# Assessing Selection Footprints Based On Breed Differentiation – An Application To Whole Genome Cattle Data

*S. Qanbari*[1], D. Gianola[2], G. Thaller[3], F. S. Schenkel[4], S. Miller[4], S. Moore[5] and H. Simianer[1]

## Introduction

Identifying recent positive selection signatures in domesticated animals could provide information on genome response to strong directional selection from domestication and artificial selection. An appealing approach for detecting selection signatures is to compare *Fst* values among loci (Wright 1951; Cockerham 1969) which provide an estimate of how much genetic variability is partitioned between, rather than within, populations. This statistic assumes that geographically variable selective forces favor different gene variants in different genomic regions. Hence, between-population allele frequency differences may be more extreme in genome regions harboring such variants. The method scans patterns of variation over many loci and considers loci in the tails of the empirical distribution as candidate targets of selection (Akey *et al*. 2002). In this study we scan the genome of a diverse set of cattle breeds to examine how various directions of positive selection may have affected the genomic pattern of those breeds, using a 50K SNP panel.

## Material and methods

Our sample population was composed of 4 cattle breeds, including 2091 Holstein (HF), 277 Brown Swiss (BS), 103 Angus (AN) and 43 Piedmontese (PI) animals. DNA from semen or blood samples was genotyped using the Illumina Bovine SNP 50K BeadChip (Matukumalli *et al*. 2009). Markers assigned to unpositioned contigs and with ≥5% missing genotypes were deleted. The final data set consisted of 40,595 common SNPs typed on 2514 animals. We estimated the $Fst = \theta$ statistic (Cockerham 1969) using a new Bayesian method proposed by Gianola *et al*. (2010).

[1] Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August University, 37075 Göttingen, Germany

[2] Department of Animal Sciences and Department of Dairy Science, University of Wisconsin-Madison, Madison, Wisconsin 53706

[3] Institute of Animal Breeding and Animal Husbandry, Christian-Albrechts-University, 24098 Kiel, Germany

[4] Centre for Genetic Improvement of Livestock, Animal and Poultry Science Department, University of Guelph, Guelph, Ontario, N1G 2W1 Canada

[5] Department of Agricultural, Food and Nutritional Science, University of Alberta, Edmonton, AB (Canada)

## Results and discussion

To determine if recent selection was responsible for the differences in allele frequencies between dairy and beef breeds, we examined *Fst* among HF and BS versus AN and PI. The distribution of posterior means of *Fst* values between dairy breeds and between beef breeds was different from that between dairy and beef breeds. Fixation index estimated between two dairy breeds, HF and BS, was 0.05 ± 0.01 and between two beef breeds, AN and PI, was 0.02 ± 0.01. Overall, the average *Fst,* comparing of dairy vs. beef breeds, was equal to 0.3 which is substantially higher than the differentiation index reported by MacEachern *et al*. (2009) between HF and AN. The higher average of *Fst* as well as the nearly similar pair-wise *Fst* within dairy and beef breeds might reflect the dominating influence of a large number of fixed SNPs in the pair-wise comparisons of breeds and groups.

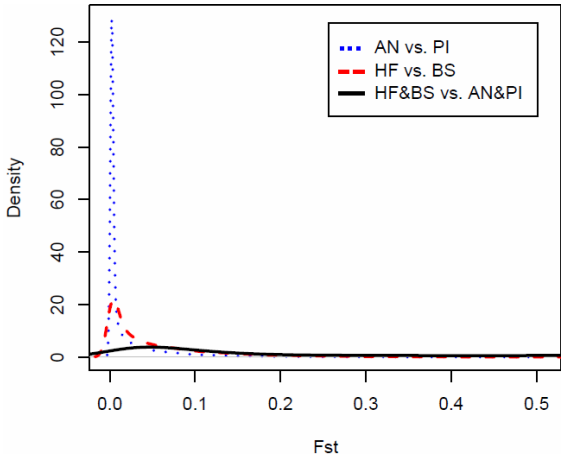


**Figure 1: Density plot shows the difference between the distribution of posterior means of *Fst* values within and across dairy and beef breeds.**

To facilitate comparisons of genomic regions among breeds and to reduce locus-to-locus variation in the inferences of selection we averaged the *Fst* values for non-overlapping windows of 500 kb across the genome. Evidence of positive selection was assumed for windows in the extreme 2.5 % of the empirical distribution, which resulted in 127 significant windows.

