ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

doi: 10.1111/age.12210

A genome-wide scan for selection signatures in Nellore cattle

A. L. Somavilla*, T. S. Sonstegard†, R. H. Higa‡, A. N. Rosa§, F. Siqueira§, L. O. C. Silva§, R. A. A. Torres Júnior§, L. L. Coutinho¶††, M. A. Mudadu, M. M. Alencar**†† and L. C. A. Regitano**††**

*Programa de Pós-Graduação em Genética e Melhoramento Animal, UNESP/FCAV, Jaboticabal, Brasil. †Bovine Functional Genomics Laboratory, ANRI, USDA-ARS, Beltsville, MD, USA. ‡Embrapa Informática Agropecuária, Campinas, Brasil. §Embrapa Gado de Corte, Campo Grande, Brasil. ¶Escola Superior de Agricultura Luiz de Queiroz (Esalq), USP, Piracicaba, Brasil. **Embrapa Pecuária Sudeste, São Carlos, Brasil. ††CNPq Fellow, Brasília, Distrito Federal, Brasil.

Summary

Brazilian Nellore cattle (*Bos indicus*) have been selected for growth traits for over more than four decades. In recent years, reproductive and meat quality traits have become more important because of increasing consumption, exports and consumer demand. The identification of genome regions altered by artificial selection can potentially permit a better understanding of the biology of specific phenotypes that are useful for the development of tools designed to increase selection efficiency. Therefore, the aims of this study were to detect evidence of recent selection signatures in Nellore cattle using extended haplotype homozygosity methodology and BovineHD marker genotypes (>777 000 single nucleotide polymorphisms) as well as to identify corresponding genes underlying these signals. Thirty-one significant regions ($P < 0.0001$) of possible recent selection signatures were detected, and 19 of these overlapped quantitative trait loci related to reproductive traits, growth, feed efficiency, meat quality, fatty acid profiles and immunity. In addition, 545 genes were identified in regions harboring selection signatures. Within this group, 58 genes were associated with growth, muscle and adipose tissue metabolism, reproductive traits or the immune system. Using relative extended haplotype homozygosity to analyze high-density single nucleotide polymorphism marker data allowed for the identification of regions potentially under artificial selection pressure in the Nellore genome, which might be used to better understand autozygosity and the effects of selection on the Nellore genome.

Keywords beef cattle, *Bos indicus*, genotyping, linkage disequilibrium, relative extended haplotype homozygosity, single nucleotide polymorphisms

Introduction

The Brazilian beef herd consists mainly of Nellore cattle (*Bos indicus*) and has been selected primarily for improved growth and fertility for the past 50 years. The genetic gain over these generations has been reported relative to genetic trends for body weight and weight gain (Laureano *et al.* 2011). However, the increase in meat consumption and exports and increasing consumer demand an even more rapid increase in productivity, fertility and product quality. Such a broad spectrum of specific phenotypic traits requires a better understanding of the Nellore genome at a population level to develop breeding strategies and tools that rapidly accelerate selection for improvement (Stella *et al.* 2010).

Typically, intensive artificial selection favors desired phenotypes in a population over a short period of time, when compared to natural selection, which has the effect of maintaining gene variations related to the selected traits through generations (Innan & Kim 2008). Among the methods available in the literature for the detection of selection signatures, extended haplotype homozygosity (EHH; Sabeti *et al.* 2002) has been used previously for the analysis of a single genetic group in the absence of information about ancestral alleles (Qanbari *et al.* 2010). The relative extended haplotype homozygosity (REHH) method was derived from the EHH method to correct for local variation in recombination rates and is determined by comparing the EHH of the tested core haplotype to the relative homozygosity on the chromosome (Sabeti *et al.*

Address for correspondence

L. C. A. Regitano, Embrapa Pecuária Sudeste, Rodovia Washington Luiz Km 234 SN SAO CARLOS, São Paulo, Brasil.
E-mail: luciana.regitano@embrapa.br

Accepted for publication 15 July 2014

2002). The aim of this study was to detect evidence for signatures of recent selection in the Nellore cattle genome. These results can be useful to better understand autozygosity and regions under selection in Nellore cattle and might shed some light on regions or groups of genes that control production traits.

Materials and methods

Samples

Data from 762 steers, descendants of 34 registered Nellore sires, were used. These sires belong to the lineages most frequently used in Brazil, according to the National Summary of Nellore by the Brazilian Association of Zebu Breeders (ABCZ) and the National Research Center for Beef Cattle. Sires were chosen to minimize the degree of kinship among them and to represent the genetic variability of the breed. Preliminary analysis using principal components did not provide evidence of population structure (data not shown). Estimated breeding values for economically important traits were not considered. Random Nellore dams from five different farms were inseminated by fixed-time artificial insemination, and the steers were born in 2007, 2008 and 2009.

Genotyping and quality control

DNA of sires was extracted from frozen semen acquired from artificial insemination centers, and DNA of steers was extracted from blood. Steer rearing and sample collection protocols were approved by Animal Care and Use Committee from Embrapa Southeast Livestock. Genotyping was carried out using the Illumina BovineHD BeadChip containing 777 962 single nucleotide polymorphisms (SNPs). After paternity corrections, only autosomal SNPs with corresponding UMD 3.1 (Zimin *et al.* 2009) genome coordinates and a GC score ≥0.2 (score related to genotyping quality) were considered for analyses. Further quality control filtering removed SNPs with a minor allele frequency (MAF) <1% and a call rate <95% per SNP. Individuals with an overall call rate <90% were also removed.

Haplotype reconstruction

A haplotype is a sequence of alleles at different loci that are on the same chromosome and can be phased (i.e., determining the non-allelic combinations more likely to be transmitted in a population, as they tend to be inherited together) using statistical methods for linkage disequilibrium (LD) such that the corresponding length and frequency can be inferred.

To perform this reconstruction, an algorithm developed for related individuals available in ALPHAPHASE 1.1 software was chosen (Hickey *et al.* 2011). Briefly, ALPHAPHASE 1.1 is a combination of two other important methods: (i) 'long-range phasing', which can infer the SNP phase (or combinations) based on other haplotypes shared by related animals (surrogate parents) and (ii) 'haplotype library imputation', which is based on the database and can increase the error correction efficiency.

The following analysis parameters were used to generate *de novo* haplotypes of varying length and were based on phasing performance of ALPHAPHASE 1.1 (Hickey *et al.* 2011): cores of 100 SNPs (consecutive chains of SNPs for which phasing is being attempted), tails of 100 SNPs (consecutive strings of SNPs immediately adjacent to either end of a core), 10 individuals analyzed before inferring a common haplotype phase, maximum disagreement of 10% among the individuals analyzed and maximum percentage of missing or incorrect genotyped SNPs of 1%.

