Genome Wide Scan for Signals of Recent Selection in Angus Beef Cattle

D. Lu[*], S. Miller[*], M. Sargolzaei[*, §], G. Vander Voort[*], T. Caldwell[*], M.K. Abo-Ismail[*], Z. Wang[¶], J. Mah[¶], G. Plastow[¶], S. Moore[¶]

Introduction

Selective sweep refers to the patterns of DNA polymorphism caused by consistent directional selection in a population, and thus is important to evolutionary biologists. To animal breeders, these patterns could be useful in the process of identifying genes underlying traits of interest. The signature of recent selection is characterized by a rapid rise in allele frequency in the selected site, and a large amount of linkage disequilibrium (LD) existing over an extended region of the chromosome (Sabeti *et al.* (2002)). This signal can be detected using highly dense panels of single nucleotide polymorphisms (SNP).

Among the techniques developed for detecting selection signatures is the extended haplotype homozygosity (EHH) statistic, suggested by Sabeti *et al.* (2002). It detects core haplotypes that are preserved across the population, and is reported to be useful for SNP data rather than sequencing data (Tang *et al.* (2007)) while not requiring ancestor alleles be known (Qanbari *et al.* (2010)), making it useful for animal breeding applications in practice.

This study was designed to investigate: first, signals of recent selection across the genome of Angus beef cattle; second, the association between the discovered signals and feed efficiency traits, namely average daily gain (ADG), residual feed intake (RFI), dry matter intake (DMI), and metabolic weight at mid-point of the test period (MMWT).

Materials and methods

Animals and phenotypic data. One thousand forty-two animals were genotyped, using the Illumina Bovine SNP50 Beadchip, including 411 bulls born 1984-2006, 585 steers born 1998-2006, and 46 heifers born 1999-2005. In terms of breeds, there were 60 Angus (AN), 43 Piedmontese (PI), 20 Simmental (SM), 17 Charolais (CH), and 902 crossbreds (C). The crossbreds consisted of 41 animals of 75-87.5% AN, 235 of 51-75% AN, 96 of 51-87.5% SM, 20 of 51-87.5% CH, and the rest, 510 animals, of other breed combinations. In terms of breed composition, the population had 38.66% AN, 24.10% SM, 11.28% PI, 9.35% CH, and 16.61% others. The animals were born at one of three University of Guelph cooperating

[*] Centre for Genetic Improvement of Livestock, University of Guelph, Canada
[§] L'Alliance Boviteq, Saint-Hyacinthe, Canada
[¶] Faculty of Agricultural, Life and Environmental Sciences, University of Alberta, Canada

herds. Cows were bred to mostly purebred sires through the extensive use of artificial insemination (AI). The animals were weaned at approximately 200 days of age to be fed under feedlot conditions. Steers were either penned and fed individually, or kept in groups of 12-15 and fed automatically using the Calan gates or Insentec feeding systems. Time of feeding and amount of feed consumed were recorded. Feed intake was analyzed for 112 day test period per animal. Feed efficiency traits analyzed in this study included ADG (kg/day), DMI (kg/day), MMWT (kg), and RFI (kg/day) that is the difference between observed DMI and expected DMI based on ADG and MMWT.

Genotypic data and statistical analyses. A total number of 56,947 SNPs were genotyped across 29 *Bos taurus* autosomal chromosomes (BTA). After filtering out SNPs with minor allele frequency less than 10%, and/or out of Hardy Weinberg equilibrium ($P<0.01$), 38,745 SNPs were utilized to reconstruct haplotypes, using fastPHASE (Scheet and Stephens (2006)). The haplotypes were then fed to SWEEP v.1.1 (Sabeti *et al.* (2002)) to detect core regions based on the EHH statistic, which is fully described by Sabeti *et al.* (2002). EHH detects the transmission of an extended haplotype without recombination. Relative EHH (REHH) is the EHH corrected for the variability in recombination rates at different regions on a chromosome. A haplotype carrying signatures of recent selection should have high frequency and high REHH (Sabeti *et al.* (2002)). To account for multiple testing at each chromosomal region, type I error rate was controlled by the false discovery rate (FDR) proposed by Benjamini and Hotchberg (1995), and a threshold of 5% was used to declare a core haplotype significant.

