doi:10.1111/j.1365-2052.2009.02016.x

A genome-wide scan for signatures of recent selection in Holstein cattle

S. Qanbari*, E. C. G. Pimentel*, J. Tetens†, G. Thaller†, P. Lichtner‡, A. R. Sharifi* and H. Simianer*

*Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August University, 37075 Göttingen, Germany. †Institute of Animal Breeding and Animal Husbandry, Christian-Albrechts-University, 24098 Kiel, Germany. ‡Helmholtz Zentrum München, German Research Center for Environmental Health, Neuherberg, Germany

Summary

The data from the newly available 50 K SNP chip was used for tagging the genome-wide footprints of positive selection in Holstein–Friesian cattle. For this purpose, we employed the recently described Extended Haplotype Homozygosity test, which detects selection by measuring the characteristics of haplotypes within a single population. To assess formally the significance of these results, we compared the combination of frequency and the Relative Extended Haplotype Homozygosity value of each core haplotype with equally frequent haplotypes across the genome. A subset of the putative regions showing the highest significance in the genome-wide EHH tests was mapped. We annotated genes to identify possible influence they have in beneficial traits by using the Gene Ontology database. A panel of genes, including *FABP3*, *CLPN3*, *SPERT*, *HTR2A5*, *ABCE1*, *BMP4* and *PTGER2*, was detected, which overlapped with the most extreme *P*-values. This panel comprises some interesting candidate genes and QTL, representing a broad range of economically important traits such as milk yield and composition, as well as reproductive and behavioural traits. We also report high values of linkage disequilibrium and a slower decay of haplotype homozygosity for some candidate regions harbouring major genes related to dairy quality. The results of this study provide a genome-wide map of selection footprints in the Holstein genome, and can be used to better understand the mechanisms of selection in dairy cattle breeding.

Keywords bovine genome, EHH test, haplotype homozygosity, linkage disequilibrium, selection signature.

Introduction

Recently, linkage disequilibrium (LD) has received considerable attention among livestock geneticists, primarily to perform genome-based selection (see e.g. Meuwissen *et al.* 2001) and to determine the actual genes responsible for variation in economically important traits (Pollinger *et al.* 2005; Daetwyler *et al.* 2008; Hayes *et al.* 2008, 2009; Prasad *et al.* 2008).

The search for genes underlying phenotypic variation can be performed in two different directions; (i) from phenotype to genome, which is performed by LD-based association mapping and may involve positional cloning of QTL or by targeting particular candidate genes identified based on homology to known genes, and (ii) from genome to phenotype, which involves the statistical evaluation of genomic data to identify likely targets of past selection. The latter approaches identify patterns of LD in or between populations, which are incompatible with the hypothesis of genetic neutrality, and these patterns are called selection signatures.

Alleles under positive selection pressure encounter a fast increase in allele frequency. For a neutral mutation, it will take many generations until the mutated allele has reached a high population frequency through drift. The LD in the vicinity of this locus will be degraded through recombination (Kimura 1983), so that frequent alleles in small degrees of LD with neighbouring loci usually reflect old mutations. A novel mutation under positive selection pressure will, however, increase rapidly in frequency, so that the

Address for correspondence

S. Qanbari, Animal Breeding and Genetics Group, Department of Animal Sciences, Georg-August University, Albrecht-Thaer-Weg 3, 37075 Göttingen, Germany.
E-mail: sqanbar@gwdg.de

Accepted for publication 24 November 2009

surrounding conserved haplotype is long, which is called a 'selective sweep' (e.g. Maynard-Smith & Haigh 1974; Nielsen 2005).

This is the background of the extended haplotype homozygosity (EHH) statistic suggested by Sabeti *et al.* (2002) for the detection of recent selection. To account for facts such as variability of recombination (Simianer *et al.* 1997), Sabeti *et al.* (2002) proposed the use of the contrast of the EHH statistic of one core haplotype vs. other haplotypes in the same position. Alternative methods for detecting selective sweeps from DNA sequence data were developed, which include Tajima's *D* (Tajima 1989) and Fay and Wu's *H*-test (Fay & Wu 2000) for selected mutations, measuring large allele-frequency differences among populations by F_{ST} (e.g. Akey *et al.* 2002), and the integrated Haplotype Score (iHS; Voight *et al.* 2006), which is an extension of the EHH statistic (Sabeti *et al.* 2002). Both *D*- and *H*-tests were designed for full-sequence data and not for genome-wide collections of pre-ascertained SNPs that are currently available in some livestock species. Recently, MacEachern *et al.* (2009) applied Fay and Wu's *H*-test to examine the positive selection between Holstein and Angus cattle, which represent opposite directional selection. They used a new metric to overcome the problems of ascertainment bias and observed significant deviations of allele frequency in the two breeds. The iHS method is more powerful than the *D*- and *H*-tests for selected mutations (Voight *et al.* 2006) but, to be applied properly, it requires the genotype of the selected mutation as well as a known ancestor allele.

Among the various statistics used for recognizing signals of positive selection from polymorphism data, the EHH test is particularly useful (Walsh *et al.* 2006; Zhang *et al.* 2006). It detects selection by measuring the characteristics of haplotypes within a single population, is applied for putative core regions and does not require definition of an ancestor allele genotype. Furthermore, being less sensitive to ascertainment bias than other approaches, it is designed to work with SNP rather than sequencing data (Tang *et al.* 2007).

Holstein–Friesian cattle have been intensively selected during the last centuries, especially in the last decades after the implementation of progeny-test-based breeding programmes in the 1960s (Skjervold & Langholz 1964). Consequently, genomic regions controlling traits of economic importance are expected to exhibit signatures of selective breeding. With the availability of large-scale SNP data, it has become possible to construct a LD map with higher resolution and to scan the genome for positions that may have been targets of recent positive selection in the Holstein–Friesian population. In this study, we report the first results of such a systematic genome scan in which (i) the region of known functional candidate genes (confirmed QTL) was checked for signatures of recent selection and (ii) positional candidate genes are reported in proximity to the genomic positions showing the most significant indications of selection.

Materials and methods

DNA samples and data preparation

Semen or blood samples were obtained from 810 German Holstein–Friesian cattle, including 469 bulls and 341 bull dams. Genomic DNA was extracted by applying a modified Miller protocol (Miller *et al.*, 1988), including dithiothreitol treatment for the semen samples. Genotyping was carried out using the Illumina Bovine SNP50 BeadChip (Matukumalli *et al.* 2009) containing a total of 54001 SNPs, with a mean neighbour marker distance of 48.75 kb. Markers were filtered to exclude loci assigned to unmapped contigs, or unpositioned according to the latest reference assembly of the bovine genome Btau 4.0 (1728), monomorphic loci (11) and loci with a minor allele frequency (MAF) <0.05 (10864).

Reconstruction of haplotypes and LD analysis

The subset of animals used in this study belongs to the total population of Holstein cattle chosen for the genomic selection programme in Germany. We neglected possible effects of founders and admixture in the population demography because of the considerable number of bulls analysed, and assumed the animals to be unrelated. For the analyses, fully phased haplotype data were required. After the aforementioned filtering process, we reconstructed haplotypes for every chromosome using default parameters in fastPHASE (Scheet & Stephens 2006). Reconstructed haplotypes were inserted into HAPLOVIEW v4.1 (Barrett *et al.* 2005) to estimate LD statistics based on pairwise r^2 and to construct the blocking pattern in the candidate regions of interest to enable selection signature analysis.

Application of the EHH test

According to the natural selection theory, regions under positive selection have frequent alleles, embedded in a long-range LD background. Accordingly, the 'core region' is defined as the region of interest in the genome, presumably characterized by the strong LD among SNPs and involves a set of 'core haplotypes'. For identifying core regions, Sweep v.1.1 (Sabeti *et al.* 2002) implements the algorithm suggested by Gabriel *et al.* (2002), defining a pair of SNPs to be in strong LD if the upper 95% confidence bound of D' is between 0.7 and 0.98. The program was set to select core regions with at least three SNPs.

