Proc. Assoc. Advmt. Anim. Breed. Genet. 20:499-502

GENOMIC REGIONS ASSOCIATED WITH DIFFERENCES IN FAT PERCENTAGE IN MILK BETWEEN HOLSTEIN AND JERSERY CATTLE

K.E. Kemper[1], B.J. Hayes[2] and M.E. Goddard[1,2]

[1]Department of Agriculture and Food Systems, Melbourne School and Land and Environment, University of Melbourne, Parkville, VIC, 3010; [2]Department of Primary Industries, AgriBio, Bundoora, VIC, 3083

SUMMARY

Holstein and Jersey cattle exhibit large phenotypic differences in milk traits such as percentage of fat in milk (fat%). However, the genetic basis for this differentiation is unknown. Past strategies have attempted to identify selection in the genome using distortions to neutral loci and then locating candidate genes in these regions. In this paper we use the predicted difference between Holstein and Jersey breeds for fat% in milk to identify genomic regions and then examine these regions for evidence of selection. We localise a small predicted breed difference in fat% to regions on chromosome 14 and 5 but find little evidence for selection in these regions.

INTRODUCTION

Long-term selection is expected to increase the frequency of favourable alleles over time and leave a selection 'signature' in the surrounding genome. Under the classic 'hitchhiker' model of Maynard-Smith and Haigh (1974), selected loci are swept rapidly to fixation and this causes a reduction in heterozygosity at neutral loci surrounding the selected mutation. This type of selection signature is found for mutations with large effects, such as the *IGF1* mutation affecting stature in dogs (Stutter *et al.* 2007). However, evidence supporting this model for polygenic traits is limited (Pritchard, Pickrell and Coop 2010). This is because polygenic traits are influenced by hundreds or thousands of loci, each with relatively small phenotypic effect. Under these conditions, selection may cause only a small increase in the frequency of favourable alleles at many loci and leave little evidence for a selection signature.

Holstein and Jersey cattle differ markedly in fat% in milk, presumably due to differences in past selection. Different selection histories should leave evidence of selection in surrounding neutral loci. In this paper we introduce a novel method for identifying regions of the genome subject to past selection. We use predictions of the effect of single nucleotide polymorphisms (SNPs) on fat% to find regions of the genome where Holsteins and Jerseys are predicted to differ in genetic value for fat%. We then examine these regions for two traditional signatures of selection – large between breed allele frequency differences at neutral markers (i.e. high F_{ST}) and reduced SNP heterozygosity in either breed.

MATERIALS & METHODS

Data. Phenotypes and genotypes for 616,350 single nucleotide polymorphisms (SNP) were available for 2,767 Holstein and 825 Jersey bulls. All genotypes were quality checked, imputed from 50K to high density (as required) and phased with BEAGLE (Browning and Browning 2007) following Erbe *et al.* (2012). Phenotypes were daughter-trait-deviations from the Australian Dairy Herd Improvement Scheme for milk and fat yield from which daughter deviations in fat% were calculated. Holstein cattle have, on average, 1 % lower fat in milk compared to Jersey.

Estimating SNP effects for fat%. The effect of each SNP on fat% (b-hat) was estimated using BayesR, fitting the mean, SNP effects and a (residual) polygenic variance following Erbe *et al.* (2012). We analysed Holstein and Jersey bulls together in an analysis with 30,000 iterations and 20,000 discarded as burn in. Fat% was standardised within breed [i.e. $(x_i - \mu)/\sigma$] to have a

mean of zero and a phenotypic standard deviation of 1 prior to estimating the SNP effects. Therefore, we estimated within breed effects for the SNPs. SNP effects were the posterior mean of 5 replicate chains.

Identifying genomic regions predicting between breed differences in fat%. The autosomes were divided into sliding windows of 250 kb, with adjacent windows separated by 50 kb. The between breed difference in fat% was calculated as:

$$\sum_i p_{(Hol)i}\,\hat{b}_i - \sum_i p_{(Jer)i}\,\hat{b}_i \qquad [1]$$

where i is the i^{th} SNP in the 250 kb window, $p_{(Hol)i}$ and $p_{(Jer)i}$ are the allele frequency of SNP i in Holsteins or Jerseys, and $\hat{b}_i$ is the estimated effect of the SNP on fat%. Thus the sign of the between breed differences predicts if Holstein (positive values) or Jersey (negative values) have a higher fat%. The top 1% of windows were selected for further investigation. Windows from the top 1% were merged into regions when windows were separated by ≤ 250 kb.