REHH analysis

Relative extended haplotype homozygosity is an extension of the EHH method. The aim of both REHH and EHH is to detect the transmission of an extended haplotype occurring without recombination. Both methodologies assess the probability that two randomly chosen chromosomes carry the same tested core haplotypes by being homozygous for all SNPs in the interval (Sabeti *et al.* 2002). The EHH of core haplotype t is defined by:

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

where s is the number of unique extended haplotypes, e is the number of samples of a particular extended haplotype and c is the number of samples of a particular core haplotype.

When the EHH test was first designed, there were no good estimates for different recombination rates, as variation in recombination may occur across the bovine genome (McKay *et al.* 2007). To correct for local variation in recombination, REHH was determined by comparing the EHH of the tested core haplotype to that of other core haplotypes present at the locus (Sabeti *et al.* 2002). REHH is calculated as:

$$\mathrm{REHH}_t = \frac{\mathrm{EHH}_t}{\overline{\mathrm{EHH}}},$$

where $\overline{\mathrm{EHH}}$ (relative homozygosity on the chromosome) is defined as the decay of EHH on all other core haplotypes combined and is computed by the following equation:

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

where n is the number of different core haplotypes.

To check the adopted methodology for agreement with a selection hypothesis, in a first inspection only specific high-priority genomic regions containing known genes related to growth, fat deposition, reproduction and immunity were considered for REHH determination and comparison (Table 1). HAPLOVIEW v4.1 (Barrett *et al.* 2005) was used

Table 1 Description of candidate genes used in applying EHH methodology.

Chr	Gene symbol	Gene name	Trait	Reference
1	*SST*	*Somatostatin*	Regulation of growth and development	Gao *et al.* (2011)
2	*MSTN*	*Myostatin*	Extended muscular development	Wiener *et al.* (2009)
3	*LEPR*	*Leptin receptor*	Growth and fat deposition	Guo *et al.* (2008)
4	*LEP*	*Leptin*	Growth and fat deposition	Guo *et al.* (2008)
5	*IGF1*	*Insulin-like growth factor 1 (somatomedin C)*	Growth and puberty	Rogberg-Muñoz *et al.* (2011) and Liron *et al.* (2012)
7	*IL4*	*Interleukin 4*	Inflammatory response	Woods & Judd (2008)
14	*FABP4*	*Fatty acid binding protein 4, adipocyte*	Intramuscular and subcutaneous fat deposition	Michal *et al.* (2006)
17	*IL2*	*Interleukin 2*	Tick resistance	Piper *et al.* (2009)
19	*GH1*	*Growth hormone 1*	Growth	Silveira *et al.* (2008)
20	*GHR*	*Growth hormone receptor*	Intramuscular and subcutaneous fat deposition	Baeza *et al.* (2011)
21	*IGF1R*	*Insulin-like growth factor 1 receptor*	Growth and puberty	Micke *et al.* (2011) and Fortes *et al.* (2012)
23	*BOLA3*	*BolA homolog 3 (E. coli)*	Tick resistance	Martinez *et al.* (2006)

Table 2 Description of single nucleotide polymorphisms (SNPs) present on the Illumina BovineHD BeadChip and haplotype blocks distributed across the genome of Nellore cattle.

Chr	SNP (*n*)	Mean distance between SNPs (kb)	Blocks (*n*)	Mean block length (kb)	Block coverage length (kb)	Maximum block length (kb)	Block length/Chr length (%)	SNPs in blocks (*n*)	SNPs in blocks/SNPs in Chr
1	36 565	4.33	4176	15.58	65082.24	177.97	0.41	20 820	0.57
2	29 945	4.57	3713	16.49	61214.78	230.56	0.45	18 251	0.61
3	27 248	4.45	3388	15.76	53378.45	251.52	0.44	16 513	0.61
4	26 618	4.53	3301	16.31	53831.80	166.13	0.45	16 208	0.61
5	25 016	4.84	2987	17.98	53699.22	395.44	0.44	15 197	0.61
6	28 227	4.23	3492	14.67	51212.96	141.13	0.43	17 008	0.60
7	25 487	4.42	3033	15.87	48128.71	303.66	0.43	14 944	0.59
8	26 097	4.34	3149	14.80	46618.14	200.06	0.41	15 213	0.58
9	24 699	4.28	2997	15.14	45366.80	150.17	0.43	14 618	0.59
10	22 346	4.67	2747	16.61	45625.75	241.41	0.44	13 420	0.60
11	24 379	4.40	3019	16.96	51208.60	260.11	0.48	15 324	0.63
12	18 740	4.86	2260	16.76	37887.46	204.71	0.42	11 269	0.60
13	17 329	4.86	2032	17.87	36312.63	245.46	0.43	10 072	0.58
14	20 514	4.09	2454	14.09	34564.78	165.81	0.41	11 708	0.57
15	19 445	4.38	2422	15.50	37535.50	281.49	0.44	11 587	0.60
16	18 498	4.42	2297	15.98	36710.29	328.50	0.45	11 221	0.61
17	16 914	4.44	2060	15.34	31610.69	162.47	0.42	9877	0.58
18	14 922	4.42	1750	16.41	28710.78	187.45	0.43	8666	0.58
19	13 478	4.75	1623	16.45	26692.06	177.48	0.42	7646	0.57
20	16 271	4.42	2025	14.51	29390.53	155.30	0.41	9497	0.58
21	16 679	4.29	2069	15.83	32745.16	424.92	0.46	10 193	0.61
22	13 280	4.61	1613	16.21	26145.55	360.33	0.43	7679	0.58
23	11 889	4.41	1377	14.76	20320.72	149.44	0.39	6296	0.53
24	14 055	4.45	1730	15.27	26421.10	175.21	0.42	8289	0.59
25	8965	4.77	1065	15.90	16928.95	187.33	0.39	4985	0.56
26	11 917	4.33	1489	14.53	21639.93	351.96	0.42	6793	0.57
27	9929	4.57	1197	13.39	16028.08	491.27	0.35	5210	0.52
28	10 293	4.49	1278	14.41	18419.95	255.63	0.40	5706	0.55
29	10 820	4.75	1311	15.57	20417.81	239.96	0.40	6052	0.56
Total	560 565		68 054		1073849.40			330 262	
Mean		4.50		15.69			0.42		0.58

to estimate pairwise LD, quantified by the square of the coefficient of correlation (r^2), which indicates the amount of information of one locus, that is, explained by another (Ardlie *et al.* 2002). SWEEP v1.1 was used to estimate haplotype blocks and to identify selection signatures (see below). To investigate the EHH for the high-priority genomic regions, the nearest SNP available on the BovineHD chip, relative to the candidate gene, was used to find the nearest haplotype block.

