



**HAL**  
open science

## Signatures of selection and environmental adaptation across the goat genome post-domestication

Francesca Bertolini, Bertrand Servin, Andrea Talenti, Estelle Rochat, Eui Soo Kim, Claire Oget, Isabelle Palhière, Alessandra Crisà, Gennaro Catillo, Roberto Steri, et al.

### ► To cite this version:

Francesca Bertolini, Bertrand Servin, Andrea Talenti, Estelle Rochat, Eui Soo Kim, et al.. Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genetics Selection Evolution*, 2018, 50 (1), pp.57. 10.1186/s12711-018-0421-y . hal-02405277

**HAL Id: hal-02405277**

**<https://hal.science/hal-02405277v1>**

Submitted on 11 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.




Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Open Access



# Signatures of selection and environmental adaptation across the goat genome post-domestication

Francesca Bertolini<sup>1,2\*</sup> , Bertrand Servin<sup>3</sup>, Andrea Talenti<sup>4</sup>, Estelle Rochat<sup>5</sup>, Eui Soo Kim<sup>6</sup>, Claire Oget<sup>3</sup>, Isabelle Palhière<sup>3</sup>, Alessandra Crisà<sup>7</sup>, Gennaro Catillo<sup>7</sup>, Roberto Steri<sup>7</sup>, Marcel Amills<sup>8</sup>, Licia Colli<sup>9,10</sup>, Gabriele Marras<sup>11</sup>, Marco Milanesi<sup>9,12</sup>, Ezequiel Nicolazzi<sup>11</sup>, Benjamin D. Rosen<sup>13</sup>, Curtis P. Van Tassell<sup>13</sup>, Bernt Gulbrandsen<sup>14</sup>, Tad S. Sonstegard<sup>6</sup>, Gwenola Tosser-Klopp<sup>3</sup>, Alessandra Stella<sup>10</sup>, Max F. Rothschild<sup>1</sup>, Stéphane Joost<sup>5</sup>, Paola Crepaldi<sup>4</sup> and the AdaptMap consortium

## Abstract

**Background:** Since goat was domesticated 10,000 years ago, many factors have contributed to the differentiation of goat breeds and these are classified mainly into two types: (i) adaptation to different breeding systems and/or purposes and (ii) adaptation to different environments. As a result, approximately 600 goat breeds have developed worldwide; they differ considerably from one another in terms of phenotypic characteristics and are adapted to a wide range of climatic conditions. In this work, we analyzed the AdaptMap goat dataset, which is composed of data from more than 3000 animals collected worldwide and genotyped with the CaprineSNP50 BeadChip. These animals were partitioned into groups based on geographical area, production uses, available records on solid coat color and environmental variables including the sampling geographical coordinates, to investigate the role of natural and/or artificial selection in shaping the genome of goat breeds.

**Results:** Several signatures of selection on different chromosomal regions were detected across the different breeds, sub-geographical clusters, phenotypic and climatic groups. These regions contain genes that are involved in important biological processes, such as milk-, meat- or fiber-related production, coat color, glucose pathway, oxidative stress response, size, and circadian clock differences. Our results confirm previous findings in other species on adaptation to extreme environments and human purposes and provide new genes that could explain some of the differences between goat breeds according to their geographical distribution and adaptation to different environments.

**Conclusions:** These analyses of signatures of selection provide a comprehensive first picture of the global domestication process and adaptation of goat breeds and highlight possible genes that may have contributed to the differentiation of this species worldwide.

\*Correspondence: fbert@iastate.edu; franb@dtu.dk

<sup>1</sup> Department of Animal Science, Iowa State University, Ames, IA 50011, USA

Full list of author information is available at the end of the article



## Background

The goat (*Capra hircus*) is considered one of the earliest domesticated livestock species. The domestication process started around 10,000 years ago in the Fertile Crescent area from a unique wild and still living ancestor, the bezoar or *Capra aegagrus* [1]. At present, there are more than one billion of goats that inhabit all types of ecological areas across the globe [2]. Compared with the other major livestock species such as pigs, cattle and sheep, goats have undergone the largest increase (+34%) in population since 2000, i.e. larger than pigs (+15%), cattle (+14%) and sheep (+14%) (<http://www.fao.org/faostat/en/>). Today, over 90% of the goats are distributed across Asia and Africa, followed by the Americas, Europe, and Oceania [3]. In the most rural areas of the world, goats are often considered the poor person's cow. In fact, goats can be used for milk, meat, fiber, and leather production [4], as well as transport. Moreover, they are easy to house and manage; goats can be raised by small families, women, and children and provide a fundamental source of food for millions of people. Several factors have contributed to the differentiation of goat breeds, which are classified mainly into two types: (i) adaptation to different breeding systems and/or purposes, i.e. in some countries, breeds have been selected for specific production traits such as milk (e.g. Saanen and Alpine), meat (e.g. Landrace and Boer) and fiber (e.g. Angora and Cashmere); and (ii) adaptation to different environments, i.e. goats have adapted to various agro-climatic conditions. In addition, goat breeds have undergone differentiation through founder effects, and the processes of admixture and genetic drift. Thus, about 600 breeds have been developed and are distributed worldwide [5]. They differ from one another in terms of many phenotypic characteristics such as size, color, horn shape and dimension, reproductive and productive traits and are adapted to a wide range of bioclimatic conditions. Directional natural and artificial selection events have left footprints across the genome, which are known as signatures of selection. Signatures of selection are defined as the reduction, elimination or change of genetic variation in genomic regions that are adjacent to causative variants in response to natural or artificial selective pressure. Such variants usually affect several traits and contribute to shaping a breed [6]. The process by which the frequency of a selectively favored variant increases in a population is termed a selective sweep. The recent development of species-specific genomic tools (such as single nucleotide polymorphism (SNP) arrays) have allowed researchers to extend whole-genome analyses to livestock species, which cover many aspects of

genetic diversity, including signatures left by selection processes [7].

In spite of the major economic importance of goats, high-throughput genomic resources for this species have become available only recently. In 2011, the International Goat Genome Consortium developed and released the first high-throughput SNP chip with more than 50,000 SNPs (Illumina CaprineSNP50 BeadChip), which was built using 10 biologically and geographically different breeds [8]. The first complete assembly of the goat genome was released in 2013 by Dong et al. [9] and a second version of the reference genome that exploits single-molecule long read sequencing (PacBio) has just been released (ARS1; [10]). With this new version, gene annotation has improved considerably and the position of the SNPs on the CaprineSNP50 BeadChip has been updated. With the availability of genomic and high-throughput SNP tools, there is an increased interest in identifying and exploring signatures of selection and the genomic diversity resulting from adaptation to environment and human selective pressure. These genomic tools were used to identify signatures of selection in circumscribed datasets. The CaprineSNP50 BeadChip was used to investigate and compare several Swiss goat breeds and genomic signatures of selection were detected in regions that affect variation in coat color, growth, and milk composition [11]. The combination of analyses of runs of homozygosity (ROH),  $F_{ST}$  (fixation index), XP-EHH (cross population extended haplotype homozygosity), and the use of Bayesian methods allowed the detection of signatures of selection in a region that contains genes related to the immune system in another mountain breed raised in the North of Italy, the Valdostana Italian goat breed [12]. Other analyses of signatures of selection were performed in a reduced number of breeds, thus detecting regions that are linked to production and reproduction traits [13] and in the Barki Egyptian goat breed, which is raised and adapted to hot/dry environments [14]. In the latter study, analyses of iHS (integrated haplotype score) and pairwise  $F_{ST}$  that identified selective sweeps led to the identification of genes related to thermotolerance, body size, energy metabolism and nervous and auto-immune response [14]. Furthermore, signatures of selection that were linked to dry and hot conditions and to metabolic traits were identified by using whole-genome sequence (WGS) data from Moroccan goat breeds through XP-CLR analysis [15]. WGS information was also successfully used to detect regions that are under different selection pressures in Chinese and Mongolian goats and are related to breeding or reproductive traits [16, 17].

Genomic changes that result from climate changes and are linked to adaptation to different environments can be

analyzed with a landscape genomic approach, which was successfully applied in other livestock species, for example in Ugandan cattle [18]. To date in goats, this approach was applied only on a reduced number of SNPs or amplified fragment length polymorphism (AFLP) markers [19, 20]. However, it would be interesting to apply it at the genomic level since it provides useful information on the environmental factors that shaped the genome.

In this work, we used different approaches to identify regions under artificial and environmental selection across the AdaptMap goat dataset, which is composed of data from more than 3000 animals that were collected worldwide and genotyped with the CaprineSNP50 Bead-Chip. For this purpose, we considered several groups of animals that were partitioned based on geographical area, as done in Colli et al. [21], production uses, available records on solid coat color and environmental variables in relation to the geographical coordinates of sampling. For the first time, we applied FLK and hapFLK analyses on goat data, which have been successfully used in analyses of high-throughput data from sheep [22, 23]. These methods increase the power of detection for signatures of selection, and they enable detection of soft or incomplete selective sweeps. Finally, we also applied landscape genomic approaches to this large goat dataset, to investigate the role of natural selection in shaping the goat genome.

