Animal: An International Journal of Animal Bioscience

Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano Reggiano cheese production system

--Manuscript Draft--

Manuscript Number:	ANIMAL-19-10695R2
Full Title:	Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano Reggiano cheese production system
Short Title:	Selection signatures in the Reggiana cattle genome
Article Type:	Research Article
Section/Category:	1. Breeding and Genetics
Keywords:	Autochthonous breed; Bos taurus; Genome; selection signature; Selection sweep.
Corresponding Author:	Luca Fontanesi, PhD University of Bologna Reggio Emilia, RE ITALY
First Author:	Luca Fontanesi, PhD
Order of Authors:	Luca Fontanesi, PhD
	Francesca Bertolini
	Giuseppina Schiavo
	Samuele Bovo
	Maria Teresa Sardina
	Salvatore Mastrangelo
	Stefania Dall'Olio
	Baldassare Portolano
Manuscript Region of Origin:	ITALY
Abstract:	Reggiana is an autochthonous cattle breed reared mainly in the province of Reggio Emilia, located in the North of Italy. Reggiana cattle (originally a triple-purpose population largely diffused in the North of Italy) are characterized by a typical solid red coat colour. About 2500 cows of this breed are currently registered to its herd book. Reggiana is now considered a dual-purpose breed even if it is almost completely dedicated to the production of a mono-breed branded Protected Designation of Origin (PDO) Parmigiano Reggiano cheese, which is the main driver of the sustainable conservation of this local genetic resource. In this study, we provided the first overview of genomic footprints that characterize Reggiana and define the diversity of this local cattle breed. A total of 168 Reggiana sires (all bulls born over 35 years for which semen was available) and other 3321 sires from three cosmopolitan breeds (Brown, Holstein and Simmental) were genotyped with the Illumina BovineSNP50 panel. ADMIXTURE analysis suggested that Reggiana breed might have been influenced, at least in part, by the other three breeds included in this study. Selection signatures in the Reggiana genome were identified using three statistical approaches based on allele frequency differences among populations or on properties of haplotypes segregating in the populations (fixation index FST; integrated haplotype score, iHS; Cross-Population Extended Haplotype Homozygosity, XP-EHH). We identified several regions under peculiar selection in the Reggiana breed, particularly on bovine chromosome (BTA) 6 in the KIT gene region, that is known to be involved in coat colour pattern distribution, and within the region of the LAP3, NCAPG and LCORL genes, that are associated with stature, conformation and carcass traits. Another already known region that includes the PLAG1 gene (BTA14), associated with conformation traits, showed a selection signature in the Reggiana cattle. On BTA18, a signal of selection included the MC1R gene, that causes the red coat colour in cattle. Other selection sweeps were in regions, with high density of QTL for milk production traits (on BTA20) and in several other large regions that might have contributed to

	shape and define the Reggiana genome (on BTA17 and BTA29). All these results, overall, indicate that the Reggiana genome might still contain several signs of its multi-purpose and non-specialized utilization, as already described for other local cattle populations, in addition to footprints derived by its ancestral origin and by its adaptation to the specialized Parmigiano Reggiano cheese production system.

Manuscript number: ANIMAL-19-10695R1
Title: Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano Reggiano cheese production system

Dear Dr. Rodriguez-Zas,

Thank you very much for sending me the reviewers' comments on the manuscript ANIMAL-19-10695R1
Entitled "Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano Reggiano cheese production system" that we have submitted to Animal Journal.

Please find below a point by point reply to their comments.

All changes into the manuscript are highlighted in reds.
We thank again the reviewer whose suggestions made it possible to improve our manuscript.
Please let me know if you need additional changes.
Thank you very much again for your editorial work.
Looking forward to receiving your evaluation.

Sincerely Yours,

Luca Fontanesi (on behalf of all co-authors)

Bologna, 25th of Sept. 2019

Reviewer #3: The authors addressed the most of my comments but the last and the most serious one remains unaddressed.

The authors responded: " The Fst pairwise analyses showed extreme differences between the compared breeds. Therefore, the breeds are different in these regions. We do not know for what alleles (the ancestral or the derived SNP alleles), but they are different. In addition, the window based approach calculated mean Fst values that provided averaged Fst differences over genomic regions and not just for one SNP position."

These statements do not contribute to address the issue nor advance the understanding of the results of the study. Good and accepted practices in the science are not being followed. If the authors do not know in which of the breeds the region was selected , then the authors should not report this analysis. If the authors' work is in the paradigm of the selective sweep regions, in this case the region with high Fst between two breeds but with low diversity in the first one and moderate diversity in the second would mean the selected region is in the first breed.
The manuscript and reply do not follow the previous straightforward logic.

Authors:
Thank you for the comments.
We could now better clarify what we stated in the first reply to the reviewers' comments.
We said that "we do not know for what alleles (the ancestral or the derived SNP alleles) but they are different…"
Actually, this is still true and might be applied to the SNPs alleles that are located in the analysed windows.
Simplifying, what Fst in windows can identify are allele/haplotype frequencies differences between the compared breeds.

We agree that it could be possible that some differences might be the combination of opposite selections acting in different ways in both compared breeds or with an effect in just one of the two breeds. However, we would better say that the stringency that we applied in the identification of Fst differences in the paired analyses cannot reach extreme values if there would be just selection acting only in the cosmopolitan breed used in the comparison (that would mean that the signature that we might identify is derived by the combination of opposite directions in the two compared breeds, a cosmopolitan breed vs the Reggiana breed).
Anyway, as based on the Fst test we do not know the direction of the signals in the Fst analyses, we modified the text following this reasoning.
In addition, the title of the related paragraph which includes the Fst results was modified: we substituted the previous title "Fixation index (F_{ST}) selection signals in the Reggiana genome" with the new title "Fixation index (F_{ST}) signals in the Reggiana vs cosmopolitan breed comparisons" (line 268)
We included a few sentences in Results related to this issue:
Lines 273-275:
"It is worth mentioning that, as the pairwise F_{ST} analyses cannot distinguish the direction of the signals, we regarded the identified signals obtained with this test as derived by regions that can differentiated the compared breeds."

And in Discussion:

Lines 410-416.
“It is also clear that the signals determined by the mFST tests cannot completely be assigned to an effect originated from the Reggiana breed only. Extreme mFST values might be also derived by forces acting on opposite direction on the compared cosmopolitan breed, thus this test could contain, in part signatures not only present in the Reggiana genome. Therefore, a combination of signals derived by other methods was also used for the general interpretation of the results, particularly when F_{ST} signals were involved.”

However, Fst analysis clearly identified a selection signature in the comparison between Reggiana and Holstein. This selection signature is in the MC1R gene region. It is well known that MC1R alleles are different in the two breeds and might be fixed or almost fixed for alternative alleles in both Reggiana and Holstein.
We also evidenced this point in the text at lines 425-427.

Moreover, final interpretation and gene enrichment analyses were obtained using signals detected by at least two methods.

Reviewer#3:
The study is expected to include a check for diversity among the breeds using the sliding window approach in the regions with high Fst.

Authors:
This is what we have already included at lines 200-201 that can answer the reviewer’s request.

“Average F_{ST} (mF_{ST}) was calculated in overlapping windows of 1 Mb with a step of 500 kb using an in-house script.”

Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano Reggiano cheese production system

Francesca Bertolini[1], Giuseppina Schiavo[2], Samuele Bovo[2], Maria Teresa Sardina[3], Salvatore Mastrangelo[3], Stefania Dall'Olio[2], Baldassare Portolano[3] and Luca Fontanesi[2]

[1] National Institute of Aquatic Resources, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

[2] Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy

[3] Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, 90128 Palermo, Italy

Corresponding author: Luca Fontanesi. Email: luca.fontanesi@unibo.it.

Short title:

Selection signatures in the Reggiana cattle genome

Abstract

Reggiana is an autochthonous cattle breed reared mainly in the province of Reggio Emilia, located in the North of Italy. Reggiana cattle (originally a triple-purpose population largely diffused in the North of Italy) are characterized by a typical solid red coat colour. About 2500 cows of this breed are currently registered to its herd book. Reggiana is now considered a dual-purpose breed even if it is almost completely dedicated to the production of a mono-breed branded Protected Designation of Origin (PDO) Parmigiano Reggiano cheese, which is the main driver of the sustainable conservation of this local genetic resource. In this study, we provided the first overview of genomic footprints that characterize Reggiana and define the diversity of this local cattle breed. A total of 168 Reggiana sires (all bulls born over 35 years for which semen was available) and other 3321 sires from three cosmopolitan breeds (Brown, Holstein and Simmental) were genotyped with the Illumina BovineSNP50 panel. ADMIXTURE analysis suggested that Reggiana breed might have been influenced, at least in part, by the other three breeds included in this study. Selection signatures in the Reggiana genome were identified using three statistical approaches based on allele frequency differences among populations or on properties of haplotypes segregating in the populations (fixation index F_{ST}; integrated haplotype score, iHS; Cross-Population Extended Haplotype Homozygosity, XP-EHH). We identified several regions under peculiar selection in the Reggiana breed, particularly on bovine chromosome (BTA) 6 in the *KIT* gene region, that is known to be involved in coat colour pattern distribution, and within the region of the *LAP3, NCAPG and LCORL* genes, that are associated with stature, conformation and carcass traits. Another already known region that includes the *PLAG1* gene (BTA14), associated with conformation traits, showed a selection signature

in the Reggiana cattle. On BTA18, a signal of selection included the *MC1R* gene, that causes the red coat colour in cattle. Other selection sweeps were in regions, with high density of QTL for milk production traits (on BTA20) and in several other large regions that might have contributed to shape and define the Reggiana genome (on BTA17 and BTA29). All these results, overall, indicate that the Reggiana genome might still contain several signs of its multi-purpose and non-specialized utilization, as already described for other local cattle populations, in addition to footprints derived by its ancestral origin and by its adaptation to the specialized Parmigiano Reggiano cheese production system.

Keywords: Autochthonous breed; *Bos taurus*; Genome; Selection signature; Selection sweep.

Implications

Reggiana cattle breed, once a multi-purposes autochthonous breed, is now used to produce a mono-breed branded Parmigiano Reggiano cheese, which is now the main driver of the sustainable conservation of this local genetic resource. This study identified selection signatures in the Reggiana genome that provided information for both almost fixed breed-specific traits (e.g. coat colours) and several other more diluted signs of its re-adaptation and more recent production shifts. It was evident that this breed still contains signs of its multi-purpose and non-specialized past utilization suggesting the need to better define a tailored selection strategy for its current main use.

Introduction

Selection signature analyses based on single nucleotide polymorphism (SNP) chip data have been carried out in cattle to identify loci under natural or artificial selection and peculiar genetic features that might be useful to describe breed specific characteristics (e.g. Flori *et al.*, 2009; Zhao *et al.*, 2015). The statistical approaches that were used for these studies are based either on the evaluation of allele frequency differences among populations or on properties of haplotypes segregating in the populations. The fixation index F_{ST} (Wright, 1951) is one of the most used allele frequency difference approaches that quantifies population differentiation. F_{ST} provides an estimate of the amount of genetic variability that exists between populations relative to that within populations. This statistic assumes that different selective forces acting on different populations may favor divergent alleles. Therefore, allele frequency differences between populations may be more extreme in the chromosome regions in which these variants are located. Among the most frequently applied haplotype-based approaches, the integrated haplotype score iHS (Voight *et al.*, 2006) is an improvement of the extended haplotype homozygosity (EHH) method and compares EHH between derived and ancestral alleles within a population. The Cross-Population Extended Haplotype Homozygosity (XP-EHH; Sabeti *et al.*, 2007) is based on both EHH and iHS but it is not calculated within populations but between populations and does not need to define ancestral and derived alleles as requested by iHS. According to their assumptions, these tests could be complementary to identify selection signatures (Gautier and Naves, 2011).

Reggiana is an autochthonous cattle breed reared mainly in the province of Reggio Emilia, located in the Emilia Romagna region, in the North of Italy. This breed is characterized by a typical red coat colour (referred as “fromentino”). Tradition dates

back the origin of the Reggiana ancestral population in the Barbaric invasion period after the fall of the Roman Empire (VI century). Historical records of the XII century indicate that a red cattle population was used by the monks to produce in the same region a typical cheese from which subsequently originated the Parmigiano Reggiano cheese, now a renowned and well-known worldwide Protected Designation of Origin (PDO) dairy product. At that time, this population was not a specialized dairy cattle as it served for work and meat production as well.

Reggiana remained one of the most numerous cattle populations in the North of Italy till the mid of the XX century (139695 heads were recorded in 1954; ANABORARE, 2019). This number decreased progressively in the following decades due to the substitution of the Reggiana cattle with more specialized and productive Holstein cattle and in the 1980s this local breed reached the minimum number of about 500 cows. Mean milk yield of Reggiana cows is about 30% lower than that of Holstein cows (Gandini *et al.*, 2007). Then a conservation program, linked to a new brand of Parmigiano-Reggiano cheese made only of Reggiana milk, started in the 1990s. The economic advantage derived by selling this mono-breed cheese made it possible to fill the production gap in terms of economic income that the Reggiana farmers had compared to the farmers who raised more productive breeds. This branded Parmigiano-Reggiano cheese reverted the decreasing trend of the Reggiana population reaching, at present, the number of about 2500 cows reared in about 180 different farms.

A selection program in Reggiana started in 1956 with the constitution of the National Association of Reggiana Cattle Breeders (Associazione Nazionale Allevatori Bovini di Razza Reggiana: ANABORARE), which officially could be considered the recognition of the Reggiana breed. The program was organized in a modern way in the

1996 with the re-definition of the herd book of the breed which designed a breeding strategy aimed to reduce inbreeding. In addition, according to the use of the milk produced by Reggiana cows, a specific estimated breeding value for cheese making objectives (Parmigiano-Reggiano yield genetic index) has been implemented to improve both milk yield and milk quality for this production (including fat percentage and protein percentage, with a preference on casein variants positively associated with rennet coagulation properties; ANABORARE, 2019).

