Proc. Assoc. Advmt. Anim. Breed. Genet. 20:455-458

COMPOSITE SIGNATURES OF DIRECTIONAL SELECTION IDENTIFIED MULTIPLE GENES FOR STATURE ON BOVINE CHROMOSOME 13 AND 14

I.A.S. Randhawa, M.S. Khatkar, P.C. Thomson and H.W. Raadsma

ReproGen - Animal Bioscience Group, Faculty of Veterinary Science, University of Sydney, Camden, NSW 2570, Australia

SUMMARY

Advances in genomic tools have made it possible to identify signatures of positive selection for complex traits in non-inbred populations. We investigated the evidence of selective sweeps for stature in 9 breeds of European *Bos taurus* by using 34,857 SNPs genotyped with an Illumina BovineSNP50 chip assay. The genotypic data were grouped in two phenotypic categories according to body size of the breeds (small-medium and medium-large). We implemented our recently developed composite index of multiple selection tests called MFR (mean fractional rank) that combines the rank distribution of three complementary test statistics to capture signatures of selection. Two strong selective sweeps were detected at loci that harbour UQCC-GDF5 and PLAG1-CHCHD7 gene pairs on chromosome 13 and 14, respectively. The two loci have previously been associated with height in humans, while PLAG1-CHCHD7 has also been reported for stature in cattle. Further investigations of the several variants in newly identified genes may help to explain the biological function of causative mutations in the diversity of bovine stature.

INTRODUCTION

Recent advances in genomic tools have facilitated studies on diverse genetic models and complex modes of their underlying inheritance in many species. Understanding the role of genetic variants in phenotypic diversity has always been challenging, and requires specific resources, tools, costs and time. Recently we developed a new method that combines multiple pieces of evidence of trait-specific selection signatures, by using the rank distribution of single nucleotide polymorphism (SNP) and haplotype-based selection tests (Randhawa *et al.* 2013).This method can be used to expand our knowledge about the genomic regions and genes controlling the diverse functions of complex traits in domestic species.

Height is a polygenic trait with high heritability in many species including cattle (Kemper and Goddard 2012). Genetic architecture of human height has been extensively investigated to find variants with major effects across the genome (Lettre *et al.* 2008; Sanna *et al.* 2008). In cattle, to date, only a few genes responsible for stature (body size) have been reported from genome-wide association studies (Pryce *et al.* 2011; Visscher and Goddard 2011; Nishimura *et al.* 2012). The known genes explain only a small proportion of the existing phenotypic variation in bovine stature (Kemper and Goddard 2012). Hence, further studies implementing new genomic tools are required to improve understanding of the genetic control of stature. To find undiscovered genetic factors, we investigated several breeds of cattle for their diversity in body size in this study.

MATERIALS AND METHODS

Data on stature in 241 animals representing nine breeds of European *Bos taurus* (Decker *et al.* 2009; Gautier *et al.* 2010) were used for this study. These breeds were selected based on the availability of the precise information on stature. The animals were genotyped with an Illumina BovineSNP50 chip assay. After quality control (MAF > 0.05) 34,857 SNPs were retained for further analysis. The animals were grouped in two phenotypic categories according to body size of their breeds (small-medium and medium-large). Breeds (sample size) selected for the small-medium group were Angus (44), Hereford (31), Limousin (35) and Romosinuano (8). Breeds

(sample size) of the medium-large sized group were Charolais (55), Chianina (8), Piedmontese (26), Romagnola (24) and Simmental (10). Imputation of missing genotypes and haplotype phasing were performed with BEAGLE 3.3 (Browning and Browning 2007). All the SNPs were mapped on UMD3.1 bovine assembly. Ancestral and derived allelic phases of these SNPs were acquired from Decker *et al.* (2009) and Matukumalli *et al.* (2009).