## Methods

Prior to the application of methods for detecting signatures of selection, we applied several filtering steps to the AdaptMap goat dataset, which originally contained 4563 animals from 144 breeds that were collected worldwide and genotyped with the CaprineSNP50 Bead-Chip. This dataset was first edited by removing mixed breeds, related animals, SNPs that were monomorphic, unmapped, mapped to sex chromosomes or with low call rate, which resulted in a working dataset of 46,654 SNPs and 3197 animals [21].

Additional filtering steps were applied depending on the analysis performed, as described below. The main analyses are summarized as follows: (1) detection of signatures of selection based on the genetic diversity of subcontinental populations or breeds, and (2) detection of signatures of selection associated with specific phenotypes, traits or external variables (e.g. annual mean temperature based on global positioning system (GPS) coordinates). In general, each investigation was conducted using two complementary approaches. For each analysis, genes within selected regions or nearby detected SNPs ( $\pm 100$  kb) were identified using the Bedtools

software [24] and the most recent version of the goat genome assembly (ARS1; [10]).

## Signatures of selection based on genetic diversity of subcontinental populations or breeds

### Signatures between and within subcontinental groups

To detect signatures of major differentiation between populations, we used the FLK [25] and hapFLK [26] approaches. Briefly, these methods account for population structure and differences in effective population size by modelling the genetic divergence between populations as derived from drift and population division. Because these methods are not completely robust to strong bottlenecks and large admixture events, subsets of populations were selected starting from the working dataset, removing admixed animals and strongly inbred populations. This was done first by identifying genetic sub-structure in the initial diversity analysis which defined sub-continental groups [21]. Then, an admixture analysis was carried out to identify and remove admixed populations within each sub-continental group, which generated sub-continental filtered groups. This last step was performed using the Treemix software [27], allowing for up to three migration events within each group. Graphical representations of the results of all Treemix analyses are in Additional file 1. The number of animals included in each sub-continental group is in Table 1. The final dataset comprised 2481 individuals that are grouped into 62 (61 *Capra hircus* + one *Capra aegagrus*) populations.

Following filtering of the data, FLK and hapFLK analyses were carried out on each of the sub-continental filtered groups, using the wild ancestor of domestic goat (Bezoar, *Capra aegagrus*) as an outgroup to root population trees. For hapFLK, the cross-validation procedure was performed with the fastPHASE software [28], which determined that 30 haplotype clusters were needed to capture haplotype diversity. For both hapFLK and FLK analyses, p-values were computed as explained in the hapFLK software documentation. False discovery rates (FDR) were estimated using the qvalue R package [29] and SNPs corresponding to an FDR of 0.15 or less were considered significant.

A genome scan for signatures of selection between sub-continental groups was also performed using FLK analysis as described by Fariello et al. [22]. Briefly, the frequency of the ancestral allele of each group was estimated from the within-group analysis. These ancestral alleles were then used to perform a new genome scan using FLK analysis. Frequencies of Bezoar alleles were used to root the population tree. Only the groups that corresponded to clear geographical clustering of goat populations were considered for this ancestral analysis,

**Table 1 Number of animals and breeds that composed the sub-continental groups after filtering steps**

Groups	Breed code	Breed name (Country)	Number
Alpines	ALP_FR	<i>Alpine (France)</i>	50
	BIO	<i>Bionda dell'Adamello</i>	24
	FSS	<i>Fosses</i>	24
	ORO	<i>Orobica</i>	22
	PTV	<i>Poitevine</i>	27
	SAA_FR	<i>Saanen (France)</i>	50
	VAL	<i>Valdostana</i>	24
	VSS	<i>Valpassiria</i>	24
Angoras	ANG_AR	<i>Angora (Argentina)</i>	50
	ANG_FR	<i>Angora (France)</i>	26
	ANG_ZA	<i>Angora (South Africa)</i>	48
	ANK	<i>Ankara</i>	18
	IRA	<i>Iranian goat (unknown)</i>	9
	KIL	<i>Kil</i>	23
Boers	BOE_AU	<i>Boer (Australia)</i>	32
	BOE_CH	<i>Boer (Switzerland)</i>	50
	BOE_US	<i>Boer (United States)</i>	29
	BOE_ZW	<i>Boer (Zimbabwe)</i>	17
Central Asia	THA	<i>Thari</i>	16
	TED	<i>Teddi</i>	47
	PAH	<i>Pahari</i>	19
	KAC	<i>Kachan</i>	19
	DDP	<i>DDP</i>	20
East Africa	ABR	<i>Abergelle</i>	49
	GAL	<i>Galla</i>	23
	GUM	<i>Gumez</i>	39
	KAR	<i>Karamonja</i>	19
	KEF	<i>Keffa</i>	44
	MAA	<i>Maasai</i>	18
	PRW	<i>Pare White</i>	19
	SEA	<i>Small East African</i>	50
	SEB	<i>Sebei</i>	21
	SNJ	<i>Sonjo</i>	20
Egypt	WYG	<i>Woyito Guji</i>	39
	BRK	<i>Barki</i>	50
	NBN_EG	<i>Nubian (Egypt)</i>	50
	OSS	<i>Oasis</i>	50
Northern Europe	SID	<i>Saidi</i>	50
	LNR_DK	<i>Landrace (Denmark)</i>	50
	LNR_FI	<i>Landrace (Finland)</i>	20
North west Africa	NRW	<i>Norwegian</i>	17
	CAM	<i>Cameroon goat</i>	37
	GUE	<i>Malagueña</i>	16
	PEU	<i>Peulh</i>	22
	RSK	<i>Red Sokoto</i>	19
	SAH	<i>Sahel</i>	15
	TAR	<i>Targui</i>	19
	WAD	<i>West African Dwarf</i>	50

i.e. the Southwestern, Southeastern, Northern and Alpine European groups, the Central Asian group and the Northwestern, Easter and Southern African groups.

#### Signatures of selection within breeds

The above-mentioned filtered dataset that was used for FLK and hapFLK analyses was also used to detect signatures of selection within the breeds using: (i) ROH and (ii) iHS statistic [30]. ROH analyses were performed using the Zannardi software [31] by considering a minimum of 15 SNPs per ROH, a minimum ROH length of 1 Mb and allowing for one heterozygous SNP within an ROH to account for the possibility of genotyping errors. For each SNP, the proportion of animals that displayed a homozygous region at that SNP was calculated. Then, this measure was transformed by its empirical quantile for all SNPs across the genome (i.e. the proportion of all SNPs with a higher or equal proportion of homozygous animals in the breed considered). The iHS analyses were performed in populations with at least 10 genotyped individuals. Within each population, phasing of individuals was performed using SHAPEIT2 [32]. For SNPs with a minor allele frequency higher than 0.05, the ancestral alleles were randomly assigned. The iHS statistics were calculated using the rehh v2.0 R package [33]. Standardized iHS values were computed in allele frequency bins of 0.05 and then further corrected using robust estimations of their mean and variance using the rlm function from the MASS R package [34]. Because assessing significance of ROH and iHS values is less robust, ROH and iHS signals were reported only for significant FLK or hapFLK signatures of selection.

#### Signatures of selection associated with specific traits or external covariates

##### External phenotypes and production traits

Reduced subsets of the dataset were investigated to detect genomic regions associated with specific phenotypes. Reductions were performed according to the availability of the information for each animal or breed included in the dataset. Therefore, two reduced panels were created corresponding to differences related to production purposes and phenotypes (solid coat colors).

**Panel 1: production purposes** A questionnaire that contained four possible production assignments (milk, meat, fiber, and leather) was circulated to all AdaptMap members who provided the samples, to obtain information on the main purpose of each breed. Pre-filtering was performed by considering only breeds with known and prevalent purposes. Since none of the breeds was described



**Table 1 (continued)**

Groups	Breed code	Breed name (Country)	Number
South Africa	DZD	<i>Dedza</i>	15
	LND	<i>Landin</i>	29
	MSH	<i>Mashona</i>	22
South eastern Europe	ARG	<i>Argentata</i>	24
	CCG	<i>Ciociara Grigia</i>	16
	DIT	<i>Di Teramo</i>	19
	GAR	<i>Garganica</i>	15
	GGT	<i>Girgentana</i>	24
South western Europe	BEY	<i>Bermeya</i>	23
	MAL	<i>Mallorquina</i>	18
	MLG	<i>Malaguena</i>	40
	MUG	<i>Murciano-Granadina</i>	20
	PYR	<i>Pyreanean</i>	26
	RAS	<i>Blanca de Rasquera</i>	20

Country is indicated when necessary for the analysis

as being solely raised for leather production, this purpose was excluded from subsequent analyses.