So far, few investigations were carried out in this breed to describe its genetic variability. After the pioneering studies of Mariani and Russo (1971) who evaluated the frequency distribution of k-casein protein variants, Caroli *et al.* (2004) analysed polymorphisms in three caseins and in beta-lactoglobulin by isoelectrofocusing on milk. Then, 20 DNA markers were analysed in candidate genes to obtain information on their allele distribution and to identify polymorphisms associated with milk production and composition traits in Reggiana sires (Fontanesi *et al.*, 2015). Polymorphisms in coat colour genes were then investigated to identify markers useful for the authentication of Reggiana branded Parmigiano-Reggiano cheese (Russo *et al.*, 2007) and to study the genetic mechanisms differentiating solid coloured (i.e. Reggiana) from spotted patterns in cattle breeds (Fontanesi *et al.*, 2010b; 2012). Bertolini *et al.* (2015, 2018) used single nucleotide polymorphism (SNP) array data obtained from Reggiana and several other cattle breeds to identify population informative markers. Mastrangelo *et al.* (2016, 2018a, 2018b) used SNP chip data obtained in Reggiana cattle for a comparative analysis of genomic inbreeding parameters, runs of homozygosity (ROH) islands and population structure with other Italian local and commercial cattle breeds. The genetic structure of this breed reflects the small size of its population, with a contemporary

effective population size of about 100 and a proportion of its autosomal genome covered by ROH of about 5%, similar to that of other local breeds of the North of Italy. The breed also clustered with several other cattle breeds of the North of Italy suggesting a general geographical influence of its genetic background (Mastrangelo *et al.*, 2018a).

In this study, we used Illumina SNP chip data and several statistical approaches based on allele frequency differences among populations and on properties of haplotypes segregating in the populations (F_{ST}, iHS and XP-EHH) to identify selection signatures in the Reggiana cattle genome that may distinguish this autochthonous breed from three cosmopolitan breeds (Holstein, Brown and Simmental) and that might be indirectly derived by its ancestral origin and by its specialized use in the Parmigiano Reggiano cheese production system.

Material and methods

Animals and genotyping data

A total of 3489 bulls of four cattle breeds (Reggiana, n. = 168; Holstein, n. = 2093; Brown, n. = 749; and Simmental, n. = 479) were genotyped with the Illumina BovineSNP50 v1 or v2 BeadChip arrays (Illumina, San Diego, CA, USA). Reggiana bulls were all sires born from 1975 to 2010 for which it was possible to obtain frozen semen in 2014. Considering that, on average, about 6-8 sires where available/approved per year over these 35 years, the analysed Reggiana bulls constituted about 70% of all bulls that were used for artificial insemination over this period in this autochthonous breed. The different numbers of analysed sires for the four breeds reflect the dimension of their respective populations.

Single nucleotide polymorphisms were used with their coordinate position on the latest assembly of the bovine genome (ARS-UCD1.2; GCA_002263795.2). Basic SNP statistics were computed with PLINK software version 1.9 (Chang *et al*., 2015). Only common SNPs across the two array versions and with a call rate ≥90% in each breed were retained for further analyses. All monomorphic SNPs across the dataset were removed. After filtering, all cattle had individual call rate of > 0.90 and no animal was therefore discarded. The dataset was imputed using Beagle 3.3.2 (Browning and Browning, 2009) and phased for the haplotype-based analyses using fastPHASE (Scheet and Stephens, 2006) using default parameters. Imputation and phasing were carried out breed by breed.

Population structure analyses

Multidimensional scaling (MDS) plots were obtained with the cluster function of PLINK software version 1.9 (Chang *et al.*, 2015). Population stratification analysis was also performed with the ADMIXTURE software (Alexander *et al.*, 2009), with number of subpopulations (K) that ranging from 1 to 29. As ADMIXTURE does not take linkage disequilibrium into consideration, and to reduce the computational time, the number of markers were reduced according the observed sample correlation coefficient using the --indep-pairwise option of PLINK (Chang *et al.*, 2015).

Selection signature analyses

Detection of selection signatures in the Reggiana cattle genome was based either on the evaluation of allele frequency differences among populations and on properties of haplotypes segregating in the populations. The applied methods included within

184 population (iHS) and between population (F_{ST} and XP-EHH) tests. Between population
185 tests were applied to identify potential sweeps that occurred in the Reggiana breed
186 compared to the other three cosmopolitan breeds (Holstein, Brown and Simmental),
187 which constitute the most numerous cattle populations in the North of Italy. The
188 threshold selected for all these analyses was settled as the 99.5th percentile of the
189 empirical distribution.

190 *Integrated haplotype score (iHS).* This statistic is applied to individual SNPs and
191 was calculated following the procedures defined by Voight *et al.* (2006) and Sabeti *et al.*
192 (2007). Information on the ancestral and derived alleles on all bovine SNPs was
193 obtained from Rocha *et al.* (2014). The *rehh* R package v 2.0.4" (Gautier *et al.*, 2017)
194 was used to calculate |iHS| for each autosomal SNP. Large positive or negative iHS
195 values indicate unusually long haplotypes carrying the ancestral or derived alleles,
196 respectively.

197 *Fixation index (F_{ST}).* Three pairwise F_{ST} analyses were performed comparing each
198 time the Reggiana breed with one of the other cosmopolitan breeds included in this
199 study. Wright's F_{ST} for each SNP was calculated with PLINK 1.9 (Chang *et al.*, 2015).
200 Average F_{ST} (mF_{ST}) was calculated in overlapping windows of 1 Mb with a step of 500 kb
201 using an in-house script. All windows that contained at least 4 SNPs were then
202 considered. Overall averaged F_{ST} was also calculated considering all SNPs in the
203 pairwise comparisons.

204 *Cross-Population Extended Haplotype Homozygosity (XP-EHH).* Three pairwise
205 XP-EHH analyses were run. The XP-EHH scores were calculated using the *rehh* R
206 package v 2.0.4 with default parameters (Gautier *et al.*, 2017) to detect alleles with
207 increased frequency to the point of fixation or near-fixation in Reggiana compared to

other analysed breed. In these pairwise analyses, the Reggiana breed was considered as the reference population. Therefore, only the extreme negative XP-EHH scores identified SNPs under selection in Reggiana but not in the other breeds. As XP-EHH searches for unusually long haplotypes, at least three consecutive SNPs should be above the threshold, rendering this analysis conservative. The threshold was determined using the log(*P*-value).

Annotation of candidate genome regions

Genes that were within the genome windows or haplotype regions identified as described above or that were ± 500 kbp from iHS signals were retrieved from the *Bos taurus taurus* genome assembly ARS-UCD1.2 (https://www.ncbi.nlm.nih.gov/assembly/GCF_002263795.1/) using the National Center of Biotechnology Information (NCBI) *Bos taurus* Annotation Release 106 (https://www.ncbi.nlm.nih.gov/genome/annotation_euk/Bos_taurus/106/). Identification of potential candidate genes for selection was obtained by comparing our results with those in the literature.

Gene enrichment analysis was performed with Enrichr (Chen *et al.*, 2013), via Fisher's exact test. Analyses run over the Gene Ontology – Biological Process (GO:BP; http://geneontology.org/), Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.kegg.jp/) and Reactome (https://reactome.org/) databases. As input, Enrichr took the whole set of genes (no. = 52) mapped within the genome regions identified by more than one method. We considered statistically enriched terms presenting: (i) at least two genes of the input set related to (at least) two different genome regions and (ii) an adjusted *p*-value < 0.05.

Results

Population descriptors

Supplementary Table S1 presents a descriptive summary of the genotyping data of the Reggiana and cosmopolitan cattle breeds. Reggiana cattle had intermediate values for both average minor allele frequency (MAF) and heterozygosity (Het), compared to all other breeds (MAF = 0.253 ± 0.145 and Het = 0.340 ± 0.153). Brown breed had the lowest values for these two measures (MAF = 0.232 ± 0.152 and Het = 0.313 ± 0.168) among the four analysed cattle breeds. Average heterozygosity distributed over all chromosomes in the four investigated breeds is reported in Supplementary Figure S1. No differences among chromosomes and breeds could be observed.

Figure 1 reports two-dimensional MDS plots obtained using the SNP chip data of the four investigated breeds. All breeds were clearly separated by the first three Coordinates (C). Reggiana sires were closer to the Brown and Simmental clouds than to the Holstein group.

The ADMIXTURE analysis plots are shown in Supplementary Figure S2. By inspecting the plot obtained with K = 5, a well-defined pattern could not be observed, suggesting that Reggiana breed can be considered a distinct genetic resource, compared to the other three breeds included in this study, and matching the MDS plot results. However, the plot obtained with K = 3 showed that Reggiana might be influenced by all three cosmopolitan breeds with a larger impact from the Simmental breed than from Brown or Holstein breeds.

Integrated haplotype score signatures in the Reggiana genome

The genome-wide distribution of |iHS| values in the Reggiana breed is shown in Figure 2. A total of 169 SNPs distributed over 18 out of 29 autosomes marked selection sweep regions in the Reggiana genome (Supplementary Table S2). BTA17 and BTA29 were the chromosomes harboring the largest number of these SNPs (44 and 60, respectively; which included three and four regions of contiguous SNPs, respectively), followed by BTA2 and BTA3 (14 SNPs each). Among the top 10 |iHS| markers, six are located on BTA29, two on BTA2, one on BTA6 and one on BTA17 (Table 1). Some of these SNPs are within or close to genes already shown to be included in selection signature regions in the cattle genome (*TENM4* and *KIRREL3*; Bertolini *et al.*, 2018) or involved in key metabolic functions (e.g. *INSIG2* and *ETNPP*). Details and annotations for all 169 |iHS| markers are reported in Supplementary Table S2.

Fixation index (F_{ST}) signals in the Reggiana vs cosmopolitan breed comparisons

Average F_{ST} values including all tested SNPs in the three between-breeds comparisons, i.e. Reggiana *vs* Brown, Reggiana *vs* Holstein and Reggiana *vs* Simmental, were 0.0938, 0.0972 and 0.0533 respectively. Figure 3 reports the Manhattan plots of the window-based pairwise genome-wide F_{ST} analyses of the Reggiana breed against all other breeds. It is worth mentioning that, as the pairwise F_{ST} analyses cannot distinguish the direction of the signals, we regarded the identified signals obtained with this test as derived by regions that can differentiated the compared breeds. F_{ST} signals were identified on 19 autosomes (Supplementary Table S3). The highest total number of 1 Mbp-outlier regions (considering all three comparisons; partially or completely overlapping or independent) was observed on BTA6 (n. = 21) and BTA5 (n. = 10).

On BTA6, eleven, one and nine regions were identified against the Brown, Holstein and Simmental breeds, respectively. Among them, two partially overlapping windows indicated a region (from positions 69.0 Mbp to 70.5 Mbp) that was in common in the Brown and Simmental comparisons. This BTA6 region contains the *KIT* gene, that is well known to be involved in coat colour pattern distribution (e.g. Fontanesi *et al.*, 2010b).

In the Reggiana *vs* Brown comparison, the genomic windows with the highest mean F_{ST} (mF_{ST}) values were on BTA11, from 67.5 Mbp to 69.0 Mbp (two partially overlapping regions, mF_{ST} = 0.47 and 0.43, respectively), and on BTA6, from 69.5 to 70.5 Mbp (mF_{ST} = 0.39) and from 78.0 to 79.0 Mbp (mF_{ST} = 0.38). The BTA11 region corresponds to one of the most extended signatures reported by Rothammer *et al.* (2013) in a Swiss dual purpose (dairy-beef) cattle breed (i.e. Original Braunvieh) and includes a few genes affecting meat and carcass traits (*CAPN14* and *PCBP1*). The first BTA6 region overlaps or is contiguous with other four windows with mF_{ST} above the threshold. As already mentioned, the *KIT* gene is contained in this large window, whereas in the second region, no gene is annotated.

The chromosome regions having the highest mF_{ST} values against the Holstein breed were located on BTA14 (positions 22.5-23.5 Mbp) and on BTA20 (positions 43.5-45.0 Mbp, in which no genes are annotated), with mF_{ST} = 0.49 and 0.48, respectively. The BTA14 region (which was also detected in the Simmental comparison) contains the *PLAG1* gene (23.33-23.38 Mbp), that has been already shown to determine pleiotropic QTL affecting body weight, stature, reproduction traits and milk production in several cattle populations (e.g. Utsunomiya *et al.*, 2017). Another region was identified on BTA4 (76.0-77.0 Mbp, mF_{ST} = 0.42) and contains SNPs that have been already reported to

differentiate cattle breeds, including Reggiana *vs* Holstein, using a random forest classification method (Bertolini *et al.*, 2015). The signal on BTA6 (windows from 37.0-38.0 Mbp) against the Holstein breed contains other genes (*LAP3, NCAPG and LCORL*) already associated to conformation and carcass traits, stature of the animals and calving easy (e.g. Takasuga, 2016). A signal was also observed on BTA18 with two overlapping regions (13.5-14.5 Mbp and 14.0-15.0 Mbp, mFst = 0.35 and 0.32, respectively) that include the *MC1R* gene, determining different coat colours in cattle. Two overlapping regions on BTA26 (22.0-23.0 and 21.5-22.5 Mbp, mF_{ST} = 0.43 and 0.41, respectively) were also identified. This chromosome portion include genes (*PAX2*, *FGF8*, *KCNIP2*, *BTRC*, *HPS6*, *ELOVL3* and *MGEA5*) already suggested to be involved in several processes determining coat colour and QTLs for meat and carcass traits, milk production traits and heat tolerance (e.g. *Hu et al.*, 2019; Macciotta *et al.*, 2017).

The highest mF_{ST} values (0.33 and 0.31) in the Reggiana *vs* Simmental comparison were again on BTA6 for the common *KIT* region (four partially overlapping regions spanning from 69.0 to 71.5 Mbp). Other mF_{ST} signals in the Simmental breed comparison were also observed on BTA7 (three regions, two of which partially overlapping), on BTA11 (three windows), on BTA16 (one region), on the same BTA18 region reported for the Holstein breed and on two overlapping windows of BTA29.