For the purpose of detecting signatures of selection in Angus, only the Angus haplotypes were used to detect core regions and estimate REHH for core haplotypes. To obtain core haplotypes for the association test, reconstructed haplotypes from 1,042 animals were used. Regions containing significant haplotypes were compared with significant regions found in the Angus, then overlapped regions were used to test for association with feed efficiency traits. A univariate animal model was used to estimate the effect of each haplotype at each region, $y_{ijk} = \mu+age+b_i+h_j+t_k+hy+a_{ijk}+\sum_{l=1}^{s}\beta_l x_l+e_{ijk}$, where y_{ijk} was the phenotype, μ the overall mean, b_i the breed effect, h_j the heterosis effect, t_k the trial treatment effect, hy the herd:year effect, a_{ijk} the additive genetic effect of individual, β_l the linear regression coefficient for the l^{th} haplotype, x_l the number of copies of the l^{th} haplotype, e_{ijk} the residual. The herd:year and additive genetic effects were random while the rest of the effects were fixed. ASREML (Gilmour *et al.* (2009)) was used to estimate the effects. A total number of 591 animals with phenotypes on the feed efficiency traits were used in this association analysis.

Results and discussion

Signatures of recent selection in Angus. The distribution of SNPs varied among the chromosomes, with BTA1 having the highest number of SNPs (2,461) and BTA22 having the fewest (682); however the average SNP intervals were relatively consistent among the chromosomes, and the overall average distance between 2 adjacent SNPs was 60kb. A total number of 4,779 core regions was discovered in the Angus population. BTA1 contained the highest number of core regions (316), while BTA28 had the fewest (72). In terms of core

length, the largest region was found on BTA5(1,526.74kb), and the smallest was on BTA29 (18.57kb). The longest average core length was 169.15kb on BTA5, and the shortest was 113.90kb on BTA28. On average each core region was 135.48kb long. Of 4,779 core regions discovered, 160 contained a total of 185 significant haplotypes (FDR<5%). Chromosome 8 had the highest number of core regions containing significant haplotypes (19), while chromosomes 22, 23, 28, 29 had only 1 significant region each, and chromosome 27 had none. Frequencies of significant haplotypes found in those regions are plotted in figure 1. The lowest frequency was 9.5%, found on BTA9. Forty haplotypes had frequency less than 20%, 125 between 20% and 50%, and 18 over 50%. Positions of significant haplotypes were entered on the Bovine Genome Sequence Assembly (Btau_4.0) to search for corresponding genes. A total of 72 functional genes were found in those regions and related to disease resistance (e.g. *GSTK1*, *IL12RB2*, *IL23R*, *SPP1*, *LYAR*), protein metabolism (e.g. *GART*, *CTSA*), fat metabolism (e.g. *PLTP*, *LPL*), cell growth (e.g. *DIO1*, *DOK1*, *EAPP*, *GPC5*, *S1PR3*), and animal behaviour (e.g. *EPHB3*, *NMUR1*).

Association analyses. There were 2,721 core regions discovered when both purebred and crossbred animals were analyzed together. Chromosome 1 had the highest number of core regions (201) whilst chromosome 28 had the least number (36). Twenty-five overlapped regions were found between the Angus and the whole population. There were 6, 4, 9 and 4 haplotypes in significant association (P<0.05) with ADG, RFI, DMI and MMWT, respectively. Three haplotypes, on chromosomes 3, 5 and 24, significantly affected both RFI and DMI. In terms of the proportion of phenotypic variance explained by haplotypes, these haplotypes each accounted for between 0.6 and 2.8% of the variation. Haplotype ACA, positioned at 83,717,076- 83,771,090Mb on BTA3, significantly reduced RFI and DMI, and explained 2.8% and 1.4% of the variation in phenotypes, respectively. This region and another, positioned at 906,239-1,032,843Mb on BTA24, are in close proximity with 2 quantitative trait loci (QTL) for RFI detected at 76.7-93.6cM (BTA3) and -1.6-6.6cM (BTA24) (Sherman et al. (2009)), respectively. Haplotype TAA at 75,606,681-75,654,771Mb on BTA13 was significantly associated with both ADG and MMWT, and located near a QTL for body weight at 66cM (Morris *et al.* (2009)). Similarly haplotype GGCTC at 45,603,947-45,738,360Mb on BTA21 was significant for DMI, and near a QTL for feed intake at 53.5cM (Nkrumah *et al.* (2007)).

