ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

FULL PAPER

doi: 10.1111/age.13113

Detecting the footprint of selection on the genomes of Murciano-Granadina goats

D. Guan*, A. Martínez†, M. G. Luigi-Sierra*, J. V. Delgado†, V. Landi†,‡, A. Castelló*,§, J. Fernández Álvarez¶, X. Such**, J. Jordana§ and M. Amills*,§

*Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. †Departamento de Genética, Universidad de Córdoba, Córdoba 14071, Spain. ‡Department of Veterinary Medicine, University of Bari "Aldo Moro", SP. 62 per Casamassima km. 3, Valenzano 70010, Italy. §Departament de Ciència Animal i dels Aliments, Facultat de Veterinària, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain. ¶Asociación Nacional de Criadores de Caprino de Raza Murciano-Granadina (CAPRIGRAN), Fuente Vaqueros, Granada 18340, Spain. **Group of Research in Ruminants (G2R), Department of Animal and Food Science, Universitat Autònoma de Barcelona (UAB), Bellaterra 08193, Spain.

Summary

Artificial selection is one of the major forces modifying the genetic composition of livestock populations. Identifying genes under selection could be useful to elucidate their impact on phenotypic variation. We aimed to identify genomic regions targeted by selection for dairy and pigmentation traits in Murciano-Granadina goats. Performance of a selection scan based on the integrated haplotype score test in a population of 1183 Murciano-Granadina goats resulted in the identification of 77 candidate genomic regions/SNPs. The most significant selective sweeps mapped to chromosomes 1 (69.86 Mb), 4 (41.80–49.95 Mb), 11 (65.74 Mb), 12 (31.24 and 52.51 Mb), 17 (34.76–37.67 Mb), 22 (31.75 Mb), and 26 (26.69–31.05 Mb). By using previously generated RNA-Seq data, we built a catalogue of 6414 genes that are differentially expressed across goat lactation (i.e. 78 days post-partum, early lactation; 216 days post-partum, late lactation; 285 days post-partum, dry period). Interestingly, 183 of these genes mapped to selective sweeps and several of them display functions related with lipid, protein, and carbohydrate metabolism, insulin signaling, cell proliferation, as well as mammary development and involution. Of particular interest are the *CSN3* and *CSN1S2* genes, which encode two major milk proteins. Additionally, we found three pigmentation genes (*GLI3*, *MC1R*, and *MITF*) co-localizing with selective sweeps. Performance of a genome-wide association study and Sanger sequencing and TaqMan genotyping experiments revealed that the c.801C>G (p.Cys267Trp) polymorphism in the melanocortin 1 receptor (*MC1R*) gene is the main determinant of the black (GG or GC genotypes) and brown (CC genotypes) colorations of Murciano-Granadina goats.

Keywords Casein, coat color, iHS test, MC1R, selective sweeps

Introduction

Selection for production and morphological traits has strongly modified the genetic composition of livestock populations (Saravanan *et al.* 2020). Increased linkage disequilibrium (LD) and reduced haplotype diversity around the selected site are typical genomic signatures left by selection (Harris *et al.* 2018). Multiple statistical methods based on diversity reduction, shifts of the site frequency spectrum, specific LD patterns, and single site or haplotype differentiation have been developed in order to detect hard selective sweeps at a genome-wide scale (Vitti *et al.* 2013; Saravanan *et al.* 2020). In livestock, the identification of genes under selection should be helpful to elucidate their biological roles as well as to pinpoint mutations with beneficial effects on traits of economic interest. Aiming to achieve this goal, multiple selection scans have been performed in goats to map selective sweeps related to pigmentation (Bertolini *et al.* 2018; Oget *et al.* 2018), milk traits (Bertolini *et al.* 2018; Oget *et al.* 2018; Zhang *et al.* 2018), litter size (Lai *et al.* 2016; E *et al.*, 2019), cashmere

Address for correspondence

M. Amills, Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Universitat Autònoma de Barcelona, Bellaterra 08193, Spain.
Email: marcel.amills@uab.cat

Accepted for publication 16 June 2021

production (Li *et al.* 2017; Zhang *et al.* 2018), and environmental adaptation (Kim *et al.* 2016; Song *et al.* 2016; Brito *et al.* 2017; Onzima *et al.* 2018), to mention a few.

Murciano-Granadina (MG) goats, which can display black or brown coats, are fundamentally devoted to milk production and cheese manufacturing (Delgado *et al.* 2017). The breeding program is currently managed by the Asociación Nacional de Caprino de Raza Murciano-Granadina (CAPRIGRAN) and aims to improve milk yield and composition as well as morphology (Delgado *et al.* 2017). So far, selection has substantially improved the milk productivity of MG goats, which currently reaches 530 kg of milk yield per lactation and 5.6% and 3.6% of fat and protein contents respectively (Delgado *et al.* 2017). The main aim of the current work was to identify genomic signatures produced by artificial selection for dairy and pigmentation traits in MG goats. To achieve this goal, we have genotyped a population of MG goats with the Goat SNP50 BeadChip and subsequently, we have investigated the footprint of selection by integrating several sources of information.

Materials and methods

Sampling of MG goats and genotyping with the Goat SNP50 BeadChip

The experimental design of this research is summarized in Fig. S1. As animal material, we have used 1197 MG goats distributed in 19 farms from the autonomous region of Andalusia. The current census of the MG breed accounts for more than 100 000 individuals, which are generally raised in semi-intensive conditions with the goal of producing cheese (Spanish Ministry of Agriculture, Fisheries and Food, https://www.mapa.gob.es/es/ganaderia/temas/zootecnia/razas-ganaderas/razas/catalogo-razas/caprine/murciano-granadina/iframe-ejemplo-arca.aspx). One of the main hallmarks of the MG breed is its extraordinary adaptation to harsh climate conditions and aridity, being able to thrive in marginal lands with scarce food resources (Spanish Ministry of Agriculture, Fisheries and Food). Blood samples were collected in EDTA K_3 coated vacuum tubes and stored at −20°C before processing. Genomic DNA was isolated using a previously reported salting-out procedure (Guan *et al.* 2020) and resuspended in 1 ml TE buffer (Tris-HCl 10 mmol/l, EDTA 1 mmol/l, pH = 8). The Goat SNP50 BeadChip (Tosser-Klopp *et al.* 2014), which contains 53 347 single nucleotide polymorphisms (SNPs), was used to genotype samples following the instructions of the manufacturer (Illumina). Data normalization and genotype calling were performed with the GENOMESTUDIO software 2.0.4 (Illumina, https://emea.illumina.com). The PLINK v1.9 software (Purcell *et al.* 2007) was used to filter out unmapped and non-autosomal SNPs, as well as those with a low call rate (<90%), low frequency (<1%), or that deviated significantly from the Hardy–Weinberg expectation ($P \leq 1 \times 10^{-5}$). We also removed samples with genotyping rates lower than 90%. After applying these filtering criteria, 47 098 SNPs and 1183 MG goats were selected to perform a selection scan.