Cross-Population Extended Haplotype Homozygosity (XP-EHH) signatures in the Reggiana genome

Results of the pairwise genome-wide XP-EHH analyses between the Reggiana breed and all other three cosmopolitan breeds are shown in the Manhattan plots included in Figure 4. Signals of selection were reported on 12 out of 29 autosomes but only for nine

of these chromosomes (BTA2, BTA5, BTA6, BTA7, BTA10, BTA13, BTA17, BTA20 and BTA29) at least three consecutive SNPs were identified. Of these signals, negative XP-EHH values (indicating a selection on the Reggiana breed genome) were identified on the following chromosomes on the three comparisons (see Supplementary Table S4 for details):

i) against the Brown breed, on BTA5 (three close regions separated by less than 1 Mbp), BTA6 (nine regions divided in five blocks separated by more than 1 Mbp), BTA13 (two regions separated by more than 1 Mbp), for a total of about 5.9 Mbp;

ii) against the Holstein breed, on BTA10 (one region) and BTA20 (nine regions divided in six blocks, separated by more than 1 Mbp), for a total of about 3.4 Mbp;

iii) against the Simmental breed, on BTA5 (four regions divided in two main blocks separated by more than 1 Mbp), BTA6 (one region) and BTA7 (three close regions, separated by less than 1 Mbp), for a total of about 4.0 Mbp.

The signals of selection on BTA5 identified against the Brown and the Simmental breeds (located in QTL regions for feed efficiency and selection signature reported in other studies) did not overlap. BTA6, summing up what observed in the different comparisons, again, showed the largest number of selection signature regions (n. = 10). This chromosome contained the region with the highest XP-EPP log value of this study (7.608, against the Simmental breed; positions from about 68.3 to 71.4 Mbp), which encompasses the *KIT* gene. The BTA20 region detected in the Reggiana *vs* Holstein analysis contained several signals of selection in regions that have a high density of QTL for several milk production traits (Hu *et al.*, 2019).

Comparative analysis of selection signatures

The diagram of Figure 5 visualizes the distribution of selection signatures obtained with the three used approaches (i.e. iHS, F_{ST} and XP-EHH) across all chromosomes. Only a small proportion of all signals overlapped among these tests. In all cases, overlapping signatures derived only by two tests. A total of 13 regions on six chromosomes (BTA6, BTA7, BTA13, BA17, BTA26 and BTA29) were identified by more than one method (Table 2). BTA6 contained the largest number of overlapping regions (n. = 6), followed by BTA13 and BTA26, with two regions each. Seven regions were detected by both F_{ST} and XP-EHH tests. Three of all these overlapping regions were congruent, that means that the pairwise results were obtained against the same breed, whereas in four cases the pairwise tests identified overlapping regions derived by the comparison of different breeds. It is however worth to mention that in the first part of the overlapping regions of BTA6 (from about 68.3 to 70.7 Mbp; Table 2), the signals observed for the Brown (F_{ST} test) and Simmental (XP-EHH) seems parts of a broader region actually captured by both methods on each breed, as deduced from Figures 3 and 4, Supplementary Tables S3 and S4. Annotation of these regions identified several candidate genes already reported by other studies to be included in selection sweeps or to be associated with several production traits in cattle (e.g. Hu *et al.*, 2019), as also mentioned above for the description of the single methods (Table 2).

Functional analysis was carried out with Enrichr among all genes (n. = 52) mapped in the genomic regions detected with at least two different approaches. This analysis over-represented a total of six functional terms when run over the Biological Process branch of the GO hierarchy (Supplementary Table S5). These terms outline different processes

involving the androgen metabolic process (putatively linked to fertility) and melanocyte differentiation (linked to coat colour). Other processes were related to the vesicle-mediated transport, the regulation of kinase activity and the regulation of transcription factor activity. Analyses over the KEGG and Reactome databases did not highlight any over-represented pathway.

Discussion

Reggiana breed is a small cattle population that can be considered a unique example of conservation of an animal genetic resource in an advanced agricultural production system, represented by the specialized dairy sector of the North of Italy. Reggiana cattle are, at present, almost completely dedicated to the production of Parmigiano Reggiano cheese. The past un-specialized purpose of this cattle population (Reggiana was a triple-purpose breed, dairy-beef-work, till the 1960'; ANABORARE, 2019) has been redefined after the constitution of its first herd book. However, signs of its un-differentiated purposes could be left behind in the genome of these animals. Then, this red breed passed through a recent genetic bottleneck that may have further contributed to shape its current genetic makeup. Oral traditions and historical records indicate that a few genetic introgressions might have occurred in the past from Brown, Simmental and Danish Red (ANABORARE, 2019). ADMIXTURE analysis and MDS plots however indicate that this breed could be considered a distinct genetic pool, compared to the most important cattle breeds that constitute the backbone of the North of Italy dairy industry. Reggiana breed is however clearly closer to Simmental cattle, a dual-purpose breed. Genetic variability of Reggiana population is similar to that of the other analysed cosmopolitan breeds (Supplementary Table S1) and its estimated effective population

size is larger or very close to that of the Holstein and Brown breeds, as previously determined (Marras *et al.*, 2015; Mastrangelo *et al.*, 2016; Mastrangelo *et al.*, 2018a).

In this study, we wanted to identify the unique genetic patterns that characterize the Reggiana breed genome, compared to that of the three most diffused cosmopolitan breeds in the same geographic area. Therefore, we genotyped with the Illumina BovineSNP50 panel all Reggiana sires for which we could get semen samples. The sires were born over a period of about 35 years and constitute the most active bulls that have been used since the recovery of the breed that started in the 1980'. Selection signatures were detected using three methods (i.e. iHS, F_{ST} and XP-EHH tests) which can potentially capture different selection sweep events or structures (González-Rodríguez *et al.*, 2016). Considering the complementarity of the applied methods, as expected, a small proportion of signals overlapped between these tests. It is also clear that the signals determined by the mFST tests cannot completely be assigned to an effect originated from the Reggiana breed only. Extreme mFST values might be also derived by forces acting on opposite direction on the compared cosmopolitan breed, thus this test could contain, in part signatures not only present in the Reggiana genome. Therefore, a combination of signals derived by other methods was also used for the general interpretation of the results, particularly when F_{ST} signals were involved.

A strong selection signal, detected with both pairwise approaches, was identified in the *KIT* gene region (well known to affect coat colour patterns; e.g. Fontanesi *et al.*, 2010b), in the comparisons against the Simmental and Brown breeds. The signal in this region against the Holstein was just below the applied threshold. This is in agreement to what we already reported by comparing a few *KIT* haplotypes in several cattle breeds having different coat colours and patterns (including Reggiana and the other three

cosmopolitan breeds included in this study; Fontanesi *et al.*, 2010a). Another signal associated with different coat colour phenotypes detected by the F_{ST} pairwise analysis between Reggiana and Holstein was observed in the *MC1R* gene region, on BTA18. In this case, even if this signal was detected only with the F_{ST} analysis, it is obvious that these two breeds in this region have extreme allele frequency differences. Holstein cattle are expected to carry the E^D allele (determining the dominant black coat colour) at high frequency whereas Reggiana cattle are fixed for the recessive "*e*" allele (determining the red coat colour) at the *Extension* locus (Russo *et al.*, 2007). The same BTA18 region reported a signal of selection in the F_{ST} analysis against the Simmental breed. As Reggiana and Simmental cattle have the same red coat colour (even if the latter has a spotted phenotype) and carry the same almost fixed genotype at the *MC1R* gene (allele "*e*" frequency in Simmental is >96%; Russo *et al.*, 2007), it seems plausible to suppose that other genetic factors may contribute to differentiate this genomic region between these two red breeds.

Other selection signatures were detected in regions containing genes (e.g. *LAP3*, *NCAPG* and *LCORL* on BTA6 and *PLAG1* on BTA14) that have been already reported to be under strong selection in cattle and shown to affect several morphological traits (Takasuga, 2016; Utsunomiya *et al.*, 2017). These regions were also described to differentiate dairy, dual-purpose and beef breeds (Gutiérrez-Gil et al., 2015).

In addition to the selection signatures identified using the methods reported in this study, other regions of the Reggiana genome might have been under selection pressure. Mastrangelo *et al.* (2018b) analysed the Reggiana genome and identified runs of homozygosity (ROH) islands in a total of eight windows of six different chromosomes (BTA1, BTA3, BTA6, BTA17, BTA26 and BTA29; see Supplementary Table S2 and

Supplementary Table S3). Three of these ROH islands overlap with the iHS signals we detected on BTA3 (positions from about 75.0 to 78.0 Mbp), BTA17 (from about 54.9 to 59.6 Mbp) and BTA29 (from about 16.1 to 22.6 Mbp) and another ROH island overlaps with an F_{ST} signal we reported against the Holstein breed on BTA6 (from about 37.0 to 38.0 Mbp).

Reggiana cows have, on average, a lower milk yield compared to that of Holstein and Brown. The dual-purpose Simmental breed has a similar average milk yield to that of the Reggiana breed. Simmental *vs* Reggiana has also an almost halved overall averaged F_{ST} value than that obtained in the Brown and Holstein breed comparisons (0.0533 against Simmental; 0.0938 against Brown; 0.0972 against Holstein). This lower differentiation level against the Simmental breed is also evident from the window based mF_{ST} analysis that showed that the regions over the 99.5th percentile had a lower average value (mean mF_{ST} = 0.179) than that observed against the Brown (mean mF_{ST} = 0.313) and Holstein (mean mF_{ST} = 0.338) breeds.

Several selection sweeps detected in the Reggiana genome are located in QTL regions for milk and production efficiency traits. It is plausible to suggest that Reggiana might have a higher frequency of the less efficient and productive haplotypes for most of these regions, in addition to a general genomic background favoring heavy carcasses and high statures (as also inferred from the iHS analysis and the XP-EHH results). Taking together all these results, it could be possible to deduce that the Reggiana breed genome might still contain several signs of its multi-purpose and non-specialized utilization, as already described for other local cattle populations (Gutiérrez-Gil et al., 2015). The signatures that might address the adaptation (or re-adaptation) to the Parmigiano Reggiano production system (which cannot be simplified or summarized

with few genetic determinants) are therefore mixed and then diluted with other signatures that should have been derived by the history of the Reggiana cattle breed. It will be interesting to further evaluate the genetic background of the Reggiana ancestral genome architecture in comparisons with other autochthonous breeds of similar ancestry or with other local selection goals.

Conclusion

This study provided the first overview of genomic footprints in the Reggiana cattle breed. Several signatures, that have been probably left behind from the ancestral unspecialized purpose of Reggiana, have contributed to differentiate this breed and testify the diversity of this cattle genetic resource. Selection sweeps were located in a few chromosome regions already known to affect coat colour and morphological traits. Several other signatures might be the results of the slow re-adaptation of this breed to its peculiar production system, at present dominated by the Parmigiano-Reggiano cheese. Being constituted by a small and close population, genetic progress of Reggiana breed towards milk yield has been limited and its genomic footprint might reflect, in general, this productive weakness even if only indirect proof could be detected with the applied methods. Other studies are needed to evaluate what could be the achievable genetic progress on milk production traits in this breed.

Acknowledgements

We thank ANABORARE and the INNOVAGEN Consortium for the collaboration in this study. This study was funded by the PSRN Dual Breeding project and it received

494 financial support from the Italian Ministry of Agriculture, Food and Forestry (MiPAAF)
495 under the project INNOVAGEN, from the PON02_00451_3133441 project, CUP:
496 B61C1200076005 funded by MIUR and from the University of Bologna RFO 2018
497 program.
498

499 **Declaration of interests**

500 The authors declare that they do not have competing interests.
501

502 **Ethics statement**

503 No ethical approval was required since only genotyping data were used in the study and
504 data were provided by the previous research programs.
505

506 **Software and data repository resources**

507 None of the data were deposited in an official repository.
508

509 **References**

510 Alexander DH, Novembre J and Lange K 2009. Fast model-based estimation of ancestry in
511 unrelated individuals. Genome Research 19, 1655-6164.
512 ANABORARE - Associazione Nazionale Allevatori Bovini di Razza Reggiana 2019. Retrieved on
513 1st June 2019 from https://www.razzareggiana.it/.
514 Bertolini F, Galimberti G, Calò DG, Schiavo G, Matassino D and Fontanesi L 2015. Combined
515 use of principal component analysis and random forests identify population-informative
516 single nucleotide polymorphisms: application in cattle breeds. Journal of Animal Breeding
517 and Genetics 132, 346-356.

Bertolini F, Galimberti G, Schiavo G, Mastrangelo S, Di Gerlando R, Strillacci MG, Bagnato A, Portolano B and Fontanesi L 2018. Preselection statistics and Random Forest classification identify population informative single nucleotide polymorphisms in cosmopolitan and autochthonous cattle breeds. Animal 12, 12-19.

Browning BL and Browning SR 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics 84, 210-223.

Caroli A, Chessa S, Bolla P, Budelli E and Gandini GC 2004. Genetic structure of milk protein polymorphisms and effects on milk production traits in a local dairy cattle. Journal of Animal Breeding and Genetics 121, 119-127.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM and Lee JJ 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience 4, s13742–015–0047–8.

Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, Clark NR and Ma'ayan A 2013. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics 128, 14.

Flori L, Fritz S, Jaffrezic F, Boussaha M, Gut I, Heath S, Foulley JL and Gautier M 2009. The genome response to artificial selection: a case study in dairy cattle. PLoS One 4, e6595.

Fontanesi L, Scotti E and Russo V 2010a. Analysis of SNPs in the *KIT* gene of cattle with different coat colour patterns and perspectives for using these markers for breed traceability and authentication of beef and dairy products. Italian Journal of Animal Science 9, e42.

Fontanesi L, Tazzoli M, Russo V and Beever J 2010b. Genetic heterogeneity at the bovine *KIT* gene in cattle breeds carrying different putative alleles at the spotting locus. Animal Genetics 41, 295-303.

