Proc. Assoc. Advmt. Anim. Breed. Genet. 20:151-154

INTEGRATED ASSEMBLY OF POSITIVELY SELECTED GENES IN CATTLE

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

In this study, we assembled genome-wide scans for selective sweeps in various breeds of cattle and constructed an integrated genomic map of positively selected genes on UMD3.1 assembly. Available studies have explored a variety of genetic diversity in the form of microsatellites, SNP genotypes and DNA sequences on animals from world-wide populations of pure bred and crossbred cattle. These studies tested for departure from neutrality using various tests, mostly based on estimates of population allele differentiation and haplotype homozygosity. Definite genomic regions harbouring genes associated with simple traits (e.g. coat colour, polledness, muscle hypertrophy etc.) have been identified through signatures of selection. The genes identified under selection for the polygenic traits (e.g. adaptation, production, reproduction, feed efficiency, immunity, behaviour etc.) have also been supported by gene networks, QTLs and genome-wide association studies. These diverse investigations highlight the advantages and limitations of the available bovine genomic resources and different methodologies and have been reviewed here.

INTRODUCTION

Recent advances in various fields of genetic research have increased the availability of high throughput molecular biology tools and analytical approaches for investigating genetic diversity of farm animals. This has led us to an early understanding of the origin of species, domestication, genetic control of adaptation and imprints for selection for health and production traits (Andersson and Georges 2004; Lenstra *et al.* 2011). Modern domesticated species are a result of positive selection for the traits of economic and social importance for efficient and sustainable production in the past ~10,000 years (Mirkena *et al.* 2010). Largely due to the diverse panel of ~ 800 breeds and mixture of factors shaping their high genetic diversity, the cattle genome has been extensively investigated for signatures of selection (Barendse *et al.* 2009; Flori *et al.* 2009; Qanbari *et al.* 2010; Stella *et al.* 2010). Here we present a survey of positively selected genes for various traits identified by many tests and data sets and integrated them on the genomic positions of UMD3.1 bovine assembly (http://www.cbcb.umd.edu/research/bos_taurus_assembly).

MATERIALS AND METHODS

Available studies have explored a variety of genetic polymorphism data in the form of microsatellites, SNP genotypes (10K and 50K Illumina's BovineSNP chip assays) and DNA sequences composed of thousands of animals of multiple populations (pure breeds and crossbred). We have selected those studies which used whole-genome high-density panels of SNP genotypes for characterization of positive selection across several major cattle breeds (Table 1). The studies which have used microsatellites, DNA sequences or restricted genotyping datasets are almost twofold of genome-wide scans (data and references are not shown) and have not been included in the present study. The populations in these studies were investigated using various methods to estimate parameters in support of historical or ongoing sweeps of beneficial mutations. An integrated genomic map of positively selected genes from previous bovine assemblies (Btau3.1 and Btau4.0) was constructed by placing them – along with unique indicators for the references, selection tests and number of reporting studies – on UMD3.1 genomic positions.

Table 1: Summary of selected studies on genome-wide scans of selection signatures in cattle

Study	Test	Data (SNPs) and genome assembly	Breeds and (samples)	Genes (N)	Selective sweeps examined
Hayes *et al.* (2009)	iHS AFD	10K (9,323) Btau3.1	4 (774)	4	Milk production
Chan *et al.* (2010)	F_{ST} EHH	10K (9,919) Btau4.0	13 (317)	33	Tropical adaptation: Tick resistance, Heat resistance, Immune system
Barendse *et al.* (2009)	F_{ST} iHS CLR	10K (8,859) Btau4.0	21 (385)	2	Residual feed intake, Beef yield (intramuscular fatness)
TBHMC (2009)	F_{ST} iHS CLR	TBHMC (37,470) Btau3.1	19 (497)	20	Domestication, Behaviour Immunity, MHC, Feed efficiency, Double Muscling, Milk yield & composition, Intramuscular fat content
Stella *et al.* (2010)	CLL	TBHMC (32,689) Btau4.0	19 (497)	55	Polledness, Coat color (Black, Piebald), Dairy production, Reproduction
Gautier *et al.* (2009)	BF	50K (36,320) Btau4.0	11 (437)	42	Adaptation (pathogens & climate), Trypanosomiasis tolerance, Immune response Nervous system, Skin and hair properties
Flori *et al.* (2009)	F_{ST}	50K (42,486) Btau4.0	3 (2803)	48	Milk production, Reproduction, Body coloration
Qanbari *et al.* (2010)	EHH REHH	50K (41,398) Btau4.0	1 (810)	44	Milk yield and composition, Reproduction, Behaviour, Dairy quality
Qanbari *et al.* (2011)	F_{ST} iHS	50K (42,600) Btau4.0	12 (3876)	26	Reproduction (fertility), Muscle formation, Feed efficiency, Productive life
Gautier and Naves (2011)	iHS Rsb	50K (44,057) Btau4.0	22 (725)	11	Reproduction, Metabolism, Immunity

