilizing phenotypes on 3,555 crossbred calves and the 15,658 SNP from the GGP_LD. SNPs surrounding significant QTL peaks were extracted to evaluate the proportion of genetic variance explained by SNPs in the QTL region.

Estimated breeding values (**EBV**) and heritability ($\mathbf{h^2}$) estimates were first calculated using an animal model with MTDFREML (Boldman *et al.*, 1995). Due to a lack of pedigree data, sires (established via genotyping) were considered unrelated and dams were unknown. The EBV for Charolais sires with 10 or more progeny were extracted from the results. The GBLUP method implemented in Golden Helix's SNP and Variation Suite (Golden Helix, Bozeman, MT) was then utilized to estimate SNP marker effects on 8,000 SNP that are common to both the GGP_LD and GG_HD, for prediction of genomic breeding values (**GBV**). A 5-fold cross validation approach was used to calculate the accuracy of the GBV. Sires with EBV were randomly allocated to one of 5 groups such that approximately an equal number of progeny were included in each group. In each training analysis, the progeny of the sires in each of the 5-fold cross validation groups were excluded for the development of the genomic prediction equation for those sires. Accuracy of the genomic breeding value for the sires was calculated as the correlation between the EBV and the GBV, divided by the average accuracy of the EBV (Neves *et al.* 2014).

RESULTS AND DISCUSSION

Collection of phenotypes at the feedlot and through the processing facility was facilitated by the use of electronic capture of all records at the processing chute, and the use of matched pair sets of visual ID and EID and the nextGen™ ear tissue sampling unit (Allflex USA, Dallas, TX) to collect a DNA sample for genotyping. A total of 4195 DNA samples were analysed of which 3269 were identified to a total of 325 single sires (77.93%). The use of two sets of 500 SNPs for sire exclusion clearly identified animals with no genotyped sire. Principal component analysis of the genomic data clearly revealed clustering of half-sib groups for groups of calves with no sire assignment. Data from calves that were not assigned to a specific sire were also used as part of the training population for the GBV. Collection of DNA from all possible sires remains one of the predominant difficulties when working with large commercial populations. The proportion of possible sires that were genotyped increased for the year 2 progeny as demonstrated by an increase in sire assignment rate to 87.5%.

The GWAS analysis identified a small number of calves with incorrect gender assignment, and correctly identified a significant LOD peak for the polled locus on Chromosome 1. Significant SNPs were identified for CWT, ADG, and FAT on chromosomes 6 and 7 (**Figure 1**), both of which have been associated with growth in beef cattle previously (Lindholm-Perry *et al.*, 2011; Saatchi *et al.*, 2014). Further analysis for CWT on SNPs surrounding the peak on chromosome 6 revealed that 10 SNP on either side of the peak accounted for 6.29% of the genetic variance with the most significant SNP accounting for 1.1%. Two SNP in close proximity to the peak on chromosome 7, and two SNP upstream accounted for 1.72% and 0.77% of the genetic variance, respectively. The identification of QTLs that are in common with those found in other studies using different breeds of cattle supports the integrity of the field phenotypic data collection in this



study.

Figure 1. Manhattan plot using 3555 hot carcass weight (CWT) records implemented with a mixed model analysis, using the Efficient Mixed-Model Association eXpedited (EMMAX) model.

Accuracies of genomic prediction using a 5-fold cross validation approach ranged from 0.13 to 0.38 and from 0.18 to 0.45 for sires with ≥10 or ≥ 15 progeny, respectively (**Table 1**). Estimates of GBV accuracy using this 5-fold cross-validation approach are likely lower than true accuracy, because a large number of calves (i.e. ≥10 calves x ~ 22 sires) were removed from the training population in each of the five iterations.

Table 1. Accuracies of EBVs of Charolais sires with at least 10 progeny records, and GBVs[1] for the same sires when using all progeny records to train the prediction equations, or when excluding the sire's own progeny from the training population using a 5-fold cross validation approach.

			Sires ≥ 10 Progeny					Sires ≥ 15 Progeny				
Trait[2]	h^2 ± SE	N_{Train}[3]	N_{Sire}[4]	EBV Acc.[5]	r_{All}[6]	$r_{5\text{-}Fold}$[7]	GBV Acc.[8]	N_{Sire}[4]	EBV Acc.[5]	r_{All}[6]	$r_{5\text{-}Fold}$[7]	GBV Acc.[8]
ADG	0.31±0.06	3392	112	0.75	0.84	0.21	0.29	65	0.78	0.88	0.29	0.37
CWT	0.32±0.06	3555	114	0.74	0.92	0.22	0.30	72	0.77	0.92	0.21	0.28
MRB	0.49±0.08	3370	111	0.80	0.92	0.31	0.38	67	0.82	0.93	0.37	0.45
REA	0.40±0.07	3370	111	0.77	0.87	0.21	0.28	67	0.8	0.89	0.27	0.33
FAT	0.49±0.08	3370	111	0.80	0.93	0.13	0.16	67	0.82	0.94	0.15	0.18