542 Fontanesi L, Scotti E and Russo V 2012. Haplotype variability in the bovine MITF gene and
543 association with piebaldism in Holstein and Simmental cattle breeds. Animal Genetics 43,
544 250-256.
545 Fontanesi L, Scotti E, Samorè AB, Bagnato A and Russo V 2015. Association of 20 candidate
546 gene markers with milk production and composition traits in sires of Reggiana breed, a
547 local dairy cattle population. Livestock Science 176, 14-21.
548 Gandini G, Maltecca C, Pizzi F, Bagnato A and Rizzi R 2007. Comparing local and commercial
549 breeds on functional traits and profitability: the case of Reggiana dairy cattle. Journal of
550 Dairy Science 90, 2004-2011.
551 Gautier M and Naves M 2011. Footprints of selection in the ancestral admixture of a New World
552 Creole cattle breed. Molecular Ecology 20, 3128-3143.
553 Gautier M, Klassmann A and Vitalis R 2017. rehh 2.0: a reimplementation of the R package rehh
554 to detect positive selection from haplotype structure. Molecular Ecology Resources 17, 78-
555 90.
556 González-Rodríguez A, Munilla S, Mouresan EF, Cañas-Álvarez JJ, Díaz C, Piedrafita J,
557 Altarriba J, Baro JÁ, Molina A and Varona L 2016. On the performance of tests for the
558 detection of signatures of selection: a case study with the Spanish autochthonous beef
559 cattle populations. Genetique Selection Evolution 48, 81.
560 Gutiérrez-Gil B, Arranz JJ, Wiener P 2015. An interpretive review of selective sweep studies in
561 Bos taurus cattle populations: identification of unique and shared selection signals across
562 breeds. Frontiers in Genetics 6, 167.
563 Hu Z-L, Park CA and Reecy JM 2019. Building a livestock genetic and genomic information
564 knowledgebase through integrative developments of Animal QTLdb and CorrDB. Nucleic
565 Acids Research 47, D701-D710.

566 Macciotta NPP, Biffani S, Bernabucci U, Lacetera N, Vitali A, Ajmone-Marsan P and Nardone A
567 2017. Derivation and genome-wide association study of a principal component-based
568 measure of heat tolerance in dairy cattle. Journal of Dairy Science 100, 4683-4697.
569 Mariani P and Russo V 1971. Distribuzione delle varianti genetiche delle caseine e della b-
570 lattoglobulina nelle vacche di razza Reggiana. Rivista di Zootecnia 44, 310-322.
571 Marras G, Gaspa G, Sorbolini S, Dimauro C, Ajmone-Marsan P, Valentini A, Williams JL and
572 Macciotta NP 2015. Analysis of runs of homozygosity and their relationship with inbreeding
573 in five cattle breeds farmed in Italy. Animal Genetics 46, 110-121.
574 Mastrangelo S, Ciani E, Ajmone-Marsan P, Bagnato A, Battaglini L, Bozzi R, Carta A, Catillo G,
575 Cassandro M, Casu S, Ciampolini R, Crepaldi P, D'Andrea M, Di Gerlando R, Fontanesi L,
576 Longeri M, Macciotta NPP, Mantovani R, Marletta D, Matassino D, Mele M, Pagnacco G,
577 Pieramati C, Portolano B, Sarti MF, Tolone M and Pilla F 2018a. Conservation status and
578 historical relatedness of Italian cattle breeds. Genetics Selection Evolution 50, 35.
579 Mastrangelo S, Sardina MT, Tolone M, Di Gerlando R, Sutera AM, Fontanesi L and Portolano B
580 2018b. Genome-wide identification of runs of homozygosity islands and associated genes
581 in local dairy cattle breeds. Animal 12, 2480-2488.
582 Mastrangelo S, Tolone M, Di Gerlando R, Fontanesi L, Sardina MT and Portolano B 2016.
583 Genomic inbreeding estimation in small populations: evaluation of runs of homozygosity in
584 three local dairy cattle breeds. Animal 10, 746-754.
585 Rocha D, Billerey C, Samson F, Boichard D and Boussaha M 2014. Identification of the putative
586 ancestral allele of bovine single-nucleotide polymorphisms. Journal of Animal Breeding
587 and Genetics 131, 483-486.
588 Rothammer S, Seichter D, Förster M and Medugorac I 2013. A genome-wide scan for signatures
589 of differential artificial selection in ten cattle breeds. BMC Genomics 14, 908.
590 Russo V, Fontanesi L, Scotti E, Tazzoli M, Dall'Olio S, Davoli R 2007. Analysis of melanocortin 1
591 receptor (*MC1R*) gene polymorphisms in some cattle breeds: their usefulness and

592 application for breed traceability and authentication of Parmigiano Reggiano cheese.
593 Italian Journal of Animal Science 6, 257-272.
594 Sabeti PC, Varilly P, Fry B, et al. 2007. Genome-wide detection and characterization of positive
595 selection in human populations. Nature 449, 913-918.
596 Scheet P and Stephens M 2006. A fast and flexible statistical model for large-scale population
597 genotype data: applications to inferring missing genotypes and haplotypic phase.
598 American Journal of Human Genetics 78, 629-644.
599 Takasuga A 2016. PLAG1 and NCAPG-LCORL in livestock. Animal Science Journal 87, 159-
600 167.
601 Utsunomiya YT, Milanesi M, Utsunomiya ATH, Torrecilha RBP, Kim ES, Costa MS, Aguiar TS,
602 Schroeder S, do Carmo AS, Carvalheiro R, Neves HHR, Padula RCM, Sussai TS, Zavarez
603 LB, Cipriano RS, Caminhas MMT, Hambrecht G, Colli L, Eufemi E, Ajmone-Marsan P,
604 Cesana D, Sannazaro M, Buora M, Morgante M, Liu G, Bickhart D, Van Tassell CP,
605 Sölkner J, Sonstegard TS and Garcia JF 2017. A PLAG1 mutation contributed to stature
606 recovery in modern cattle. Scientific Reports 7, 17140.
607 Voight BF, Kudaravalli S, Wen X and Pritchard JK 2006. A map of recent positive selection in
608 the human genome. PLoS Biology 4, e72.
609 Wright S 1951. The genetical structure of populations. Annals of Eugenetics15, 323-354.
610 Zhao F, McParland S, Kearney F, Du L and Berry DP 2015. Detection of selection signatures in
611 dairy and beef cattle using high-density genomic information. Genetics Selection Evolution
612 47, 49.

Table 1. List of the top 10 integrated haplotype score measures |iHS| for the single nucleotide polymorphism (SNP) markers and their closest genes, with information on the chromosome (BTA) position. A complete list is reported in Supplementary Table S2.

SNP	BTA	Position	|iHS|	Closest genes on both SNP sides	Distance (bp)[1]
ARS-BFGL-NGS-39422	29	16996267	3.868	*TENM4*	0
BTB-01391891	2	54536305	3.833	*KYNU - HIGD1A*	647722 - 443471
Hapmap40017-BTA-65421	29	31971753	3.697	*ETS1 - FLI1*	21920 - 212503
ARS-BFGL-NGS-39172	29	36048959	3.679	*TMEM45B*	0
ARS-BFGL-NGS-52511	29	30103989	3.638	*KIRREL3 - ENSBTAG00000050013*	222182 - 448314
Hapmap49404-BTA-100549	2	70190436	3.637	*INSIG2 - ENSBTAG00000050695*	403503 -138266
ARS-BFGL-NGS-9657	17	46895925	3.561	*PIWIL1 - FZD10*	133818 - 36824
BTB-00247622	6	16367079	3.532	*ENSBTAG00000049691 – ETNPPL*	451494 - 186958
ARS-BFGL-NGS-18412	29	28560818	3.516	*TMEM218 - PKNOX2*	11801 - 249050
Hapmap58618-rs29012371	29	32801728	3.502	*ENSBTAG00000055310 - JAM3*	87554 - 192247

[1] Zero indicates that the SNP is within the reported gene. Two distances are reported when the SNP is between the indicated genes.

Table 2. Selection sweeps identified by more than one test in the Reggiana chromosomes (BTA) and annotated genes in these regions.

Tests[1]	BTA (Start-End) [2]	Annotated genes
F_{ST} (Brown); XP-EHH (Simmental)	6 (68331252-70500000)	*RF00568, GSX2, RF00026, RF00026, USP46, RASL11B, CHIC2, KIT, SCFD2, FIP1L1, LNX1, PDGFRA*
F_{ST} (Brown); XP-EHH (Brown)	6 (70431058-70500000)	-
F_{ST} (Simmental); XP-EHH (Simmental)	6 (68331252-71428675)	*RF00568, GSX2, RF00026, USP46, RASL11B, CHIC2, KIT, KDR, SRD5A3, TMEM165, PDCL2, EXOC1L, CEP135, SCFD2, FIP1L1, LNX1, PDGFRA, CLOCK, NMU, EXOC1*
F_{ST} (Simmental); XP-EHH (Brown)	6 (70431058-70716954)	*RF00026, KDR*
F_{ST} (Simmental); XP-EHH (Brown)	6 (91500000-91800062)	*SOWAHB, SEPT11, SHROOM3*
F_{ST} (Simmental); XP-EHH (Brown)	6 (92383295-92430962)	*CNOT6L*
F_{ST} (Simmental); XP-EHH (Simmental)	7 (43047351-43105247)	*C2CD4C, MIER2, THEG*
XP-EHH (Holstein); iHS	13 (44978611-45082428)	*PITRM1*
XP-EHH (Holstein); iHS	13 (45936412-46049129)	*ADARB2*
F_{ST} (Holstein); iHS	17 (18961407-19000000)	-
F_{ST} (Holstein); iHS	26 (21500000-21534491)	-
F_{ST} (Holstein); iHS	26 (22661478-23500000)	*HPS6, RF00099, PITX3, NFKB2, FBXL15, TRIM8, CYP17A1, CYP17A1, LDB1, NOLC1, ELOVL3, PSD, CUEDC2, MFSD13A, ACTR1A, ARL3, WBP1L, CYP17A1, ARMH3, PPRC1, GBF1, SUFU, SFXN2*
F_{ST} (Brown); iHS	29 (27381130-27500000)	-

[1] Selection signature detection methods are reported, including the breed used in the fixation index (F_{ST}) or Cross population extended haplotype homozygosity (XP-EHH) comparisons.

[2] Chromosome positions are given in bp on the cattle reference genome for that chromosome (BTA). Regions are identified by combining positions of selection signatures derived by the different approaches. Integrated haplotype score (iHS) signal borders were defined as ± 500 kb from the detected single nucleotide polymorphisms.

List of figure captions

Figure 1 Multidimensional scaling (MDS) plots of the four investigated cattle breeds obtained with the single nucleotide polymorphism chip data. The plot on the left shows the distribution of the first (C1) and the second (C2) coordinates. The plot on the right shows the distribution of the first (C1) and the third (C3) coordinates.

Figure 2 Plot of the integrated haplotype score (iHS) analysis on the Reggiana breed. The |iHS| value corresponding to the bottom of the 99.5th percentile distribution was = 2.754 and is indicated with the red line in the Manhattan plot.

Figure 3 Manhattan plots showing the results of the mean fixation index F_{ST} (mF_{ST}) analyses against the Brown (a), Holstein (b) and Simmental (c) breeds. The red line in each plot represents the bottom of the 99.5th percentile distribution that is equal to 0.287, 0.279 and 0.154 for the comparisons against the Brown, Holstein and Simmental breeds, respectively.

Figure 4 Cross population extended haplotype homozygosity (XP-EHH) analyses for Brown (a), Holstein (b) and Simmental (c) against the Reggiana. For each figure, the red line represents the bottom of the 99.5th percentile distribution.

Figure 5 Genomic footprint map of the Reggiana breed, including selection signatures obtained with the three used approaches (integrated haplotype score, iHS; fixation index, F_{ST}; Cross-Population Extended Haplotype Homozygosity, XP-EHH).