AFD: Allele Frequency Difference, $\boldsymbol{F_{ST}}$: Fixation index, **BF**: Bays Factor, **TBHMC**: The Bovine HapMap Consortium, **iHS**: Integrated Haplotype-homozygosity Score, **CLR**: Composite of Likelihood Ratios, **CLL**: Composite of Log Likelihood, **EHH**: Extended Haplotype Homozygosity, **REHH**: Relative EHH, **Rsb**: a measure of across population haplotype homozygosity using single locus EHH

RESULTS AND DISCUSSION

A total of 285 genes declared as candidates under selection were assembled, of which only 11 genes (9 twice and 2 thrice) were identified in multiple populations (Table 1). The integrated map contains 272 genes underlying 236 selection regions of the bovine genome (Figure 1). At least 26 selection regions identified by different studies were less than 1 Mb apart . This discrepancy may either be due to different versions of gene annotation or the nature of selection test capturing slightly different patterns of genetic diversity shaped by selection, or could be due to different genetic factors. Evidence of selection was based on the measures of population differentiation, the allele frequency spectrum, linkage disequilibrium (LD) and haplotype structures. The most common tests used to analyse genomic regions under selection were estimates of population differentiation (F_{ST}) and haplotype homozygosity (EHH and iHS).

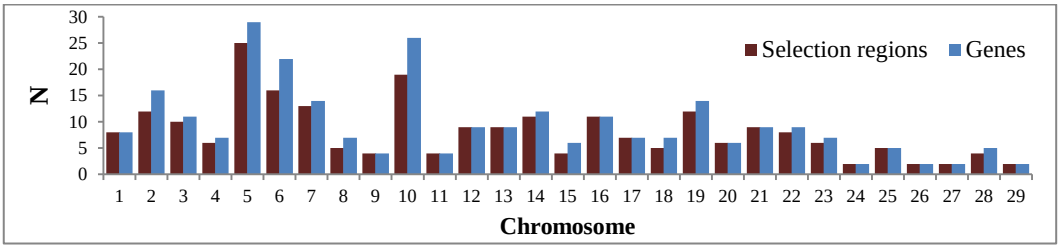

Figure 1: Distribution of chromosome-wise selection regions and genes in cattle genome.

Bovine chromosomes (BTA) 5, 6 and 10 have the highest number of identified selection regions and genes (Figure 1), whereas, BTA-2, 6 and 14 contain important candidate genes linked to various phenotypic traits in cattle (Figure 2). Cattle breeds undergoing directional or divergent selection for specific traits have shown a lack of concordance for genomic regions under selection when measured by different selection tests (Qanbari *et al.* 2011). Breed-wise sample composition, SNP panels and their density might have contributed to the differences in the results across studies (Barendse *et al.* 2009). Overall, poor concordance among studies and, selection tests within and across studies, especially in similar populations indicate the limitation of the available data sets and lack of power of selection tests. Signatures of selection harbouring genes associated with simple traits have been easily identified at the explicit genomic regions using outlier loci by applying simple genome-wide threshold strategies. For example, genes harbouring genetic mutations of major effect that control simple traits in cattle include; the polled gene on BTA-1 for absence of horns (Stella *et al.* 2010), MSTN on BTA-2 for double muscles (TBHMC 2009) and MC1R on BTA-18 for coat colour (Flori *et al.* 2009; Stella *et al.* 2010).