[1]Genomic breeding value (GBV) accuracy estimates were calculated on Charolais sires with at least 10 or at least 15 progeny records; [2]ADG = average daily gain from feedlot arrival to final implant (μ= 135 days); CWT= carcass weight; MRB= marbling scored by camera; REA= ribeye area scored by camera; FAT= external fat thickness in adjusted yield grade units; [3]Number of crossbred calves with associated phenotypes and genotypes used to train the prediction equations; [4]Number of Charolais sires with ≥ 10 and 15 progeny respectively; [5]Average accuracy of estimated breeding values (EBV); [6]Pearson's correlation between EBV and GBV, r(EBV,GBV), when all crossbred calves are included in the training; [7]Pearsons correlation between EBV and GBV, r(EBV,GBV), for 5-fold cross-validation, where progeny from one sire group were excluded for the prediction of GBVs for that respective sire group; [8]Accuracies calculated as the Pearson's correlation between the EBV and the 5-fold cross-validated GBV, divided by the average accuracy

of the EBV.

Accuracy of genomic prediction when using phenotypes is affected by heritability of the trait, quality of the phenotypic data, number of animals in the training population for each trait, marker density, and statistical prediction methodology. Thus, we anticipate improved accuracy to result from increases in the number of sires and phenotyped calves from subsequent calf drops, as well as future work to impute genotypes to greater density (Marchini and Howie, 2010), and implement Bayesian prediction methodology (Fernando *et al*., 2014).

The impediments to the adoption of genomic technology in the beef cattle industry include the need for large training populations, the lack of a national breeding objective that includes and appropriately weights varying economic drivers in the different sectors of the beef cattle industry, and the difficulty of obtaining phenotypes from the whole supply chain. Much of the value derived from selection at the seedstock sector is realized by downstream supply chain partners (e.g. processing sector). Frequently there is no price signalling back to the seedstock producer making investments in phenotyping and genotyping to improve genetic progress in these traits, and this market failure impacts the commercial viability of any genetic technology (Van Eenennaam *et al*., 2011).

Vertically-integrated enterprises have the opportunity to develop their own breeding objective, and derive all of the value associated with genetic improvements across the various sectors of the beef industry, and hence are ideally situated to fully realise the potential of genomic information (Van Eenennaam and Drake, 2012). One advantage that vertically-integrated beef operations have when developing their breeding objective is the opportunity to include non-conventional traits. They are more likely to have ready access to records of economically relevant traits (e.g. feedlot feed requirements; survival to market endpoint) with very high relative economic value (Van Eenennaam and MacNeil, 2011), or related indicator traits (e.g. disease treatment/death records).

It is envisioned that at the end of this three-year collaborative project accurate GBVs will have been developed for traits of economic importance to this large vertically-integrated beef cattle enterprise for their Charolais terminal sire seedstock herd, and the value proposition associated with the multiple potential uses of the genomic information and phenotypic information being collected as a part of this project will have been evaluated.

ACKNOWLEDGEMENTS

The authors acknowledge the excellent field data collection by Tommie L. Crockett, Joel A. Ontiveros, and Nicole Sykes.

REFERENCES

Boldman K.G., Kriese L.A., Van Vleck L.D., Van Tassell C.P. and Kachman S.D. 1995. A manual for use of MTDFREML. A set of programs to obtain estimates of variances and covariances. U.S. Department of Agriculture, Agricultural Research Service. Washington, DC.

Fernando R.L., Dekkers J.C., and Garrick D.J. (2014) *Genet. Sel. Evol*. **46**:50.

Lindholm-Perry A.K., Sexten A.K., Kuehn L.A., Smith T.P., King D.A., *et al*. (2011) *BMC Genet*. **14**:103.

Marchini, J. and Howie, B. (2010). *Nat. Rev. Genet.* **11**:499.

Neves H.H., Carvalheiro R., O'Brien A.M., Utsunomiya Y.T., do Carmo A.S., **et al**. (2014) *Genet. Sel. Evol.* **46**:17.

Saatchi, M., Schnabel R.D., Taylor J.F., and Garrick, D.J. (2014) *BMC Genomics* **15**: 442.

Van Eenennaam A.L., and Drake D.J. (2012) *Livest. Prod. Sci.* **52**:185.

Van Eenennaam, A.L. and MacNeil M.D. (2011) pp.61-68 in "Proceedings of the Beef Improvement Federation 43rd Annual Research Symposium". Bozeman, MO.

Van Eenennaam A.L., van der Werf J.H.J. and Goddard M.E. (2011) *J. Anim. Sci.* **89**:307.