The analysis was performed with the mean fractional rank (MFR) method, explained in the companion paper (Randhawa *et al.* 2013), in which we combined results from commonly used 3 tests i.e., population differentiation (F_{ST}), change in derived allele frequency (ΔDAF) and across population extended haplotype homozygosity (XP-EHH) to capture evidence for selection from SNP data across multiple populations. The $-\log_{10}(p\text{-value})$ of MFR statistics were smoothed by averaging over SNPs within 1 Mb sliding windows centered at each SNP and their genome-wide top 0.1% of were used to declare the SNPs as significant. Clusters of significant SNPs were identified as the genomic regions under selection and their positions (± 0.5 Mb) were investigated to report the candidate genes under selection.

RESULTS AND DISCUSSION

Figure 1 shows the genome-wide map of the smoothed MFR scores from comparing a panel of small-medium against medium-large body size cattle breeds. Two regions of strong selective sweeps were detected which harbour multiple gene pairs on *Bos taurus* autosomes (BTA) 13 and 14 (Figure 1, Table 1). The two regions show an enrichment of high scores based on F_{ST} and XP-EHH as depicted in Figure 2. Simultaneously, an additional prominent peak at BTA1 – which is close to the significance threshold (Figure 1) – is localized at the *POLL* locus (Allais-Bonnet *et al.* 2013). This can be explained by the existence of strong secondary phenotype diversity for polledness across two breed groups, see Randhawa *et al.* (2013) for polled against horned breeds panel analysis. MFR analyses of individual breed pair data ($n \geq 24$) with contrasting body size confirmed both candidate loci on BTA13 and BTA14, however, these identified a higher number of additional peaks, likely breed-specific or spurious, than combined panels (results not shown).

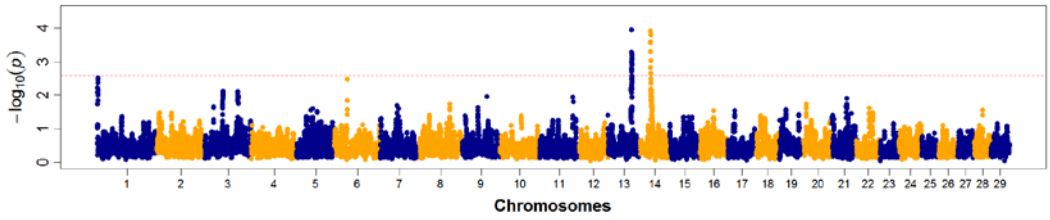

Figure 1: Genome-wide Manhattan plot of smooth $-\log_{10}(p\text{-value})$ of the Mean Fractional Ranks (MFR). Dashed (red) line indicates the top 0.1% threshold of significance.

UQCC-GDF5 locus. On BTA13, a 1.8 Mb selective sweep was localized where the ubiquinol-cytochrome c reductase complex chaperone (UQCC) and growth differentiation factor 5 (GDF5) genes are located at 65.233–65.344 Mb positions on UMD3.1 assembly of cattle (Table 1). UQCC is involved in growth control network in a number of mammalian species; along with several other genes it initiates and promotes morphogenesis and skeletal growth. GDF5 is involved in bone growth and its mutations are associated with several disorders in human skeletal development. Common variants in these two genes have been associated with variation in human height (Sanna *et al.* 2008) and strong signals of recent selection have also been identified at the GDF5 locus in European and East Asian human populations (Voight *et al.* 2006). Ensembl searches show that

UQCC and GDF5 genes have three and two mis-sense mutations, respectively (Table 1). The functional role of the putative variants underlying UQCC-GDF5 locus is unknown in cattle.

PLAG1-CHCHD7 locus. On BTA14 a 1.0 Mb selective sweep was localized where the pleiomorphic adenoma gene 1 (PLAG1) and coiled-coil-helix-coiled-coil-helix domain containing 7 (CHCHD7) genes are located at 25.007–25.059 Mb positions in cattle (Table 1). PLAG1 is consistently rearranged in salivary gland adenomas and its activation results in up regulation of target genes. CHCHD7 has no known function. Both genes have less obvious connections to body size, however, they have been considered either being in strong linkage disequilibrium with the actual causal alleles in other genes or they might indirectly regulate height via different pathways (Lettre *et al.* 2008). Previously, these two genes have been associated with height in humans (Lettre *et al.* 2008) and stature in cattle (Karim *et al.* 2011; Pryce *et al.* 2011; Nishimura *et al.* 2012). Ensemble reports detailed only two synonymous variants in PLAG1. Additional exonic variants propagating at low frequency or that have been fixed in some breeds can be identified by sequencing diverse breeds. Exploring gene networks involving PLAG1-CHCHD7 locus can further help understand the (direct / indirect) role of these genes in the diversity of stature in cattle.