Extended haplotype homozygosity and REHH statistics were calculated using SWEEP 1.1 on haplotypes generated by ALPHAPHASE, which comprised different blocks when at least 95% of the SNP pairs were in strong LD, that is, exhibited little evidence of recombination (Gabriel *et al.* 2002). Tests were applied at a distance of 0.25 cM (about 250 kb) on both sides of the haplotypes. To determine the REHH significance, SWEEP 1.1 was used with default parameters, and the following steps were performed. Haplotypes were parsed into 20 frequency bins (e.g., <5%, 5 to <10%, until ≥95%) and compared with other haplotypes equally frequent on the chromosome. REHH values were transformed into a logarithmic scale ($-\log_{10}$) to obtain an approximate normal distribution. After that, probability distribution was estimated, as were the corresponding means and standard deviations. *P*-values were obtained by SWEEP 1.1 by the number of standard deviations in a normal distribution.

Comparison of significant REHH regions to genes and QTL regions

All REHH regions of significance ($P < 0.001$) were used, and these search areas were extended by 250 kb on both sides of the significant haplotypes. This distance is shorter than that used for Holstein cattle (Qanbari *et al.* 2010) because of the lower extent of LD observed in this Nellore population (data not shown). GBROWSE 1.71 tool loaded with the UMD 3.1 (Zimin *et al.* 2009) bovine genome was used to manually identify genes. Selection signatures that overlapped QTL related to economically important traits in beef cattle available in CattleQTLdb database (Hu & Reecy 2007; release 17, April 20, 2012) were also investigated. Putative selection signatures regions that were entirely within the QTL were considered.

Database for Annotation, Visualization and Integrated Discovery (DAVID; Dennis *et al.* 2003) was used to retrieve functional classification (GO annotation) for the set of identified genes.

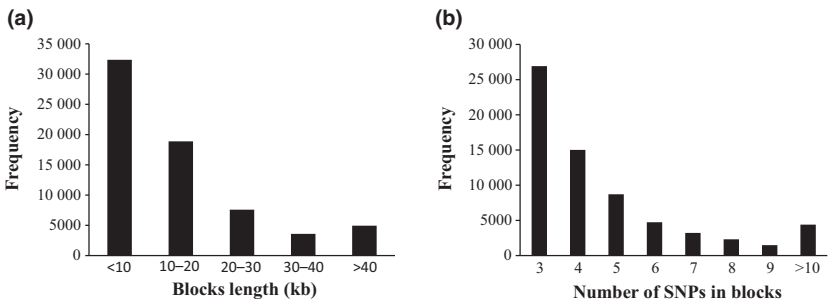

Figure 1 Distribution of haplotype blocks length (a) and number of single nucleotide polymorphisms forming blocks (b) in the Nellore cattle genome.

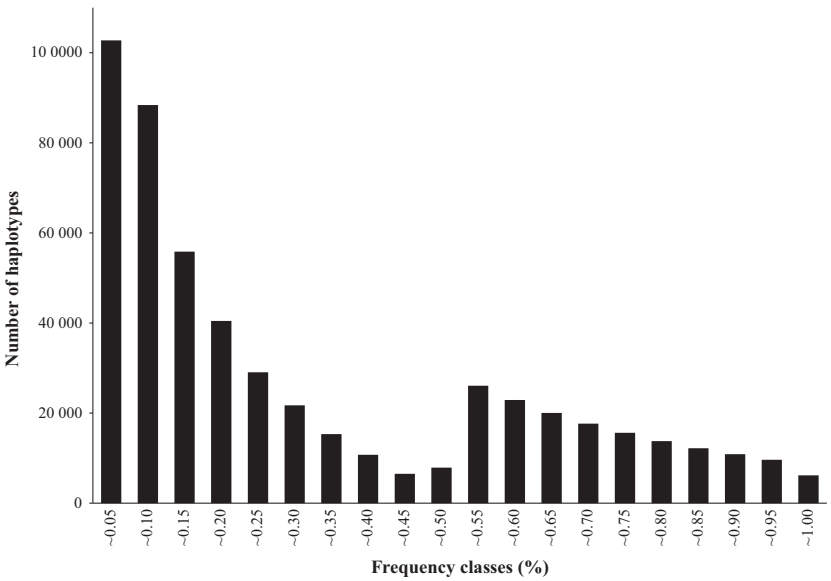

Figure 2 Distribution of haplotypes in different frequency classes in Nellore cattle.

Table 3 Extended haplotype homozygosity analysis in candidate genes.

Gene symbol	Chr	Gene position (bp)	Closest single nucleotide polymorphism name & position (bp)	Block position (bp)	Haplotype frequency (%)[a]	EHH[b]		REHH[b]		REHH[b] *P*-value	
SST	1	80 250 205–80 251 648	*rs134230636* (80272087)	80 272 087–80 300 957	*H1*: 67	0.13	0.11	12.69	3.93	**0.001**	**0.09**
					H2: 25	0.01	0.07	0.07	0.24	1.00	0.89
MSTN	2	6 213 566–6 220 196	*rs109995530* (6225434)	6 225 434–6 228 548	*H1*: 50	0.17[c]	–	4.33[c]	–	0.12[c]	–
					H2: 47	0.04[c]	–	0.23[c]	–	0.96[c]	–
LEPR	3	80 071 689–80 147 000	*rs137065305* (80107536)	80 107 536–80 146 148	*H1*: 56	0.16	0.14	7.24	3.21	**0.02**	0.28
					H2: 37	0.02	0.04	0.14	0.31	0.94	0.63
LEP	4	93 249 874–93 266 624	*rs109414345* (93250208)	93 245 659–93 257 549	*H1*: 69	0.11	0.10	1.67	3.86	0.57	0.12
					H2: 18	0.07	0.03	0.67	0.25	0.38	0.87
IGF1	5	66 532 879–66 604 699	*rs133523553* (66536500)	66 529 557–66 566 541	*H1*: 63	0.07	0.10	0.96	0.95	0.73	0.90
					H2: 18	0.07	0.08	1.04	0.76	0.20	0.29
IL4	7	22 993 178–23 001 067	*rs42522202* (23004080)	23 004 080–23 016 916	*H1*: 77	0.06	0.07	1.24	0.89	0.63	0.82
					H2: 26	0.05	0.08	0.81	1.13	0.24	0.11
FABP4	14	46 833 665–46 838 053	*rs137697868* (46823262)	46 821 487–46 823 262	*H1*: 86	0.07	0.07	0.74	3.16	0.65	**0.03**
					H2: 14	0.10	0.02	1.35	0.32	0.14	0.86
IL2	17	35 618 115–35 622 861	*rs137281525* (35619076)	35 583 312–35 637 592	*H1*: 94	0.05	0.05	0.49	0.49	0.87	0.87
					H2: 03	0.07	0.09	1.47	1.67	0.27	0.22
GH1	19	48 768 618–48 772 014	*rs135233632* (48772161)	48 765 934–48 772 161	*H1*: 72	0.09	0.09	1.12	0.55	0.77	0.97
					H2: 13	0.10	0.24	1.10	2.78	0.15	**0.01**
GHR	20	31 890 736–32 064 200	*rs109136815* (31891078)	31 889 529–31 894 246	*H1*: 71	0.09	0.09	1.21	1.08	0.69	0.75
					H2: 28	0.07	0.08	0.83	0.92	0.20	0.15
IGF1R	21	7 967 724–8 268 932	*rs137322287* (7962990)	7 951 484–7 962 990	*H1*: 60	0.11	0.13	1.27	1.59	0.76	0.63
					H2: 33	0.09	0.08	0.78	0.62	0.18	0.29
BOLA3	23	25 472 161–25 476 885	*rs108963203* (25475120)	25 475 120–25 481 868	*H1*: 78	0.08	0.09	2.43	0.37	0.29	0.28
					H2: 12	0.03	0.03	2.45	0.32	0.74	0.80