To evaluate how LD decays across the genome, we performed the EHH test (Sabeti *et al.* 2002). This test is based on the contrast of a core haplotype with a combination of high frequency and extended homozygosity with other core haplotypes at the same locus. EHH is the probability that two randomly chosen haplotypes carrying the candidate core haplotype are homozygous for the entire interval spanning

doi:10.1111/j.1365-2052.2009.02016.x

Table 1 Summary of genome-wide marker and core region (CR) distribution in Holstein cattle.

Chr	SNP (*n*)	Chr length (Mbp)	Mean distance (kb)	No. CR (*n*)	Mean CR length (Kb)	Coverage CR length[1] (kb)	Max CR length (kb)	CR length/Chr length[2]	CR SNPs[3] (*n*)	Max CR SNPs (*n*)	CR SNPs/SNP[4]
1	2641	161.06	61.0	265	134.2 ± 93.5	35 372.8	795.7	0.22	946	11	0.36
2	2149	140.63	65.4	207	125.1 ± 89.3	25 901.7	696.3	0.18	733	9	0.34
3	2037	127.91	62.8	209	136.8 ± 112.3	28 593.4	908.5	0.22	746	7	0.37
4	1999	124.13	62.1	191	124.5 ± 82.9	23 646.5	490.2	0.19	683	9	0.34
5	1718	125.80	73.2	164	142.4 ± 93.5	23 214.8	523.5	0.18	579	11	0.34
6	2044	122.54	60.0	217	124.7 ± 87.1	27 075.6	517.7	0.22	784	12	0.38
7	1767	112.06	63.4	171	136.9 ± 94.7	23 281.9	610.6	0.21	630	10	0.36
8	1849	116.91	63.2	187	136.9 ± 87.5	25 598.6	566.6	0.22	675	9	0.37
9	1623	108.07	66.6	127	132.9 ± 96.3	16 879.7	588.0	0.16	455	8	0.28
10	1713	106.20	62.0	171	125.6 ± 172.9	21 488.6	2212.8	0.20	592	7	0.35
11	1813	110.17	60.8	174	120.1 ± 89.7	20 900.2	661.3	0.19	607	9	0.33
12	1320	85.28	64.6	104	126.5 ± 102.0	13 164.1	661.7	0.15	359	7	0.27
13	1396	84.34	60.4	138	121.9 ± 70.1	16 824.7	446.4	0.20	483	8	0.35
14	1356	81.32	60.0	127	131.8 ± 89.9	16 743.4	546.0	0.21	454	10	0.33
15	1365	84.60	62.0	114	122.8 ± 87.3	14 008.1	686.4	0.17	392	6	0.29
16	1251	77.82	62.2	127	136.5 ± 151.0	17 341.6	1331.7	0.22	473	14	0.38
17	1284	76.45	59.5	110	111.6 ± 65.4	12 276.7	378.8	0.16	373	8	0.29
18	1100	66.12	60.1	98	117.8 ± 74.9	11 548.2	689.7	0.17	332	6	0.30
19	1108	65.21	58.9	89	125.8 ± 81.2	11 202.8	586.8	0.17	309	9	0.28
20	1252	75.71	60.5	121	122.1 ± 73.5	14 776.7	483.4	0.20	417	8	0.33
21	1093	69.17	63.3	94	114.0 ± 61.5	10 720.8	349.6	0.15	315	6	0.29
22	1009	61.83	61.3	92	112.7 ± 52.3	10 373.6	279.5	0.17	316	6	0.31
23	871	53.33	61.2	62	105.8 ± 71.5	6559	502.6	0.12	207	8	0.24
24	1013	64.95	64.1	85	125.0 ± 88.6	10 631.1	457.5	0.16	296	8	0.29
25	810	44.02	54.3	71	100.3 ± 51.8	7121.5	273.9	0.16	242	6	0.30
26	849	51.73	60.9	70	124.6 ± 96.3	8724.5	719.3	0.17	239	8	0.28
27	798	48.73	61.1	57	108.8 ± 69.5	6204.6	436.4	0.13	184	5	0.23
28	779	46.00	59.0	41	112.4 ± 65.0	4608.1	314.5	0.10	135	5	0.17
29	847	51.98	61.4	58	126.6 ± 77.6	7343.9	399.7	0.14	195	6	0.23
Total	40 854	2544.07	61.91	3741	123.7 ± 87.2	447 827.2	2212.8	0.18	13 151	14	0.32

[1]Total length covered by core regions.

[2]The proportion of total core region lengths on chromosome length.

[3]Number of SNPs forming core regions.

[4]The proportion of total number of SNPs forming core regions on number of SNPs used.

doi:10.1111/j.1365-2052.2009.02016.x

the core region to a given locus (Sabeti *et al.* 2002). The EHH of a tested core haplotype t is

$$EHH_t = \frac{\sum_{i=1}^{s} \binom{e_{ti}}{2}}{\binom{c_t}{2}},$$

where c_t is the number of samples of a particular core haplotype t, e_{ti} is the number of samples of a particular extended haplotype i and s is the number of unique extended haplotypes.

It has been observed in many experimental organisms (reviewed in Lichten & Goldman 1995; Petes 2001) that various chromosomal regions have higher (or lower) recombination rates than would be expected on the basis of the genome average recombination rate (~1 cM/Mb). Simianer *et al.* (1997) demonstrated that this variability is also prevalent in the bovine genome, and that recombination probabilities even differ between families. Regions with high (or low) recombination fractions are called hot (or cold) spots. Accordingly, the LD would be stronger in recombination cold spots than in recombination hot spots, which raises the possibility that a larger LD statistic may rather be as a result of low recombination rates in a particular region and not necessarily because of recent positive selection. The 'Relative Extended Haplotype Homozygosity' (REHH) statistic proposed by Sabeti *et al.* (2002) corrects EHH for the variability in recombination rates. It is computed by $EHH_t/\overline{EHH}$, with $\overline{EHH}$ defined as the decay of EHH on all other core haplotypes combined and is calculated as:

$$\overline{EHH} = \frac{\sum_{j=1,j\neq t}^{n} \left[\sum_{i=1}^{s} \binom{e_i}{2}\right]}{\sum_{i=1,j\neq t}^{n} \binom{c_i}{2}},$$

where n is the number of different core haplotypes.

To determine the empirical significance of REHH values, we ordered haplotypes into 20 bins according to their frequency and compared the REHH for each common haplotype in a candidate region with all equally frequent haplotypes. *P*-values were obtained by log-transforming the REHH in the bin to achieve normality, and calculating the mean and the standard deviation. As such, core haplotypes with extreme REHH in the distribution were considered significant, as suggested by Sabeti *et al.* (2002).

Results

Marker and core haplotype statistics

A total of 41 398 (76.66%) markers passed the filtering criteria and, excluding chromosome X, 40 854 SNPs were included in the final analysis. This subset of markers covers 2544.1 Mbp of the genome (Btau 4.0 assembly), with an average of 61.91 kbp adjacent the marker spacing. For the SNPs analysed in this study, the average MAF was 0.28 ± 0.15. Table 1 presents a descriptive summary of genome-wide marker and haplotype distribution in the data set. A total of 3741 core regions spanning 472 127.2 kbp (18.55%) of the genome were detected.

Mean core region length was estimated as 123.7± 87.2 kb, with a maximum of 2212.8 kb. There were 265 core regions spanning 35 372.8 kb in chromosome 1, and 41 core regions covering 4608.1 kb in chromosome 28. These were the largest and smallest haplotypic structures in the genome. For each chromosome, the proportion of length covered by core regions vs. total length as well as the number of SNPs forming core regions vs. the total number of SNPs are given in Table 1. The distribution of the size of core regions is depicted in Fig. 1. Overall, 13 151 SNPs (31.19%) participated in forming core regions, with a range of 3–14 SNPs per tract.