Figure 1

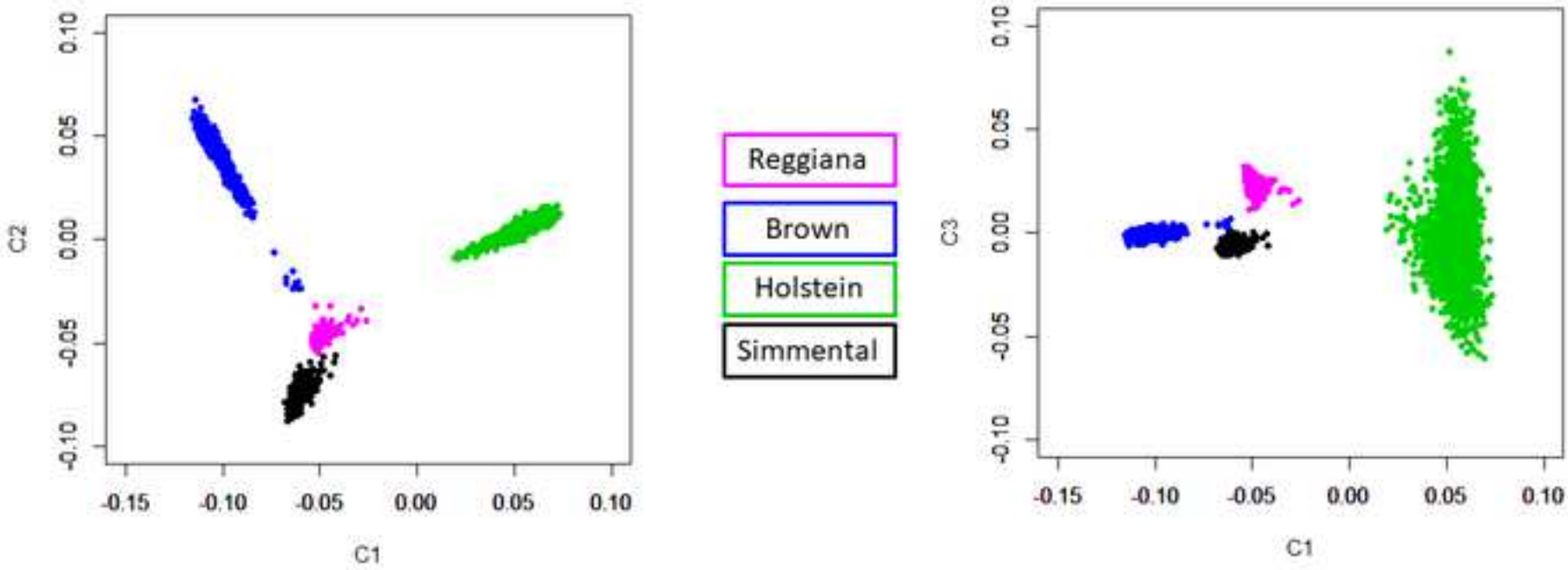


Figure 2

Click here to access/download;Figure;Figure 2 HQ.tif

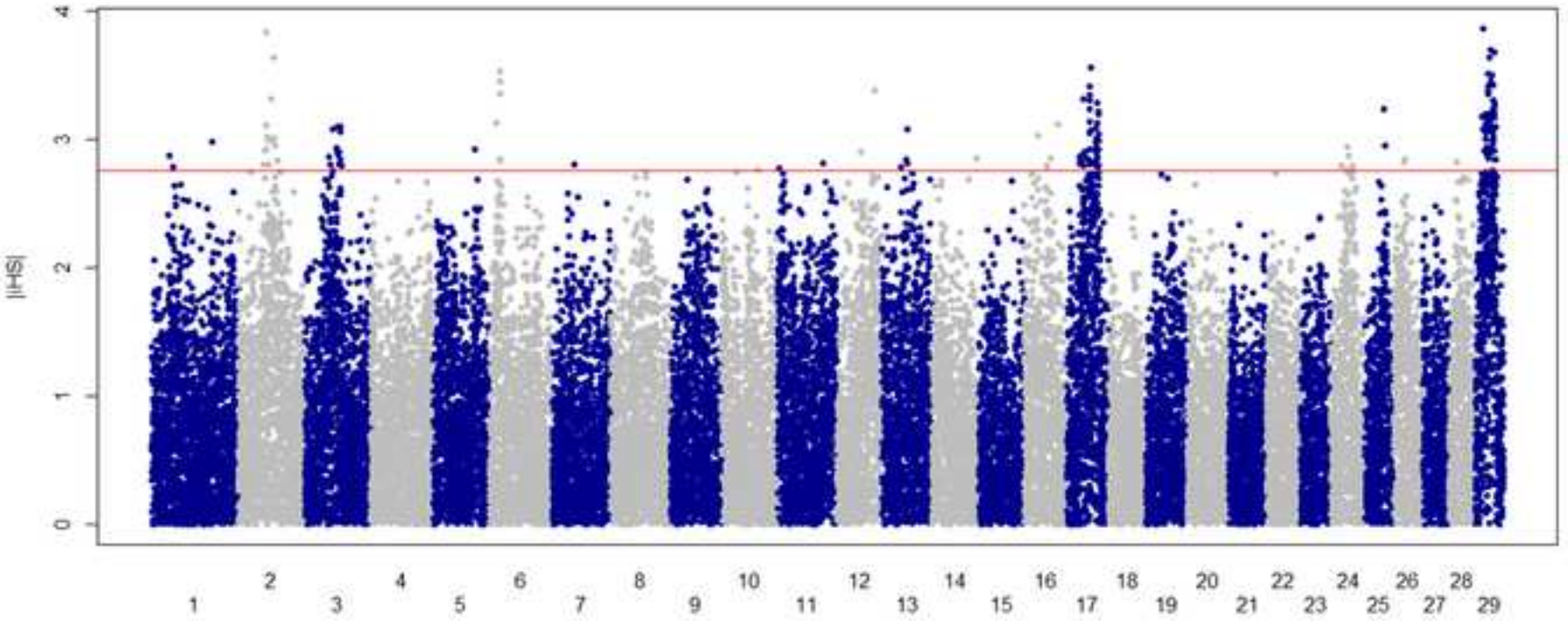


Figure 3

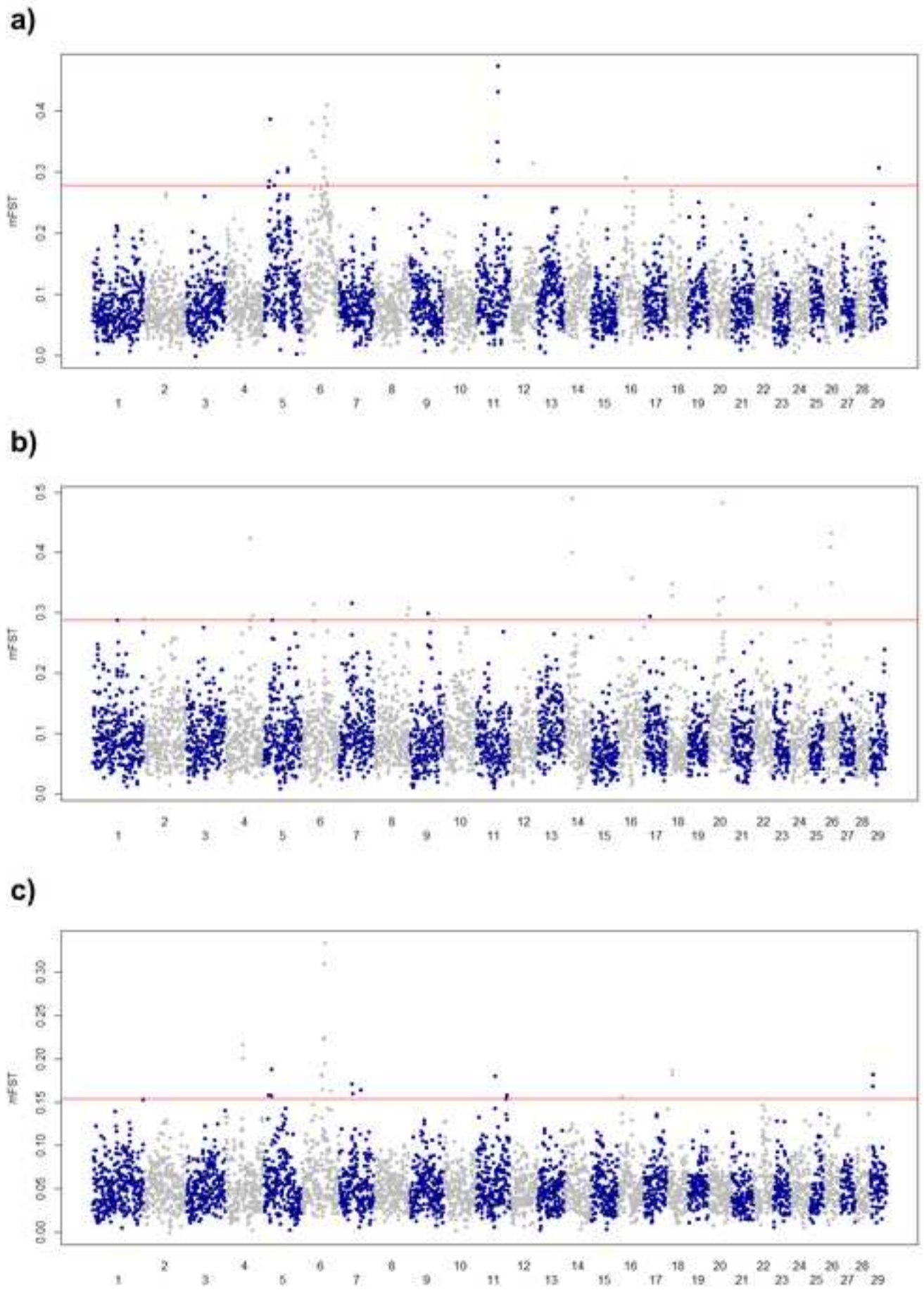


Figure 4

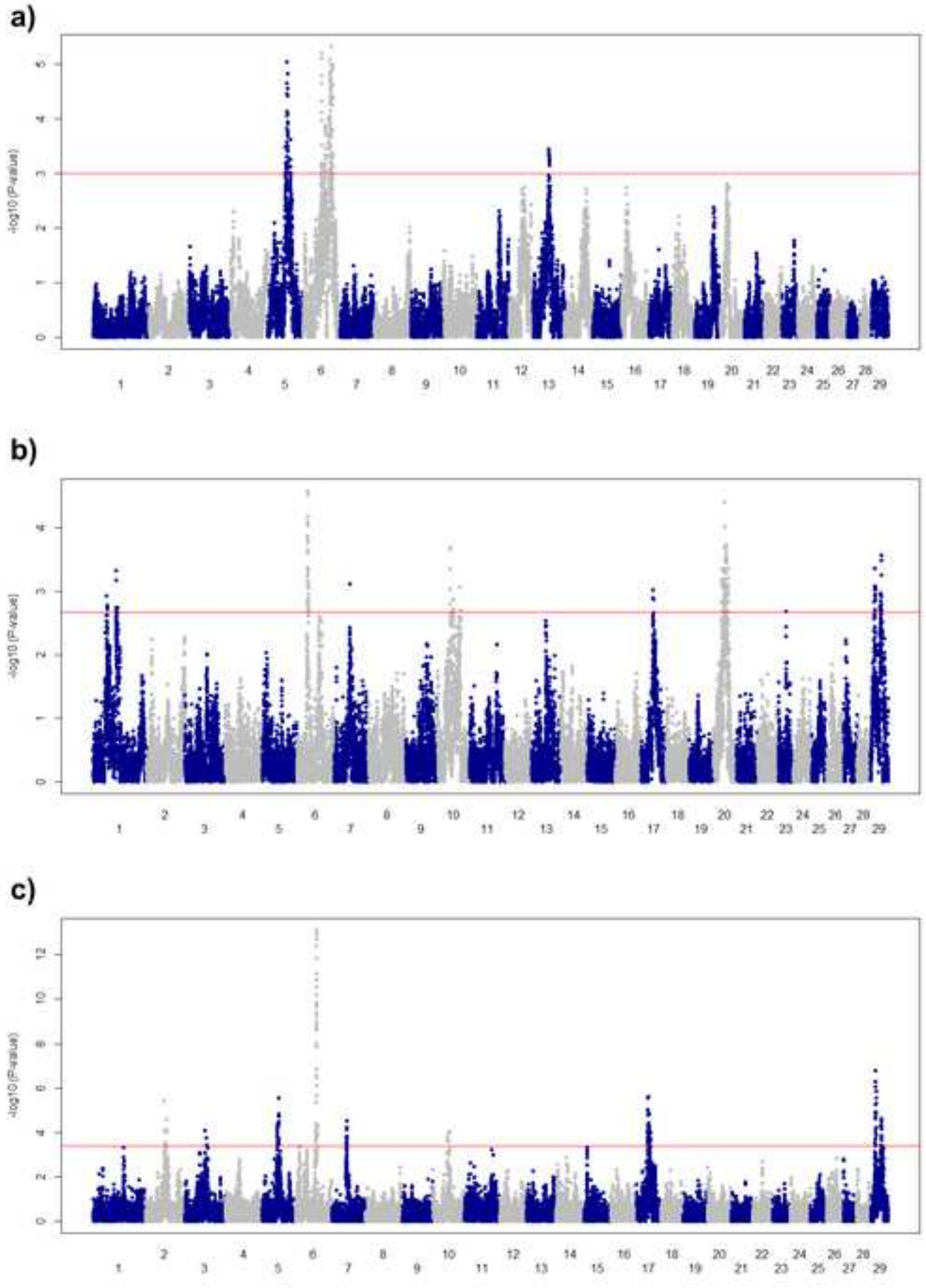
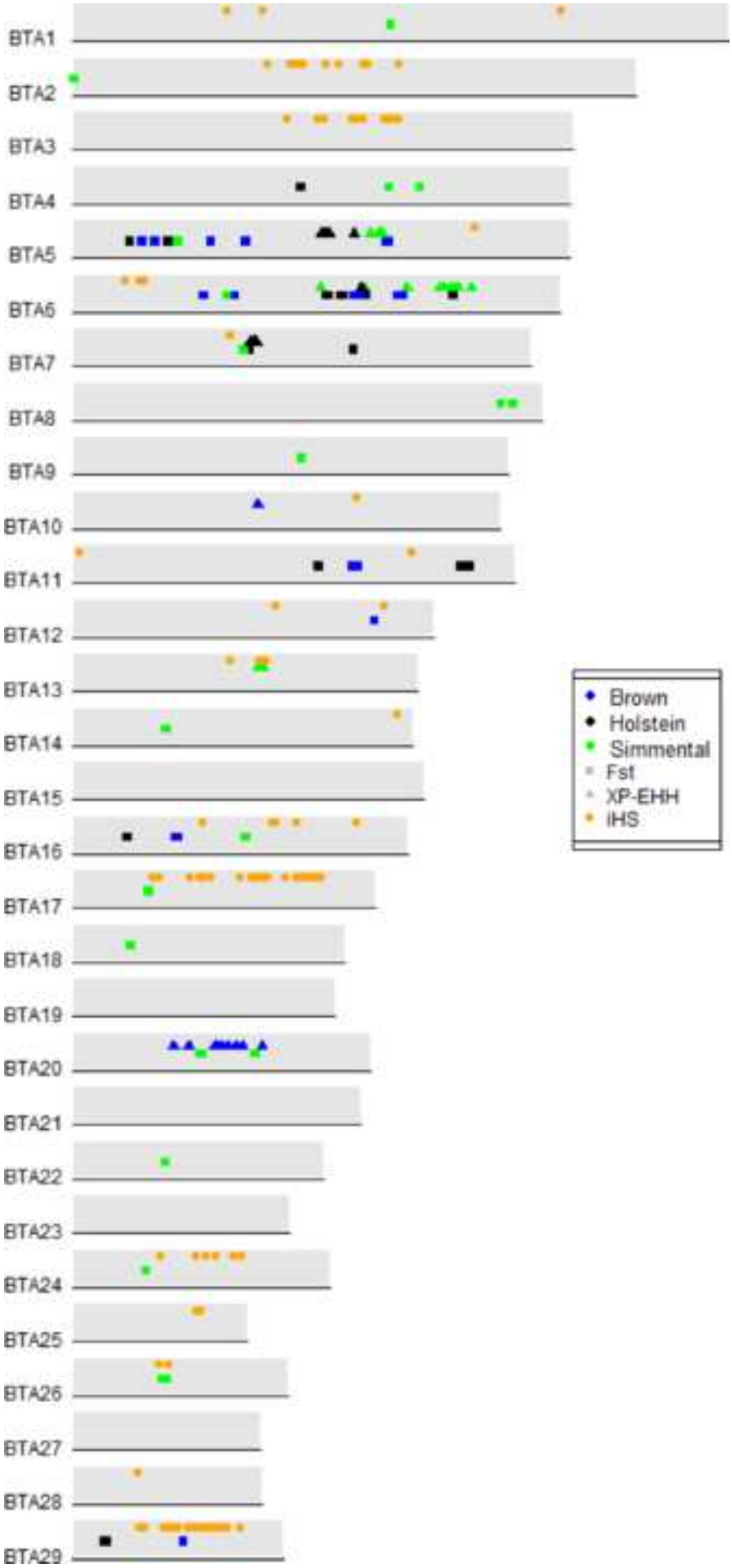