Testing genomic regions for evidence of selection. For each of the regions identified above, we added ±125 kb of flanking sequence and calculated the mean of Wright's F_{ST} and the mean haplotype heterozygosity. Wright's F_{ST} measures the degree of allele frequency change between two populations and it was calculated per SNP following Weir and Cockerham (1984) as:

$$F_{ST} = \frac{\left(\overline{p_j^2} - \bar{p}^2\right)}{\bar{p}(1-\bar{p})} \qquad [2]$$

where p_j is the allele frequency in either Holstein or Jersey, $\bar{p}$ is the average allele frequency from Holsteins and Jerseys at the locus and $\overline{p_j^2}$ is the mean of the squared frequencies. The haplotype heterozygosity was calculated by dividing phased genotypes from BEAGLE into non-overlapping 30 SNP haplotypes and calculating 1-freq(homozygotes) in either Holstein or Jersey. P-values were determined by sampling 1000 random regions (of equal size as the observation), calculating F_{ST} and heterozygosity for these regions and determining the proportion of the sampled regions less than (or equal to) the observed F_{ST} or greater than (or equal to) the observed heterozygosity. Hence, $P < 0.05$ when the observed value was in the top (F_{ST}) or bottom (heterozygosity) 5 % of sampled regions. Finally, we re-calculated the between breed effect for fat% with [1] for the merged regions to avoid double counting of SNP from overlapping windows.

RESULTS

Between breed differences for fat%. For all SNP, the predicted difference in fat% was - 0.04 SD (i.e. - 0.002 %), implying a lower fat% for Holstein compared to Jersey cattle. This predicted between breed difference is much smaller than the observed phenotypic difference, probably because the SNP effects (b-hat) were estimated within breed. However, the direction of the between breed effect was consistent with phenotypic observations. Therefore, across all loci, Holstein cattle have a slightly higher frequency of alleles with negative effect on fat% than Jersey.

There were 510 windows identified from the top 1% of windows contributing to the between breed differences in fat%. The effects (per window) from the top 1 % had effects of between 0.003 and 0.36 SD. The 510 windows were consolidated into 110 genomic regions of up to 21 windows, from 250 kb to 1.4 Mbp.

Most (6/8) regions with between breed effects > 0.01 SD predict a slightly higher fat% in Holsteins than Jerseys (Table 1) but the largest effects on BTA5 and BTA14 predict a lower fat% in Holstein. These two regions potentially cause the lower fat% in Holstein, relative to Jersey.

For the measures of selection across all locations, the mean F_{ST} between Holstein and Jersey was 0.07 and haplotype heterozygosity was higher for Holsteins (mean heterozygosity = 0.84) compared to Jersey (mean heterozygosity = 0.75). However, the regions identified with breed differences in fat% showed little evidence for selection in the form of high F_{ST} or low heterozygosity (Table 1). In particular, the regions identified as contributing a large relative

increase in fat% for Jersey on BTA5 and 14 show do not show genomic evidence of selection for either measures of selection (Figure 1).

Table 1. Regions with large breed differences in fat% of milk for Holstein and Jersey cattle. Reported is the location, average F_{ST} and heterozygosity for each region

BTA	Region location (Mbp)	F_{ST}		Het. Holstein		Het. Jersey		Avg. effect (SD) (Hol-Jer)
3	15.25-15.7	0.146	*	0.726	*	0.43	*	0.037
3	16.55-17.2	0.046		0.817		0.78		0.026
5	93.35-94.35	0.099		0.885		0.78		-0.179
13	46.05-47.45	0.098		0.737	*	0.782		0.025
14	1.6-2.5	0.054		0.814		0.822		-0.358
14	2.6-3.15	0.058		0.815		0.725		0.020
19	42.6-43.2	0.076		0.852		0.747		0.020
20	33.9-34.8	0.095		0.735	*	0.754		0.063

$^{\#}P \leq 0.1$; $^{*}P \leq 0.05$

Several regions have previously been identified within Holsteins as associated with fat% in milk. Notably, the two regions with the extreme between breed differences for fat% on BTA14 and 5 contain the well-known *DGAT1* mutation (~1.8 Mbp; Grishart *et al.* 2004) and a region previously associated with fat% by Cole *et al.* (2011).

One of the largest regions associated with between breed differences in fat% was on BTA20 (30.9 - 32.3 Mbp), surrounding the growth hormone receptor gene (*GHR*, ~32 Mbp). This gene has been previously identified as associated with milk yield and composition (Blott *et al.* 2003). However, there was almost no predicted difference in fat% between the breeds over this region because it contained windows that predicted a higher fat% in Holstein and other windows predicting a high fat% in Jerseys and when summed across the whole region the effects tended to cancel out.

DISCUSSION

Large predicted between breed differences occur when a difference in allele frequency between Holstein and Jersey coincided with a large estimated SNP effect for fat% (i.e. see eq. [1]). The regions with the largest between breed differences in fat% were on BTA5 and 14, where previous studies have also identified genetic markers associated with fat%. However, we did not observe evidence of selection through increased F_{ST} or reduced heterozygosity in these regions. This could be because selection has not caused a big enough change in allele frequency between Holstein and Jersey or because the causative mutation is very old. If the favoured mutation is old, the linkage disequilibrium on the selected haplotype may have broken down (through mutation and recombination) or it could have existed on multiple haplotypes prior to selection.