Detecting signatures of selection in the MG breed

To detect selective sweeps in the MG population, we employed the integrated haplotype score (iHS) algorithm implemented in the REHH 2.0 software (Gautier *et al.* 2017). The iHS algorithm measures the observed decay of extended haplotype homozygosity summed over both directions away from a given SNP (Voight *et al.* 2006). Large negative iHS scores indicate that a derived allele has swept up in frequency, while large positive iHS values at nearby SNPs would imply either that the ancestral allele hitchhiked with the selected site or that selection switched to favor an ancestral allele (Voight *et al.* 2006). In this study, we assumed as ancestral the reference allele by using the 'polarize_vcf = FALSE' option implemented in the REHH package (Gautier *et al.* 2017). Genome-wide iHS values were adjusted to make sure that they approximately follow a normal distribution with mean 0 and variance 1 (Voight *et al.* 2006; Gautier & Vitalis 2012; Gautier *et al.* 2017). Following Gautier *et al.* (2017), statistical significance was inferred as follows:

$$\text{piHS} = -\log_{10}(1 - 2|\Phi(\text{iHS}) - 0.5|).$$

where Φ(iHS) represents the Gaussian cumulative distribution function and piHS is a two-sided *P*-value, in a $-\log_{10}$ scale, associated with the null hypothesis (H_0 = selective neutrality). The false discovery rate (FDR) approach was used to correct for multiple testing, and the significance threshold was set to a *q*-value ≤0.05 (Benjamini & Hochberg 1995). Following Guan *et al.* (2020), significant SNPs were grouped into genomic regions when their physical positions were <1 Mb apart. Genes located within regions under selection were retrieved with the BEDTOOLS software (Quinlan & Hall 2010) by using the ARS1 goat genome (Bickhart *et al.* 2017) as a reference. Functional enrichment analysis for the set of genes co-localizing with putative selective sweeps was carried out with the DAVID Bioinformatics Resources 6.8 database (Huang *et al.* 2008, 2009) by considering goat background genes.

Establishing a catalogue of genes related with goat lactation and pigmentation and analyzing their co-localization with selective sweeps

Milk yield and composition are complex traits determined by the action of multiple genes, the identities of which are mostly unknown. To facilitate the interpretation of the selection scan data, we have established a catalogue of

genes related with lactation. This goal has been achieved by reanalyzing RNA sequencing (RNA-Seq) data generated in a previous work (Guan *et al.* 2020). In the study of Guan *et al.* (2020), the differential expression of mRNAs transcribed in the mammary gland of goats in the beginning (T1, day 78) and ending (T2, day 216) of lactation as well as in the dry period (T3, day 285) was investigated by setting, as thresholds of significance, a q-value ≤ 0.05 and an absolute fold-change >2.82. In the current work, we have used less stringent thresholds of significance, i.e. $q \leq 0.05$ and an fold-change >1.5, in order to increase the chances of finding differentially expressed (DE) genes co-localizing with selective sweeps (Fig. S1). Briefly, we have filtered out features with read counts below 10 based on the expression matrix of protein-coding genes generated by the FEATURECOUNTS v1.5.0-p1 tool (Liao *et al.* 2014). Subsequently, differential expression analysis was carried out with the DESEQ2 software (Love *et al.* 2014). After correcting for multiple testing with the FDR approach (Benjamini & Hochberg 1995), a list of DE genes was generated by taking into account the thresholds of significance mentioned above. Moreover, we have built a catalogue of 80 genes with well-known roles in pigmentation by consulting the Color Genes database (http://www.espcr.org/micemut/, accessed 27 February 2020). Once these two catalogues of genes associated with lactation and pigmentation were established, we investigated their overlap with the sets of genes mapping to selective sweeps (please see the previous section).

Performance of a genome-wide association study for pigmentation traits in MG goats

The selection scan highlighted the existence of several pigmentation genes co-localizing with selective sweeps. Following Bâlteanu *et al.* (2021), we sought to confirm whether any of these genes is involved in the coloration of MG goats by carrying out a genome-wide association study (GWAS). Coat color (black or brown) of 529 MG goats chosen at random was phenotyped by visual assessment (Fig. S1). These 529 goats also had Goat SNP50 BeadChip genotypes. Unmapped and non-autosomal SNPs, as well as those with a low call rate (<90%) and low frequency (<1%) were filtered out with the PLINK v1.9 software (Purcell *et al.* 2007). After applying these filtering criteria, 43 240 SNPs were selected to perform a GWAS analysis with the GEMMA software (version 0.98.1), which implements the Genome-wide Efficient Mixed Model Association algorithm (Zhou & Stephens 2012). The following statistical model was used:

$$Y = x\beta + u + \varepsilon.$$

where Y is a vector of phenotypic values coded as 1 (black, $N = 391$) or 2 (brown, $N = 138$); x is a n-vector of marker genotypes harbored by each individual; β is the effect size of the marker (allele substitution effect); u is a n-vector of random effects with a n-dimensional multivariate normal distribution $(0, \lambda\tau^{-1}K)$, being τ^{-1} the variance of the residual error, λ the ratio between the two variance components, and K a $n \times n$ relatedness matrix derived from 42 793 valid SNPs. Lastly, ϵ represents a vector of errors. In this model, we did not include the fixed factors defined by Guan *et al.* (2020) because they are not expected to have any effect on coat color. The GEMMA software contrasts the alternative hypothesis (H_1: $\beta \neq 0$) against the null hypothesis (H_0: $\beta = 0$) by carrying out likelihood ratio tests for each marker. Besides, the relatedness matrix, which is constructed by taking into account all genome-wide SNPs, is employed to correct for population structure. Multiple testing was implemented through an FDR approach (Benjamini & Hochberg 1995), and a q-value ≤ 0.05 was established as threshold of significance in the GWAS.

Sanger sequencing of the MC1R gene

Two pairs of primers were designed with the Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/index.cgi) in order to amplify the coding region of the *MC1R* gene (Table S1) in 9 black and 13 brown MG individuals. PCRs contained 50 ng genomic DNA, 1× BIOTAQ PCR buffer (Bioline), 200 μmol/l of dNTPs, 0.2 μmol/l of each primer, 1.5 mmol/l $MgCl_2$, and 0.65 units of BIOTAQ DNA Polymerase (Bioline). Nuclease-free water was added to a final volume of 25 μl. The thermal cycle was as follows: a hot-start step at 95°C for 2 min, followed by 34 cycles of 95°C for 45 s (denaturation), 60°C for 45 s (annealing), and 72°C for 45 s (extension), plus a final extension at 72°C for 5 min. Five μl of the PCR were mixed with 1.5 μl of a mixture containing 1.13 μl PCR buffer 1× (composition for 1 ml: 100 μl PCR Gold Buffer 10× + 100 μl $MgCl_2$ 25 mmol/l + 800 μl H_2O), 0.12 μl Exonuclease I (20 units/μl, Thermo Fisher Scientific) and 0.25 μl FastAP Thermosensitive Alkaline Phosphatase (1 unit/μl, Thermo Fisher Scientific). This mixture was incubated at 37°C for 15 min plus 85°C for 15 min. Purified amplicons were sequenced with the BIGDYE TERMINATOR CYCLE SEQUENCING KIT v1.1 (Applied Biosystems). Sequencing reactions were run on an ABI 3730 DNA analyzer (Applied Biosystems). Finally, sequences were viewed and aligned with the MOLECULAR EVOLUTIONARY GENETICS ANALYSIS software (MEGA X; Kumar *et al.* 2018).