[a]Tests were performed on the two most frequent haplotypes of each block (H1 and H2).
[b]Extended haplotype homozygosity (EHH), relative extended haplotype homozygosity (REHH) and *P*-values were calculated for both sides of the core haplotype.
[c]The haplotype could only be extended to the left.
Bold type indicates statistical significance at the 10% level.

© 2014 Stichting International Foundation for Animal Genetics, **45**, 771–781

Results

Markers and haplotype blocks

Using the BovineHD assay results for 34 sires and 755 steers (789 animals), a core set of 560 565 SNPs covering 2522.54 Mbp of the UMD 3.1 (Zimin *et al.* 2009) genome assembly was used for selection signature analysis after quality control filtering of the SNPs. The average pairwise space between consecutive SNPs was 4.5 kb, whereas the mean MAF (±SE) was 0.26 (±0.0002), a result similar to that reported for Nellore (0.20; The Bovine HapMap Consortium 2009). A total of 68 054 haplotype blocks were generated to cover 1.07 Mbp (42%) of the 29 autosomes. Mean length (±SD) of these haplotype blocks was 15.69 kb (±17.94 kb) with a relatively large standard deviation caused by the wide variation in block length ranging from 0.65 to 491.27 kb (Table 2).

More than 50 000 blocks had a maximum length of 20 kb, and more than 40 000 blocks were formed by up to four SNPs (Fig. 1). Estimated LD for blocks ranging between 10 and 25 kb was $r^2 = 0.28$ (data not shown). Low-frequency haplotypes (up to 10%) were the most common (Fig. 2).

REHH analysis

Previous studies in cattle have shown an association between candidate genes and traits of commercial interest (Table 1). To assess whether these genes underwent recent selection, we investigated the presence of selection signatures in regions that contain some of these genes (Table 3). Recent selection signatures were detected for *growth hormone 1* (*GH1*), *leptin receptor* (*LEPR*), *somatostatin* (*SST*) and *fatty acid binding protein 4, adipocyte* (*FABP4*) genes (Table 3), whereas no evidence was observed for immune-related genes (*IL2*, *IL4* and *BOLA3*). Homozygosity around candidate genes did not extend beyond 0.2 cM (Fig. 3), suggesting that the LD in the Nellore breed is lower than that reported for Holstein cattle (Qanbari *et al.* 2010). The block length around candidate genes ranged from 1 to 38 kb, also suggesting a low LD extent (Fig. 4).

Only haplotypes with a frequency higher than 30% were included in the subsequent steps to rule out possible events of recombination. A total of 193 944 tests were performed, and more than 200 possible regions harboring recent potential selection signatures were identified ($P < 0.001$; Table 4 and Table S1). Because it is not a simple process to directly compare results that used different genome assembly coordinates to derive genome position, a manual inspection suggested there were 24 similar significant regions of variation in allele frequency among this Nellore population and different populations of dairy and beef cattle (Qanbari *et al.* 2011).

To visualize the selection signatures distribution, the transformed *P*-values of REHH were plotted against the haplotype position on each chromosome (Fig. 5). Only 31 regions ($P < 0.0001$) were used to investigate overlapping QTL related to reproductive traits, growth, efficiency, meat quality, fatty acids profile and immunity (Table 4). Among all regions, 19 contain QTL described previously (Table S2). In addition, 545 functional candidate genes were identified in regions harboring recent selection signatures.

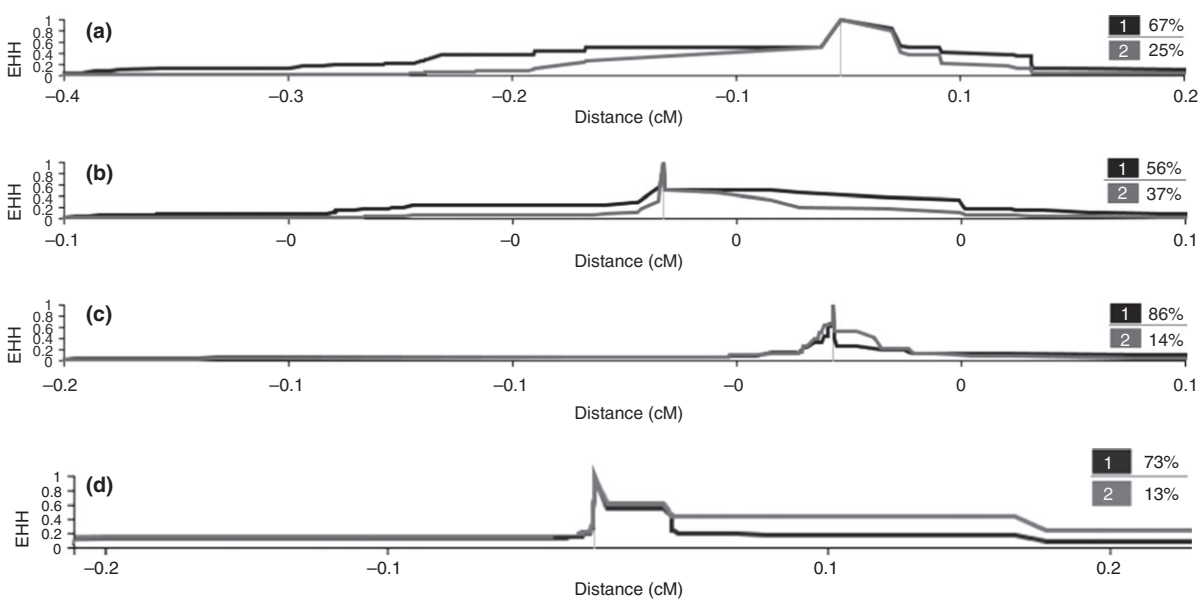

Figure 3 Decay of extended haplotype homozygosity for the *SST* (a), *LEPR* (b), *FABP4* (c) and *GH1* (d) genes. The two most frequent haplotypes (1 and 2) of each block are shown.