Then, another filtering step was performed on the working dataset to remove breeds that are raised for more than one purpose by using an in-house script that discarded animals belonging to different purpose groups that shared the same coordinates of the first component  $\pm 1.5$  of the overall quantile distribution (see Additional file 2: Figure S1). The final filtered dataset included 192 animals (three breeds) for fiber, 241 animals (12 breeds) for meat and 818 animals (23 breeds) for milk. Details on breeds and sample sizes are in Table 2a. Three types of analyses were carried out on the final filtered dataset: ROH,  $F_{ST}$  and XP-EHH. ROH analyses were performed for each separate purpose group using the Zannardi software [31] by considering a minimum of 15 SNPs per ROH, a minimum ROH length of 1 Mb and allowing for one heterozygous SNP within an ROH to account for the possibility of genotyping errors, as previously mentioned. The  $F_{ST}$  and XP-EHH analyses were carried out by comparing each group against all the others and using the script described by Talenti et al. [12] in which 1-Mb windows with an overlap of 500 kb were considered for window-based  $F_{ST}$  and the Selscan software [35] for XP-EHH. The results were normalized with the norm normalization tool included in the software suite.

For all three analyses (ROH, window-based  $F_{ST}$  and XP-EHH), the top 0.5% of SNPs based on marker or window distribution was retained as relevant. The Bedtools software [24] was used to find consensus regions between two or all three approaches. Only regions detected by

at least two approaches were considered for further analyses.

**Panel 2: solid coat colors** Pictures representative of the breeds for each animal of the working dataset were provided by the AdaptMap members and were inspected to find common and unique patterns across each breed. Considering the high variability of patterns and the lack of availability of high-quality pictures for some breeds, only those with confident solid coat colors were considered. Three groups were created: a Black group (79 animals and three breeds), a White group (375 animals and six breeds), and a Red group (93 animals and three breeds). Breeds and sample sizes are in Table 2b. For each group, a  $F_{ST}$  analysis was carried out by considering the following pairwise comparisons: White vs. Black and Red, Black vs. White and Red vs. White. These comparisons were performed using the script described by Talenti et al. [12] for which 500-kb windows with overlaps of 250 kb were considered. For each comparison, the top four windows that contained at least four SNPs, corresponding to the 0.9996 percentile of the overall distribution were considered. The genes within these windows were screened using the coat color gene website of the International Federation of Pigment Cell Society (<http://www.espcr.org/micemut>) to investigate their direct or indirect associations with coat colors in mice and shades of skin color in humans.

SNPs that were located on either side of candidate genes for both purpose and coat color panels ( $\pm 5$  Mb=short and  $\pm 10$  Mb=long) were used as sets of variables with the same animals for canonical discriminant analysis (CDA), which was performed with the CANDISC procedure implemented in SAS-stat software (SAS Institute, Inc., vers.9.4). These analyses allowed the identification of the SNPs that contributed most to the discrimination between groups. These contributions were summarized as canonical functions (CAN), which were linear combinations of the original variables (discriminant SNPs). Visual inspection of the CAN1 vs. CAN2 scatter plot of the CAN1 and CAN2 values for each SNP was used to pinpoint the precise regions associated with the separation between groups.

#### **Signatures of selection associated with bioclimatic environmental variables**

To study the influence of the environment on the distribution of adaptive genetic variation, a landscape genomic approach using the Samβada software [36] and  $F_{ST}$  analyses were applied to identify genotypes that are significantly associated with environmental variables. For this purpose, we used a reduced subset of samples with

**Table 2** List of filtered animals and breeds for comparisons of breeds with different production purposes<sup>a</sup> and coat colors<sup>b</sup>

Purpose <sup>a</sup>	Breed code	Breed name	Number	Coat color <sup>b</sup>	Breed code	Breed name	Number	
Milk	ALP	<i>Alpine (Camosciata delle Alpi)</i>	150	White	ANG	<i>Angora</i>	131	
	ARG	<i>Argentata</i>	24		ANK	<i>Ankara</i>	18	
	ASP	<i>Aspromontana</i>	23		CAS	<i>Nicastrese</i>	44	
	BIO	<i>Bionda dell'Adamello</i>	24		CRP	<i>Carpathian</i>	14	
	CCG	<i>Ciociana Grigia</i>	16		GAL	<i>Galla</i>	23	
	CRS	<i>Corse</i>	29		SAA	<i>Saanen</i>	145	
	DIT	<i>Di Teramo</i>	19		Black	DDP	<i>DDP</i>	20
	GAR	<i>Garganica</i>	15			KIL	<i>Kil</i>	23
	LNR	<i>Landrace</i>	85			KLS	<i>Killis</i>	36
	Fiber	MLG	<i>Malagueña</i>		40	Red	BEY	<i>Bermeya</i>
		MLS	<i>Maltese Sarda</i>	12	RME		<i>Rossa Mediter-ranea</i>	30
		MLT	<i>Maltese</i>	16	MLG	<i>Malaguena</i>	40	
		MUG	<i>Murciano-Granadina</i>	20				
		NIC	<i>Nicastrese</i>	20				
		NRW	<i>Norwegian</i>	17				
		ORO	<i>Moroccan</i>	22				
		PTV	<i>Poitevine</i>	27				
		PVC	<i>Provencale</i>	17				
		RME	<i>Rossa Mediterranea</i>	30				
	SAA	<i>Saanen</i>	142					
SAR	<i>Sarda</i>	27						
TOG	<i>Toggenburg</i>	19						
VSS	<i>Valpassiria</i>	24						
Meat	BOE	<i>Boer</i>	138					
	BRI	<i>Bari</i>	25					
	LOP	<i>Local_Pothohari</i>	13					
	RAN	<i>Murciano-Granadina</i>	2					
	TED	<i>Teddi</i>	47					
	THA	<i>Thari</i>	16					

known geographic coordinates of the sampling point. This reduced dataset contained 2661 animals from 28 countries (Table 3). Finally, an additional filtering step was applied before performing the  $F_{ST}$  analyses (see below).

**Landscape genomic analyses** The environmental conditions were characterized using the bioclimatic variables from the WorldClim database (<http://www.worldclim.org/>) (see Additional file 3: Table S1). These data were available worldwide at a resolution of one arc-second and represented an average of the conditions from years 1950 to 2000. The values of these variables were extracted for the coordinates of each sampling point using the QGIS

2.14.7 software and were centered-reduced. A buffer area around each sampling point was generated to integrate, into the model, the environmental variability of the area that could influence an individual. Here, a radius of 5 km was selected which assumed that the goat could move in a circular area of 10 km in diameter, centered on the sampling point. Within that area, in order to consider only the area corresponding to the potential home range of the goats, a land cover discrimination based on the Global Land Cover 2000 classification was applied [37]. On this basis, the areas corresponding to artificial surfaces (urban land cover) and water bodies were removed. For the remaining areas, the bioclimatic variables were retrieved and for each one, seven statistical measures (minimum,

**Table 3 List of breeds (name, code and number of animals per breed) with available GPS coordinates used for landscape genomic analysis**