TaqMan genotyping of a missense mutation in the *MC1R* gene

To further confirm the causality of the c.801C>G mutation, we used a Custom TaqMan SNP Genotyping Assay (Applied Biosystems) to genotype 49 black and 41 brown individuals. These 90 individuals were randomly selected from the population used in the GWAS analysis (see above). The description of TaqMan probes is reported in

Table S2. Five samples with GG ($N = 1$), CC ($N = 2$), and GC ($N = 2$) genotypes ascertained by Sanger sequencing were used as positive controls. Genotyping reactions were carried out in a final volume of 15 μl containing 1× TaqMan Universal PCR Master Mix (Applied Biosystems), 1× TaqMan Custom Genotyping Assay designed for rs669694251 (Applied Biosystems) and 18 ng genomic DNA. Real-time PCRs were performed in 96-well reaction plates and they were run in a 7900-HT Real Time PCR system (Applied Biosystems). The thermal profile was: 50°C for 2 min, 95°C for 10 min, and 40 cycles of 95°C for 15 s and 60°C for 1 min. Genotypes were obtained using the Genotyping Analysis Module of the APPLIED BIOSYSTEMS ANALYSIS Software accessible in the Thermofisher Cloud (https://www.thermofisher.com/es/es/home/digital-science.html, accessed March 20, 2020).

Results

Identification of putative selective sweeps in a population of MG goats

By making a selection scan based on the iHS score algorithm, we identified 77 genomic regions/SNPs that reached the threshold of significance (Fig. 1, Tables 1 & S3). Among them, we observed a highly significant signal on chromosome 4 (41.80–49.95 Mb, iHS = 10.30, $q = 3.38 \times 10^{-20}$), which co-localized with 72 genes (Table S3). In this selective sweep, the most significant SNP (rs268259933) lies close to the zinc finger protein 804B (*ZNF804B*) gene, and there are also several genes related with metabolism, such as adenylyl cyclase type 1 (*ADCY1*), insulin-like growth factor-binding proteins 1 (*IGFBP1*) and 3 (*IGFBP3*). As shown in Fig. 1 and Table 1, three additional selective sweeps of considerable size were observed on chromosome 6 (86.08–86.21 Mb, iHS = −5.44, $q = 4.08 \times 10^{-5}$), chromosome 16 (40.09–51.83 Mb, iHS = 5.23, $q = 1.05 \times 10^{-4}$), and chromosome 17 (34.76–37.67 Mb, iHS = 5.34, $q = 6.38 \times 10^{-5}$). Interestingly, the genomic region on chromosome 6 (86.08–86.21 Mb) overlaps the casein gene cluster (Table S3). The most significant SNPs on chromosomes 16 and 17 map close to the solute carrier family 45 member 1 (*SLC45A1*) gene and to the transient receptor potential cation channel subfamily C member 3 (*TRPC3*) gene respectively (Table S3). In addition, we found selective sweeps on chromosomes 18 (16.64–19.46 Mb, iHS = −5.76, $q = 7.19 \times 10^{-6}$), and 22 (31.75 Mb, iHS = −4.81, $q = 7.45 \times 10^{-4}$), as shown in Fig. 1, Tables 1 & S3. As indicated in Table S4, functional enrichment analysis of genes mapping to selective sweeps did not identify significant pathways/terms after correction for multiple testing ($q > 0.05$), but the pathway 'oocyte meiosis' was the most significant one at the nominal level ($P = 8.98 \times 10^{-4}$).

Co-localization between selective sweeps and genes differentially expressed across goat lactation

Selection scans often generate long list of genes mapping to selective sweeps. Such data can be difficult to interpret because the functional relationship between genes and phenotypes under selection is usually unknown. To shed some light into this issue, we have built a catalogue of genes that are expressed in the caprine mammary gland and DE across lactation. By reanalyzing RNA-Seq data generated by Guan *et al.* (2020), we were able to identify 410 (T1/T2), 5368 (T1/T3), and 4437 (T2/T3) DE genes. In total, 6414 genes showed differential expression across comparisons, and 183 of them co-localized with selective sweeps (Table S5). In this reduced list of 183 genes, we found loci related with: lipid (abhydrolase domain-containing protein 2, *ABHD2*; perilipin 1, *PLIN1*; stearoyl-CoA desaturase, *SCD*), casein (casein α_{S2}, *CSN1S2*; casein κ, *CSN3*), and carbohydrate (glucokinase, *GCK*; enolase 1, *ENO1*) metabolism; insulin signaling (phosphatidylinositol-4-phosphate

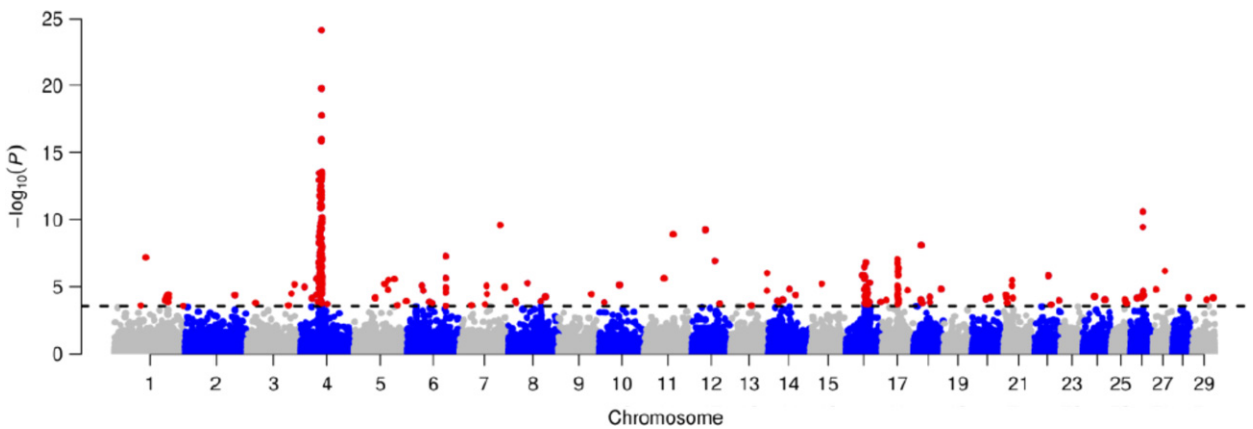

Figure 1 Selection scan in a population of Murciano-Granadina goats. The *x*- and *y*-axes indicate genomic coordinates and negative $\log_{10}$ *P*-values respectively. The dashed line indicates the threshold of significance ($q \leq 0.05$). Significant SNPs are marked in red.

Table 1 List of selective sweeps identified with the iHS-based algorithm in Murciano-Granadina goats (the full list of selective sweeps is provided in Table S3).