This is a preliminary study which aimed to investigate if selection for a polygenic trait could be associated with regions of the genome. Our approach first identified regions where within breed QTL segregate with different SNP allele frequencies in Holstein and Jersey cattle, and then tested these regions for evidence of selection. This approach is similar to humans studies, where height-associated SNP were found at different frequencies in European populations and the allele frequency differences were attributed to selection (Turchin *et al.* 2012). However, although we identified some regions which could contribute to between breed differences, we found no evidence for selection surrounding these loci. Our approach could be improved by using different measure of selection (such as extended haplotype heterozygosity) or a different method to identify QTL responsible for between breed differences. For example, our analysis may be weak when the

linkage disequilibrium between the QTL and SNP varies between Holstein and Jersey. It is also likely that selection and drift has driven alternative alleles underlying between breed differences to fixation (or extreme frequencies) in our two populations. Such regions could be identified from studies with crossbred cattle.

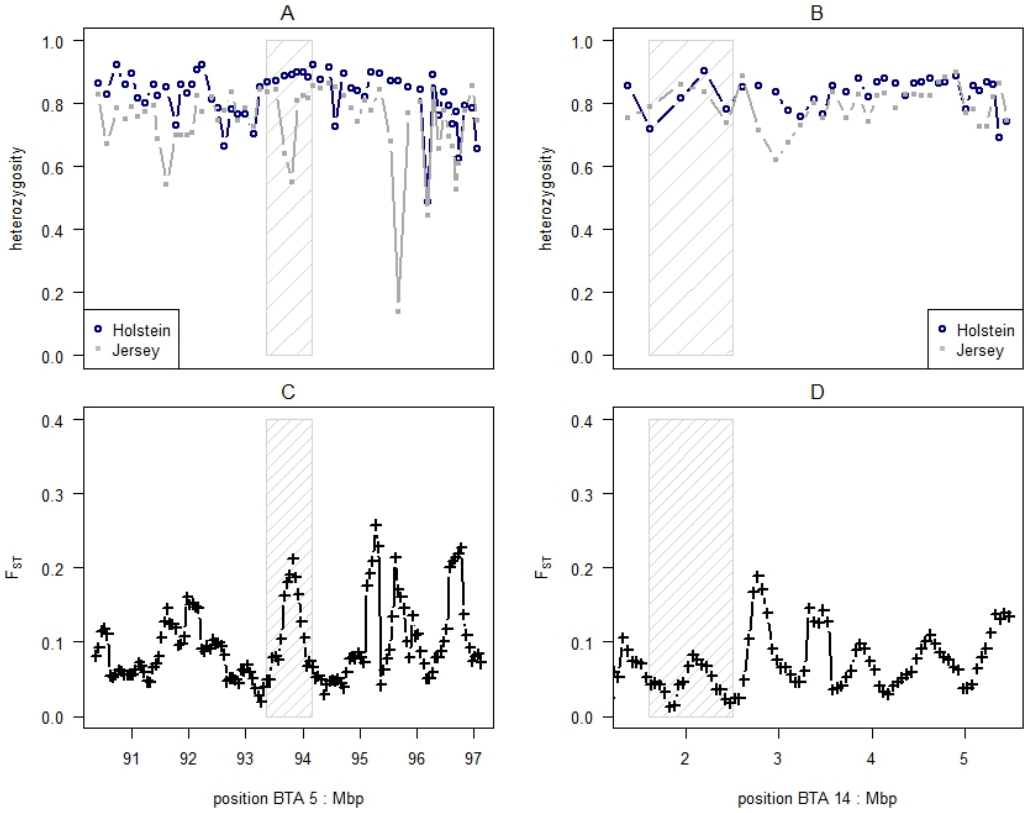

Figure 1. Heterozygosity (A, B) and F_{ST} (C, D) for regions with large between breed differences for fat% on BTA 5 (A, C) and 14 (B, D). F_{ST} is averaged over 250 kb windows

REFERENCES

Browning S.R. and Browning B.L. (2007) *Am. J. Hum. Genet.* **81:** 1084.
Cole J.B., Wiggans G.R. Ma L., Sonstegard T.S., *et al.* (2011) *BMC Genomics* **12:** 408.
Blott S., Kim J.J., Moisio S., Schmidt-Kuntzel A., Cornet A. *et al.* (2003) *Genetics* **163:** 253.
Erbe M., Hayes B.J., Matukmalli L.K., Goswami S., *et al.* (2012) *J. Dairy Sci.* **95:** 4114.
Grisart B., Farnir F., Karim L., Cambisano N., Kim J.J., *et al.* (2004) *Nat. Acad. Sci.* **101:** 2398.
Prichard J.K. Pickrell J.K. and Coop G. (2010) *Curr. Biol.* **20:** R208.
Stutter N.B., Bustamante C.D., Chase K., Gray M.M., Zhao L. *et al.* (2007) *Science* **316:** 112.
Maynard-Smith J.M. and Haigh J. (1974) *Genet Res.* **23**: 23.
Turchin, M.C., Chiang C.W.K., Palmer C.D. *et al.* (2012) *Nat. Genetics* **44:** 1015.
Weir B.S. and Cockerham C.C. (1984) *Evolution* **38:** 1358.