Breed code	Breed name	Number	Breed code	Breed name	Number	Breed code	Breed name	Number
ABR	<i>Abergelle</i>	49	GHA	<i>Ghazalia</i>	4	NSJ	<i>Nsanje</i>	6
ALB	<i>Alpine x Boer</i>	5	GOG	<i>Gogo</i>	12	OIG	<i>Old Irish Goat</i>	11
ALP	<i>Alpine (Camosciata delle Alpi)</i>	146	GUE	<i>Guera</i>	16	ORO	<i>Orobica</i>	22
AND	<i>Nganda</i>	6	GUM	<i>Gumez</i>	39	OSS	<i>Oasis</i>	50
ANG	<i>Angora</i>	80	IRA	<i>Iranian goat</i>	9	PAF	<i>Pafuri</i>	4
ARG	<i>Argentata</i>	24	JAT	<i>Jattan</i>	15	PAH	<i>Pahari</i>	19
ARR	<i>Traditional Arran</i>	8	JON	<i>Jonica</i>	11	PAL	<i>Palmera</i>	14
ASP	<i>Aspromontana</i>	23	KAC	<i>Kachan</i>	19	PAT	<i>Pateri</i>	27
BAB	<i>Barbari</i>	16	KAM	<i>Kamori</i>	38	PEU	<i>Peulh</i>	22
BAR	<i>Barcha</i>	4	KAR	<i>Karamonja</i>	19	PRW	<i>Pare White</i>	19
BAW	<i>Balaka-Ulongwe</i>	12	KEF	<i>Keffa</i>	44	PTV	<i>Poitevine</i>	27
BEY	<i>Bermeya</i>	23	KES	<i>Koh-e-sulmani</i>	13	PVC	<i>Provençale</i>	15
BEZ	<i>Bezoar</i>	7	KIG	<i>Kigezi</i>	4	PYR	<i>Pyrenean</i>	26
BIO	<i>Bionda dell'Adamello</i>	24	LGW	<i>Lilongwe</i>	3	RAN	<i>Rangeland</i>	50
BLB	<i>Bilberry</i>	10	LND	<i>Landin</i>	29	RAS	<i>Blanca de Rasquera</i>	20
BOE	<i>Boer</i>	108	LNR	<i>Landrace</i>	50	RME	<i>Rossa Mediterranea</i>	30
BRI	<i>Bari</i>	25	LOH	<i>Lohri</i>	17	RSK	<i>Red Sokoto</i>	19
BRK	<i>Barki</i>	50	LOP	<i>Local Pothohari</i>	13	SAA	<i>Saanen</i>	106
BUT	<i>Bugituri</i>	31	MAA	<i>Maasai</i>	18	SAH	<i>Sahel</i>	15
CAM	<i>Cameroon goat</i>	37	MAL	<i>Mallorquina</i>	18	SDN	<i>Soudanaise</i>	22
CAN	<i>Caninde</i>	23	MAN	<i>LaMancha</i>	3	SEA	<i>Small East African</i>	50
CAS	<i>Cashmere</i>	44	MAU	<i>Maure</i>	13	SEB	<i>Sebei</i>	21
CCG	<i>Ciociana Grigia</i>	1	MEN	<i>San Clemente</i>	19	SHL	<i>Sahel</i>	19
CHA	<i>Chappar</i>	9	MLG	<i>Malagueña</i>	24	SID	<i>Saidi</i>	50
CRE	<i>Creole</i>	49	MLY	<i>Malya</i>	11	SNJ	<i>Sonjo</i>	20
CRO	<i>Local Cross</i>	5	MOR	<i>Moroccan goat</i>	10	SOF	<i>Sofia</i>	22
CRP	<i>Carpathian</i>	14	MOX	<i>Moxoto'</i>	23	SOU	<i>SudOuest</i>	8
CRS	<i>Corse</i>	29	MSH	<i>Mashona</i>	22	TAP	<i>Tapri</i>	22
DDP	<i>DDP</i>	20	MTB	<i>Matebele</i>	22	TAR	<i>Targui</i>	19
DIA	<i>Diana</i>	14	MUB	<i>Mubende</i>	18	TED	<i>Teddi</i>	47
DJA	<i>Djallonke</i>	10	MUG	<i>Murciano-Granadina</i>	20	THA	<i>Thari</i>	16
DRA	<i>Draa</i>	4	NAI	<i>Naine</i>	14	THY	<i>Thyolo</i>	9
DZD	<i>Dedza</i>	15	NBN	<i>Nubian</i>	63	TOG	<i>Toggenburg</i>	20
FSS	<i>Fosses</i>	24	NDA	<i>Noire de l'Atlas</i>	4	TUN	<i>Tunisian</i>	21
GAL	<i>Galla</i>	23	NGD	<i>Nganda</i>	11	VAL	<i>Valdostana</i>	24
GAR	<i>Garganica</i>	15	NOR	<i>Nord</i>	4	WAD	<i>West African Dwarf</i>	49
GAZ	<i>Gaza</i>	4	NRW	<i>Norwegian</i>	17	WYG	<i>Woyito Guji</i>	39

maximum, mean, standard deviation, range, median and mode) were computed.

The Samβada software [36] was used to compute the parallel processing of multiple univariate logistic regressions between each genotype versus each environmental variable. All univariate models were computed and filtered considering a significance threshold of 0.05 before Bonferroni correction. Then, a second step of filtering was applied to retrieve only the SNPs

for which at least two genotypes were significantly associated with an environmental variable and showed simultaneously: (1) a very strong effect of the environment on the genotype (absolute value of the  $\beta_1$  regression coefficient higher than the 0.99 percentile of all absolute values of  $\beta_1$  of the significant models), (2) a high statistical significance of the association (G and Wald scores higher than the 0.95 percentile of all G and Wald scores of the significant models) and (3) a strong



goodness of fit of the models, with an Akaike's information criterion (AIC) lower than the 0.10 percentile of all AIC of the significant models. The Enrichr database [38, 39] was used for the genes detected with the Samβada analyses to identify the major biological processes that involved the identified genes (GO biological process) by considering significant clusters with a P value lower than 0.05.

**$F_{ST}$  analyses** Each animal of the working dataset with a known GPS geographic location was assigned to a Köppen climate group (Tropical, Dry, Temperate, and Cold; [40]) through the website climate-data.org. Then, we applied the following filters: (i) for each climate group, only breeds with at least 10 animals allocated were considered, and (ii) if the same breed was in two or more groups only the animals that belonged to the groups of known breed origin were considered and the others were discarded. The assignment of each individual to a Köppen group and the subsequent filtering of animals provided a dataset of 1689 animals (141 for the Tropical group, 796 for the Dry group, 632 for the Temperate group and 120 for the Continental group), with no redundant breeds across the groups; assignments are summarized in Table 4. MDS (multidimensional scaling) of the filtered animals was performed using the Plink 1.9 software [41]. Single SNP  $F_{ST}$  was performed by comparing each group to the remaining groups merged together. The top 20 SNPs of each comparison, corresponding to the 0.9995 percentile of the distribution were considered and compared with the previous results of the landscape genomics analysis that had been filtered based on a significance threshold of 0.05 before Bonferroni correction. Only the selected  $F_{ST}$  SNPs with a G score or Wald score and  $\beta_1$  regression coefficient higher than 0.99 quantile were retained for further analysis. Allele frequencies of the SNPs that were shared by different groups were calculated and compared using the Plink 1.9 software [41].

## Results

### Selection of goat populations for the analysis of signatures of selection based on worldwide genetic diversity

A dataset of goat breeds ensuring robust modelling was developed prior to FLK and hapFLK analyses. This dataset included 61 breeds, which overall represented all the genetic diversity present in the complete dataset (Fig. 1). The genetic diversity in this set of breeds mirrored their geographical origin and was consistent with a radiation from the domestication center. Thus, each population was assigned to one of the 11 sub-continental groups, and each sub-continental group was analyzed independently. In Fig. 1, the root of the population tree, which is located at the center, corresponded to the ancestral population

**Table 4 Breed composition and number of animals according to the Köppen group classification**

Köppen group	Breed code	Breed name	Number	
Tropical	CAM	Cameroon goat	11	
	NAI	Naine	14	
	WAD	West African Dwarf	15	
	SEA	Small East African	16	
	CAN	Caninde	23	
	MOX	Moxoto'	23	
	GUM	Gumez	39	
	Dry	CHA	Chapparr	9
		KES	Koh-e-sulmani	13
		LOP	Local Pothohari	13
		MAU	Maure	13
		JAT	Jattan	15
		SAH	Sahel	15
		BAB	Barbari	16
GUE		Guera	16	
THA		Thari	16	
LOH		Lohri	17	
MUG		Murciano-Granadina	17	
KAC		Kachan	19	
PAH		Pahari	19	
TAR		Targui	19	
DDP		DDP	20	
MTB		Matebele	22	
PEU		Peulh	22	
SDN		Soudanaise	22	
TAP		Tapri	22	
BRI		Bari	25	
PAT	Pateri	27		
BUT	Bugituri	31		
KAM	Kamori	38		
TED	Teddi	47		
ABR	Abergelle	49		
BRK	Barki	50		
OSS	Oasis	50		
RAN	Rangeland	50		
SID	Saidi	50		
NBN	Nubian	54		
Temperate	BLB	Bilberry	10	
	TOG	Toggenburg	10	
	JON	Jonica	11	
	MAL	Mallorquina	12	
	MLS	Maltese sarda	12	
	NRW	Norwegian	12	
	PAL	Palmera (Canaria breed)	12	
	LNR	Landrace	13	
	OIG	Old Irish Goat	13	

**Table 4 (continued)**

Köppen group	Breed code	Breed name	Number
	MLT	<i>Maltese</i>	14
	BEY	<i>Bermeya</i>	15
	GAR	<i>Garganica</i>	15
	CCG	<i>Ciociera Grigia</i>	16
	MSH	<i>Mashona</i>	16
	PVC	<i>Provençale</i>	16
	ASP	<i>Aspromontana</i>	17
	NIC	<i>Nicastrese</i>	17
	DIT	<i>Di Teramo</i>	19
	TUN	<i>Tunisian</i>	21
	VAL	<i>Valdostana</i>	21
	ORO	<i>Orobica</i>	22
	VSS	<i>Valpassiria</i>	22
	GGT	<i>Girgentana</i>	23
	ARG	<i>Argentata</i>	24
	FSS	<i>Fosses</i>	24
	ANG	<i>Angora</i>	25
	PYR	<i>Pyrenean</i>	26
	PTV	<i>Poitevine</i>	27
	SAR	<i>Sarda</i>	27
	CRS	<i>Corse</i>	28
	MLG	<i>Malagueña</i>	29
	RME	<i>Rossa Mediterranea</i>	30
	BOE	<i>Boer</i>	33
Continental	CRP	<i>Carpathian</i>	10
	BIO	<i>Bionda dell'Adamello</i>	17
	SAA	<i>Saanen</i>	46
	ALP	<i>Alpine (Camosciata delle Alpi)</i>	47

of all goat breeds, *i.e.* the ancestral domesticated population. Extant populations radiated from this ancestral population with populations that were geographically closer to the domestication center also usually genetically closer. The populations that were closest to the ancestor belonged to the “Angora” group, which originated from Turkey. A little further, in their own sub-group, there was a set of breeds from Pakistan (“Central Asia” group). The other two large sub-trees consisted of populations from Africa and Europe, which were split further into sub-continental groups.