Chromosome	Region (Mb)	SNP	iHS value	*P*	*q*	Key genes
4	41.80–49.95	rs268259933	10.3	7.32E-25	3.38E-20	*ZNF804B, ADCY1, IGFBP1, IGFBP3, GCK, OGDH, HUS1, GLI3*
5	90.12	rs268282714	4.69	2.72E-06	1.19E-03	*PIK3C2G*
6	86.08–86.21	rs268293092	−5.44	5.47E-08	4.08E-05	*CSN3, CSN1S2*
7	60.11–60.89	rs268262914	−4.44	8.86E-06	3.39E-03	*EGR1, CDC25C*
16	40.09–51.83	rs268275948	5.23	1.68E-07	1.05E-04	*SLC45A1, ENO1, PIK3CD, PGD, B3GALT6, TNFRSF8, TNFRSF9, TNFRSF18*
17	34.76–37.67	rs268275466	5.34	9.24E-08	6.38E-05	*TRPC3*
18	16.64–19.46	rs268267832	−5.76	8.24E-09	7.19E-06	*MC1R*
21	5.85–8.98	rs268257442	4.11	3.89E-05	1.12E-02	*IGF1R*
21	19.88–21.86	rs268283170	4.65	3.27E-06	1.40E-03	*ANPEP, ABHD2, PLIN1*
22	31.75	rs268246615	−4.81	1.51E-06	7.45E-04	*MITF*
25	31.54–35.43	rs268242461	−3.91	9.33E-05	2.22E-02	*MDH2*
26	26.69–31.05	rs268273064	−6.68	2.40E-11	4.11E-08	*SCD, NFKB2*

3-kinase catalytic subunit type 2 γ, *PIK3C2G*; phosphatidylinositol-4,5-bisphosphate 3-kinase catalytic subunit δ, *PIK3CD*); Krebs cycle and pentose phosphate pathway (malate dehydrogenase 2, *MDH2*; oxoglutarate dehydrogenase, *OGDH*; phosphogluconate dehydrogenase, *PGD*); cell proliferation (cell division cycle 25C, *CDC25C*; HUS1 checkpoint clamp component, *HUS1*); and mammary development and involution (alanyl aminopeptidase, membrane, *ANPEP*; early growth response 1, *EGR1*; insulin like growth factor 1 receptor, *IGF1R*; insulin like growth factor binding protein 3, *IGFBP3*; nuclear factor kappa B subunit 2, *NFKB2*; TNF receptor superfamily members 8, 9 and 18, *TNFRSF8*, *TNFRSF9* and *TNFRSF18*).

Genomic architecture of coat color in MG goats

With regard to coat color, we generated a catalogue of 80 genes with well-known functions on pigmentation (http://www.espcr.org/micemut/, accessed 27 February 2020) and explored their co-localization with selective sweeps. We found that the GLI family zinc finger 3 (*GLI3*), melanocortin 1 receptor (*MC1R*) and melanocyte-inducing transcription factor (*MITF*) pigmentation genes map to selective sweeps on chromosomes 4 (41.80–49.95 Mb), 18 (16.64–19.46 Mb) and 22 (31.75 Mb) respectively. To further clarify the genomic architecture of pigmentation in MG goats, we performed a GWAS analysis comprising 529 MG goats with available color phenotypes (Fig. 2a,b). We detected 27 SNPs on chromosome 18 (12.18–22.30 Mb) showing significant associations with coloration (Fig. 2c), while no significant signal was observed for chromosomes 4 and 22. The chromosome 18 marker showing the highest statistical significance was rs268287597 (chromosome 18: 15 887 136 bp, $q = 1.34 \times 10^{-19}$, Fig. 2c), which lies close to the *MC1R* gene (chromosome 18: 16 104 986–16 105 939 bp). The quantile-quantile plot illustrates that part of the *P*-values observed in the GWAS deviate from the expected *P*-values drawn from a theoretical Chi-square distribution (Fig. 2d).

Based on these results, we decided to sequence the coding region of the *MC1R* gene. By doing so, we identified two missense mutations: c.748G>T (p.Val250Phe) and c.801C>G (p.Cys267Trp). These two polymorphisms have been already registered in the Ensembl database (https://www.ensembl.org) with identifiers rs657434682 and rs669694251 respectively. Visual inspection of the sequences revealed a single G peak (GG genotype) at position 748 in nine black and 10 brown goats, while only three brown goats showed a double GT peak at this location (Table S6). Moreover, no goat displayed a single T peak (TT genotype). Altogether, these results rule out c.748G>T (p.Val250Phe) as a causal mutation for the black/brown coat color of MG goats. With regard to c.801C>G, black goats showed either a single G peak (GG genotype, $N = 1$) or a GC double peak (GC genotype, $N = 8$), while all brown goats ($N = 13$) displayed a single C peak (CC genotype, Fig. 3a, Table S6).

By performing a TaqMan genotyping experiment in 49 black and 41 brown individuals, we found that all brown goats were CC and all black goats were either GG or GC (Fig. 3b), indicating that the G-allele has probably a dominant effect over the C-allele. Moreover, a second GWAS was carried out by using a SNP data set comprising 43 240 SNPs contained in the Goat SNP50 BeadChip plus the rs669694251 (c.801C>G) marker. The GEMMA software (Zhou & Stephens 2012) was employed to investigate the association between these 43 241 markers and color phenotypes in 49 black and 41 brown MG goats ($N = 90$) genotyped with the TaqMan assay. In Fig. 3c, it can be appreciated that marker rs669694251 displays the most significant association with coat color (c.801C>G, $q = 2.91 \times 10^{-25}$). In this subset of 90 individuals, we