EHH test in candidate genes

The first step of our analysis focused on ten genes or gene clusters that are well known to be related to dairy qualities

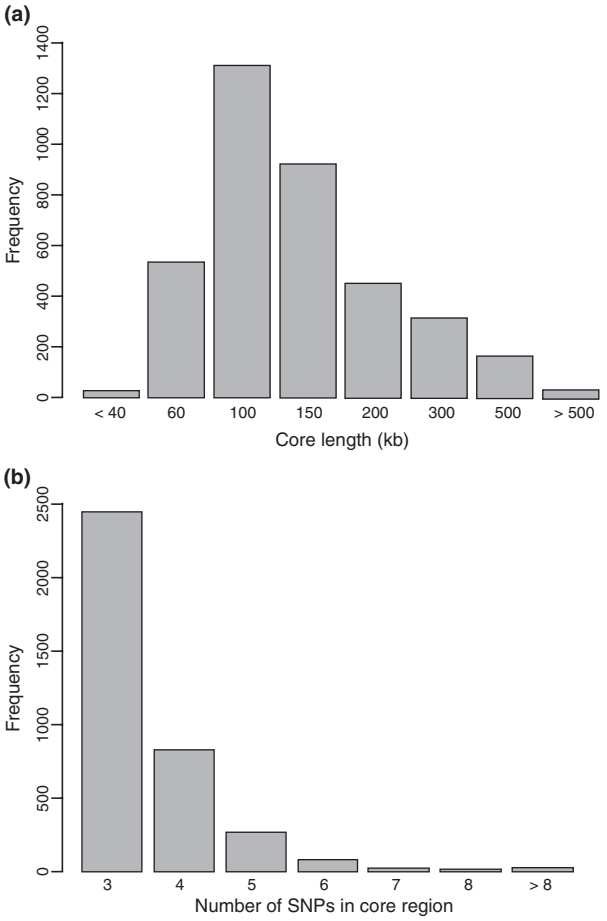

Figure 1 Distribution of the length of core regions (a) and the number of SNPs forming the core regions (b) in the Holstein genome.

doi:10.1111/j.1365-2052.2009.02016.x

and, therefore, were assumed to be potentially under recent selection. Table 2 gives the names, details and test statistics for the chosen panel. For these candidate genes, we calculated REHH as a measure of LD surrounding a haplotype of interest. REHH values much greater than 1 indicate increased homozygosity of a haplotype compared with all other core haplotypes in the genome. REHH was calculated at 1 cM distance on both the upstream and downstream sides from a core for all the possible cores present. We chose this length because of the longer extent of LD in cattle compared with that in humans, in which the considered length is commonly around 250 kb (Sabeti *et al.* 2002; Yu *et al.* 2005).

The results of the EHH test for the casein cluster show that *P*-values for the core haplotype 1 (frequency = 47%) exceeded the 99th percentile when REHH was plotted against the haplotype frequency (Figs 2.a1 and 4). *P*-values calculated for core haplotype 1 in the upstream and downstream directions are both 0.01, which indicates a clear signal of recent selection. In the case of the *DGAT1* gene, the second most frequent haplotype (frequency = 30%) showed the highest REHH in the core region when plotted up to 1 cM from the candidate region in the downstream direction. As shown in Figs 2.a2 and 3, haplotype homozygosity extended up to 1 cM only in the downstream direction for this core region. This is because of

Table 2 Summary statistics of the extended haplotype homozygosity (EHH) test for selection signature in candidate genes.

Candidate region	Chr	Closest SNP name & position (bp)	Core position	Hap freq (%)	EHH	REHH[1]	REHH[1] *P*-value
DGAT1	14	ARS-BFGL-NGS-4939	236 533–443 936	H1: 53	–/0.19	–/0.32	–/0.83[2]
		443 936		H2: 30	–/0.70	–/3.47	–/**0.06**
Casein	6	Hapmap24184-BTC-070077	88 350 095–88 427 760	H1: 48	0.63/0.54	3.69/3.61	**0.01/0.01**
Cluster		88 391 612		H2: 38	0.15/0.13	0.24/0.24	0.95/0.95
GH	19	ARS-BFGL-NGS-73805	49 523 705–49 690 250	H1: 31	0.21/0.17	0.42/0.35	0.86/0.90
		49 652 377		H2: 24	0.41/0.38	1.35/1.60	0.92/0.94
GHR	20	UA-IFASA-8974	33 908 597–34 080 608	H1: 54	0.72/0.86	1.62/1.76	**0.10/0.08**
		33 908 597		H2: 25	0.19/0.24	0.25/0.28	0.98/0.97
SST	1	ARS-BFGL-NGS-38958	81 283 582–81 376 956	H1: 34	0.76/0.84	3.16/2.44	**0.03/0.07**
		81 376 956		H2: 30	0.27/0.44	0.49/0.72	0.80/0.62
IGF-1	5	ARS-BFGL-NGS-116459	71 073 539–71 381 565	H1: 32	0.35/0.24	1.10/0.80	0.38/0.55
		71 169 823		H2: 31	0.22/0.20	0.52/0.47	0.76/0.82
ABCG2	6	BTA-22850-no-rs	37 135 014–37 374 911	H1: 35	0.19/0.18	0.50/0.45	0.76/0.79
		37 374 911		H2: 21	0.29/0.25	1.10/0.95	0.53/0.61
Leptin	4	ARS-BFGL-NGS-34894	95 715 500–95 825 044	H1: 79	0.14/0.15	0.37/0.38	0.45/0.42
		95 715 500		H2: 9	0.33/0.34	2.25/2.22	0.39/0.40
LPR	3	ARS-BFGL-NGS-74572	85 129 366–85 176 769	H1: 49	0.08/0.11	0.25/0.27	0.92/0.90
		85 569 203		H2: 37	0.33/0.43	3.04/3.03	**0.04/0.04**
PIT-1	1	DPI-55	35713131–36 085 241	H1: 32	0.22/0.24	0.65/0.62	0.67/0.69
		35 756 434		H2: 20	0.18/0.20	0.40/0.35	0.92/0.94

[1]REHH and *P*-values are presented for upstream and downstream sides from each core haplotype respectively.
[2]As shown in Figs 2.A2 and 3, haplotype homozygosities were extended up to 1 cM only in the downstream direction for this core region.
Values in bold represent core regions with haplotypes extended longer than expected.

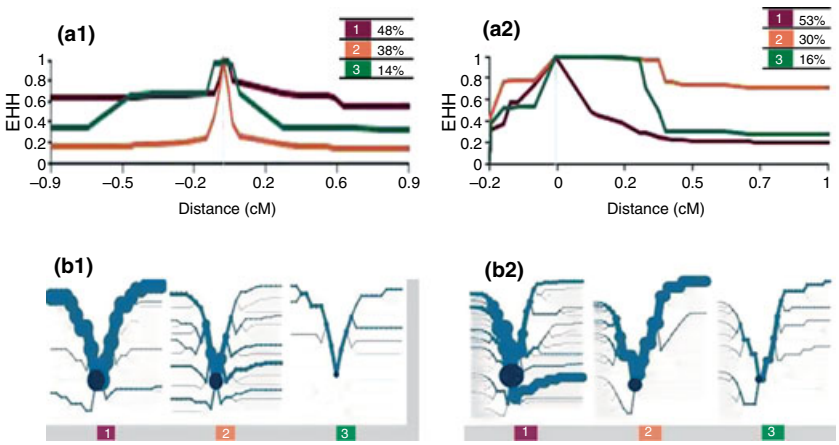

Figure 2 a1 and a2: EHH vs. distance plots for the casein cluster (1) and *DGAT1* (2) core regions, showing decay of haplotype homozygosity as a function of distance for the three most frequent haplotypes. Legends represent the core haplotype frequencies. b1 and b2: Haplotype bifurcation plots of three core haplotypes for the casein cluster and *DGAT1* regions respectively.

doi:10.1111/j.1365-2052.2009.02016.x

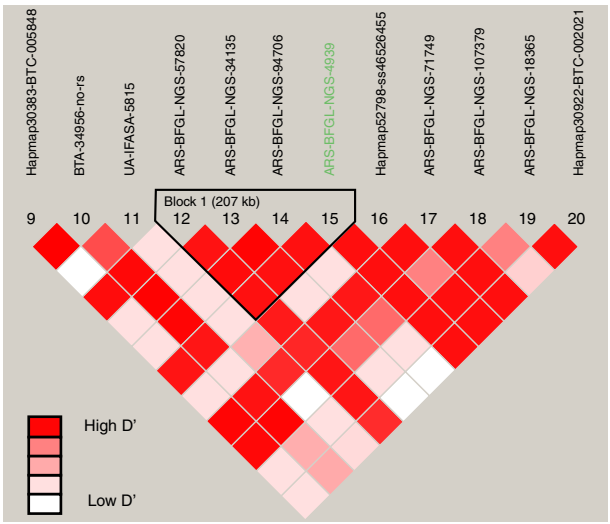

Figure 3 A graphical representation of pairwise D' for the *DGAT1* region calculated and visualized using Haploview. The SNP shown in green represents the closest SNP to the *DGAT1* gene and is involved in the block structure of length 207 kb.