Figure 5



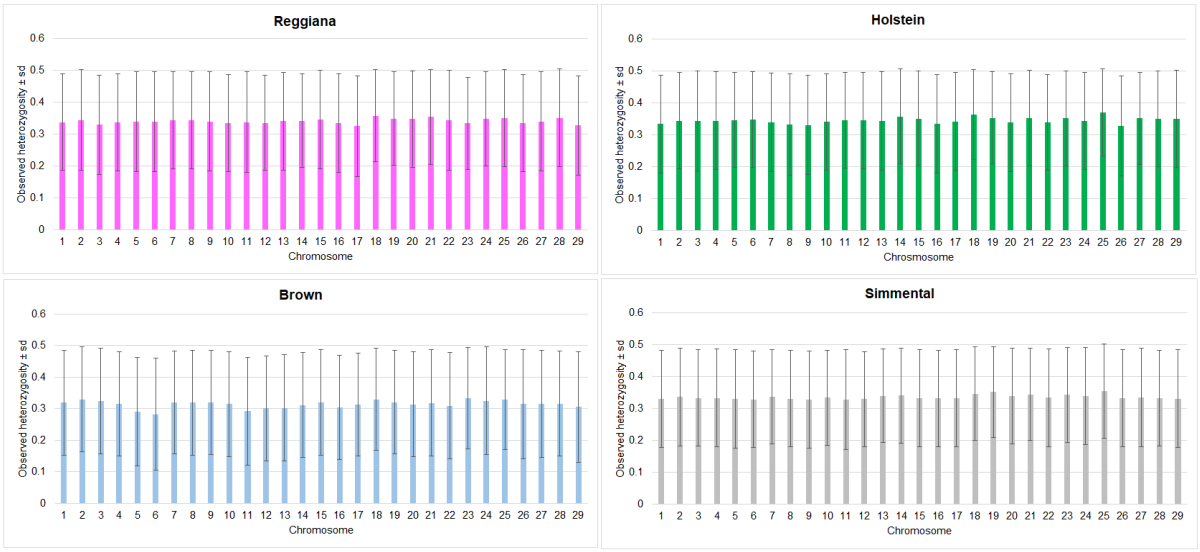
Supplementary Material for:

Comparative selection signature analyses identify genomic footprints in Reggiana cattle, the traditional breed of the Parmigiano Reggiano cheese production system

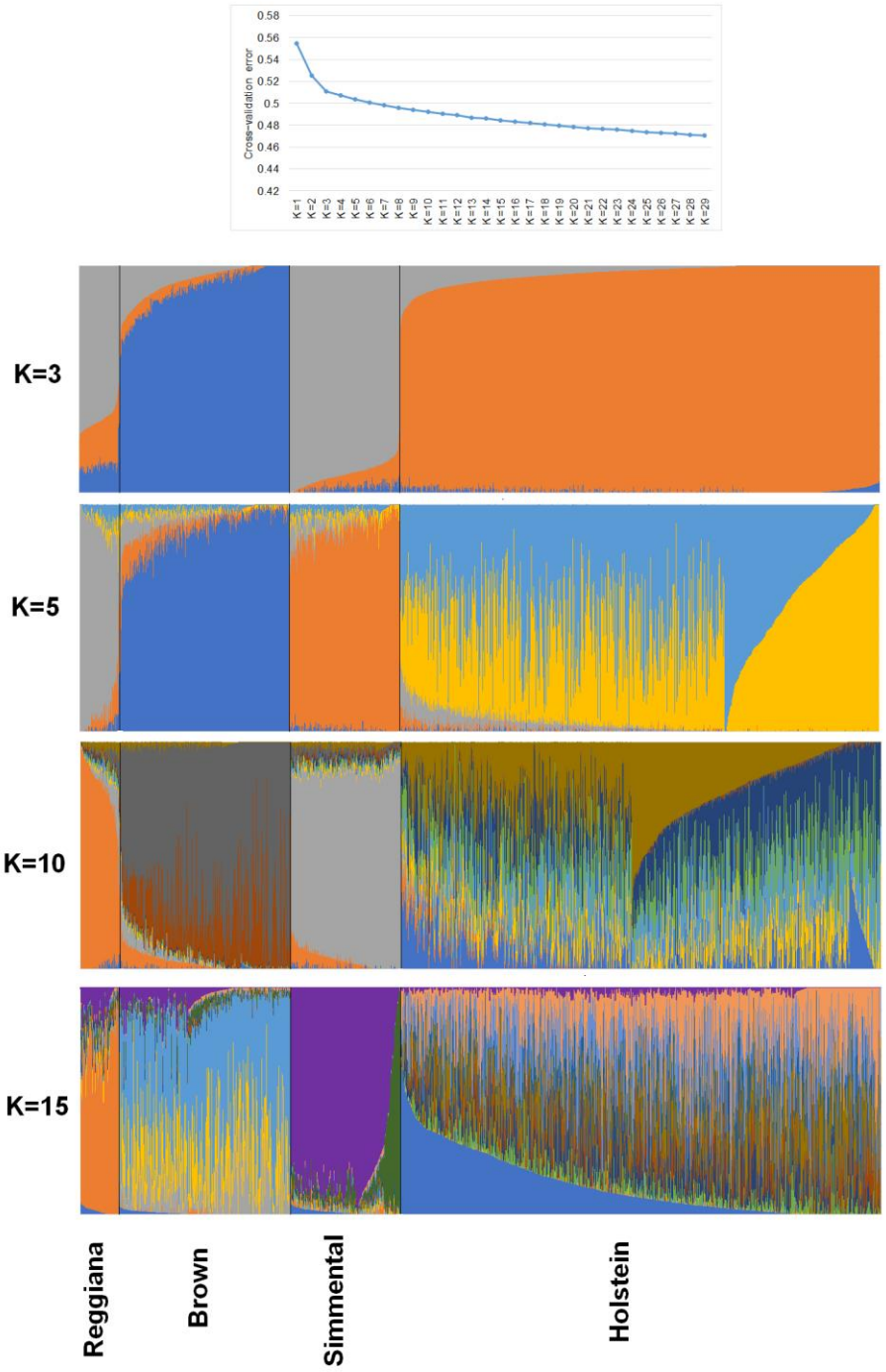
Francesca Bertolini, Giuseppina Schiavo, Samuele Bovo, Maria Teresa Sardina, Salvatore Mastrangelo, Stefania Dall'Olio, Baldassare Portolano and Luca Fontanesi

Animal Journal

Supplementary Figure S1. Average observed heterozygosity (± standard deviation) for each chromosome in the four analysed cattle breeds.



Supplementary Figure S2. Results of the ADMIXTURE analysis. Cross validation error with K from 1 to 29 (top) and plot distribution with K=3, 5, 10 and 15 of the considered breeds.



Supplementary Table S1. Summary of genotyping data: minor allele frequency (MAF) and observed heterozygosity (Het) and their standard deviation (SD) in Reggiana, Brown, Holstein and Simmental cattle breeds.

Breed	Average MAF ± SD	Average Het ± SD
Reggiana	0.253 ± 0.145	0.340 ± 0.153
Brown	0.232 ± 0.152	0.313 ± 0.168
Holstein	0.259 ± 0.146	0.344 ± 0.152
Simmental	0.251 ± 0.145	0.334 ± 0.151

Supplementary Table S2. Results of the integrated haplotype score (iHS) analysis in the Reggiana breed including the top 99.5th percentile single nucleotide polymorphisms (SNPs). Reported information includes bovine chromosome (BTA) position of the SNP, the SNP name, the |iHS| value and the gene symbol of the annotated genes ± 200 kbp from the SNP.

BTA[1]	SNP[2]	Position	|iHS| value	Annotated genes
1	DPI-27	37381473	2.874	*EPHA3*
1	ARS-BFGL-NGS-114234	45929427	2.785	*TRMT10C, ZBTB11, SENP7, PCNP*
1	BTB-00052487	117704023	2.988	*TSC22D2*
2	ARS-BFGL-NGS-32858	47013329	2.804	*LYPD6B*
2	Hapmap41178-BTA-120553	52511673	2.916	*GTDC1*
2	BTB-01412441	53545307	2.925	*ARHGAP15*
2	BTB-01391891	54536305	3.833	-
2	BTB-00183384	54634481	3.114	-
2	Hapmap60963-rs29015781	55779254	3.020	*LRP1B*
2	Hapmap48387-BTA-55128	61253063	2.988	*CXCR4*
2	Hapmap41674-BTA-88236	61300454	2.803	*CXCR4, DARS*
2	ARS-BFGL-NGS-41523	64151134	3.322	*NCKAP5*
2	Hapmap49404-BTA-100549	70190436	3.636	-
2	Hapmap35220-BES9_Contig365_495	70532188	2.992	*EN1*
2	ARS-BFGL-NGS-1606	70931242	3.013	*C1QL2, STEAP3*
2	ARS-BFGL-NGS-16745	71437395	2.949	*TMEM177, CFAP221*
2	UA-IFASA-2241	78787839	2.837	*RF00612, GYPC*
3	ARS-USMARC-Parent-AY842474-rs29003226	51817697	2.864	*CDC7*
3	ARS-BFGL-NGS-35164	51841394	2.810	*CDC7*
3	INRA-451	59220453	3.082	*MCOLN2, LPAR3*
3	Hapmap53609-rs29011253	60566649	2.760	-
3	BTB-00133369	67338807	2.935	*PIGK, AK5*
3	BTA-94549-no-rs	68497585	2.926	-
3	BTB-00135094	70009576	3.097	*TYW3, CRYZ*

3	BTA-10440-no-rs	70057563	2.885	*TYW3, CRYZ*
3 (ROH)	ARS-BFGL-NGS-32439	75251155	2.902	*LRRC7*
3 (ROH)	ARS-BFGL-NGS-8612	75279283	2.806	*LRRC7*
3 (ROH)	INRA-142	76452556	2.842	-
3 (ROH)	BTB-00137261	77599781	3.102	*RF00026, GADD45A, GNG12*
3 (ROH)	BTB-00137287	77652594	3.058	*RF00026, GADD45A, GNG12*
3 (ROH)	BTB-01168089	78669044	2.792	*TCTEX1D1, SGIP1*
5	ARS-BFGL-NGS-5720	96995660	2.921	*GPRC5D, APOLD1, HEBP1, GPRC5A, DDX47*
6	BTA-108507-no-rs	12478100	3.135	-
6	ARS-BFGL-NGS-107549	16334539	3.457	-
6	BTB-00247622	16367079	3.532	-
6	Hapmap44568-BTA-77505	16407876	2.832	-
6	ARS-BFGL-NGS-80568	17056293	3.353	*LEF1*
6	ARS-BFGL-NGS-45046	17498903	2.844	*PAPSS1*
7	ARS-BFGL-NGS-35666	38039999	2.803	*RNF44, EIF4E1B, FAF2, CDHR2, SNCB, TSPAN17*
10	ARS-BFGL-NGS-57077	68564596	2.762	*PELI2*
11	Hapmap48122-BTA-91937	1818651	2.773	*MALL, MAL, NPHP1*
11	ARS-BFGL-NGS-105586	81697133	2.816	*FAM49A*
12	BTB-01544419	49153380	2.900	-
12	Hapmap24871-BTA-157401	75360273	3.390	*RF00026, FARP1*
13	UA-IFASA-7733	37955994	2.791	*DSTN, BANF2, BFSP1, RRBP1*
13	ARS-BFGL-NGS-18246	45034339	2.844	*PITRM1*
13	ARS-BFGL-NGS-35887	46089659	3.083	*ADARB2*
13	ARS-BFGL-NGS-107916	46808920	2.817	*DIP2C*
14	ARS-BFGL-NGS-40495	78423753	2.852	-
16	Hapmap49866-BTA-114054	31458337	3.032	*SMYD3*
16	Hapmap41252-BTA-39046	48366828	2.796	-
16	Hapmap60283-rs29014986	48816898	2.784	-