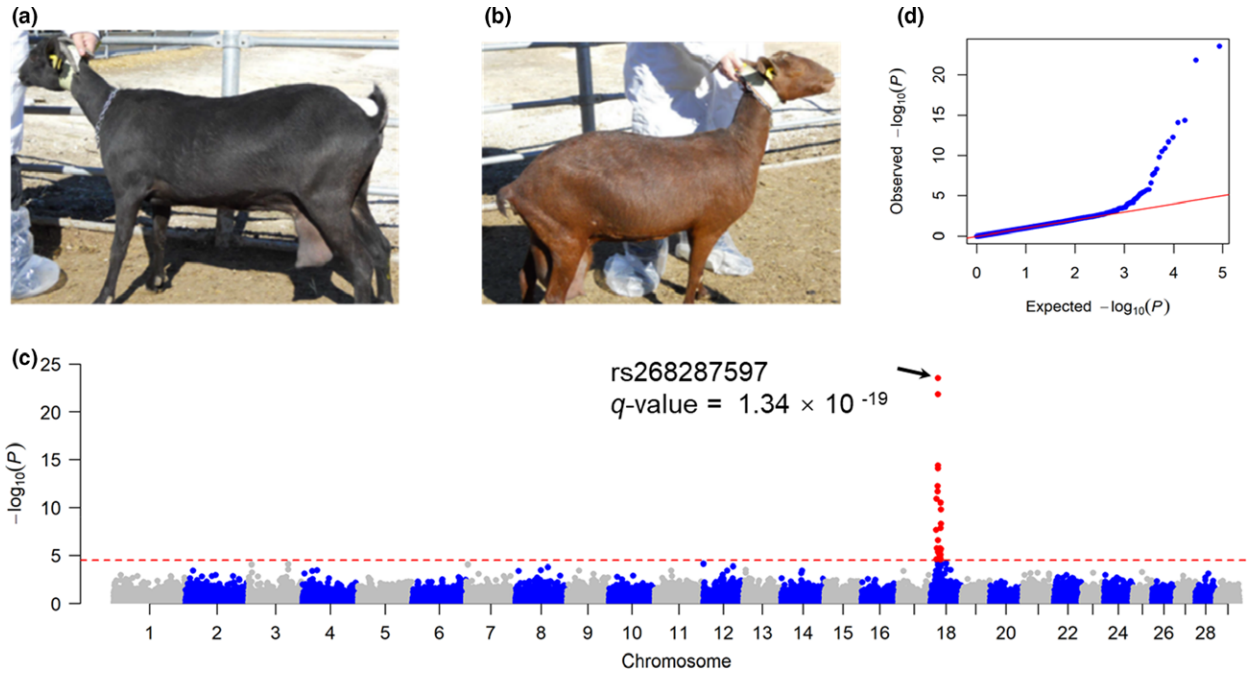

Figure 2 Murciano-Granadina goats with black (a) and brown (b) pigmentation patterns, which were respectively coded as 1 and 2 in the genome-wide association study. (c) Manhattan plot depicting the associations between coat color and Goat SNP50 BeadChip genotypes from 391 black and 138 brown Murciano-Granadina goats. Negative $\log_{10}$ *P*-values (*y*-axis) of the associations between SNPs and phenotypes are plotted against the genomic location of each SNP marker (*x*-axis). Markers on different chromosomes are denoted with different colors. The horizontal dashed line indicates statistical significance after correction for multiple testing by using the false discovery rate approach reported by Benjamini & Hochberg (1995). The arrow indicates the leading SNP that shows the highest association with phenotype (rs268287597, $q = 1.34 \times 10^{-19}$). (d) Quantile-quantile plot of the expected versus observed *P*-values in the genome-wide association study analysis for pigmentation.

also found that the rs669694251 SNP is in high disequilibrium with the leading SNP rs268287597 identified in the first GWAS analysis ($r^2 = 0.30$, $D' = 0.81$).

Discussion

Selective sweeps encompass genes related with mammary metabolism and lactation

The most significant selective sweep ($q = 3.38 \times 10^{-20}$) mapped to chromosome 4 (41.80–49.95 Mb) and harbored 72 genes. The leading SNP was located downstream of the *ZNF804B* gene, which has been associated with dry matter intake (de Oliveira *et al*. 2014) and milk oleic acid content in cattle (Mychaleckyj *et al*. 2018), as well as with fat deposition in sheep (Xu *et al*. 2017). Moreover, in this chromosome 4 region, we detected several genes that are DE in lactating and non-lactating goats. For instance, the *GCK* gene encodes an enzyme that catalyzes the formation of glucose-6-phosphate which happens to be the rate-limiting step in glucose metabolism (Matschinsky & Wilson 2019). This molecule is a primary glucose sensor in mammals, being responsible for the regulation of glucose homeostasis (Matschinsky & Wilson 2019). Another DE gene of interest was *OGDH*, which encodes oxoglutarate dehydrogenase, an enzyme participating in the Krebs cycle, one of the main pathways by which acetyl-CoA derived from lipids, sugars, and proteins is oxidized to generate NADH, GTP, and $FADH_2$ as well as the building blocks needed for macromolecule synthesis (Martínez-Reyes & Chandel 2020). On chromosome 4 (41.80–49.95 Mb), we also detected the *IGFBP3* gene, which modulates the involution of the mammary gland (Neuenschwander *et al*. 1996; Flint *et al*. 2001) as well as lipid and carbohydrate metabolism (Kim 2013).

The iHS scan also captured a significant signal ($q = 4.08 \times 10^{-5}$) on a chromosome 6 region (86.08–86.21 Mb) co-localizing with the *CSN1S2* and *CSN3* genes. Casein genes synthesize the major fraction (>80%) of the milk proteome and their variability has been associated with milk composition in MG (Guan *et al*. 2020) and Alpine and Saanen (Martin *et al*. 2017) goats. By using a candidate gene approach, Caravaca *et al*. (2008, 2009, 2011) found that the BB and EE casein α_{S1} (*CSN1S1*) genotypes are associated with CSN1S1 protein synthesis and curdling rate respectively. Moreover, the AB and BB *CSN3* genotypes have been associated with higher milk protein and casein contents (Caravaca *et al*. 2009). A recent study carried out by Pizarro Inostroza *et al*. (2019) supported the strong association of the *CSN1S1* and *CSN3* genotypes with milk

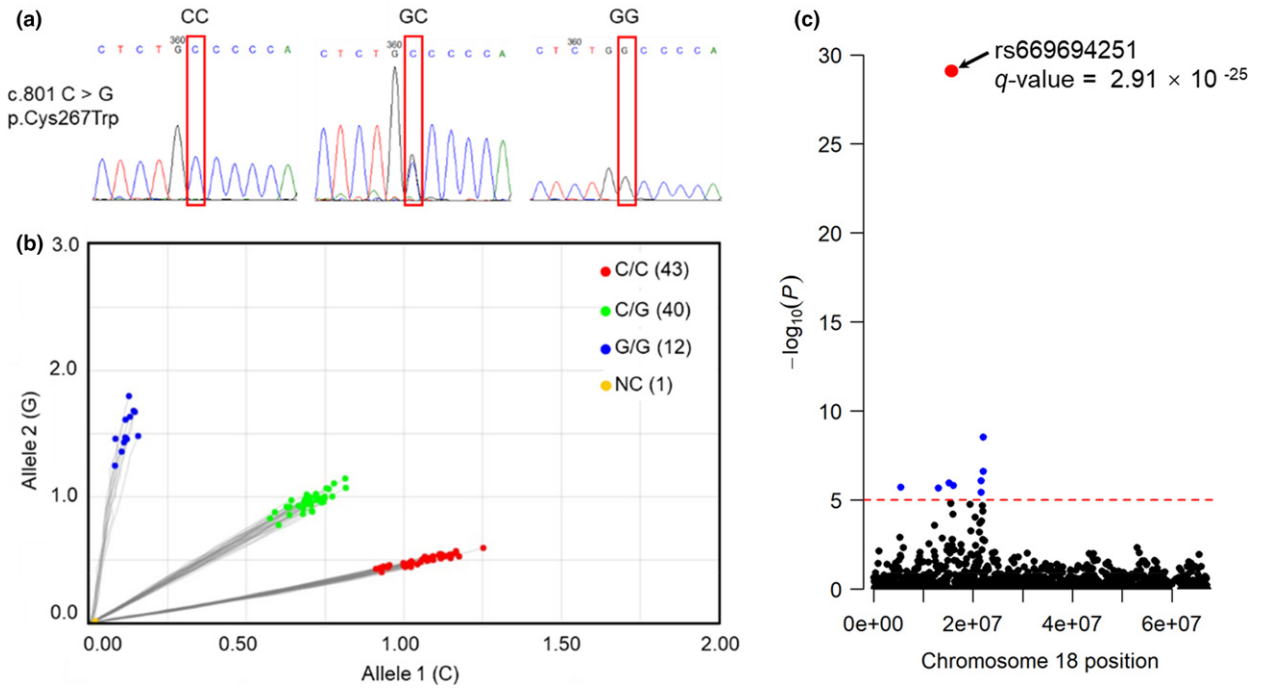

Figure 3 (a) Sequencing electropherograms showing the region containing the c.801C>G, (p.Cys267Trp) SNP in individuals with CC, GC and GG genotypes. (b) Cluster plots of TaqMan genotyping results obtained with the Genotyping Analysis Module implemented in the ThermoFisher Cloud computing application (Applied Biosystems). The horizontal and vertical axes correspond to alleles C and G respectively. The dots with red, green, and blue colors represent CC, CG and GG genotypes respectively. The negative control is indicated by an orange dot. Five samples with known genotypes (GG, $N = 1$; CC, $N = 2$; and GC, $N = 2$) assessed through Sanger sequencing were used as positive controls. (c) Manhattan plot depicting associations between coat color (41 brown and 49 black Murciano-Granadina goats) and the genotypes of marker rs669694251 (red dot) plus 1134 additional Goat SNP50 BeadChip markers mapping to goat chromosome 18. The dashed line represents the negative $\log_{10}$ *P*-value defining the threshold of significance ($q \leq 0.05$) after correcting for multiple testing with a false discovery rate approach (Benjamini & Hochberg 1995). Significant SNPs are indicated with blue dots.

composition traits in MG goats. In the light of these results, our finding that artificial selection has targeted casein genes in MG goats makes sense. Indeed, casein genotypes have been used since 2012 in the breeding program of the MG breed as selection criteria (Delgado *et al.* 2017).