Discussion

The choice to investigate effects of selection in Nellore cattle was based on the importance of zebu breeds for being the main source of beef raised in subtropical environments. Genotyping on a high-density platform permitted genome-wide coverage using available markers. In Holstein animals genotyped for a panel of 50 000 SNPs, 702 regions were identified ($P < 0.05$; Qanbari *et al.* 2010). The analysis of more than 500 000 SNPs (with an average pairwise space of 4.5 kb) in the present study may have influenced the number of haplotype blocks formed resulting in a broader coverage of the genome. This finer intermarker resolution potentially allowed the identification of a larger number of significant regions.

Many haplotypes with low frequency (up to 10%) were observed (Fig. 2), and this may be a consequence of strong bottlenecks, as discussed by Tajima (1989). Considerable

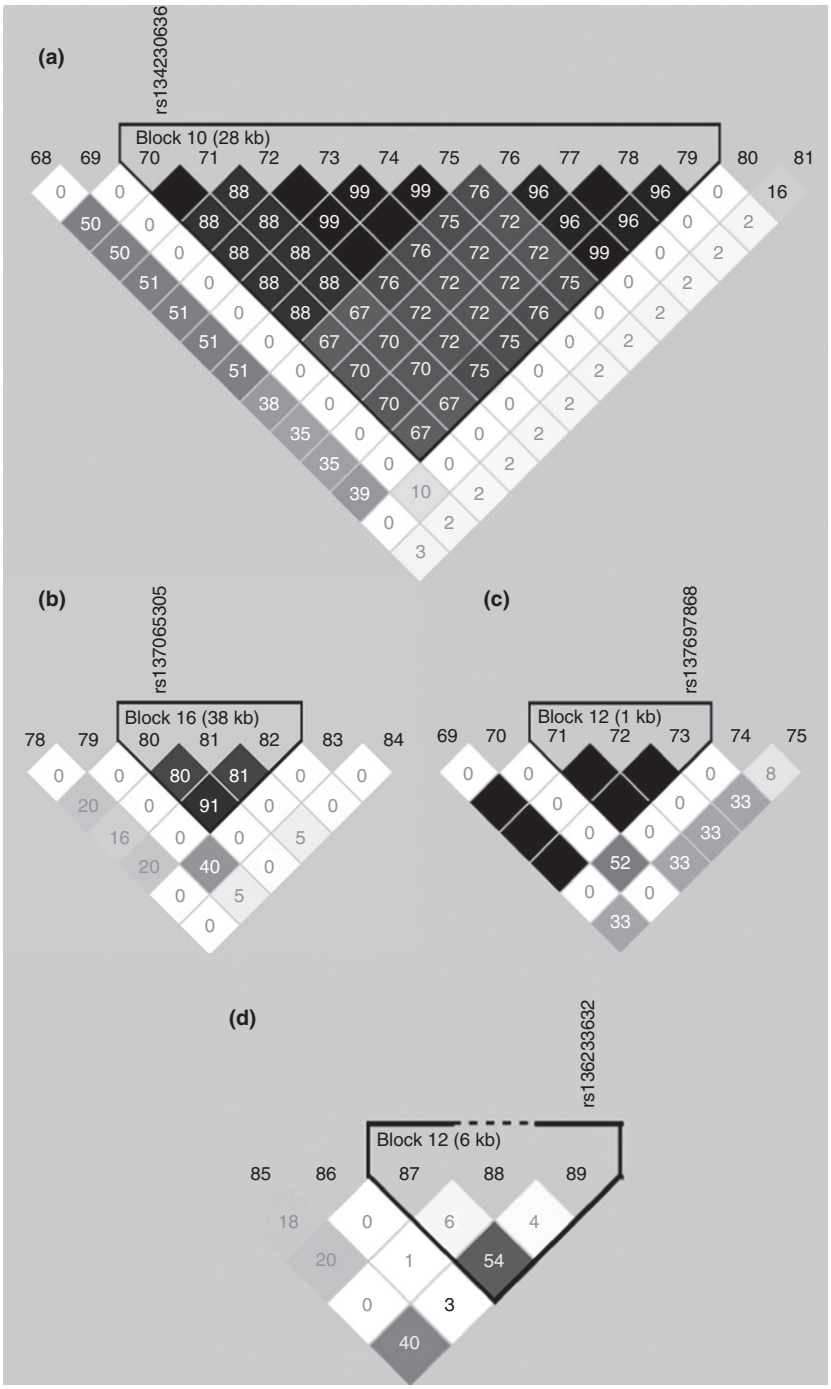

Figure 4 Graphic representation of pairwise r^2 statistic for the blocks corresponding to selection signatures in the *SST* (a), *LEPR* (b), *FABP4* (c) and *GH1* (d) genes. White, $r^2 = 0$; black, $r^2 = 1$. Circled single nucleotide polymorphisms are located near or inside the candidate gene.

Table 4 Results of whole genome extended haplotype homozygosity test for selection signatures in Nellore cattle.

Chr	Tests on blocks (*n*)	*P*-value < 0.01 (*n*)	*P*-value < 0.001 (*n*)	*P*-value < 0.0001 (*n*)
1	10 741	113	22	4
2	10 510	105	7	1
3	9762	107	15	1
4	9417	107	10	0
5	8429	90	10	2
6	9644	125	13	3
7	8729	91	11	0
8	9483	92	14	2
9	8989	91	13	3
10	8060	86	10	0
11	8262	96	14	0
12	6349	66	8	2
13	6202	65	8	4
14	7108	78	10	0
15	6656	65	8	0
16	6631	85	9	1
17	5939	67	1	0
18	4615	36	2	0
19	4526	50	6	1
20	5746	76	14	2
21	5724	54	4	0
22	4405	41	2	0
23	4264	33	2	0
24	5162	69	15	0
25	3164	37	5	2
26	4286	38	5	1
27	3667	43	4	0
28	3774	48	5	2
29	3700	49	5	0
Total	19 3944	2103	252	31

bottleneck events were observed in the Nellore breed in the last decades; for example, based on LD, it is estimated that the Ne (effective population size) of the Nellore breed has decreased from about 400 to less than 100 animals over the last 100 generations (Villa-Angulo *et al.* 2009). In Brazil, a decrease from 85 to 68 animals has been observed across four generations between 1979 and 1998 (Faria *et al.* 2009).