16	Hapmap26379-BTA-130999	53997980	2.860	-
16	Hapmap49429-BTA-107409	68461967	3.124	*PDPN, PRDM2*
17	BTB-01087937	19461407	2.865	*SLC7A11*
17	Hapmap51443-BTA-40619	20682264	2.814	-
17	BTB-01731152	28148532	2.822	*RF00100*
17	BTB-01869986	30566646	3.319	-
17	BTB-01426795	30798193	2.887	-
17	ARS-BFGL-NGS-74608	31986086	2.798	-
17	BTB-01308307	33433050	2.900	-
17	BTA-122399-no-rs	40371968	2.859	*RXFP1*
17	BTB-00676954	43133674	2.848	-
17	Hapmap29721-BTA-131409	44455186	2.806	*PGAM5, RF00026, CHFR, ANKLE2, PXMP2, GOLGA3, POLE*
17	ARS-BFGL-NGS-27620	44586221	3.048	*PGAM5, RF00026, P2RX2, LRCOL1, ANKLE2, PXMP2, POLE, FBRSL1*
17	ARS-BFGL-NGS-105739	44725907	3.057	*FBRSL1*
17	ARS-BFGL-NGS-103650	44812935	3.342	*GALNT9*
17	ARS-BFGL-NGS-86713	45036186	3.304	*NOC4L, RF00562, bta-mir-6520, DDX51, GALNT9, EP400*
17	BTB-00678060	45302168	2.929	*bta-mir-2285af-2, MMP17, SFSWAP*
17	BTA-41003-no-rs	45331881	2.824	*bta-mir-2285af-2, MMP17, SFSWAP*
17	Hapmap38384-BTA-117953	45517219	3.143	-
17	ARS-BFGL-NGS-79176	45561948	3.237	-
17	Hapmap28761-BTA-159815	45637897	2.825	-
17	ARS-BFGL-BAC-27022	45697242	3.420	-
17	ARS-BFGL-NGS-9657	46895925	3.561	*FZD10*
17	ARS-BFGL-NGS-115236	47192942	3.068	*RF00026, RF00100, RFLNA, NCOR2*
17	Hapmap39519-BTA-85553	51369492	2.824	-
17	Hapmap49158-BTA-41145	53987189	2.924	*CAMKK2, P2RX4, P2RX7, IFT81*
17	BTB-00679561	54482037	2.922	*PPTC7, TCTN1, PPP1CC, RAD9B, HVCN1*

17	ARS-BFGL-NGS-5696*	54765150	3.000	*MYL2, CCDC63, CUX2*
17	Hapmap43572-BTA-41227*	54798284	2.797	
17	ARS-BFGL-NGS-21400*	54862080	2.908	
17 (ROH)	ARS-BFGL-BAC-34676*	54925511	2.806	
17 (ROH)	Hapmap48751-BTA-41232	55767778	2.829	*TMEM233, CIT*
17 (ROH)	BTA-25636-no-rs	55916308	2.976	*CCDC60*
17 (ROH)	ARS-BFGL-NGS-54784	56479267	3.087	-
17 (ROH)	BTB-00680019	56987958	2.853	*TAOK3*
17 (ROH)	BTB-00680348*	57375842	2.878	*RF00026, KSR2, NOS1*
17 (ROH)	ARS-BFGL-NGS-112404*	57495561	3.128	
17 (ROH)	ARS-BFGL-NGS-116162*	57562662	2.913	
17 (ROH)	ARS-BFGL-NGS-10055*	57716043	2.930	
17 (ROH)	ARS-BFGL-NGS-27713	57937352	3.291	*TESC, NOS1, FBXO21FBXW8*
17 (ROH)	BTB-00681880*	58741798	3.214	*MED13L*
17 (ROH)	BTB-00681839*	58796978	3.044	
17 (ROH)	ARS-BFGL-NGS-54448*	58989148	2.991	
17 (ROH)	ARS-BFGL-NGS-75591	59530888	2.800	-
17 (ROH)	ARS-BFGL-NGS-1369	59560283	2.924	-
17	BTB-01095540	59953298	3.201	-
24	Hapmap38797-BTA-99366	21431369	2.800	*GALNT1*
24	ARS-BFGL-NGS-53865	29888368	2.767	*CHST9*
24	BTA-57840-no-rs	32183082	2.941	*HRH4, IMPACT*
24	BTB-00886759	34511917	2.875	*ABHD3, MIB1*
24	Hapmap57118-rs29009938	38934184	2.769	*EPB41L3*
24	ARS-BFGL-NGS-98552	40627809	2.798	*PTPRM*
25	ARS-BFGL-NGS-42870	29938832	3.235	-
25	Hapmap23619-BTC-057878	31203678	2.957	*RF00026*
26	BTB-00930720	21034491	2.830	*RF00026, ERLIN1, DNMBP, CHUK, CPN1*

26	ARS-BFGL-NGS-119202	23161478	2.849	*MFSD13A, ACTR1A, SUFU*
28	ARS-BFGL-NGS-41944	15601752	2.830	*CCDC6, ANK3*
29 (ROH)	ARS-BFGL-NGS-118102	16069820	3.182	-
29 (ROH)	ARS-BFGL-NGS-2529	16569370	3.077	*bta-mir-708*
29 (ROH)	ARS-BFGL-NGS-39422	16996267	3.867	*TENM4*
29 (ROH)	ARS-BFGL-NGS-1508	17291020	3.178	*TENM4*
29 (ROH)	BTB-01012731	22198059	2.919	*GAS2*
29 (ROH)	BTA-106551-no-rs	22648358	3.197	*ANO5*
29	BTB-00934783	23661968	3.050	-
29	Hapmap43319-BTA-65094	24529656	3.116	*PRMT3*
29	BTB-01013468	24721139	3.005	*NAV2*
29	Hapmap41325-BTA-65112	24898216	2.885	*NAV2*
29	ARS-BFGL-NGS-9185	24940506	3.077	*bta-mir-449d*
29	ARS-BFGL-NGS-39535	25169363	3.180	*NAV2*
29	ARS-BFGL-NGS-91937	25338021	2.944	*RF00408*
29	ARS-BFGL-NGS-56290	25466622	3.145	.
29	ARS-BFGL-NGS-12494	25632177	2.883	*E2F8*
29	Hapmap45305-BTA-65247	27881130	2.859	*OR8B4*
29	ARS-BFGL-NGS-94355	28154066	3.373	*PANX3, TBRG1, NRGN, ESAM, SIAE, SPA17, VSIG2, MSANTD2*
29	BTB-01017247*	28440651	2.928	*CCDC15, HEPACAM,*
29	ARS-BFGL-NGS-37244*	28518134	3.356	*RF00100, ROBO3, ROBO4, SLC37A2, TMEM218*
29	ARS-BFGL-NGS-102700*	28539785	3.083	
29	ARS-BFGL-NGS-18412	28560818	3.516	*SLC37A2, TMEM218*
29	UA-IFASA-5034	28684366	2.872	-
29	ARS-BFGL-NGS-23652	28731093	3.411	*PKNOX2*
29	ARS-BFGL-NGS-25532	29117496	3.082	*SSLP1, PATE1, bta-mir-2285ce, PATE2, PATE3, RF00099, STT3A, CHEK1, ACRV1*

29	ARS-BFGL-NGS-105093	29536191	2.895	*SRPRA, RPUSD4, FAM118B, FOXRED1*
29	ARS-BFGL-NGS-17769	29928524	2.930	*KIRREL3*
29	ARS-BFGL-NGS-29938	29953458	3.429	*KIRREL3*
29	ARS-BFGL-NGS-52511*	30103989	3.638	-
29	Hapmap40782-BTA-65467*	30157180	3.149	
29	BTA-65463-no-rs*	30188256	3.184	
29	ARS-BFGL-NGS-109714	31764186	2.933	*ETS1*
29	BTB-01020010	31875920	3.165	*ETS1*
29	Hapmap40017-BTA-65421	31971753	3.697	*ETS1*
29	ARS-BFGL-NGS-87575*	32078660	3.455	*FLI1, KCNJ1, KCNJ5, FLI1, KCNJ1, ARHGAP32*
29	ARS-BFGL-NGS-12309*	32128256	3.210	
29	UA-IFASA-7219*	32253717	3.111	
29	Hapmap58618-rs29012371	32801728	3.502	-
29	BTB-01023253	32940189	2.905	*JAM3*
29	Hapmap49699-BTA-65589	33406139	3.084	-
29	Hapmap60712-rs29014894	33436030	2.989	-
29	ARS-BFGL-NGS-24911	33487878	2.909	-
29	ARS-BFGL-NGS-34282	33704537	3.060	-
29	ARS-BFGL-NGS-36490	34592350	3.303	*OPCML*
29	UA-IFASA-9704	34687068	3.022	*NTM*
29	ARS-BFGL-NGS-12285	34751748	2.976	*NTM*
29	Hapmap53268-rs29022154	34808241	3.001	*NTM*
29	UA-IFASA-6129	34835983	3.236	*NTM*
29	ARS-BFGL-NGS-28392*	35237784	3.426	*RF00619*
29	ARS-BFGL-NGS-115969*	35373548	3.039	
29	ARS-BFGL-NGS-89027*	35393862	2.754	
29	BTB-01027202*	35564539	3.085	
29	ARS-BFGL-NGS-39172	36048959	3.679	*NFRKB, PRDM10, TMEM45B*

29	ARS-BFGL-NGS-2990	36088900	3.211	*NFRKB, PRDM10, TMEM45B*
29	ARS-BFGL-NGS-29244	36130086	3.291	*NFRKB, PRDM10, TMEM45B*
29	ARS-BFGL-NGS-4431	36619214	3.258	*ADAMTS8, ADAMTS15*
29	ARS-BFGL-NGS-17583	36932109	3.173	*SNX19*
29	ARS-BFGL-NGS-111280	36986707	2.845	*MS4A8, SNX19, MS4A18*
29	ARS-BFGL-NGS-56408	37168472	2.916	*MS4A15, PTGDR2, MS4A8, MS4A10, CCDC86, TMEM109, TMEM132A, MS4A18, PRPF19*
29	ARS-BFGL-NGS-104963	40438138	3.100	*FTH1, FADS1, FADS3, RAB3IL1, BEST1, FADS2*
29	ARS-BFGL-NGS-110249	40565850	2.845	*FTH1, BEST1, INCENP*

[1] Chromosome regions overlapping runs of homozygosity (ROH) islands reported by Mastrangelo *et al.* (2018b) in the Reggiana cattle breed are indicated with "(ROH)".
[2] Consecutive SNPs are indicated with an asterisk "*") in the SNP column.

Supplementary Table S3. Results of the pairwise fixation index F_{ST} analysis of Reggiana *vs* the three cosmopolitan breeds (Brown, Holstein and Simmental). Reported information includes the bovine chromosome (BTA) position (start and end nucleotide position on the chromosome) of the top 99.5th percentile of the mean F_{ST} values in 1 Mbp sliding windows and the annotated genes in the corresponding chromosome regions. Overlapping or adjacent windows were merged. However, the total number of all sliding windows (overlapping or partially overlapping) is reported in the text.

BTA[1]	Starting window position	End window position	mFST	Annotated genes in the genomic windows
Brown				
5	16500000	17500000	0.286	-
5	19500000	20500000	0.387	*RF00001, ATP2B1*
5	33000000	34000000	0.279	*AMIGO2, SLC38A2, PCED1B, SLC38A4, SLC38A1*
5	41500000	42500000	0.300	*bta-mir-2428, RF00026, ABCD2, KIF21A, CPNE8*
5	75500000	77000000	0.307	*SSTR3, bta-mir-1835, ELFN2, RF00026, RF00026, ALG10, IL2RB, RAC2, CYTH4, CARD10, USP18, C1QTNF6, MFNG, SYT10, TMPRSS6, USP18, SYT10*
6	31000000	32500000	0.335	*RF00026, GRID2*
6	39000000	40000000	0.326	*SLIT2*
6	67500000	70500000	0.307	*bta-mir-4449, CHIC2, CWH43, DCUN1D4, FIP1L1, GSX2, KIT, LNX1, LRRC66, OCIAD1, OCIAD2, PDGFRA, RASL11B, RF00026, RF00568, SCFD2, SGCB, SPATA18, USP46,*
6	79000000	80500000	0.279	*RF00001, TECRL, RF00001, TECRL, BMP10, GKN1, RF00100, PROKR1, GKN3P, GKN2, NFU1, ANTXR1, GFPT1, AAK1, ARHGAP25*
11	67000000	69500000	0.350	*ANTXR1, ANXA4, ASPRV1, C11H2orf42, CAPN13, CAPN14, EHD3, GALNT14, GFPT1, GMCL1, LCLAT, MXD1, NFU1, PCBP1, PCYOX1, SNRNP27, SNRPG, TIA1, AAK1*
12	72500000	73500000	0.315	*DZIP1, CLDN10, DNAJC3, UGGT2, HS6ST3*
16	24500000	26000000	0.290	*RF00096, DUSP10*
29	26500000	27500000	0.308	*M-SAA3.2, OR8D4, RF00026, OR4D5, RF00056, OR10S1, OR10G6, OR10D3, TMEM225, VWA5A*
Holstein				
1	76500000	77500000	0.287	*CLDN16, CLDN1, IL1RAP, TMEM207, P3H2, TP63*

2	0	1000000	0.291	*RF00026, LGSN, NIPA1, OCA2, HERC2*
4	76000000	77000000	0.424	*bta-mir-2420, RF00392, RF00392, RF00411, RF00026, PURB, bta-mir-4657, TMED4, IGFBP3, IGFBP1, ADCY1, RAMP3, TBRG4, CCM2, MYO1G, H2AFV, PPIA, ZMIZ2, DDX56, NPC1L1, OGDH, NUDCD3*
4	83500000	84500000	0.296	*RF00001*
5	25000000	26000000	0.288	*MUCL1, GLYCAM1, GPR84, bta-mir-148b, HNRNPA1, SMUG1, HOXC4, HOXC5, METAP2, USP44, PPP1R1A, PDE1B, GTSF1, ITGA5, COPZ1, NFE2, CBX5, NCKAP1L, ZNF385A*
6 (ROH)	37000000	38000000	0.314	*MED28, DCAF16, LAP3, FAM184B, NCAPG, LCORL*
7	41000000	42000000	0.315	*OR6F1, OR11L1, RF00001, OR2M4, TRIM58, OR2W3, OR2L13*
8	103000000	104000000	0.297	*MIR455, ORM1, RF00416, RF00560, KIF12, AKNA, ATP6V1G1, TNFSF15, ZNF618, COL27A1, WHRN, TMEM268*
8	106000000	107000000	0.307	*RF00413, ASTN2*
9	55000000	56000000	0.299	*RF00100, RF00026*
14	22000000	23500000	0.457	*SOX17, RP1, LYPLA1, XKR4, MRPL15RPS20, RF01277, RF00003, MOS, TGS1, CHCHD7, SDR16C5, SDR16C6, XKR4, LYN, PLAG1, TMEM68*
16	41500000	42500000	0.358	*KIAA2013, RF00020, NPPB, NPPA, RF02158, RF02157, RF02156, bta-mir-12050, FBXO2, TNFRSF1B, TNFRSF8, MIIP, MFN2, PLOD1, CLCN6, MTHFR, DRAXIN, MAD2L2, FBXO6, FBXO44, DISP3, UBIAD1, ANGPTL7, MTOR, AGTRAP*
17	18000000	19000000	0.294	*NDUFC1, MGARP, RF00026, MGST2, SETD7, RAB33B, NOCT, MAML3, NAA15, ELF2,*
18	13500000	15000000	0.347	*ACSF3, ANKRD11, APRT, bta-mir-2327, CBFA2T3, CDH15, CDK10, CDT1, CHMP1A, CPNE7, CTU2, CYBA, DBNDD1, DEF8, DPEP1, FANCA, GALNS, GAS8, MC1R, MVD, PIEZO1, RF00003, RF00324, RNF166, RPL13, SHCBP1, SLC22A31, SNAI3, SPATA2L, SPG7, SPIRE2, TCF25, TRAPPC2L, TUBB3, VPS35, VPS9D1, ZC3H18, ZFPM1, ZNF276, ZNF469*
20	30500000	32000000	0.296	*RF00026, RF00017, PAIP1, TMEM267, CCL28, HMGCS1, NIM1K, FGF10, NNT, C20H5orf34, RF00302, bta-mir-12004, ZNF131, SELENOP, CCDC152, GHR*