The most significant SNP mapping to the chromosome 16 selective sweep (40.09–51.83 Mb, $q = 1.05 \times 10^{-4}$) lies close to the solute carrier family 45 member 1 (*SLC45A1*) gene, which encodes a glucose transporter (Vitavska & Wieczorek 2013). In this region, we also found the β-1,3-galactosyltransferase 6 (*B3GALT6*) gene, which mediates the transfer of galactose. Both glucose and galactose are necessary for the synthesis of lactose (Hennet 2002). Besides, this region contains several metabolic loci that are DE in lactating and non-lactating goats (Guan *et al.* 2020). One of these genes encodes phosphogluconate dehydrogenase, the second dehydrogenase in the pentose phosphate shunt (Berdis & Cook 1993). This biochemical route is used by cells to synthesize NADPH, a molecule essential for the reductive biosynthesis of lipids (Giacomini *et al.* 2020). We also detected the *ENO1* gene, which yields a glycolytic enzyme (Kang *et al.* 2008), and *PIK3CD*, a locus involved in the regulation of insulin signaling (Du *et al.* 2018).

Despite its modest statistical significance ($q = 1.4 \times 10^{-3}$), the chromosome 21 (19.88–21.86 Mb) selective sweep is also worth to mention because it contains two genes that are DE in milking and dried goats i.e. *PLIN1*, which is involved in the synthesis of lipid droplets (Itabe *et al.* 2017), and *ANPEP*, a proteolytic enzyme (Turner 2004). Another significant chromosome 21 selective sweep (5.85–8.98 Mb, $q = 0.01$) encompasses the *IGF1R* gene, which is also DE in lactating and dried goats (Guan *et al.* 2020). This locus encodes a receptor binding insulin-like growth factor 1, a key mediator of mammary terminal end bud and ductal formation during development and also an essential player in mammary gland function and maintenance (Christopoulos *et al.* 2015). On chromosome 26 (26.69–31.05 Mb), we also detected a highly significant selective sweep ($q = 4.11 \times 10^{-8}$). Genes of interest mapping to this region are *SCD*, which encodes an enzyme catalyzing the synthesis of monounsaturated fatty acids (Miyazaki & Ntambi 2003), and *NFKB2*, which is a key regulator of milk production (Connelly *et al.* 2010). Noteworthy, both loci are DE in lactating and non-lactating goats (Guan *et al.* 2020).

About the genomic architecture of pigmentation in MG goats

Several genes with well-defined roles in pigmentation (http://www.espcr.org/micemut/, accessed 27 February 2020) co-localized with the selective sweeps identified in our study. The *GLI3* gene, for instance, maps to the highly significant chromosome 4 selective sweep (41.8–49.95 Mb) and its inactivation leads to hypopigmentation due to a loss of melanoblast specification (Matera *et al*. 2008). Another gene of interest is *MITF*, which co-localizes with the chromosome 22 (31.75 Mb) selective sweep. The MITF transcription factor regulates the expression of numerous pigmentation genes to promote melanocyte differentiation (Kawakami & Fisher 2017) and its polymorphism has been associated with white spotting phenotypes in horses (Hauswirth *et al*. 2012), and dogs (Baranowska Körberg *et al*. 2014) as well as with piebaldism in cattle (Fontanesi *et al*. 2012). A third pigmentation gene of obvious interest is *MC1R*, which maps to a highly significant ($q = 7.19 \times 10^{-6}$) selective sweep on chromosome 18 (16.64–19.46 Mb). Importantly, the melanocyte receptor encoded by this gene is a major regulator of the synthesis of melanin (Wolf Horrell *et al*. 2016).

Given that our study identified several pigmentation genes which might have been targeted by selection, we aimed to elucidate their potential involvement in the coat coloration of MG goats. To achieve this goal, we carried out a GWAS comprising 529 MG goats with color phenotypes. We identified a single peak surpassing the threshold of significance ($q \leq 0.05$) on chromosome 18. This region contains the *MC1R* gene, thus confirming one of the most relevant results of the selection scan. Importantly, a previous study based on a candidate gene approach reported a strong association between *MC1R* genotype (c.801C>G, p.Cys267Trp) and coat color in 28 MG goats (Fontanesi *et al*. 2009).

Subsequent Sanger sequencing and TaqMan genotyping experiments confirmed that the coat color of MG goats is fully explained by *MC1R* c.801C>G genotype (all brown goats are CC and all black goats are either GG or GC), which is consistent with the results reported by Fontanesi *et al*. (2009). The G-allele (Trp) might be a dominant gain-of-function mutation located in the extracellular loop 3 of the MC1R. We hypothesize that its mechanism of action might involve the disruption of a disulfide bond between Cys_{267} and Cys_{275}, thus altering the three-dimensional structure and activity of the MC1R protein towards an increased production of eumelanin (Wolf Horrell *et al*. 2016), but measuring the activity of the Cys_{267} and Trp_{267} MC1R alleles in cultured melanocytes would be essential to confirm this. Dominant mutations increasing eumelanin synthesis have been previously reported in pigs, cattle, and sheep (Klungland *et al*. 1995; Kijas *et al*. 1998; Våge *et al*. 1999).

In conclusion, by integrating a selection scan and other sources of information we have been able to identify several genes, with important roles in metabolism and lactation, which might have been targeted by selection for dairy traits. One of the most compelling cases would be that of the *CSN3* and *CSN1S2* genes, which are involved in casein synthesis. Moreover, we have demonstrated that several pigmentation genes co-localized with selective sweeps but, according to our data and previous results (Fontanesi *et al*. 2009), the coat coloration of MG goats fundamentally depends on *MC1R* genotype.

Acknowledgements

This research was funded by the European Regional Development Fund (FEDER)/Ministerio de Ciencia e Innovación – Agencia Estatal de Investigación/Project Reference grant: PID2019-105805RB-I00 and by the CERCA Programme/Generalitat de Catalunya. We also acknowledge the support of the Spanish Ministry of Economy and Competitivity for the Center of Excellence Severo Ochoa 2020–2023 (CEX2019-000902-S) grant awarded to the Centre for Research in Agricultural Genomics (CRAG, Bellaterra, Spain). Dailu Guan was funded by a PhD fellowship from the China Scholarship Council (CSC). María Gracia Luigi-Sierra was funded with a fellowship Formación de Personal Investigador (BES-C-2017-079709) awarded by the Spanish Ministry of Economy and Competitivity. The authors are indebted to the Asociación Nacional de Criadores de Caprino de Raza Murciano-Granadina (CAPRIGRAN), and more specifically to Miguel García, Teresa Novoa and Manolo Delgado, for providing the blood samples and color phenotypes of goats analysed in the current study. We also want to acknowledge Drs. Juan Manuel Serradilla (Universidad de Córdoba), Baltasar Urrutia (Instituto Murciano de Investigación y Desarrollo Agrario y Alimentario), and Juan Carrizosa (Instituto Murciano de Investigación y Desarrollo Agrario y Alimentario) for providing pictures of Murciano-Granadina goats.