Relative extended haplotype homozygosity analysis of *LEPR*, *SST*, *GH1* and *FABP4* genes was significant. The *LEPR* gene is associated with energy balance, fat deposition and growth traits in *Bos taurus* animals (Guo *et al.* 2008). The *somatostatin* (*SST*) gene, encoding a protein hormone that negatively regulates *GH1* (Martin & Millard 1986), has been related to QTL for traits such as marbling, loin eye area and subcutaneous fat thickness in Angus cattle (Morsci *et al.* 2006). In addition, SNPs in this gene are associated with height and length in *Bos taurus* (Gao *et al.* 2011). The *FABP4* gene plays an important role in lipid metabolism and adipocyte homeostasis in cattle (Hoashi *et al.* 2008) and is involved in lipid transport during lipolysis (Shen *et al.* 1999). Two polymorphisms in the *FABP4* gene have been associated with marbling and subcutaneous fat deposition in *Bos taurus* animals (Michal *et al.* 2006).

Based on the candidate gene analysis, there was no significance for immune system genes, probably because these genes undergo natural selection or genetic drift based on breed pedigree structure or are under balancing selection (Porto-Neto *et al.* 2013). This fact prevents a rapid increase in allele frequencies (Fay & Wu 2000) and is likely to render these genes undetectable by the method used.

The distribution of the inferred selection signatures was not uniform across the genome (Fig. 5), a finding that is consistent with the hypothesis that traits controlled by different regions are under artificial selection. Furthermore, there are many different complex traits under natural and artificial selection that are controlled by many genes and interactions not evenly distributed across the genome.

The number of selection signatures ($P < 0.001$) per chromosome ranged from one (Chr17) to 22 (Chr1; Table 4 and Table S1). QTL related to reproductive traits were identified on chromosomes 1, 3, 9, 20 and 24 in Holstein cattle (Sahana *et al.* 2010), and 15 regions identified in the present study overlapped with 11 of those loci. In addition, SNPs located on chromosomes 3 and 8 have been associated with residual feed intake in a composite population of *Bos taurus* and *Bos indicus* (Bolormaa *et al.* 2011).

Among the genes found in regions of selection signatures, 58 are involved in biological processes related to growth, muscle and adipose tissue metabolism, reproduction or immunity (Table S3). Some immune-related genes participate in the development and viability of follicles and oocytes and in pregnancy maintenance in mammals.

The *PTAFR* (*platelet-activating factor receptor*) gene is involved in regulation of granulosa cell alteration cycle in bovine ovaries (Viergutz & Löhrke 2007). The *PDGFB* (*platelet-derived growth factor beta polypeptide*) gene and its receptors are expressed in the corpus luteum of rats and inhibit secretion of steroids stimulated by luteinizing hormone (LH; Taylor 2000). These genes have been suggested to be involved in the migration and proliferation of theca cells and in microvascularization of the luteal parenchyma after the LH-dependent phase (Sleer & Taylor 2007). Their expression in the endometrium is associated with recruitment of macrophages, which are important for immune regulation, tissue remodeling, angiogenesis and apoptosis during the postpartum period (Oliveira *et al.* 2010). The *RPS6KA5* (*ribosomal protein S6 kinase*, 90 kDa, *polypeptide* 5) gene has been suggested to play a role in reprogramming during the blastocyst stage of bovine embryos obtained by nuclear transfer (Smith *et al.* 2005). The *PAX6* (*paired box 6*) gene is expressed in oocytes of adult cows and is also a potential candidate associated with oocyte viability and early embryo development (Dorji *et al.* 2012). The *POLB* [*polymerase* (*DNA directed*), *beta*] gene is expressed after *in vitro* maturation of bovine oocytes (Dalbiès-Tran & Mermillod 2003), and this regulation is mediated by estradiol, conferring viability on oocytes (Murdoch & Van Kirk 2001).

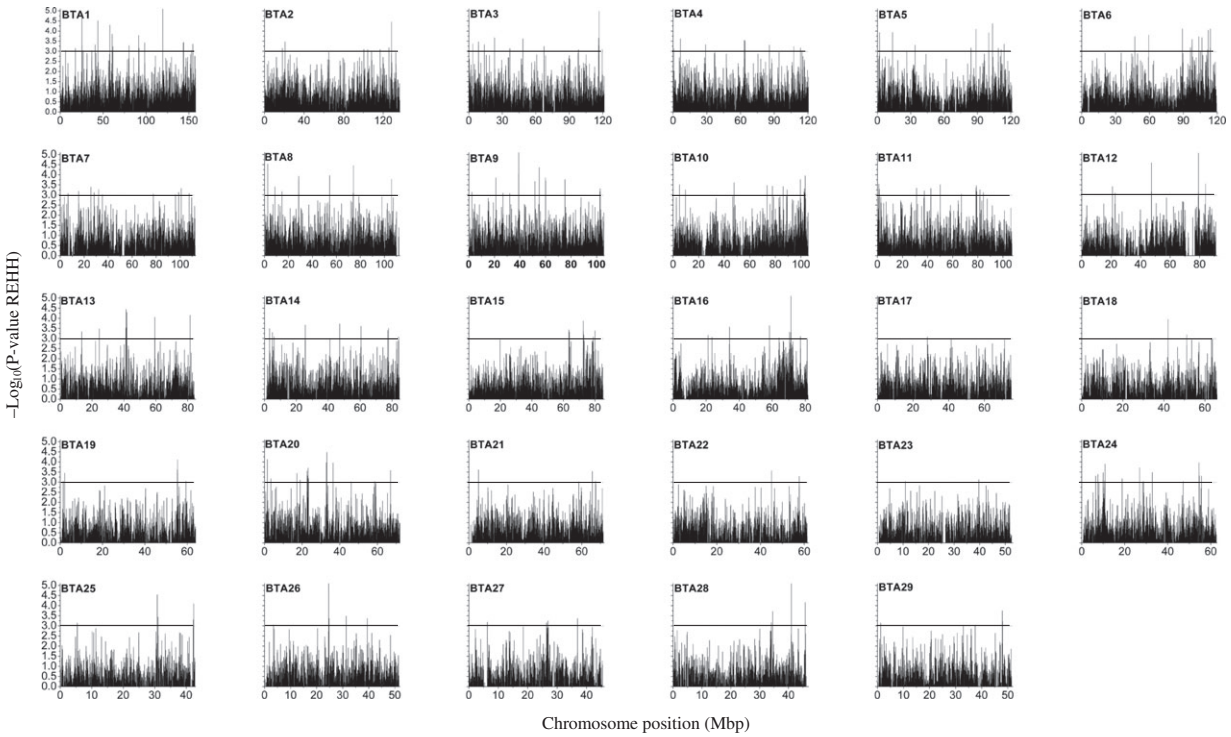

Figure 5 Genome-wide distribution of *P*-values for haplotypes with frequency ≥0.30 in Nellore cattle genome. Horizontal lines indicate significant regions ($P < 0.001$).

Considering the genetic gains observed in reproductive traits, such as age at first calving, of Nellore cattle in Brazil (Laureano *et al.* 2011), it is likely that the homozygosity around these genes is a consequence of selection programs. Further studies of reproductive traits should explore the regions that harbor these genes.