Table 1: Summary of selection regions and number of genetic variants in candidate genes

BTA: region (Mb)	Candidate genes		Illumina 50K SNPs (n)	Genetic variants (n) from Ensembl data			
	Gene ID	Location (Mb)		5´UTR	Intronic	Exonic	3´UTR
13:63.9-65.7	UQCC	65.233–65.327	3 (intronic)	2	$273+1^{SR}$	$3^{MS}+1^{SN}$	6
	GDF5	65.340–65.344	-	-	10	$2^{MS}+2^{SN}$	-
14:24.4-25.4	PLAG1	25.007–25.009	-	-	4	2^{SN}	-
	CHCHD7	25.052–25.059	-	-	$21+1^{SD}$	-	-

UTR: Untranslated region, SR: Splice region, MS: Mis-sense, SN: Synonymous, SD: Splice donor

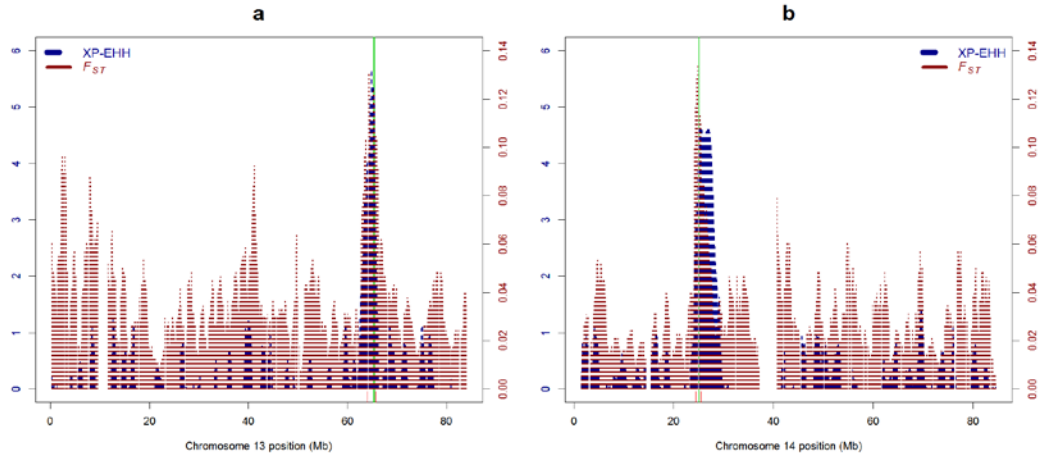

Figure 2: Plot of averaged population differentiation (F_{ST}) and across population extended haplotype homozygosity (XP-EHH) tests between the groups of small-medium and medium-large body sized breeds on a) BTA13 and b) BTA14. Vertical green lines show genic locations and red bars at bottom show candidate regions of significant Mean Fractional Ranks (MFR).

CONCLUSION

By implementing new tools for discovering selection signatures, we demonstrated the localization of candidate genes of major effects on development, skeletal growth and stature in cattle. Our results showed that the complementary signals from constituent statistics of MFR at candidate loci notably improved the resolution of MFR signals in the candidate regions. In addition, the strategy of using multi-breed panels has also contributed towards minimizing the breed-specific unique patterns of diversity in the SNP data. Further investigations of the several non-synonymous variants in the newly identified genes may help to explain the biological function of these mutations in the diversity of bovine stature. Combining selection signature analyses with genome-wide association studies can further improve the fine-mapping of causal mutations controlling stature.