Data availability statement

The dataset used to perform GWAS and selection scans is accessible at Figshare (https://doi.org/10.6084/m9.figshare.11999823), and *MC1R* sequences have been deposited in the GenBank database (accession codes: MT186757–MT186778).

References

Bâlteanu V.A., Cardoso T.F., Amills M., Luigi-Sierra M.G., Egerszegi I., Anton I. & Zsolnai A. (2021) Red and blond Mangalitza pigs display a signature of divergent directional selection in the *SLC45A2* gene. *Animal Genetics* **52**, 66–77.

Baranowska Körberg I., Sundström E., Meadows J.R.S. *et al.* (2014) A simple repeat polymorphism in the MITF-M promoter is a key regulator of white spotting in dogs. *PLoS One* **9**, e104363.

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate – a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B-Statistical Methodology* **57**, 289–300.

Berdis A.J. & Cook P.F. (1993) Overall kinetic mechanism of 6-phosphogluconate dehydrogenase from *Candida utilis*. *Biochemistry* **32**, 2036–40.

Bertolini F., Servin B., Talenti A. *et al.* (2018) Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution* **50**, 57.

Bickhart D.M., Rosen B.D., Koren S. *et al.* (2017) Single-molecule sequencing and chromatin conformation capture enable *de novo* reference assembly of the domestic goat genome. *Nature Genetics* **49**, 643–50.

Brito L.F., Kijas J.W., Ventura R.V., Sargolzaei M., Porto-Neto L.R., Cánovas A., Feng Z., Jafarikia M. & Schenkel F.S. (2017) Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics* **18**, 229.

Caravaca F., Amills M., Jordana J. *et al.* (2008) Effect of α_{s1}-casein (*CSN1S1*) genotype on milk CSN1S1 content in Malagueña and Murciano-Granadina goats. *Journal of Dairy Research* **75**, 481–4.

Caravaca F., Ares J.L., Carrizosa J. *et al.* (2011) Effects of α_{s1}-casein (*CSN1S1*) and κ-casein (*CSN3*) genotypes on milk coagulation properties in Murciano-Granadina goats. *Journal of Dairy Research* **78**, 32–7.

Caravaca F., Carrizosa J., Urrutia B., Baena F., Jordana J., Amills M., Badaoui B., Sánchez A., Angiolillo A. & Serradilla J.M. (2009) Short communication: effect of α_{S1}-casein (*CSN1S1*) and κ-casein (*CSN3*) genotypes on milk composition in Murciano-Granadina goats. *Journal of Dairy Science* **92**, 2960–4.

Christopoulos P.F., Msaouel P. & Koutsilieris M. (2015) The role of the insulin-like growth factor-1 system in breast cancer. *Molecular Cancer* **14**, 43.

Connelly L., Barham W., Pigg R., Saint-Jean L., Sherrill T., Cheng D.-S., Chodosh L.A., Blackwell T.S. & Yull F.E. (2010) Activation of nuclear factor kappa B in mammary epithelium promotes milk loss during mammary development and infection. *Journal of Cellular Physiology* **222**, 73–81.

de Oliveira P.SN., Cesar A.SM., do Nascimento M.L. *et al.* (2014) Identification of genomic regions associated with feed efficiency in Nelore cattle. *BMC Genetics* **15**, 100.

Delgado J.V., Landi V., Barba C.J., Fernández J., Gómez M.M., Camacho M.E., Martínez M.A., Navas F.J. & León J.M. (2017) Murciano-Granadina Goat: A Spanish local breed ready for the challenges of the twenty-first century. In: *Sustainable Goat Production in Adverse Environments: Volume II: Local Goat Breeds* (Ed. by J. Simões & C. Gutiérrez), pp. 205–19. Springer International Publishing, Cham, Switzerland.

Du X., Li X., Chen L. *et al.* (2018) Hepatic miR-125b inhibits insulin signaling pathway by targeting PIK3CD. *Journal of Cellular Physiology* **233**, 6052–66.

Flint D.J., Tonner E., Knight C.H., Whitelaw CBA, Webster J., Barber M. & Allan G. (2001) Control of mammary involution by insulin-like growth factor binding proteins: role of prolactin. *Livestock Production Science* **70**, 115–20.

Fontanesi L., Beretti F., Riggio V., Dall'Olio S., González E.G., Finocchiaro R., Davoli R., Russo V. & Portolano B. (2009) Missense and nonsense mutations in melanocortin 1 receptor (*MC1R*) gene of different goat breeds: association with red and black coat colour phenotypes but with unexpected evidences. *BMC Genetics* **10**, 47.

Fontanesi L., Scotti E. & Russo V. (2012) Haplotype variability in the bovine *MITF* gene and association with piebaldism in Holstein and Simmental cattle breeds. *Animal Genetics* **43**, 250–6.

Gautier M., Klassmann A. & Vitalis R. (2017) rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Molecular Ecology Resources* **17**, 78–90.

Gautier M. & Vitalis R. (2012) rehh: an R package to detect footprints of selection in genome-wide SNP data from haplotype structure. *Bioinformatics* **28**, 1176–7.

Giacomini I., Ragazzi E., Pasut G. & Montopoli M. (2020) The pentose phosphate pathway and its involvement in cisplatin resistance. *International Journal of Molecular Sciences* **21**, 937.

Guan D., Landi V., Luigi-Sierra M.G. *et al.* (2020) Analyzing the genomic and transcriptomic architecture of milk traits in Murciano-Granadina goats. *Journal of Animal Science and Biotechnology* **11**, 35.

E G.-X., Zhao Y.-J. & Huang Y.-F. (2019) Selection signatures of litter size in Dazu black goats based on a whole genome sequencing mixed pools strategy. *Molecular Biology Reports* **46**, 5517–23.

Harris A.M., Garud N.R. & DeGiorgio M. (2018) Detection and classification of hard and soft sweeps from unphased genotypes by multilocus genotype identity. *Genetics* **210**, 1429–52.

Hauswirth R., Haase B., Blatter M. *et al.* (2012) Mutations in *MITF* and *PAX3* cause "Splashed White" and other white spotting phenotypes in horses. *PLoS Genetics* **8**, e1002653.

Hennet T. (2002) The galactosyltransferase family. *Cellular and Molecular Life Sciences CMLS* **59**, 1081–95.

Huang D.W., Sherman B.T. & Lempicki R.A. (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57.

Huang D.W., Sherman B.T. & Lempicki R.A. (2009) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Research* **37**, 1–13.

Itabe H., Yamaguchi T., Nimura S. & Sasabe N. (2017) Perilipins: a diversity of intracellular lipid droplet proteins. *Lipids in Health and Disease* **16**, 83.

Kang H.J., Jung S.K., Kim S.J. & Chung S.J. (2008) Structure of human α-enolase (hENO1), a multifunctional glycolytic enzyme. *Acta Crystallographica Section D: Biological Crystallography* **64**, 651–7.