This study also identified genes involved in the transport and/or binding of calcium ions. Some of these genes have been associated with economically important traits in beef cattle. The *CACNA1C* (*calcium channel, voltage-dependent, L type, alpha 1C subunit*) gene has been shown to be significantly correlated with meat tenderness and juiciness in Charolais cattle (Bernard *et al.* 2007). The *CALM1* [*calmodulin 1* (*phosphorylase kinase, delta*)] gene was found to be more expressed in adipose cells after induction of adipogenesis (Tan *et al.* 2006) and also in the Psoas major muscle when compared to the Flexor digitorum (Moreno-Sánchez *et al.* 2010). Given that Nellore cattle are not selected for meat tenderness, one possible explanation is that these genes might exert pleiotropic effects on other traits under selection in breeding programs, which could explain the selection signatures identified. The *COL8A1* (*collagen, type VIII, alpha 1*) gene was classified by DAVID as related to muscle development and was previously found under a significant selection sweep in Nellore cattle genome (Utsunomiya *et al.* 2013). Therefore, these genes and their metabolic pathways would be good candidates for studies investigating meat quality in Nellore cattle.

The quantity and profiles of observed selection signatures are a possible consequence of the effects that genetic breeding programs have had on the genome of the Nellore cattle. The identification of regions under artificial selection pressure may guide searches for regions or groups of genes that control traits of interest, such as meat quality and reproduction. In a subsequent step, polymorphisms and haplotypes related to selection signatures can be used in association and validation studies involving different Nellore subpopulations to permit the genomic selection of superior animals with the genotyping of reduced SNP panels.

Acknowledgements

We thank the CNPq for providing fellowships to Maurício Mello de Alencar and Luciana Correia de Almeida Regitano and for providing a scholarship to Adriana Luiza Somavilla.

References

Ardlie K.G., Kruglyak L. & Seielstad M. (2002) Patterns of linkage disequilibrium in the human genome. *Nature Reviews Genetics* **3**, 299–309.

Baeza M.C., Corva P.M., Soria L.A. *et al.* (2011) Genetic markers of body composition and carcass quality in grazing Brangus steers. *Genetics and Molecular Research* **10**, 3146–56.

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

Bernard C., Cassar-Malek I., Le Cunff M., Dubroeucq H., Renand G. & Hocquette J.-F. (2007) New indicators of beef sensory quality revealed by expression of specific genes. *Journal of Agricultural and Food Chemistry* **55**, 5229–37.

Bolormaa S., Hayes B.J., Savin K., Hawken R., Barendse W., Arthur P.F., Herd R.M. & Goddard M.E. (2011) Genome-wide association studies for feedlot and growth traits in cattle. *Journal of Animal Science* **89**, 1684–97.

Dalbiès-Tran R. & Mermillod P. (2003) Use of heterologous complementary DNA array screening to analyze bovine oocyte transcriptome and its evolution during *in vitro* maturation. *Biology of Reproduction* **68**, 252–61.

Dennis G., Sherman B., Hosack D., Yang J., Gao W., Lane H.C. & Lempicki R. (2003) DAVID: Database for annotation, visualization, and integrated discovery. *Genome Biology* **4**, P3.

Dorji D., Ohkubo Y., Miyoshi K. & Yoshida M. (2012) Gene expression differences in oocytes derived from adult and prepubertal Japanese black cattle during *in vitro* maturation. *Reproduction in Domestic Animals* **47**, 392–402.

Faria F.J.C., Filho A.E.V., Madalena F.E. & Josahkian L.A. (2009) Pedigree analysis in the Brazilian zebu breeds. *Journal of Animal Breeding and Genetics* **126**, 148–53.

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

Fortes M.R.S., Li Y., Collis E., Zhang Y. & Hawken R.J. (2012) The IGF1 pathway genes and their association with age of puberty in cattle. *Animal Genetics* **44**, 91–5.

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

Gao L., Zan L.S., Wang H.B., Hao R.J. & Zhong X. (2011) Polymorphism of *somatostatin* gene and its association with growth traits in Chinese cattle. *Genetics and Molecular Research* **10**, 703–11.

Guo Y., Chen H., Lan X., Zhang B., Pan C., Zhang L., Zhang C. & Zhao M. (2008) Novel SNPs of the bovine *LEPR* gene and their association with growth traits. *Biochemical Genetics* **46**, 828–34.

Hickey J.M., Kinghorn B.P., Tier B., Wilson J.F., Dunstan N. & van der Werf J.H.J. (2011) A combined long-range phasing and long haplotype imputation method to impute phase for SNP genotypes. *Genetics Selection Evolution* **43**, 12–24.

Hoashi S., Hinenoya T., Tanaka A., Ohsaki H., Sasazaki S., Taniguchi M., Oyama K., Mukai F. & Mannen H. (2008) Association between fatty acid compositions and genotypes of *FABP4* and *LXR-alpha* in Japanese Black cattle. *BMC Genetics* **9**, 84.

Hu Z.L. & Reecy J.M. (2007) Animal QTLdb: beyond a repository. *Mammalian Genome* **18**, 1–4.

Innan H. & Kim Y. (2008) Detecting local adaptation using the joint sampling of polymorphism data in the parental and derived populations. *Genetics* **179**, 1713–20.

Laureano M.M.M., Boligon A.A., Costa R.B., Forni S., Severo J.L.P. & Albuquerque L.G. (2011) Estimativas de herdabilidade e tendências genéticas para características de crescimento e reprodutivas em bovinos da raça Nelore [Estimates of heritability and genetic trends for growth and reproduction traits in Nelore cattle]. *Arquivo Brasileiro de Medicina Veterinária e Zootecnia* **63**, 143–52.

Liron J., Prando A., Fernandez M., Ripoli M., Rogberg-Munoz A., Goszczynski D., Posik D., Peral-Garcia P., Baldo A. & Giovambattista G. (2012) Association between *GNRHR*, *LHR* and *IGF1* polymorphisms and timing of puberty in male Angus cattle. *BMC Genetics* **13**, 26.

Martin J.B. & Millard W.J. (1986) Brain regulation of growth hormone secretion. *Journal of Animal Science* **63**, 11–26.

Martinez M.L., Machado M.A., Nascimento C.S. *et al.* (2006) Association of BoLA-DRB3.2 alleles with tick (*Boophilus microplus*) resistance in cattle. *Genetics and Molecular Research* **5**, 513–24.

McKay S.D., Schnabel R.D., Murdoch B.M. *et al.* (2007) Whole genome linkage disequilibrium maps in cattle. *BMC Genetics* **8**, 74–85.