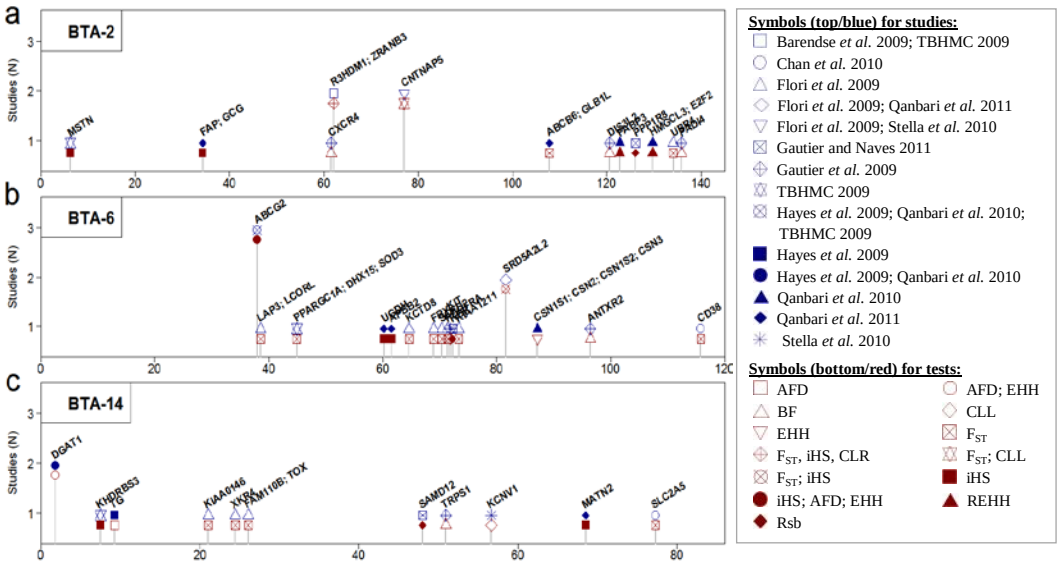

Figure 2: Genes underlying selection regions identified in various number (N) of studies by particular selection test(s) on a) BTA-2, b) BTA-6 and c) BTA-14.

The interpretation of selection signatures for complex traits is constrained by many factors, such as; limited availability of phenotypic records, variable selection pressure on polygenic alleles and inability of tests to capture selection by using conventional outlier loci approaches. The genes underlying the regions under selection for the polygenic traits have been generally linked to the phenotypic diversity in each study (see Table 1, e.g. adaptation, milk production, feed efficiency, reproduction, immune response, behaviour etc.) and in a few instances have also been supported by gene networks, QTL studies and genome-wide association studies.

Overall, the survey of genome-wide scans of selection illustrates several successful discoveries by using within and across population data sets of variable marker density. On the other hand, the disadvantages of previously available low-resolution and incomplete bovine genome maps might have provided restrictive insights. Hence, remapping previous results to the recently annotated UMD3.1 assembly and careful inspection along with new neighbouring genes can be useful. Meta-analysis of combined data from these studies can further improve the power for such analysis. Relative performance of several selection tests, as described above, has also shown differences in their power to localize a range of selection signals at varying magnitudes. A combination of multiple selection tests (Grossman *et al.* 2010; Randhawa *et al.* 2013) can be a robust approach to localize and fine-map selection regions, and link underlying genetic variation with phenotypic diversity. Moreover, the strength of signatures of selection can be improved by combining data sets and animals from multiple breeds which are phenotypically alike for the target traits.

REFERENCES

Andersson L. and Georges M. (2004) *Nature Reviews Genetics.* **5**:202.

Barendse W., Harrison B.E., Bunch R.J., Thomas M.B. and Turner L.B. (2009) *BMC Genomics.* **10**:178.

Chan E.K.F., Nagaraj S.H. and Reverter A. (2010) *Anim. Genet.* **41**:467.

Flori L., Fritz S., Jaffrézic F., Boussaha M., Gut I., Heath S., Foulley J.-L. and Gautier M. (2009) *PLoS ONE.* **4**:e6595.

Gautier M., Flori L., Riebler A., Jaffrezic F., Laloe D., Gut I., Moazami-Goudarzi K. and Foulley J.-L. (2009) *BMC Genomics.* **10**:550.

Gautier M. and Naves M. (2011) *Mol. Ecol.* **20**:3128.

Grossman S.R., Shylakhter I., Karlsson E.K., Byrne E.H., Morales S., Frieden G., Hostetter E., Angelino E., Garber M., Zuk O., Lander E.S., Schaffner S.F. and Sabeti P.C. (2010) *Science.* **327**:883.

Hayes B.J., Chamberlain A.J., Maceachern S., Savin K., McPartlan H., MacLeod I., Sethuraman L. and Goddard M.E. (2009) *Anim. Genet.* **40**:176.

Lenstra J.A., Groeneveld L.F., Eding H., Kantanen J., Williams J.L., Taberlet P., Nicolazzi E.L., Sölkner J., Simianer H., Ciani E., Garcia J.F., Bruford M.W., Ajmone-Marsan P. and Weigend S. (2011) *Anim. Genet.* **43**:483.

Mirkena T., Duguma G., Haile A., Tibbo M., Okeyo A.M., Wurzinger M. and Solkner J. (2010) *Livestock Science.* **132**:1.

Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S., Moore S., Thaller G. and Simianer H. (2011) *BMC Genomics.* **12**:318.

Qanbari S., Pimentel E.C.G., Tetens J., Thaller G., Lichtner P., Sharifi A.R. and Simianer H. (2010) *Anim. Genet.* **41**:377.

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

Stella A., Ajmone-Marsan P., Lazzari B. and Boettcher P. (2010) *Genetics.* **185**:1451.

The Bovine HapMap Consortium. (2009) *Science.* **324**:528.