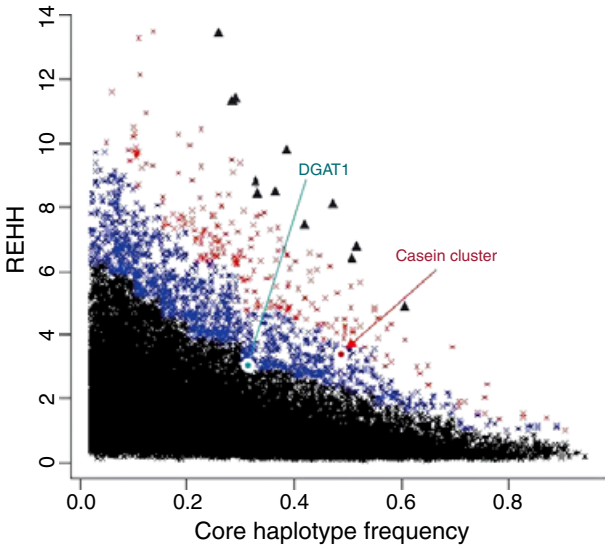

Figure 4 Distribution of REHH vs. haplotype frequencies in the Holstein genome. REHH was calculated at 1 cM distances in the upstream and downstream directions for all possible core haplotypes. Core haplotypes with *P*-values lower than 0.05 and 0.01 are presented in blue and red respectively. The panel of 12 core haplotypes displaying the lowest *P*-values ($P < 0.001$) is represented by triangles. Values representing *DGAT1* and the casein cluster are indicated.

the position of *DGAT1*, which is located at approximately 400 kb on BTA14 and also the lower LD observed in the upstream direction. This analysis also showed significant *P*-values for core haplotype 2 of the *Leptin Receptor* gene (*LPR*) and core haplotype 1 of the *Somatostatin* gene (*SST*), and approached significance for the *Growth Hormone Receptor* gene (*GHR*).

Whole genome screen for selection signatures

For all 3741 core regions, a total of 28 323 EHH tests with an average of 7.57 tests per core region were calculated. To find outlying core haplotypes, we calculated REHH at 1 cM distances on both the upstream and downstream sides. Figure 4 shows the distribution of REHH values vs. haplotype frequencies. Corresponding *P*-values are indicated by the use of different symbols. Based on the selection signature theory, a beneficial allele undergoing positive selection is fixed or is going to be fixed in the population. Hence core haplotypes harbouring these alleles would have a high frequency. Taking this into consideration, we skipped core haplotypes with frequency <25% and plotted the $-\log_{10}$ of the *P*-values associated with REHH against the chromosomal position to visualize the chromosomal distribution of outlying core haplotypes (Fig. 5). It is evident that these signals are non-uniformly distributed across chromosomes and chromosome segments, with a substantial overrepresentation on parts of chromosomes 10, 2 and 13.

Table 3 presents the genome-wide statistics of the selection signature test, including the number of tests and outlying core haplotypes for each chromosome. Of 12 435 tests on core haplotypes with frequency ≥0.25, in total 161 tests displayed outlying peaks on a threshold level of 0.01. Bovine chromosomes 6 and 14, which harbour known genes and QTL for several economically important traits (Stone *et al.* 1999; Mosig *et al.* 2001; MacNeil & Grosz 2002; Casas *et al.* 2003; Li *et al.* 2004; Ashwell *et al.* 2005; Nkrumah *et al.* 2007), showed eight and two outliers respectively. The number of peaks rises to 41 and 14 respectively when the threshold is set to $P < 0.05$.

We examined the conformity of the distribution of Tukey's outliers, with outlying core haplotypes defined on the threshold level of 0.01. Figure 6 displays box plots of the distribution of −log10 (*P*-values) within each bin of core haplotype frequency. To fit the distribution of −log10 (*P*-values), the threshold defining outliers (1%) displayed in the box plots was set to Q1-3*IQR and Q3 + 3*IQR, where IQR is the interquartile range and Q1 and Q3 are the first and third quartiles respectively. It is evident that the extreme outliers appear in the moderate bins of haplotype frequencies.

Mapping positively selected regions to genome annotations

A summary of statistics for 12 positively selected core regions presenting the lowest *P*-values of REHH test is presented in Table 4. Corresponding genes were identified using the map viewer option and by aligning the core positions to the fourth draft of the bovine genome sequence assembly (Btau 4.0). We extended core regions in both directions up to 1 cM as the length of the core domains.

doi:10.1111/j.1365-2052.2009.02016.x

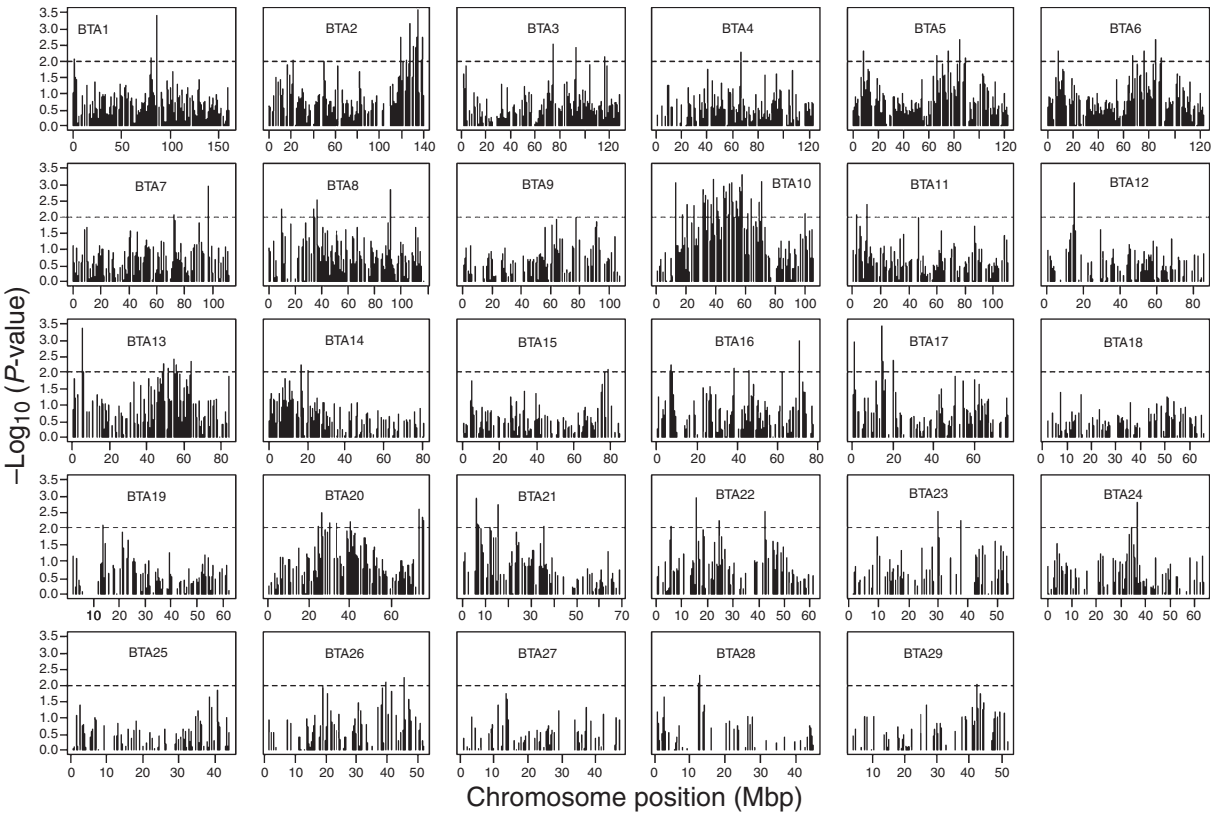

Figure 5 Genome-wide map of *P*-values for core haplotypes with frequency ≥0.25. Dashed lines display the threshold level of 0.01.