REFERENCES

Allais-Bonnet A., Grohs C., Medugorac I., Krebs S., Djari A., Graf A., Fritz S., Seichter D., Baur A., Russ I., Bouet S., Rothammer S., Wahlberg P., Esquerré D., Hoze C., Boussaha M., Weiss B., Thépot D., Fouilloux M.-N., Rossignol M.-N., van Marle-Köster E., Hreiðarsdóttir G.E., Barbey S., Dozias D., Cobo E., Reversé P., Catros O. *et al*. (2013) *PLoS ONE.* **8**:e63512.

Browning S.R. and Browning B.L. (2007) *Am. J. Hum. Genet.* **81**:1084.

Decker J.E., Pires J.C., Conant G.C., McKay S.D., Heaton M.P., Chen K., Cooper A., Vilkki J., Seabury C.M., Caetano A.R., Johnson G.S., Brenneman R.A., Hanotte O., Eggert L.S., Wiener P., Kim J.-J., Kim K.S., Sonstegard T.S., Tassell C.P.V., Neibergs H.L., McEwan J.C., Brauning R., Coutinho L.L., Babar M.E., Wilson G.A., McClure M.C., Rolf M.M., Kim J., Schnabel R.D. and Taylor J.F. (2009) *Proc. Natl. Acad. Sci.* **106**:18644.

Gautier M., Laloë D. and Moazami-Goudarzi K. (2010) *PLoS ONE.* **5**:e13038.

Karim L., Takeda H., Lin L., Druet T., Arias J.A.C., Baurain D., Cambisano N., Davis S.R., Farnir F., Grisart B., Harris B.L., Keehan M.D., Littlejohn M.D., Spelman R.J., Georges M. and Coppieters W. (2011) *Nat. Genet.* **43**:405.

Kemper K.E. and Goddard M.E. (2012) *Hum. Mol. Genet.* **21**:R45.

Lettre G., Jackson A.U., Gieger C., Schumacher F.R., Berndt S.I., Sanna S., Eyheramendy S., Voight B.F., Butler J.L., Guiducci C., Illig T., Hackett R., Heid I.M., Jacobs K.B., Lyssenko V., Uda M., Boehnke M., Chanock S.J., Groop L.C., Hu F.B., Isomaa B., Kraft P., Peltonen L., Salomaa V., Schlessinger D., Hunter D.J., Hayes R.B., Abecasis G.R., Wichmann H.E., Mohlke K.L. and Hirschhorn J.N. (2008) *Nat. Genet.* **40**:584.

Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F., Heaton M.P., O'Connell J., Moore S.S., Smith T.P.L., Sonstegard T.S. and Tassell C.P.V. (2009) *PLoS ONE.* **4**:e5350.

Nishimura S., Watanabe T., Mizoshita K., Tatsuda K., Fujita T., Watanabe N., Sugimoto Y. and Takasuga A. (2012) *BMC Genet.* **13**:40.

Pryce J.E., Hayes B.J., Bolormaa S. and Goddard M.E. (2011) *Genetics.* **187**:981.

Randhawa I.A.S., Thomson P.C., Khatkar M.S. and Raadsma H.W. (2013) *Proc. Assoc. Advmt. Anim. Breed. Genet.* **20**:(in press).

Sanna S., Jackson A.U., Nagaraja R., Willer C.J., Chen W.-M., Bonnycastle L.L., Shen H., Timpson N., Lettre G., Usala G., Chines P.S., Stringham H.M., Scott L.J., Dei M., Lai S., Albai G., Crisponi L., Naitza S., Doheny K.F., Pugh E.W., Ben-Shlomo Y., Ebrahim S., Lawlor D.A., Bergman R.N., Watanabe R.M., Uda M., Tuomilehto J., Coresh J., Hirschhorn J.N., Shuldiner A.R., Schlessinger D., Collins F.S., Smith G.D., Boerwinkle E., Cao A., Boehnke M., Abecasis G.R. and Mohlke K.L. (2008) *Nat. Genet.* **40**:198.

Visscher P.M. and Goddard M.E. (2011) *Nat. Genet.* **43**:397.

Voight B.F., Kudaravalli S, Wen X and Pritchard J. K. (2006) *PLoS Biol.* **4**:e72.