To identify differentiated windows between dairy and beef genomic background pair-wise *Fst* comparisons denoted as HF–AN, HF–PI, BS–AN and BS–PI were examined and plotted across the genome. Overall 29% of the genomic windows with a very large differentiation index (>0.3) overlapped in the four breed comparisons (Figure 2). BTA1 with 100 windows and BTA25 with 23 windows presented the largest and smallest amount of differentiation in the genome.

Annotation of the genes underlying the regions with extreme *Fst* does not appear to find many strong candidates for positive selection, with perhaps the exception of SMCP and

FGF1 genes. As an explanation, we suggest that selection may have been on genes that were not considered the primary targets of selection so far. The extreme peaks were mostly observed in presumed gene deserts, which may reflect selection acting on uncharacterized regulatory regions or simply fixation of non-coding DNA by genetic drift. This observation is consistent with the reports of Flori *et al*. (2009), and Gu *et al*. (2009) which reported poor gene content regions in genome wide analyses of cattle and Thoroughbred horse, respectively, using the *Fst* statistic. Thus, these results in combination with the observations from Voight *et al*. (2006), Carlson *et al*. (2005) and Wang *et al*. (2006) on human population data suggest that non-coding regions may have been an important substrate for adaptive evolution.

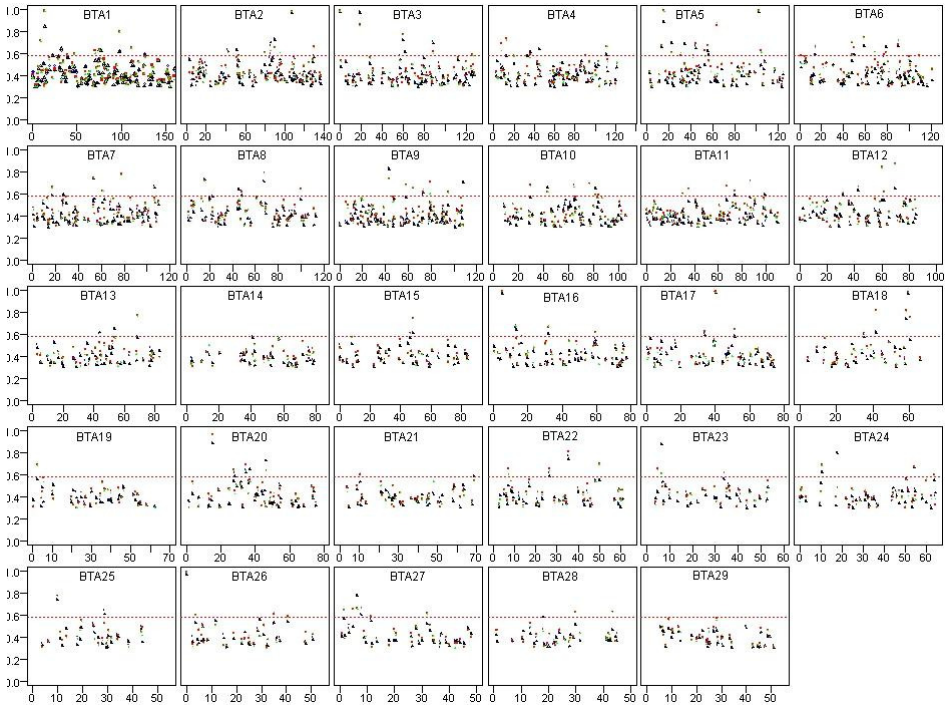


**Figure 2: Windows with *Fst* >0.3 in all pair-wise comparisons, indicating the genomic position of the most diverse regions between dairy and beef breeds. Blue, black, red, and green dots represent *Fst* values for HF–AN, HF–PI, BS–AN, and BS–PI, respectively, in each window. Dashed lines display the threshold level of 0.025.**