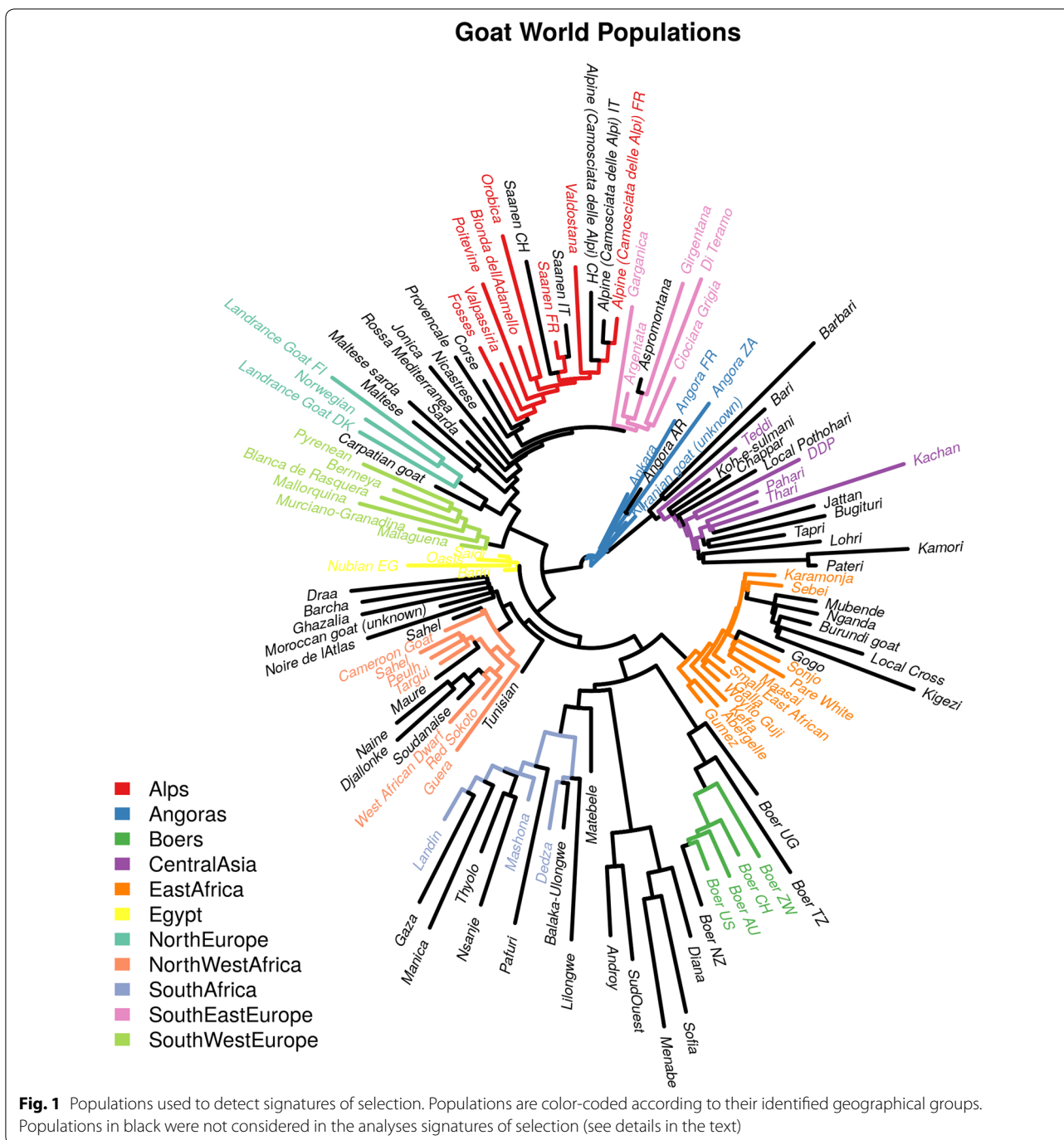
#### *Signatures of selection participating in the adaptive differentiation of goat populations*

The FLK and hapFLK analyses detected 67 signatures of selection among the 10 population groups. Figures S2 and S3 (see Additional file 2: Figures S2 and S3) show the Manhattan plots for the FLK and hapFLK analyses and Figure S4 (see Additional file 2: Figure S4) shows the overview of the genomic distribution of all signatures of selection. The complete list of signatures of selection is in Table S2 (see Additional file 3: Table S2). Graphical representations of FLK, hapFLK, ROH and iHS signals for all signatures of selection are in Additional file 4. The northern European group was the only group for which no significant signatures of selection were detected. The populations that constitute this group display long terminal branches in the population tree (Fig. 1), which indicates that they have all experienced a strong reduction in population size. Such a reduction in size creates extensive genetic drift, which alone can explain the genetic differentiation of these populations. Hence, such extensive genetic drift makes the detection of signatures of selection difficult, which can explain the lack of power observed for the northern European group.

Most of the signatures detected were specific to one population. A total of 13 genomic regions were shared between at least two groups (see Additional file 3: Table S2). These regions were annotated by (i) listing the genes that were located  $\pm 100$  kb near the highest FLK or hapFLK signal and (ii) identifying the populations that showed an elevated ROH and/or iHS signal within the region.

A signature of selection on chromosome 5 between 30 and 40 Mb encompasses the *ADAMTS20* (*ADAM metalloproteinase with thrombospondin type 1 motif 20*) gene (see Additional file 2: Figure S5 and Additional file 4). This genomic region matched with a ROH signal in the Pyrenean goat population of the southwestern European group and with an iHS signal in the Argentata dell'Etna breed of the southeastern European group. It also matched with ROH and iHS signals in the Sahel, Peulh and Thari breeds of the northwestern Africa group. These breeds were genetically quite homogeneous and clustered within the same population in the genetic diversity study, thus, they were combined for the ROH analysis. An iHS signal was present in the Murciano-Granadina population, but did not match with the position of the *ADAMTS20* gene, thus it may represent another selection target.

A second signature selection on chromosome 6 harbors the *KIT* (*KIT proto-oncogene receptor tyrosine kinase*) gene in three population groups (see Additional file 2: Figure S6 and Additional file 4). A ROH signal was



detected in the Kacchan breed of the central European group, the Abergelle breed of the east European group and the Argentata dell’Etna breed of the southeastern European group.

A third selection signature on chromosome 13 matched with the position of the *ASIP* (*Agouti signaling protein*) gene (see Additional file 2: Figure S7 and Additional file 4). In Pakistani breeds, a clear ROH signal,

corresponding to a large fixed haplotype was detected in the Kacchan population sample from Pakistan and in several breeds of the Alpine group (Alpine, Poitevine and Valdostana). The most significant FLK signal in the Alpine group corresponded to the SNP that was closest to *ASIP*, however, the highest FLK signal in central Asian breeds did not correspond to the same SNP and the region spanned several genes.

Another relevant signal was detected on chromosome 1 between 110 and 130 Mb in the Alpine and southwestern European groups (see Additional file 2: Figure S8 and Additional file 4). This region seemed to match with an FLK signal in the Alpine Valdostana breed, which presents a high proportion of homozygous individuals for this region. An iHS signal also on chromosome 1, at ~108 Mb was found in the Ciociara Grigia Italian breed of the southeastern European group (see Additional file 2: Figure S26). Another signal was detected on chromosome 6 close to the cluster of casein genes *CSN1S1*, *CSN1S2* and *CSN2*, in the groups of Alpine breeds and Eastern African populations (see Additional file 2: Figure S9 and Additional file 4). The signature of selection detected for the Alpine populations clearly points to the casein genes while for the Eastern African populations, the FLK signals seem to point to a different cluster of genes that encode glucuronosyltransferase enzymes.