A subset of genes and EST regions were annotated for each core region. We screened this list for the most interesting candidate genes in each core region presenting top peaks. Interestingly, some regions overlapped with genes previously suggested as being under selection. For example, on chromosome 2q45, a core haplotype harbouring the *Fatty Acid-Binding Protein 3 muscle and heart* (*FABP3*) gene showed a strong signature of selection (*P*-value <0.0006). FABP3 is involved in gene networks driving bovine milk fat synthesis during the lactation cycle and plays a key role in the regulation of the channelling of fatty acids towards copious milk fat synthesis in the bovine mammary gland (Bionaz & Loor 2008). Another strong signature of selection on chromosome 12 matches the *SPERT* (*spermatid-associated protein*) and 5-*hydroxytryptamine (serotonin) receptor 2A* (*HTR2A*) genes. HTR2A 5 acts in serotoninergic pathways that are involved in economically important bovine gastrointestinal (GI) motility disorders, such as displaced abomasum and caecal dilatation/dislocation (Reist *et al.* 2003). It was also suggested that variants of the *HTR2A* gene are related to behavioural disorders in humans (Khait *et al.* 2005) and aggressiveness in canines (Peremans *et al.* 2003). This point looks more interesting when we compare the temperament of modern cattle breeds, which have been bred during the last decades, with native cattle breeds worldwide.

We found an unexpectedly high number of outliers on chromosome 10. One of the core regions representing a strong signal ($P < 0.0007$) harbours the *Calpain3* (*CPN3*) gene (Barendse *et al.* 2009). Another strong peak ($P < 0.0008$) on chromosome 10 is associated with the *bone morphogenetic protein4* (*BMP4*) gene, which is involved in the bone morphogenetic protein-signalling system, present in bovine antral follicles, and plays a role in the development and functioning of follicles (Fatehi *et al.* 2005). The other signal ($P < 0.0008$) observed on chromosome 10 is in the vicinity of the *prostaglandin E receptor 2-subtype EP2* (*PTGER2*) gene. EP2 is the major cAMP-generating PGE (2) receptor that is expressed and regulated in the bovine uterus during the oestrous cycle and early pregnancy (Arosh *et al.* 2003). It should be noticed that EHH for 5 of 6 core regions on chromosome 10 was estimated as >0.95 when plotted up to 1 cM. EHH extended up to at least 2 cM in both directions and spanned a larger number of candidate genes, which could have been the targets of recent artificial selection in these regions (Table 5).

We also explored three QTL databases available online (http://genomes.sapac.edu.au/bovineqtl/index.html, http://www.animalgenome.org/QTLdb/cattle.html, http://www.vetsci.usyd.edu.au/reprogen/QTLMap/) to identify any overlapping of the outlying core regions with published QTL in dairy and beef cattle. Table 6 lists the traits, approximate

doi:10.1111/j.1365-2052.2009.02016.x

Table 3 Summary statistics of whole genome extended haplotype homozygosity tests.

Chr	Tests on CH[1] (*n*)	*P*-value <0.05 (*n*)	*P*-value <0.01 (*n*)
1	865	17	4
2	678	58	17
3	686	15	3
4	626	11	1
5	531	37	8
6	695	41	8
7	540	13	2
8	605	41	6
9	430	14	0
10	552	123	45
11	590	15	2
12	335	9	2
13	476	51	8
14	439	14	2
15	390	8	2
16	423	29	6
17	365	25	6
18	340	3	0
19	311	5	1
20	400	51	11
21	336	31	7
22	300	27	7
23	213	18	4
24	290	8	4
25	248	5	0
26	234	14	2
27	195	5	0
28	146	5	2
29	196	9	1
Total	12 435	702	161

[1]The number of tests on core haplotypes (both sides) with frequency ≥0.25.

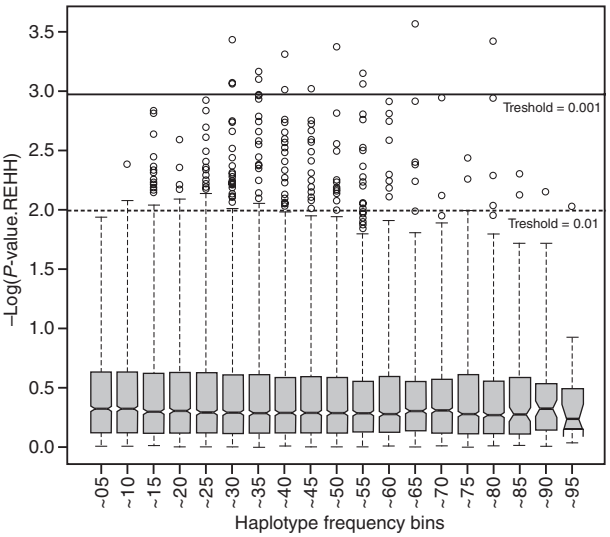

Figure 6 Box plot of the distribution of *P*-values in core haplotype frequency bins of 5% difference. Core haplotypes with *P*-values lower than 0.01 and 0.001 are separated with dashed and continuous threshold lines respectively.

position and reported population of the overlapping QTL for each core region. In the majority of cases, we found an overlap between the core regions presenting top P-values and those that had previously been identified to be harbouring beef or dairy QTL. An interesting feature of this comparison is that the majority of these QTL have been reported in Holstein populations.

Discussion

Holstein–Friesian cattle, the world's highest producing dairy animals, are believed to have been artificially selected since a few thousand years ago (Bradley & Cunningham 1998). Therefore, identifying the regions that have been subjected to selective breeding would facilitate the identification of genes related to traits of interest or biological relevance. A genome-wide map of selection events will also help us to better understand the mechanisms of selection in artificially selected populations. Unfortunately, robust inferences of recent positive selection from genomic data are difficult because of the confounding effects of population demographic history. For example, both positive selection and an increase in population size may lead to an excess of low-frequency alleles in a population relative to what is expected under a standard neutral model, i.e. a constant-size, randomly mating population in mutation-drift equilibrium (Akey *et al.* 2004). Therefore, rejection of the standard neutral model usually cannot be interpreted as unambiguous evidence for recent selection. In contrast to human populations, the strength of artificial selection is supposed to be much more pronounced than that of natural selection on fitness related traits. Therefore, it is reasonable to hypothesize that targets of artificial selection will be easier to find in domesticated livestock populations than in non-domesticated populations (Biswas & Akey 2006).

In this study, we employed the long-range haplotype test, which detects selection by measuring the characteristics of haplotypes within a single population. We mapped a subset of the putative regions, identified by extreme *P*-values across the genome (Fig. 5), and used this information to annotate genes that may be under selection pressure. The identified genes reflect a series of pathways, such as steroid metabolism, regulation of transcripts, transportation and other functional categories. For most genes associated with signals of selection, a biological link to traits such as milk yield and composition, reproduction and behaviour, which are known to be under selection, can be hypothesized. However, these results need to be confirmed by further studies.

Applying the EHH test to our data revealed 161 regions exhibiting footprints of recent positive selection at a

doi:10.1111/j.1365-2052.2009.02016.x

Table 4 Summary statistics for 12 core haplotypes showing the lowest *P*-values after the relative extended haplotype homozygosity (REHH) test.