20	43500000	45000000	0.482	-
22	22000000	23000000	0.341	*RF02196, LRRN1, SETMAR*
24	17500000	18500000	0.313	*RF00026*
26	21500000	23500000	0.408	*RF00156, SEMA4G, MRPL43, TWNK, KAZALD1, TLX1, LBX1, FGF8, NPM3, RF00001, HPS6, RF00099, PITX3, NFKB2, FBXL15, bta-mir-146b, TRIM8, CYP17A1, CYP17A1, SLF2, LZTS2, SFXN3, POLL, DPCD, KCNIP2, LDB1, NOLC1, ELOVL3, PSD, CUEDC2, MFSD13A, ACTR1A, ARL3, WBP1L, CYP17A1, PAX2, BTRC, FBXW4, OGA, ARMH3, PPRC1, GBF1, SUFU, SFXN2*
Simmental				
4	54500000	56000000	0.201	*GPR85, PPP1R3A, BMT2, LSMEM1, IFRD1, ZNF277, TMEM168*
5	13500000	14500000	0.158	*RF00026, SLC6A15*
5	22500000	24000000	0.157	*RF00026, PLEKHG7, UBE2N, MRPL42, SOCS2, EEA1, NUDT4, CRADD, PLXNC1*
6	61000000	62500000	0.182	*SHISA3, bta-mir-2285cs, RF00100, BEND4, ATP8A1, GRXCR1, RF00100, GRXCR1*
6	64500000	66000000	0.165	*COX7B2, GABRG1, GABRA2, GABRA4, GABRB1*
6	69000000	71500000	0.310	*GSX2, RF00026, RF00026, CHIC2, FIP1L1, LNX1, PDGFRA, KIT, KDR, SRD5A3, TMEM165, PDCL2, CLOCK, EXOC1L, CEP135, NMU, EXOC1*
6	91500000	92500000	0.164	*SOWAHB, SEPT11, CCNI, CCNG2, CXCL13, CNOT6L, SHROOM3*
7	42000000	43500000	0.171	*ARID3A, AZU1, BSG, C2CD4C, CDC34, CFD, ELANE, FGF22, FSTL3, GZMM, HCN2, KISS1R, LYPD8, MADCAM1, MED16, MGC137030, MIER2, MISP, ODF3L2, OR2G6, OR2T1, OR2T11, OR2T27, OR2T6, PALM, PGBD2, PLPP2, PLPPR3, POLRMT, PRSS57, PRTN3, PTBP1, R3HDM4, RF00026, RNF126, SH3BP5L, SHC, THEG, TPGS1, ZNF672, ZNF692*
7	67500000	68500000	0.164	*SGCD, TIMD4*
11	59000000	60000000	0.180	*C11H2orf74, AHSA2, USP34*
11	93500000	94500000	0.154	*OR1J2, OR1N2, OR1N1, OR1Q1, OR1B1, OR1L1, OR1L3, OR5C1, OR1K1, PDCL, RF00594, RF00579, ZBTB6, ZBTB26, GPR21, RF00026, RC3H2, RABGAP1, STRBP*
11	95500000	96500000	0.159	*bta-mir-181b-2, RF00026, RPL35, RF00264, HSPA5, RF00026, RF00020, ADGRD2, NR5A1, OLFML2A, WDR38,*

				ARPC5L, GOLGA1, RABEPK, NR6A1, SCAI, PPP6C, GAPVD1, MAPKAP1
16	13000000	14000000	0.157	RF00001, RGS18
18	13500000	15000000	0.185	*ZNF469, CYBA, MVD, SNAI3, bta-mir-2327, CDT1, APRT, TRAPPC2L, SLC22A31, RPL13, RF00324, CPNE7, CHMP1A, CDK10, SPATA2L, VPS9D1, MC1R, DBNDD1, RF00003, ZC3H18, CTU2, RNF166, PIEZO1, CBFA2T3, ACSF3, CDH15, SPG7, DPEP1, ZNF276, FANCA, SPIRE2, TCF25, TUBB3, DEF8, GAS8, SHCBP1, VPS35, ZFPM1, GALNS, ANKRD11ZNF469, CYBA, MVD, SNAI3, bta-mir-2327, CDT1, APRT, TRAPPC2L, SLC22A31, RPL13, RF00324, CPNE7, CHMP1A, CDK10, SPATA2L, VPS9D1, MC1R, DBNDD1, RF00003, ZC3H18, CTU2, RNF166, PIEZO1, CBFA2T3, ACSF3, CDH15, SPG7, DPEP1, ZNF276, FANCA, SPIRE2, TCF25, TUBB3, DEF8, GAS8, SHCBP1, VPS35, ZFPM1, GALNS, ANKRD11*
29	7500000	9000000	0.182	*DENR, RAB38, TMEM135, FZD4, PRSS23, ME3, TMEM135*

[1]Chromosome regions overlapping runs of homozygosity (ROH) islands reported by Mastrangelo *et al.* (2018b) in the Reggiana cattle breed are indicated with "(ROH)".

Supplementary Table S4. Results of the pairwise Cross-Population Extended Haplotype Homozygosity (XP-EHH) analysis of Reggiana *vs* the three cosmopolitan breeds (Brown, Holstein and Simmental). Reported information includes the bovine chromosome (BTA) position [start and end nucleotide position on the chromosome determined by the corresponding single nucleotide polymorphism (SNP) positions], the number of SNPs in the window, the averaged XP-EPP value, the averaged log value and the annotated genes in the reported chromosome regions.

BTA	Start position	End position	Start SNP	End SNP	N. of SNPs	Averaged XP-EHH	Averaged log value	Annotated genes
Brown								
5	71914501	72264476	ARS-BFGL-NGS-111053	ARS-BFGL-NGS-79121	9	-3.833	3.919	*RF00407, RF00598, LARGE1*
5	73581692	74079648	ARS-BFGL-NGS-100454	BTA-73985-no-rs	12	-3.783	3.835	*RASD2, TOM1, HMOX1, MCM5, MB, RBFOX2*
5	74740570	75224515	ARS-BFGL-NGS-118891	ARS-BFGL-NGS-117321	12	-3.590	3.487	*RF00026, EIF3D, MYH9, TXN2, FOXRED2, IFT27, PVALB, CACNG2*
6	59716766	60085027	Hapmap50098-BTA-76549	ARS-BFGL-NGS-112982	11	-3.927	4.100	*APBB2*
6	70431058	70716954	Hapmap23983-BTC-070420	BTB-00263209	10	-3.617	3.532	*RF00026, RF00026, KDR*
6	80685724	80858441	Hapmap26275-BTC-043486	Hapmap24320-BTC-043265	5	-3.713	3.690	*EPHA5*
6	88075494	89359390	Hapmap39947-BTA-77207	BTB-01458572	18	-3.696	3.669	*RF00100, CXCL8, CXCL5, CXCL2, CXCL3, GRO1, EPGN, COX18, ALB, AFP, AFM, RASSF6, EREG, ANKRD17, MTHFD2L*
6	89732280	89935361	ARS-BFGL-NGS-2935	ARS-BFGL-NGS-83505	8	-4.114	4.423	*PARM1*
6	90260850	91800062	BTB-01496160	BTA-77154-no-rs	30	-4.070	4.346	*RCHY1, RF00003, CXCL9, CXCL10, CXCL11, STBD1, RF00026, SOWAHB, THAP6, G3BP2, PPEF2, NAAA, NUP54, CCDC158, 11-Sep, SCARB2, CDKL2, USO1, SDAD1, ART3, SHROOM3*

6	92383295	92430962	Hapmap52160-rs29020798	BTB-00270310	3	-3.526	3.378	*CNOT6L*
6	93019845	93233475	ARS-BFGL-NGS-66691	Hapmap53128-rs29022916	5	-3.426	3.213	*FRAS1*
6	96162959	96424013	Hapmap36567-SCAFFOLD30438_8760	Hapmap48078-BTA-77495	9	-3.985	4.195	*RF00156*, *RASGEF1B*
13	44978611	45082428	ARS-BFGL-NGS-23830	BTA-99048-no-rs	4	-3.541	3.399	*PITRM1*
13	45936412	46049129	ARS-BFGL-NGS-101531	Hapmap42872-BTA-22214	4	-3.455	3.259	*ADARB2*
Holstein								
10	44727501	44820482	ARS-BFGL-NGS-97032	Hapmap51024-BTA-67203	5	-3.481	3.312	*GNG2*
20	24479790	24565655	Hapmap53329-rs29023196	ARS-BFGL-NGS-108866	3	-3.370	3.128	*SNX18*
20	28062228	28317303	BTA-50190-no-rs	ARS-BFGL-NGS-31598	6	-3.284	2.993	*PARP8*
20	34228714	34460986	Hapmap54938-rs29013720	BTA-50400-no-rs	6	-3.215	2.881	-
20	34710584	34965270	BTA-15204-no-rs	ARS-BFGL-BAC-31754	5	-3.257	3.076	*RF00004*, *RF00001*
20	35666579	36739131	ARS-BFGL-NGS-34478	BTB-00780480	21	-3.456	3.274	*RF00416*, *RF00560*, *RF00026*, *RF00026*, *GDNF*, *LIFR*, *EGFLAM*, *WDR70*
20	37388845	37911470	Hapmap53888-rs29021190	Hapmap49835-BTA-104494	11	-3.286	2.995	*SLC1A3*
20	39181409	39722083	ARS-BFGL-NGS-115190	ARS-BFGL-NGS-73590	14	-3.453	3.264	*RAD1*, *RF00003*, *BRIX1*, *TTC23L*, *C1QTNF3*, *NAJC21*, *RAI14*
20	41025188	41104184	ARS-BFGL-NGS-63070	Hapmap43599-BTA-50578	4	-3.410	3.195	*SUB1*
20	45677327	45971800	ARS-BFGL-NGS-37203	Hapmap58446-rs29021863	7	-3.333	3.069	-
Simmental								
5	60241477	60299953	ARS-BFGL-NGS-7741	ARS-BFGL-NGS-110018	3	-3.876	3.975	*CCDC38*, *AMDHD1*, *HAL*
5	60983725	61058294	ARS-BFGL-NGS-44773	Hapmap39777-BTA-73723	3	-3.998	4.196	*CFAP54*

5	61907183	62128735	ARS-BFGL-NGS-115187	ARS-BFGL-NGS-100699	5	-3.847	3.928	-
5	67994388	68293598	ARS-BFGL-NGS-119231	ARS-BFGL-NGS-33119	10	-4.107	4.419	*CHST11*
6	68331252	71428675	Hapmap49432-BTA-107930	ARS-BFGL-NGS-37727	44	-5.433	7.608	*bta-mir-4449, RF00568, GSX2, RF00026, RF00026, RF00026, RF00026, RF00026, USP46, RASL11B, CHIC2, KIT, KDR, SRD5A3, TMEM165, PDCL2, EXOC1L, CEP135, SCFD2, FIP1L1, LNX1, PDGFRA, CLOCK, NMU, EXOC1*
7	43047351	43105247	ARS-BFGL-NGS-112360	ARS-BFGL-NGS-74330	3	-3.886	4.001	*C2CD4C, MIER2, THEG*
7	43715046	43792866	ARS-BFGL-NGS-22438	ARS-BFGL-NGS-109750	4	-3.812	3.867	*C7H19orf24, EFNA2, PWWP3A*
7	44326829	44403367	Hapmap49311-BTA-78907	ARS-BFGL-NGS-69626	3	-3.724	3.709	*SOWAHA*

Supplementary Table S5. Gene enrichment analysis over the Gene Ontology (GO) – Biological Process resource.

Term[1]	Description[2]	Overlap[3]	*p*-value[4]	Genes[5]
GO:0006892	post-Golgi vesicle-mediated transport	4/59	8.10E-03	*CHIC2, ARL3, GBF1, EXOC1*
GO:0030318	melanocyte differentiation	2/8	2.28E-02	*KIT, HPS6*
GO:0006702	androgen biosynthetic process	2/11	2.28E-02	*SRD5A3, CYP17A1*
GO:0043549	regulation of kinase activity	4/102	2.28E-02	*PDGFRA, LDB1, KIT, KDR*
GO:0051091	positive regulation of sequence-specific DNA binding transcription factor activity	5/216	2.28E-02	*TRIM8, KIT, PPRC1, CLOCK, NFKB2*
GO:0008209	androgen metabolic process	2/21	4.42E-02	*SRD5A3, CYP17A1*

[1]Identifier retrieved from the GO resource.
[2]Brief explanation of the functional term.
[3]Number of input genes associated to the functional term over the number of genes directly associated to the functional term.
[4]Adjusted *p*-value.
[5]Genes of the input set associated to the functional term.