A complex signature of selection on chromosome 6 was identified between 25 and 50 Mb (see Additional file 2: Figure S10 and Additional file 4) and probably results from multiple signals in different genes and different groups. In the Egyptian group, the gene closest to the highest hapFLK signal was *LCORL* (*ligand dependent nuclear receptor corepressor like*). This region clearly matched with an extended region of shared ROH in the Nubian goat population from Egypt (see Additional file 4). Among the southwestern European populations, the Bermeya population presented an iHS signal at the same locus. This region in Egyptian populations overlaps with other signals detected in the southwestern and southeastern European groups. It contains the *ABCG2* (*ATP binding cassette subfamily G member 2*)

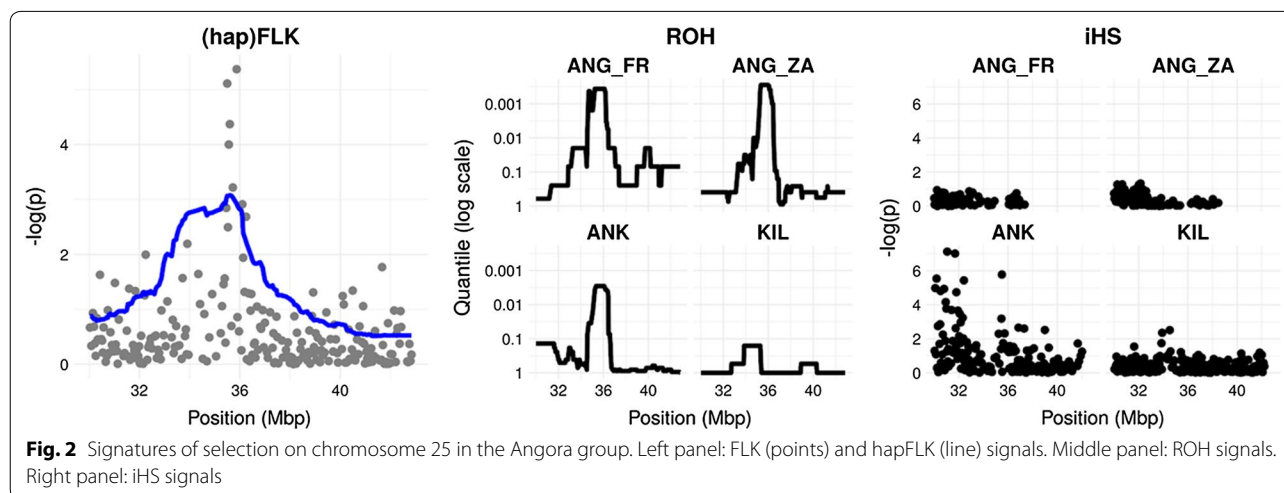
gene and matched with an iHS signal in the Murciano-Granadina breed of the southwestern European group and in the Argentata dell'Etna breed of the southeastern European group. In both population groups, the hapFLK signal is much stronger than the FLK signal and no clear ROH signal was observed in the breeds of these two groups. A possible explanation for these two observations is that the signature of selection is due to a soft selective sweep rather than a hard sweep.

Another signature of selection was found on chromosome 12 in the southwestern European group and the central Asian group (see Additional file 2: Figure S11 and Additional file 4). This region contained the *RXFP2* (*relaxin/insulin like family peptide receptor 2*) gene. The two populations that showed a clear signature of selection in this region were the Thari goat from Central Asia (iHS signal) and the Blanca de Rasquera Spanish breed (ROH signal).

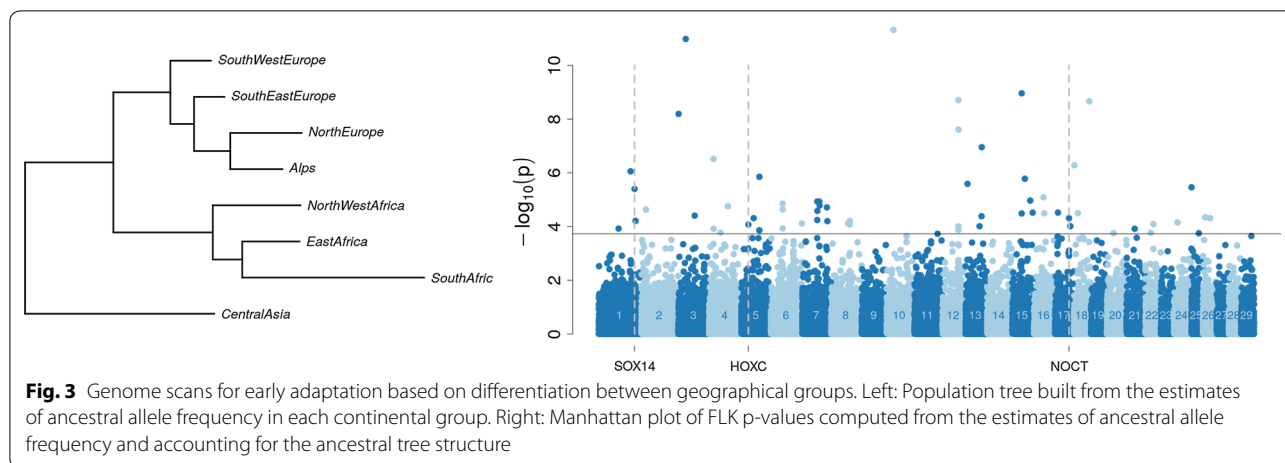
Finally, a region on chromosome 25 (35.50-35.88 Mb) was detected in the FLK analysis in the Angora group and confirmed by ROH and iHS analyses (Fig. 2).

#### Signatures of early adaptation

FLK analysis on the filtered geographical dataset based on the estimated ancestral allele frequency of geographical groups detected 62 SNPs with evidence for outlying differentiation among groups (Fig. 3) and (see Additional file 3: Table S3). The phylogenetic tree seemed to confirm the relationship between sub-continental blocks. Among the 62 SNPs, two signatures of selection were found on chromosome 1 close to the *SOX14* (*SRY-box 14*) gene and on chromosome 16 close to the *NOCT* (*nocturnin*) gene. Another signature of selection





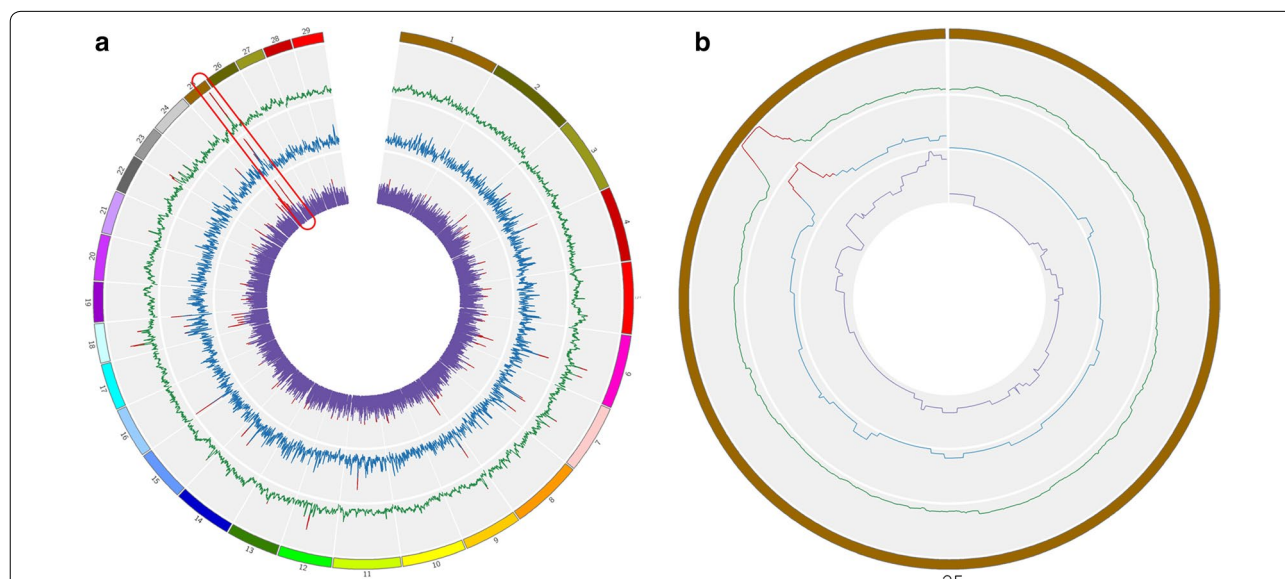


on chromosome 5 is located within the *HOXC* (*homeobox C*) gene cluster (see Additional file 3: Table S3).