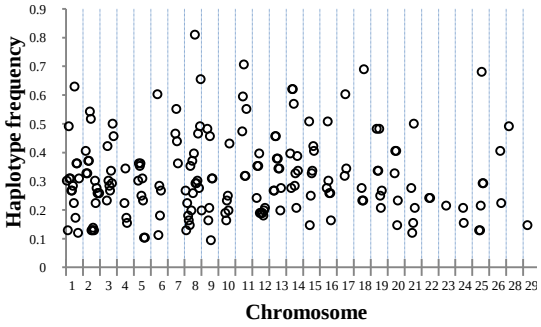

Figure 1: Distribution of significant core haplotypes (FDR<5%) across the genome

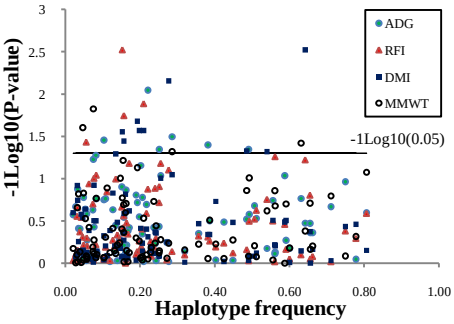

Figure 2. Distribution of P-values

In terms of distribution of significant haplotypes over frequencies (figure 2), most of the haplotypes in significant association with traits had frequencies between 0.10 and 0.30. This may suggest that the animals involved in this study have not been under intensive selection pressure for the traits analyzed. It could also be attributed to most of the animals being crossbred, and each breed has been developed for different purposes. This could also be the main reason for the number of overlapped regions being low compared to the number of significant regions found in the Angus.

Conclusion

This study reports for the first time a genome wide scan for signatures of recent selection in Angus and crossbred beef cattle, using the Illumina Bovine SNP50 Beadchip, and implementing the discovered regions in an association analysis. One hundred sixty regions were found as selection footprints in the Angus, and 25 of them were preserved across the whole population used in this study. Of these significant and overlapped regions, 18 haplotypes were in significant association with the feed efficiency traits. The results from this study would facilitate the identification of genes affecting traits of interest, and thus are important to the beef industry.

Acknowledgements

Funding from the Ontario Cattlemen's Association, the Canadian Beef Cattle Research Council, The Agriculture Adaptation Council and the Ontario Ministry of Agriculture Food and Rural Affairs. Special thanks are addressed to Dr. Flavio Schenkel for valuable advice on the statistical analysis, and Dr. Manhong Ye for technical assistance.

References

Benjamini, Y. and Hotchberg, Y. (1995). *J. R. Stat. Soc. Ser. B*, 57:289-300.

Gilmour, A.R., Gogel, B.J., Cullis, B.R. *et al.* (2009). *ASReml User Guide Release 3.0.*

Morris, C.A., Pitchford, W.S., Cullen, N.G. *et al.* (2009). *Anim. Genet.,* 40(5):648-654.

Nkrumah, J.D., Sherman, E.L., Li, C. *et al*. (2007). *J. Anim. Sci.*, 85(12):3170-3181.

Qanbari, S., Pimentel, E.C.G., Tetens, J. *et al.* (2010). *Anim. Genet.,* 9999:1365-2052.

Sabeti, P.C., Reich, D.E., Higgins, J.M. *et al.* (2002). *Nature*, 419:832-837.

Scheet, P. and Stephens, M. (2006). *American J. Hum. Genet.*, 78:629-644.

Sherman, E.L., Nkrumah, J.D., Li, C. *et al.* (2009). *J. Anim. Sci.,* 87(1):37-45.

Tang, K., Thronton, K.R., and Stoneking, M. (2007). *PLoS Biology*, 5:e171.