We examined the validity of the *Fst* analysis by testing some candidate major genes in our data set. The results revealed *Fst* values larger than expected (P < 10%) for regions harboring the Casein cluster, *GHR*, *STS*, *LP, IGF-1,* and *MYST* genes which are supposed to be targets of artificial selection. Two regions on BTA2 and BTA5 in the vicinity of *ZRANB3, R3HDM1,* and *WIF1* genes, known to affect feed efficiency and mammalian mesoderm

segmentation, respectively (Bovine HapMap consortium 2009), matched to the outlier *Fst* windows in our study.

Overall, we found a modest overlap between the results of previous genome-wide studies and our scan for selection signatures, with some noticeable exceptions. Different hypotheses can be proposed to explain these incongruities. A possible reason could be due to the fact that most studies report only the most significant results. Therefore, the results presented in this study are probably a conservative estimate of overlap between studies. Moreover, the false positive rate in genome-wide scans for selection is likely to be high. Finally, although sliding window analyses facilitate inferences of selection by reducing locus-to-locus variation, the size of the window is often subjectively determined which can influence the final results and interpretations.

## Conclusions

In this study a genomic scan based on site frequency data revealed adaptive differentiation in a substantial proportion of the cattle genome. A total of 127 regions putatively subject to recent positive selection were detected. Overall the overlap between the identified regions with previous studies was modest. Clearly, many challenges remain for this type of study, including the development of efficient methods to differentiate the effects of drift and selection, identifying the causal genes driving the signature of selection observed across large genomic regions, and functional characterization of the genes suspected as targets of selection. Our results may be of future interest for identifying signatures of recent positive artificial selection in cattle breeds and as additional evidence for polymorphisms that show associations with beef or milk traits.

## Acknowledgements

This study is part of the project FUGATO-plus GenoTrack and was financially supported by the German Ministry of Education and Research, BMBF, the Förderverein Biotechnologieforschung e.V. (FBF), Bonn, and Lohmann Tierzucht GmbH, Cuxhaven. SQ thanks the H. Wilhelm Schaumann Stiftung Hamburg for financial support.

## References

Akey, J. M., Zhang, G., Zhang, K. *et al*. (2002). *Genome Res*., 12: 1805-1814.
Carlson, C. S., Thomas, D. J., Eberle, *et al*. (2005). *Genome Res.,* 15: 1553–1565.
Cockerham, C. C. (1969). *Evolution,* 23: 72–84.
Flori, L., Fritz, S., Jaffrezic, F., *et al*. (2009). *PLoS One*, 4: e6595.
Gianola, D., Simianer, H., and Qanbari, S. (2010). *Genet. Res*., (In press).
Gu, J., Orr, N., Park, S. D., *et al*. (2009). *PLoS ONE,* 4: e5767.
Matukumalli, L. K., Lawley, C. T., Schnabel, R. D., et al. (2009). *PLoS ONE,* 4: e5350
MacEachern, S**.,** Hayes, B**.,** McEwan, J., and Goddard, M. (2009). *BMC Genom*., 10**:** 181.
The Bovine HapMap Consortium. (2009). Science, 324: 528-532
Wang, E. T., Kodama, G., Baldi, P., Moyzis, R. K. (2006). *PNAS.,* 103: 135–140.
Wright, S. (1951). *Ann. Eugen*., 15: 323-54.
Voight, B. F., Kudaravalli, S., Wen, X., *et al*. (2006). *PLoS Biol*., 4: e72.