Kawakami A. & Fisher D.E. (2017) The master role of microphthalmia-associated transcription factor in melanocyte and melanoma biology. *Laboratory Investigation* **97**, 649–56.

Kijas JMH, Wales R., Törnsten A., Chardon P., Moller M. & Andersson L. (1998) Melanocortin receptor 1 (*MC1R*) mutations and coat color in pigs. *Genetics* **150**, 1177–85.

Kim E.S., Elbeltagy A.R., Aboul-Naga A.M., Rischkowsky B., Sayre B., Mwacharo J.M. & Rothschild M.F. (2016) Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* **116**, 255–64.

Kim H.-S. (2013) Role of insulin-like growth factor binding protein-3 in glucose and lipid metabolism. *Annals of Pediatric Endocrinology & Metabolism* **18**, 9–12.

Klungland H., Vage D.I., Gomez-Raya L., Adalsteinsson S. & Lien S. (1995) The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mammalian Genome* **6**, 636–9.

Kumar S., Stecher G., Li M., Knyaz C. & Tamura K. (2018) MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Molecular Biology and Evolution* **35**, 1547–9.

Lai F.-N., Zhai H.-L., Cheng M. *et al*. (2016) Whole-genome scanning for the litter size trait associated genes and SNPs under selection in dairy goat (*Capra hircus*). *Scientific Reports* **6**, 38096.

Li X., Su R., Wan W. *et al*. (2017) Identification of selection signals by large-scale whole-genome resequencing of cashmere goats. *Scientific Reports* **7**, 15142.

Liao Y., Smyth G.K. & Shi W. (2014) featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–30.

Love M.I., Huber W. & Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology* **15**, 550.

Martin P., Palhiere I., Maroteau C. *et al*. (2017) A genome scan for milk production traits in dairy goats reveals two new mutations in *Dgat1* reducing milk fat content. *Scientific Reports* **7**, 1872.

Martínez-Reyes I. & Chandel N.S. (2020) Mitochondrial TCA cycle metabolites control physiology and disease. *Nature Communications* **11**, 102.

Matera I., Watkins-Chow D.E., Loftus S.K., Hou L., Incao A., Silver D.L., Rivas C., Elliott E.C., Baxter L.L. & Pavan W.J. (2008) A sensitized mutagenesis screen identifies Gli3 as a modifier of Sox10 neurocristopathy. *Human Molecular Genetics* **17**, 2118–31.

Matschinsky F.M. & Wilson D.F. (2019) The central role of glucokinase in glucose homeostasis: a perspective 50 years after demonstrating the presence of the enzyme in islets of Langerhans. *Frontiers in Physiology* **10**, 148.

Miyazaki M. & Ntambi J.M. (2003) Role of stearoyl-coenzyme A desaturase in lipid metabolism. *Prostaglandins, Leukotrienes and Essential Fatty Acids* **68**, 113–21.

Mychaleckyj J.C., Nayak U., Colgate E.R. *et al*. (2018) Multiplex genomewide association analysis of breast milk fatty acid composition extends the phenotypic association and potential selection of *FADS1* variants to arachidonic acid, a critical infant micronutrient. *Journal of Medical Genetics* **55**, 459–68.

Neuenschwander S., Schwartz A., Wood T.L., Roberts C.T. Jr, Hennighausen L. & LeRoith D. (1996) Involution of the lactating mammary gland is inhibited by the IGF system in a transgenic mouse model. *The Journal of Clinical Investigation* **97**, 2225–32.

Oget C., Servin B. & Palhière I. (2018) Genetic diversity analysis of French goat populations reveals selective sweeps involved in their differentiation. *Animal Genetics* **50**, 54–63.

Onzima R.B., Upadhyay M.R., Mukiibi R., Kanis E., Groenen M.A.M. & Crooijmans R.P.M.A. (2018) Genome-wide population structure and admixture analysis reveals weak differentiation among Ugandan goat breeds. *Animal Genetics* **49**, 59–70.

Pizarro Inostroza M.G., Landi V., Navas González F.J., León Jurado J.M., Martínez Martínez M.D.A., Fernández Á.J. & Delgado Bermejo J.V. (2019) Non-parametric association analysis of additive and dominance effects of casein complex SNPs on milk content and quality in Murciano-Granadina goats. *Journal of Animal Breeding and Genetics* **137**, 407–22.

Purcell S., Neale B., Todd-Brown K. *et al*. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics* **81**, 559–75.

Quinlan A.R. & Hall I.M. (2010) BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–2.

Saravanan K.A., Panigrahi M., Kumar H., Bhushan B., Dutt T. & Mishra B.P. (2020) Selection signatures in livestock genome: a review of concepts, approaches and applications. *Livestock Science* **241**, 104257.

Song S., Yao N., Yang M. *et al*. (2016) Exome sequencing reveals genetic differentiation due to high-altitude adaptation in the Tibetan cashmere goat (*Capra hircus*). *BMC Genomics* **17**, 122.

Tosser-Klopp G., Bardou P., Bouchez O. *et al*. (2014) Design and characterization of a 52K SNP Chip for goats. *PLoS One* **9**, e86227.

Turner A.J. (2004) Membrane alanyl aminopeptidase. In: *Handbook of Proteolytic Enzymes* (2nd edn) (Ed by A.J. Barrett, N.D. Rawlings & J.F. Woessner), pp. 289–94. Academic Press, London, UK.

Våge D.I., Klungland H., Lu D. & Cone R.D. (1999) Molecular and pharmacological characterization of dominant black coat color in sheep. *Mammalian Genome* **10**, 39–43.

Vitavska O. & Wieczorek H. (2013) The SLC45 gene family of putative sugar transporters. *Molecular Aspects of Medicine* **34**, 655–60.

Vitti J.J., Grossman S.R. & Sabeti P.C. (2013) Detecting natural selection in genomic data. *Annual Review of Genetics* **47**, 97–120.

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

Wolf Horrell E.M., Boulanger M.C. & D'Orazio J.A. (2016) Melanocortin 1 receptor: structure, function, and regulation. *Frontiers in Genetics* **7**, 95.

Xu S.-S., Ren X., Yang G.-L. *et al*. (2017) Genome-wide association analysis identifies the genetic basis of fat deposition in the tails of sheep (*Ovis aries*). *Animal Genetics* **48**, 560–9.

Zhang B., Chang L., Lan X. *et al*. (2018) Genome-wide definition of selective sweeps reveals molecular evidence of trait-driven domestication among elite goat (*Capra* species) breeds for the production of dairy, cashmere, and meat. *GigaScience* **7**, giy105.

Zhou X. & Stephens M. (2012) Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**, 821–4.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Figure S1 Description of the experimental design.

Table S1 Primers used in the amplification of the coding region of the melanocortin 1 receptor gene.

Table S2 Custom TaqMan probes used to genotype the rs669694251 polymorphism.

Table S3 List of selective sweeps detected with the iHS statistic in Murciano-Granadina goats.

Table S4 List of enriched pathways/terms using the set of genes co-localizing with putative selective sweeps.

Table S5 List of genes mapping to selective sweeps and showing differential expression in lactating (T1 and T2) and non-lactating (T3) goats.

Table S6 Genotypes of the rs657434682 and rs669694251 SNP markers assessed through the visual inspection of sequencing electropherograms.