Chr	Position	Core length (kb)	Hap freq	EHH	REHH	REHH *P*-value	Gene/ EST (*n*)	Candidate gene	Function	Reports in bovine
2	127 125 963–127 172 772	46.81	0.33	0.61	8.66	0.00068	12	*FABP3*	Regulation of the channelling of fatty acids towards copious milk fat synthesis in bovine mammary gland	Bionaz & Loor (2008)
2	134 666 758–134 761 842	95.08	0.61	0.47	5.07	0.00027	16	*HMGCL3*	Hydroxymethylglutaryl-coa lyase activity	Jiang *et al.* (2008)
								E2F2	Activating Transcription Factor-2 in skeletal growth control	Luvalle *et al.* (2003)
10	13 146 429–13 225 603	79.17	0.28	0.97	11.21	0.00087	31	*PTGER2*	Prostaglandin E receptor activity	Arosh *et al.* (2003)
								LCTL	Lactase-like protein	
10	38 264 640–38 625 718	361.08	0.51	0.98	6.74	0.00071	21	*CAPN3*	Calcium-dependent cysteine-type endopeptidase activity and protein binding	Barendse *et al.* (2008)
10	48 942 782–49 031 850	89.07	0.41	0.57	7.58	0.00095	12	*RORA*	Steroid hormone receptor activity, transcription factor activity and zinc ion binding	
10	51 073 231–51 138 335	65.1	0.51	0.80	6.45	0.00087	17	*GCNT3*	Transferase activity, transferring glycosyl groups	
								LIPC	LPL is a key enzyme in catabolism of plasma lipoprotein (TGs)	
10	57 638 141–57 773 467	135.33	0.39	0.96	9.76	0.00049	13	*CYP19*	Conversion of androgen to oestrogen	
10	70 455 224–70 552 188	96.96	0.33	0.91	8.40	0.00079	21	*BMP4*	Development and functioning of follicles and oocyte maturation	Fatehi *et al.* (2005)
12	14 556 717–14 658 840	102.12	0.29	0.79	11.26	0.00085	11	*SPERT*	Spermatid-associated protein	Reist *et al.* (2003)
								5HTR2A	G-protein coupled receptor activity	
13	5 082 478–5 148 264	65.79	0.48	0.99	8.03	0.00043	1	*BTBD3*	Proteins with a bric- à -brac, tramtrack, broad-complex/Poxvirus zinc fingers domain play role in DNA binding, regulation of gene transcription and organization of macromolecular structures	
16	70 812 261–71 003 946	191.69	0.36	0.79	8.42	0.00097	17	*HSD11B1*	KEGG pathway: Androgen and oestrogen metabolism, C21-Steroid hormone metabolism	
								LPGAT1	Acyltransferase	
17	13 973 226–14 208 603	235.38	0.26	0.81	13.41	0.00037	7	*ABCE1*	Transmembrane proteins	

doi:10.1111/j.1365-2052.2009.02016.x

Table 5 A list of candidate genes located near the peak regions on chromosome 10.

Gene	Position (bp)
PYGL *phosphorylase, glycogen, liver*	43 866 028–43 990 507
L2HGDH *L-2-hydroxyglutarate dehydrogenase*	43 275 277–43 326 310
TRIP4 *thyroid hormone receptor interactor 4*	45 955 859–46 007 749
LACTB *lactamase, beta*	47 246 369–47 264 257
CA12 *carbonic anhydrase XII*	46 990 214–47 052 877
BMP4 *bone morphogenetic protein 4*	67 159 768–68 659 191
CGRRF1 *cell growth regulator with ring finger domain 1*	68 781 739–68 809 111
CDKN3 *cyclin-dependent kinase inhibitor 3*	68 622 652–68 639 430
GCH1 GTP *cyclohydrolase 1*	69 125 665–69 182 772
SOCS4 *suppressor of cytokine signalling 4*	69 288 425–69 302 353
NAT12 *N-acetyltransferase 12 (GCN5-related, putative*	71 796 688–71 818 386
TCF12 *transcription factor 12*	54 015 561–54 205 361
GRINL1A *glutamate receptor, ionotropic, N-methyl D-aspartate-like 1A*	53 121 124–53 128 327
LIPC *lipase, hepatic*	52 220 965–52 415 726

threshold level of 0.01. We observed that other haplotypes present in this region display a shorter extent of homozygosity, indicating abundant historical recombination. Therefore, the long stretch of homozygosity observed in this region presumably is not simply because of a low local recombination rate but probably reflects the combination of strong and recent selective pressure, pushing beneficial mutations rapidly towards high frequency with long conserved haplotypes surrounding them. The test across the whole genome of the Holstein breed revealed a signal on position 62.27 Mbp on chromosome 2, which is close to the one reported by Barendse *et al.* (2009) and the Bovine HapMap Consortium (2009) to be related with feed efficiency traits in a set of cattle breeds. There was also a cluster of strong signals on chromosome 6 (position 88.35 Mbp) and chromosome 25 (position 30.24 Mbp), confirming the signatures related to multiple beef traits (Barendse *et al.* 2009) and the *ZNF187* gene (The Bovine HapMap consortium 2009) respectively.

We examined the validity of EHH analysis by testing some major candidate genes in our data set. The results revealed a longer than expected range of LD in core regions harbouring the casein cluster, *DGAT1*, *GHR*, *STS* and *LPR* genes, which are thought to affect milk yield and composition traits in Holstein cattle. This observation is in agreement with the results of Hayes *et al.* (2009) who suggested signatures of selection in the vicinity of *GHR* and *DGAT1* genes as revealed by allele-frequency differences. The long-range LD consistency observed in this study is also congruent with the reports of Grisart *et al.* (2001) and Marques *et al.* (2008) who used EHH plots to evaluate extended long-range LD around *DGAT1*. The long range of LD observed for the second most frequent core haplotype in the *DGAT1* region in previous reports is confirmed in our study. We examined the distribution of this haplotype in 146 animals for which the *DGAT1* genotype was available. As shown in Table 7, there is an almost perfect association of the GGCG haplotype with the Lysine variant at *DGAT1*.

A substantial proportion of the analysed candidate genes showed *P*-values ≤0.10, which supports the validity of our approach. However, some of the candidate genes such as *ABCG2* did not meet our definition of positively selected genes but may have nonetheless been targets of selection (Hayes *et al.* 2008, 2009; The Bovine HapMap consortium 2009). Different hypotheses can be proposed to explain these incongruences. The disparity shown may have arisen because of a possible higher initial frequency of beneficial alleles (Innan & Kim 2004). Such an allele might, for example, be imported into a breed through crosses with other breeds. In such a case, selection may have started from a moderate initial frequency, and beneficial alleles may be included in diverse haplotypes. The density of the markers is also critical for the power of such studies, and could be a source of discrepancy. Comparing the average marker spacing with mean core length and number of SNPs forming cores reveals that core regions are more likely to appear in regions of higher marker density. Thus a new SNP chip with increased marker density will allow for the more efficient identification of core haplotypes. Furthermore, with a denser marker map, a larger proportion than the 18.5% of the mapped genome would be assigned to core regions. Although the effect of marker density on the distribution of REHH values is not clear (Zhang *et al.* 2006), a denser map would allow a more reliable and comprehensive screening of the genome for signatures of selection. The incongruences can also result from the complex genomic interactions or lack of power, given the sample size available for this study.

Conclusion

Our results provide a genome-wide map of selection footprints in Holstein genome. Many of the regions showing top *P*-values seem to play important roles in economically important traits in dairy cattle and can now serve as starting points for formulating biological hypotheses. We also reported high values of LD and a slower decay of haplotype homozygosity for some candidate regions harbouring major genes related to dairy quality. Other candidate regions do not show such signals, which may be a result of statistical or biological factors. Additional studies are needed to confirm and refine our results. These may comprise within-population studies with larger

doi:10.1111/j.1365-2052.2009.02016.x

doi:10.1111/j.1365-2052.2009.02016.x

Table 6 Reported QTL near the core regions with the lowest *P*-values.