**Signatures of selection associated with external phenotypes and production traits**

**Production purposes** The analyses performed on the group of “fiber-producing” goat breeds detected 18 regions on 11 chromosomes with ROH, 88 regions and/or SNPs on 27 chromosomes with XP-EHH and 24 regions

on 11 chromosomes with  $F_{ST}$  (Fig. 4). Among these, six regions on chromosomes 6, 18 and 25, including 34 genes, were detected by all three methods (see Additional file 3: Table S4). The region on chromosome 25 (34.69–36.43 Mb) showed the highest values in all the three analyses (Fig. 4a). Among the three breeds that compose the “fiber-producing” goat breed group, only Angora and Ankara showed a signature of selection in this region with none being detectable in Cashmere (Fig. 4b). This signa-



**Fig. 4** ROH,  $F_{ST}$  and XP-EHH for fiber (a), and detail of the ROH analyses on chromosome 25 for the breeds that compose the group of “fiber-producing” goat breeds: *Angora*, *Ankara* and *Cashmere* (b). (a) The three analyses are shown with different plot colors, within the most external squared-based circle, where each color represent a chromosome (chromosome number outside the squares): green (external) = ROH; blue (middle) =  $F_{ST}$ ; violet (internal): XP-EHH. For the three analyses, the regions above the threshold are marked in red. The region of high homozygosity (chromosome 25: 35,240,726–36,394,939 bp) is highlighted in red. (b) The three different breeds are labelled with different colors: green (external) = *Angora*; blue (middle) = *Ankara*; violet (internal): *Cashmere*. For the three breeds, the part corresponding to the 35–36 Mb region is marked in red when above the threshold



ture of selection overlapped with the regions detected by FLK/hapFLK, ROH and iHS analyses within the Angora group (Fig. 2) and contained 24 genes, including the *CUX1* (*cut like homeobox 1*) and the *PLOD3* (*procollagen-llysine,2-oxoglutarate 5-dioxygenase 3*) genes. The window-based  $F_{ST}$  analysis between Cashmere and Angora breeds pointed out several regions on chromosomes 2, 5, 6, 8, 9, 12, 13, 14, 17, 20, 25, 28 (data not shown). The ROH analyses on the Cashmere individuals highlighted several major regions on chromosomes 2, 10, 22 (47–48 Mb), 3, 5 and 8 (data not shown).

The analyses performed on the group of “meat-producing” goat breeds detected 25 regions on eight chromosomes with ROH, 121 regions and/or SNPs on 27 chromosomes with XP-EHH and 24 regions on 16 chromosomes with  $F_{ST}$  (see Additional file 2: Figure S12). Among these regions, two regions on chromosomes 3 and 18 were detected by at least two of the three methods and contained 18 genes [see Additional file 3: Table S5], including *TSHB* (*thyroid stimulating hormone beta*), *NRAS* (*neuroblastoma RAS viral oncogene homolog*) and *AMPD1* (*adenosine monophosphate deaminase 1*), all on chromosome 3.

The analyses performed for the group of “milk-producing” goat breeds detected 11 regions on eight chromosomes with ROH, 286 regions and/or SNPs in all autosomes with XP-EHH and 24 regions on 15 chromosomes with  $F_{ST}$  (see Additional file 2: Figure S13). Among these regions, four were on chromosomes 11, 13 and 14, which included 20 genes that were detected by at least two of the three methods (see Additional file 3: Table S6), among which the *EFEMP1* (*EGF containing fibulin like extracellular matrix protein 1*) gene on chromosome 11. The ROH analysis revealed a region on chromosome 6 (75–120 Mb) that contains the cluster of casein genes, although it was not confirmed by the  $F_{ST}$  and XP-EHH analyses (data not shown).

The CDA, which was performed on the chromosomal regions surrounding the genes previously identified for the three groups, confirmed the results found in previous analyses. The region on chromosome 25 (34.69–36.43 Mb) that contained the *CUX1* and *PLOD3* genes could separate individuals in the group of “fiber-producing” goat breeds from the other two groups on the CAN1 variable, and the SNPs that best separate the three groups were in the middle of these regions (see Additional file 2: Figure S14). The region on chromosome 3 that included the *TSHB*, *NRAS* and *AMPD1* genes could separate the group of “meat-producing” goat breeds from the remaining two groups on the CAN2 variable when considering the highest peak in the middle of the region (see Additional file 2: Figure S15). Finally, the region on chromosome 6 that contains

the casein cluster could separate the “milk-producing” breeds from the other two groups on the CAN1 variable, with a high peak near to the *CSN1S1* (*alpha S1 casein*) gene (see Additional file 2: Figure S16).

**Coat color** The distribution on the MDS plot of the single breeds using the medium-density SNP chip overlapped partially between the three groups of coat color. This probably reflects geography rather than coat color, with the Middle Eastern, Asian (Pakistan) and African goats being separated from all European goats on the first component (see Additional file 2: Figure S17). The Angora and Ankara (White) breeds and the Kil and Kilis (Black) breeds, which all originate from Turkey formed two clusters. For the “coat color” groups, we detected regions on chromosomes 5 (70.0–70.5 Mb), 13 (53.0–53.5 Mb), and 18 (15.50–16.25 Mb) in the Black vs. White comparison, regions on chromosomes 5 (36.25–36.75 Mb), 9 (11.5–12 Mb), and 13 (53–53.5 Mb and 62.75–63.25 Mb) in the White vs. Black and Red comparison, and regions on chromosomes 8 (27.0–27.5 Mb), 22 (2.25–3.0 Mb) and 29 (39.25–39.75 Mb) in the Red vs. White comparison. The list of genes within these regions is in Table S7 (see Additional file 3: Table S7) and includes *ADAMTS20* and *TIMP3* (*TIMP metalloproteinase inhibitor 3*) on chromosome 5, *SOX18* (*SRY-box 18*) and *ASIP* on chromosome 13, and *MC1R* (*melanocortin 1 receptor*) on chromosome 18. CDA on the *MC1R* gene showed that SNPs present on either side of this gene could distinguish the solid Red goats from the solid Black and White animals that were close on the CAN1 (see Additional file 2: Figure S18), which disagrees with the above result that indicated that *MC1R* could separate black and white individuals. Similarly, SNPs located near the *ASIP* gene seemed to be able to distinguish the solid Red from the solid black and white groups (see Additional file 2: Figure S19). In the CDA, SNPs in the region surrounding the *ADAMTS20* gene could distinguish and separate all three groups on the CAN1 variable, and particularly the white group (see Additional file 2: Figures S20).

## Adaptation to environment

### Landscape genomics analysis

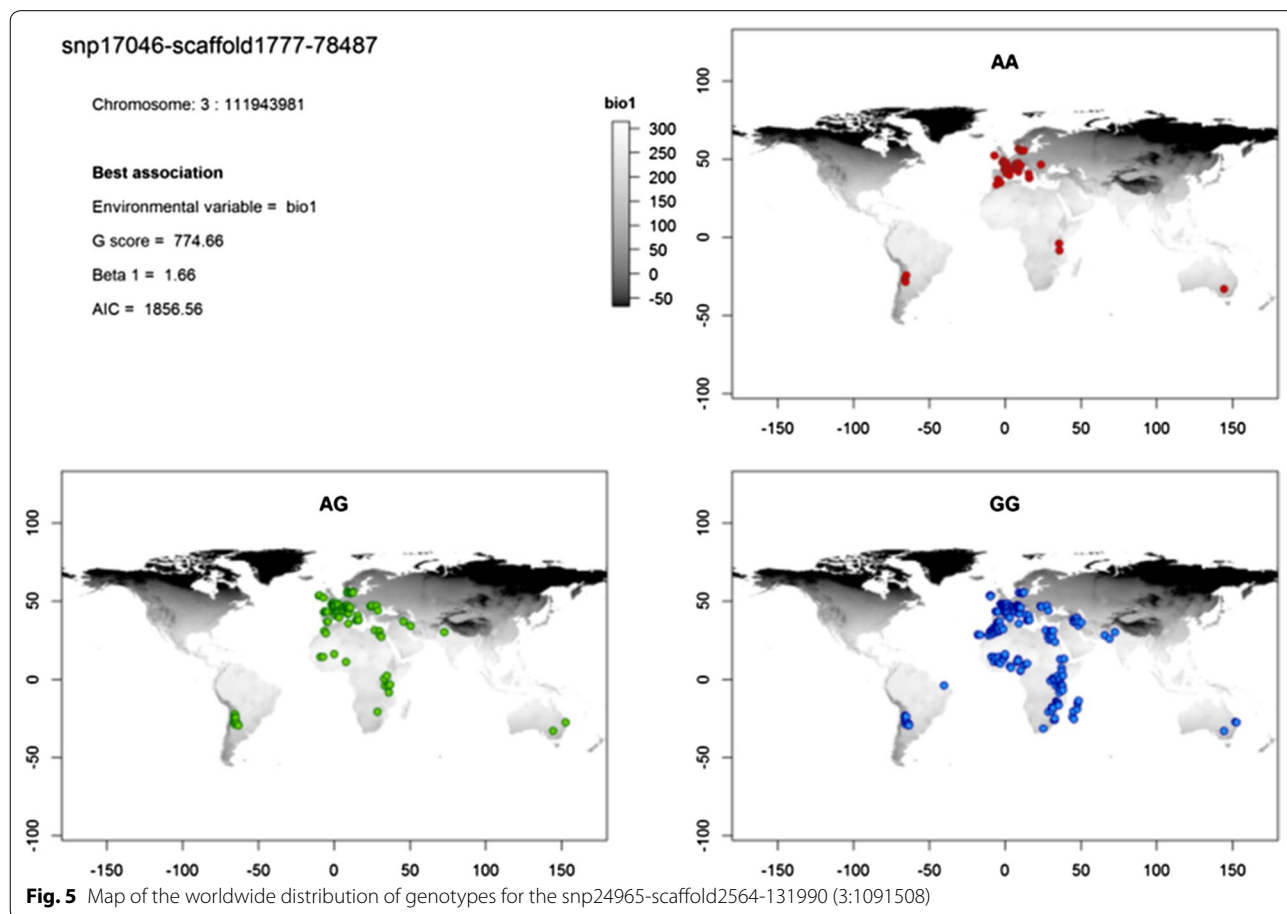
The Samβada results showed that more than 80% of the SNPs appear in at least one significant association between a genotype and a bioclimatic variable. In addition, 57 SNPs were involved in associations that respected the second series of filtering criteria (values for the  $\beta_1$  coefficient, G score, Wald score > 0.99 quantile, AIC criterion < 0.1 quantile, and at least two genotypes associated with at least one environmental variable) (see Additional file 3: Table S8). The most significant