Michal J.J., Zhang Z.W., Gaskins C.T. & Jiang Z. (2006) The bovine *fatty acid binding protein 4* gene is significantly associated with marbling and subcutaneous fat depth in Wagyu × Limousin F2 crosses. *Animal Genetics* **37**, 400–2.

Micke G.C., Sullivan T.M., McMillen I.C., Gentili S. & Perry V.E.A. (2011) Protein intake during gestation affects postnatal bovine skeletal muscle growth and relative expression of IGF1, IGF1R, IGF2 and IGF2R. *Molecular and Cellular Endocrinology* **332**, 234–41.

Moreno-Sánchez N., Rueda J., Carabaño M., Reverter A., McWilliam S., González C. & Díaz C. (2010) Skeletal muscle specific genes networks in cattle. *Functional & Integrative Genomics* **10**, 609–18.

Morsci N.S., Schnabel R.D. & Taylor J.F. (2006) Association analysis of *adiponectin* and *somatostatin* polymorphisms on BTA1 with growth and carcass traits in Angus cattle. *Animal Genetics* **37**, 554–62.

Murdoch W.J. & Van Kirk E.A. (2001) Estrogenic upregulation of DNA polymerase beta in oocytes of preovulatory ovine follicles. *Molecular Reproduction and Development* **58**, 417–23.

Oliveira L.J., McClellan S. & Hansen P.J. (2010) Differentiation of the endometrial macrophage during pregnancy in the cow. *PLoS One* **5**, e13213.

Piper E.K., Jonsson N.N., Gondro C., Lew-Tabor A.E., Moolhuijzen P., Vance M.E. & Jackson L.A. (2009) Immunological profiles of *Bos taurus* and *Bos indicus* cattle infested with the cattle tick, *Rhipicephalus (Boophilus) microplus*. *Clinical and Vaccine Immunology* **16**, 1074–86.

Porto-Neto L.R., Sonstegard T.S., Liu G.E., Bickhart D.M., Da Silva M.V.B., Machado M.A., Utsunomiya Y.T., Garcia J.F., Gondro C. & Van Tassell C.P. (2013) Genomic divergence of zebu and taurine cattle identified through high-density SNP genotyping. *BMC Genomics* **14**, 876–87.

Qanbari S., Pimentel E.C.G., Tetens J., Thaller G., Lichtner P., Sharifi A.R. & Simianer H. (2010) A genome-wide scan for signatures of recent selection in Holstein cattle. *Animal Genetics* **41**, 377–89.

Qanbari S., Gianola D., Hayes B., Schenkel F., Miller S., Moore S., Thaller G. & Simianer H. (2011) Application of site and haplotype-frequency based approaches for detecting selection signatures in cattle. *BMC Genomics* **12**, 318.

Rogberg-Muñoz A., Melucci L., Prando A., Villegas-Castagnasso E.E., Ripoli M.V., Peral-García P., Baldo A., Añon M.C. & Giovambattista G. (2011) Association of bovine chromosome 5 markers with birth and weaning weight in Hereford cattle raised under extensive conditions. *Livestock Science* **135**, 124–30.

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.

Sahana G., Guldbrandtsen B., Bendixen C. & Lund M.S. (2010) Genome-wide association mapping for female fertility traits in Danish and Swedish Holstein cattle. *Animal Genetics* **41**, 579–88.

Shen W.J., Sridhar K., Bernlohr D.A. & Kraemer F.B. (1999) Interaction of rat hormone-sensitive lipase with adipocyte lipid-binding protein. *Proceedings of the National Academy of Sciences of the United States of America* **96**, 5528–32.

Silveira L.G.G., Furlan L.R., Curi R.A. *et al.* (2008) *Growth hormone 1* gene (*GH1*) polymorphisms as possible markers of the production potential of beef cattle using the Brazilian Canchim breed as a model. *Genetics and Molecular Biology* **31**, 874–9.

Sleer L.S. & Taylor C.C. (2007) Platelet-derived growth factors and receptors in the rat corpus luteum: localization and identification of an effect on luteogenesis. *Biology of Reproduction* **76**, 391–400.

Smith S.L., Everts R.E., Tian X.C., Du F., Sung L.-Y., Rodriguez-Zas S.L., Jeong B.-S., Renard J.-P., Lewin H.A. & Yang X. (2005) Global gene expression profiles reveal significant nuclear reprogramming by the blastocyst stage after cloning. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 17582–7.

Stella A., Ajmone-Marsan P., Lazzari B. & Boettcher P. (2010) Identification of selection signatures in cattle breeds selected for dairy production. *Genetics* **185**, 1451–61.

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

Tan S., Reverter A., Wang Y., Byrne K., McWilliam S. & Lehnert S. (2006) Gene expression profiling of bovine *in vitro* adipogenesis using a cDNA microarray. *Functional & Integrative Genomics* **6**, 235–49.

Taylor C.C. (2000) Platelet-derived growth factor activates porcine thecal cell phosphatidylinositol-3-kinase-Akt/PKB and ras-extracellular signal-regulated kinase-1/2 kinase signaling pathways via the platelet-derived growth factor-β receptor. *Endocrinology* **141**, 1545–53.

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

Utsunomiya Y.T., Perez O'Brien A.M., Sonstegard T.S., Van Tassell C.P., do Carmo A.S., Meszaros G., Solkner J. & Garcia J.F. (2013) Detecting loci under recent positive selection in dairy and beef cattle by combining different genome-wide scan methods. *PLoS One* **8**, e64280.

Viergutz T. & Löhrke B. (2007) Role of the platelet-activating factor and its receptor in the proliferative regulation of bovine ovarian granulosa cells. *Cell Proliferation* **40**, 949–60.

Villa-Angulo R., Matukumalli L.K., Gill C.A., Choi J., Van Tassell C.P. & Grefenstette J.J. (2009) High-resolution haplotype block structure in the cattle genome. *BMC Genetics* **10**, 19–31.

Wiener P., Woolliams J.A., Frank-Lawale A., Ryan M., Richardson R.I., Nute G.R., Wood J.D., Homer D. & Williams J.L. (2009) The effects of a mutation in the *myostatin* gene on meat and carcass quality. *Meat Science* **83**, 127–34.

Woods A.M. & Judd A.M. (2008) Interleukin-4 increases cortisol release and decreases adrenal androgen release from bovine adrenal cells. *Domestic Animal Endocrinology* **34**, 372–82.

Zimin A.V., Delcher A.L., Florea L. *et al.* (2009) A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biology* **10**, R42.1–R42.10.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1. Description of selection signatures regions ($P < 0.001$) in Nellore cattle genome.

Table S2. Reported QTL close to regions harboring selection signatures ($P < 0.0001$) regions in Nellore cattle genome.

Table S3. Genes related to muscle and adipose tissue metabolism, growth, reproduction and immunity near the selection signatures regions in Nellore cattle genome.