					Reported statistic		
Chr	Start – End (bp)	Trait	QTL position (cM)	Population	F-ratio	*P*-value	Reference
2	127 125 963–127 172 772	Marbling score	126	Multi-breed beef		Significant	MacNeil & Grosz (2002)
		Fat yield	115–130	German Holstein		0.01	Harder *et al.* (2006)
		Birth weight	115–128	Multi-breed beef		0.014	Grosz and MacNeil (2001)
2	134 666 758–134 761 842						
10	13 146 429–13 225 603	Milk yield	11–20	German Holstein	2.3		Thomsen *et al.* (2000)
		Protein percent	19	Israel Holstein		0.01	Mosig *et al.* (2001)
		Carcass weight	0–30	Multi-breed beef	12.0		Casas *et al.* (2003)
		Marbling score	0–28	Multi-breed beef	11.0		Casas *et al.* (2003)
10	38 264 640–38 625 718						
10	[48 942 782–51 138 335]	Non-return rate	48	German Holstein		Significant	Kühn *et al.* (2003)
		SCC	49	German Holstein		0.02	Kühn *et al.* (2003)
		Body depth	46	USA Holstein	3.06		Ashwell *et al.* (2005)
10	57 638 141–57 773 467	Protein Percent	55	Israel Holstein		0.02	Mosig *et al.* (2001)
10	70 455 224–70 552 188	Carcass trait	60–79	Multi-breed beef		Suggestive	MacNeil & Grosz (2002)
		Udder depth	68.1	Holstein		0.02	Boichard *et al.* (2003)
		Calving ease	73.1	US Holstein	21.78		Schnabel *et al.* (2005)
		Milk yield	69.0	Canadian Holstein		0.05	Plante *et al.* (2001)
		SCC	73.9	US Holstein	13.18		Schnabel *et al.* (2005)
		Teat placement	68.1	Holstein		0.02	Boichard *et al.* (2003)
		Udder cleft	68.1	Holstein		0.02	Boichard *et al.* (2003)
		Teat length	68.1	US Holstein	14.03		Schnabel *et al.* (2005)
		Protein percent	73	Israel Holstein		0.03	Mosig *et al.* (2001)
12	14 556 717–14 658 840	Milk yield	21	Finnish Ayrshire		0.006	Viitala *et al.* (2003)
		Protein yield	21	Finnish Ayrshire		0.02	Viitala *et al.* (2003)
		Protein percent	21	Israel Holstein		0.01	Mosig *et al.* (2001)
13	5 082 478–5 148 264	Dairy form	0–9	USA Holstein	2.82		Ashwell *et al.* (2005)
16	70 812 261–71 003 946	Hot carcass weight	54–77	Wagyu × Limousin		Significant	Alexander *et al.* (2007)
		Udder depth	61–72	USA Holstein	3.28		Ashwell *et al.* (2005)
17	13 973 226–14 208 603	Rump angle	0–30	Holstein, Normande and Montbeliarde		0.005	Boichard *et al.* (2003)

Table 7 Distribution of the GGCG haplotype under selection vs. *DGAT1* genotypes A and L, which represent the alanine and lysine variants respectively.

	GGCG/GGCG	GGCG/–	–/–
LL	9	0	0
AL	0	66	1
AA	0	3	67

sample size and increased SNP density, comparative studies of geographically separated populations with identical or diverse breeding goals and a detailed functional characterization of the candidate regions identified to be under recent directed selection.

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., Jin L. & Shriver M.D. (2002) Interrogating a high-density SNP map for signatures of natural selection. *Genome Research* **12,** 1805–14.

Akey J.M., Eberle M.A., Rieder M.J., Carlson C.S., Shriver M.D., Nickerson D.A. & Kruglyak L. (2004) Population history and natural selection shape patterns of genetic variation in 132 genes. *PLoS Biology* **2,** e286.

Alexander L.J., Snelling W.M. & Macneil M.D. (2007) Quantitative trait loci with additive effects on growth and carcass traits in a Wagyu & Limousin F_2 population. *Animal Genetics* **38,** 413–6.

Arosh J.A., Banu S.K., Chapdelaine P., Emond V., Kim J.J. *et al.* (2003) Molecular cloning and characterization of bovine prostaglandin E2 receptors *EP2* and *EP4*: expression and regulation in endometrium and myometrium during the estrous cycle and early pregnancy. *Endocrinology* **144,** 3076–91.

Ashwell M.S., Heyen D.W., Weller J.I., Ron M., Sonstegard T.S., Van Tassell C.P. & Lewin H.A. (2005) Detection of quantitative trait loci influencing conformation traits and calving ease in Holstein–Friesian cattle. *Journal of Dairy Science* **88,** 4111–9.

Barendse W., Harrison B., Bunch R.J. & Thomas M.B. (2008) Variation at the *Calpain 3* gene is associated with meat tenderness in zebu and composite breeds of cattle. *BMC Genetics* **9,** 41.

Barendse W., Harrison B., Bunch R.J., Thomas M.B. & Turner L.B. (2009) Genome wide signatures of positive selection: the comparison of independent samples and the identification of regions associated to traits. *BMC Genomics* **10,** 178.

Barrett J.C., Fry B., Maller J. & Daly M.J. (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21,** 263–5.

Bionaz M. & Loor J. (2008) Gene networks driving bovine milk fat synthesis during the lactation cycle. *BMC Genomics* **9,** 366.

Biswas S. & Akey J.M. (2006) Genomic insights into positive selection. *Trends in Genetics* **22,** 437–46.

Boichard D., Grohs C., Bourgeois F., Cerqueira F., Faugeras R. *et al.* (2003) Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution* **35,** 77–101.

Bradley D.G. & Cunningham E.P. (1998) Genetic aspects of domestication. In: *The Genetics of Cattle* (Ed. by R. Fries & A. Ruvinski), pp. 15–32. CAB International, Oxon, UK.

Casas E., Shackelford S.D., Keele J.W., Koohmaraie M., Smith T.P. & Stone R.T. (2003) Detection of quantitative trait loci for growth and carcass composition in cattle. *Journal of Animal Science* **81,** 2976–83.

Daetwyler H.D., Schenkel F.S., Sargolzaei M. & Robinson J.A.B. (2008) A genome scan to detect quantitative trait loci for economically important traits in Holstein cattle using two methods and a dense single nucleotide polymorphism map. *Journal of Dairy Science* **91,** 3225–36.

Fatehi A.N., Hurk R.V.D., Colenbrander B., Daemen A., Van Tol H. *et al.* (2005) Expression of bone morphogenetic protein2 (BMP2), BMP4 and BMP receptors in the bovine ovary but absence of effects of BMP2 and BMP4 during IVM on bovine oocyte nuclear maturation and subsequent embryo development. *Theriogenology* **63,** 872–89.

Fay J.C. & Wu C.I. (2000) Hitchhiking under positive Darwinian selection. *Genetics* **155,** 1405–13.

Gabriel S.B., Schaffner S.F., Nguyen H. *et al.* (2002) The structure of haplotype blocks in the human genome. *Science* **296,** 2225–9.

Grisart B., Coppieters W., Farnir F. *et al.* (2001) Positional candidate cloning of a QTL in dairy cattle: identification of a missense mutation in the bovine *DGAT1* gene with major effect on milk yield and composition. *Genome Research* **12,** 222–31.

Grosz M.D. & MacNeil M.D. (2001) Putative quantitative trait locus affecting birth weight on bovine chromosome 2. *J Animal Science* **79,** 68–72.

Harder B., Bennewitz J., Reinsch N., Thaller G., Thomsen H. *et al.* (2006) Mapping of quantitative trait loci for lactation persistency traits in German Holstein dairy cattle. *Journal of Animal Breeding & Genetics* **123,** 89–96.

Hayes B.J., Lien S., Nilsen H., Olsen H.G., Berg P. *et al.* (2008) The origin of selection signatures on bovine chromosome 6. *Animal Genetics* **39,** 105–11.

Hayes B.J., Chamberlain A.J.., Maceachern S., Savin K., Mcpartlan H. *et al.* (2009) A genome map of divergent artificial selection between *Bos taurus* dairy cattle and *Bos taurus* beef cattle. *Animal Genetics* **40,** 176–84.

Innan H. & Kim Y. (2004) Pattern of polymorphism after strong artificial selection in a domestication event. *Proceedings of the National Academy of Sciences of the United States of America* **101,** 10667–72.

Jiang Z., Michal J.J., Kunej T. & Macneil M.D. (2008) Associations of nucleus encoded mitochondrial genes with carcass, meat and eating quality in beef cattle. *Proceeding of Plant & Animal Genomes XVI Conference*, January 12–16, San Diego, CA.

Khait V.D., Huang Y., Zalsman G., Oquendo M.A., Brent D.A. *et al.* (2005) Association of serotonin 5-HT2A receptor binding and the T102C polymorphism in depressed and healthy Caucasian subjects. *Neuropsychopharmacology* **30,** 166–72.

Kimura M. (1983) *The Neutral Theory of Molecular Evolution*. Cambridge University Press, Cambridge, MA.

doi:10.1111/j.1365-2052.2009.02016.x

Kühn Ch., Bennewitz J., Reinsch N. *et al.* (2003) Quantitative trait loci mapping of functional traits in the German Holstein cattle population. *Journal of Dairy Science* **86,** 360–8.

Li C., Basarab J., Snelling W.M., Benkel B., Kneeland J., Murdoch B., Hansen C. & Moore S.S. (2004) Identification and fine mapping of quantitative trait loci for backfat on bovine chromosomes 2, 5, 6, 19, 21 and 23 in a commercial line of *Bos taurus*. *Journal of Animal Science* **82,** 967–72.

Lichten M. & Goldman A.S.H. (1995) Meiotic recombination hotspots. *Annual Review of Genetics* **29,** 423–44.

Luvalle P., Ma Q. & Beier F. (2003) The Role of Activating Transcription Factor-2 in Skeletal Growth. *The Journal of Bone and Joint Surgery* **85,** 133–6.

MacEachern S., Hayes B., McEwan J. & Goddard M. (2009) An examination of positive selection and changing effective population size in Angus and Holstein cattle populations (*Bos taurus*) using a high density SNP genotyping platform and the contribution of ancient polymorphism to genomic diversity in domestic cattle. *BMC Genomics* **10,** 181.

MacNeil M.D. & Grosz M.D. (2002) Genome-wide scans for QTL affecting carcass traits in Hereford × composite double backcross populations. *Journal of Animal Science* **80,** 2316–24.

Marques E., Schnabel R., Stothard P., Kolbehdari D., Wang Z. *et al.* (2008) High density linkage disequilibrium maps of chromosome 14 in Holstein and Angus cattle. *BMC Genetics* **9,** 45.

Matukumalli L.K., Lawley C.T., Schnabel R.D., Taylor J.F., Allan M.F. *et al.* (2009) Development and characterization of a high density SNP genotyping assay for cattle. *PLoS ONE* **4,** e5350.

Maynard-Smith J. & Haigh J. (1974) The hitch-hiking effect from a favourable gene. *Genetical Research* **23,** 23–35.

Meuwissen T.H.E., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157,** 1819–29.

Miller S.A., Dykes D.D. & Polesky H.F. (1988) A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16,** 1215.

Mosig M.O., Lipkin E., Khutoreskaya G., Tchourzyna E., Soller M. & Friedmann A. (2001) A whole-genome scan for quantitative trait loci affecting milk protein percentage in Israeli-Holstein cattle, by means of selective milk DNA pooling in a daughter design, using an adjusted false discovery rate criterion. *Genetics* **157,** 1683–98.

Nielsen R. (2005) Molecular signatures of natural selection. *Annual Review of Genetics* **39,** 197–218.

Nkrumah J.D., Sherman E.L., Li C., Marques E., Crews D.H. Jr, Bartusiak R., Murdoch B., Wang Z., Basarab J.A. & Moore S.S. (2007) Primary genome scan to identify putative QTL for feedlot growth rate, feed intake and feed efficiency of beef cattle. *Journal of Animal Science* **85,** 3170–81.

Peremans K., Audenaert K., Coopman F., Blanckaert P., Jacobs F. *et al.* (2003) Estimates of regional cerebral blood flow and 5-HT2A receptor density in impulsive, aggressive dogs with 99mTc-ECD and 123I-5-I-R91150. *European Journal of Nuclear Medicine and Molecular Imaging* **30,** 1538–46.

Petes T.D. (2001) Meiotic recombination hot spots and cold spots. *Nature Review Genetics* **2,** 360–9.

Plante Y., Gibson J.P., Nadesalingam J., Mehrabani-Yeganeh H., Lefebvre S. *et al.* (2001) Detection of quantitative trait loci affecting milk production traits on 10 chromosomes in Holstein cattle. *Journal of Dairy Science* **84,** 1516–24.

Pollinger J.P., Bustamante C.D., Fledel-Alon A., Schmutz S., Gray M.M. & Wayne R.K. (2005) Selective sweep mapping of genes with large phenotypic effects. *Genome Research* **15,** 1809–19.

Prasad A., Mckay S.D., Murdoch B., Stothard P., Kolbehdari D. *et al.* (2008) Linkage disequilibrium and signatures of selection on chromosomes 19 and 29 in beef and dairy cattle. *Animal Genetics* **39,** 597–605.

Reist M., Pfaffl M.W., Morel C., Meylan M., Hirsbrunner G. *et al.* (2003) Quantitative mRNA analysis of eight bovine 5-HT receptor subtypes in brain, abomasum, and intestine by real-time RT-PCR. *Journal of Receptor Signal Transduction* **23,** 271–87.

Sabeti P.C., Reich D.E., Higgins J.M. *et al.* (2002) Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419,** 832–7.

Scheet P. & Stephens M. (2006) A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *American Journal of Human Genetics* **78,** 629–44.

Schnabel R.D., Taylor J.F. & Ashwell M.S. (2005) Whole-genome scan to detect QTL for milk production, conformation, fertility and functional traits in two US Holstein families. *Animal Genetics* **36,** 408–16.

Simianer H., Szyda J., Ramon G. & Lien S. (1997) Evidence for individual and between family variability of the recombination rate. *Mammalian Genome* **8,** 830–5.

Skjervold H. & Langholz H.J. (1964) Factors affecting the optimum structure of A.I. breeding in dairy cattle. *Zeitschrift Tierzüchtung und Züechtungsbiologie* **80,** 26–40.

Stone R.T., Keele J.W., Shackelford S.D., Kappes S.M. & Koohmaraie M. (1999) A primary screen of the bovine genome for quantitative trait loci affecting carcass and growth traits. *Journal of Animal Science* **77,** 1379–84.

Tajima F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123,** 585–95.

Tang K., Thornton K.R. & Stoneking M. (2007) A new approach for using genome scans to detect recent positive selection in the human genome. *PLoS Biology* **5,** e171.

The Bovine HapMap consortium (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324,** 528–32.

Thomsen H., Xu N.R.N., Looft C., Grupe S., Kühn C. *et al.* (2000) A male bovine linkage map for the ADR granddaughter design. *Journal of Animal Breeding & Genetics* **117,** 289–306.

Viitala S.M., Schulman N.F., De Koning D.J., Elo K., Kinos R. *et al.* (2003) Quantitative trait loci affecting milk production traits in Finnish Ayrshire dairy cattle. *Journal of Dairy Science* **86,** 1828–36.

Voight B.F., Kudaravalli S., Wen X. & Pritchard J.K. (2006) A map of recent positive selection in the human genome. *PLoS Biology* **4,** e72.

Walsh E., Sabeti P., Hutcheson H.B., Fry B., Schaffner S.F. *et al.* (2006) Searching for signals of evolutionary selection in 168 genes related to immune function. *Human Genetics* **119,** 92–102.

Yu F., Sabeti P.C., Hardenbol P., Fu Q., Fry B. *et al.* (2005) Positive selection of a pre-expansion CAG repeat of the human *SCA2* gene. *PLoS Genetics* **1,** e41.

Zhang C., Bailey D.K., Awad T., Liu G., Xing G. *et al.* (2006) A whole genome long-range haplotype (WGLRH) test for detecting imprints of positive selection in human populations. *Bioinformatics* **22,** 2122–8.

doi:10.1111/j.1365-2052.2009.02016.x