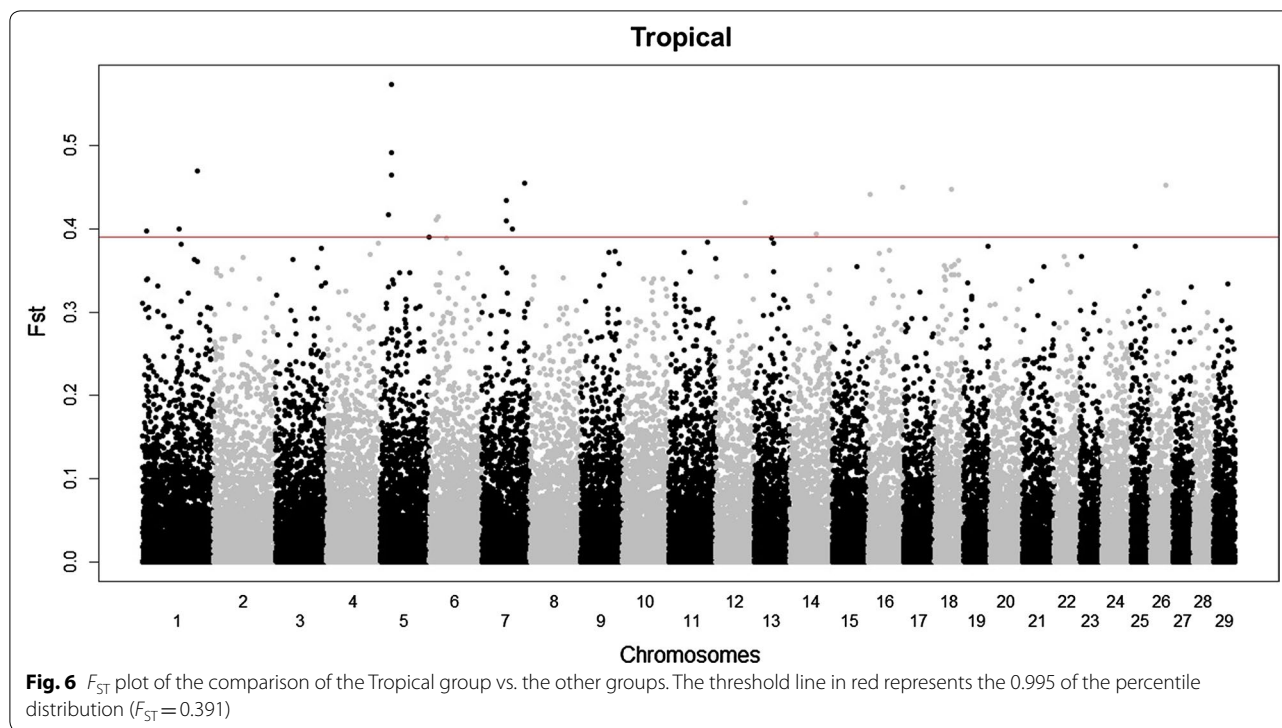
associations (highest G score) obtained with these SNPs involved the environmental variable related to annual mean temperature (bio1) for 49 SNPs, mean temperature of the coldest quarter (bio11) for three SNPs, mean diurnal range (bio2) for three SNPs, precipitation of the driest month (bio14) for one SNP and isothermality (bio3) for one SNP. The results obtained with the sampling point variables were very similar to those obtained with mode, maximum, mean and median values computed within the buffer area, whereas the associations with range, minimum and standard deviation were less significant. One of the strongest associations was observed between the CC genotype of SNP snp24965-scaffold2564-131990 located on chromosome 3:1091508 and the mean annual temperature (bio1). This association had the highest G and Wald scores of the models filtered using the criteria above and the highest Efron score of all significant models (using a threshold 0.05 before Bonferroni correction). The spatial distribution of the genotypes for this SNP showed that the AA genotype is only present in Europe or in southern regions of high altitude. The AC genotype showed a similar trend, even if it is slightly more frequent in the southern regions, while the CC genotype was observed in the

whole area of study (Fig. 5). This SNP is located close to the *pre-B cell leukemia homeobox 1 (PBX1)* gene. The list of genes that are located near ( $\pm 100$  kb) each of the 57 SNPs is in Additional file 3: Table S9 and the geographical distribution of the genotypes for the remaining 56 SNPs is in the Additional file 5. Analysis of the biological processes of the genes located in the vicinity of these SNPs (see Additional file 3: Table S10) highlighted genes that are linked to several pathways, such as the insulin and glucose signaling pathway and metabolism [(*IGF2 (insulin-like growth factor 2)*), glycogen metabolism, lipid biosynthetic processes, oxidative stress [*GPR37L1 (prostanoid receptor GPR37L1)* and *INS (insulin)* genes] and regulation of vasodilation. Two genes are involved in circadian rhythms regulation: *RAI1 (retinoic acid induced 1)* and *TH (tyrosine hydroxylase)*.

*F<sub>ST</sub>* and landscape genomic analyses of climatic associations

The MDS analysis (see Additional file 2: Figure S21) showed an overlap between the Köppen climate groups, with Tropical and Continental being the smallest and most clustered groups. The *F<sub>ST</sub>* plots are in Fig. 6 and Figures S22 to S24 (see Additional file 2: Figures S22, S23





and S24) and indicate for each analysis the number of SNPs above the selected threshold. Twenty SNPs for each comparison met the threshold of the quantile  $>0.9995$ , among which, 13 are shared between the Dry and Temperate/Continental groups with opposite major alleles between these groups (see Additional file 3: Table S11). Interestingly, the Tropical group did not have any common SNPs with the other groups, which reduced the list from 80 to 65 SNPs. When these 65 SNPs were compared with the Samβada results, they all showed either high G scores, Wald scores or  $\beta_1$  regression coefficients (quantile values considering only those obtained from the significant models  $>0.99$  of the empirical distribution) for at least one genotype in several environmental variables. These SNPs are summarized in Table S12 (see Additional file 3: Table S12). In the landscape genomic analysis, nine of these SNPs showed G scores, Wald scores and  $\beta_1$  regression coefficients that were all higher than the 0.999 quantile. In total, 197 genes were detected in the regions ( $\pm 100$  kb) close to these 65 SNPs (see Additional file 3: Table S13). For the Tropical group, a cluster of *HOXC* genes (*HOXC4*, *HOXC5*, *HOXC6*, *HOXC8*, *HOXC9*, *HOXC10*, *HOXC11*, *HOXC12* and *HOXC13*) that are located near the three SNPs on chromosome 5 was detected. Analysis of the genes for the four Köppen climate groups revealed several genes such as *SOX17* (*transcription factor SOX-17*) for the Dry and Temperate groups, *CLYBL* (*citrate lyase subunit beta-like protein*)

for the Dry and Continental groups, *LYPLA1* (*lysophospholipase 1*), *ATP6VIH* (*vacuolar ATPase*) and *RGS20* (*regulator of G-protein signaling 20*) for the Dry and Temperate groups and, with a different SNP, also in the Tropical group, *CAPN10* (*calpain 10*) and *RNPEPL1* (*arginyl aminopeptidase (aminopeptidase B)-like 1*) were found for the Continental group only but also detected in the landscape genomic analyses and in the FLK analyses for signature of early adaptation.

## Discussion

Detection of signatures of selection from genotype data alone is possible either by searching for genomic regions that show high levels of differentiation between populations or by looking for regions of low genetic diversity within a population. In this work, both approaches were used because they detect different kinds of selection events. Selection events that lead to the rapid fixation of an initially rare variant, i.e. hard sweeps, should lead to signals that can be detected by both approaches. However, in this case, within-population methods are more powerful than differentiation-based methods, but differentiation-based approaches can detect a greater diversity of selection signals, such as selection on standing variation (soft sweeps) or diversifying selection.

Our analyses were not limited to searching for signals of selection within or among populations. We also exploited other available information, such as differences

in production purposes, phenotypes, and geographical/bioclimate coordinates to elucidate the mechanisms of selection. The detection of adaptive loci in the genome is an essential part of studies on environmental adaptation, since it can help understanding which regions of a genome and therefore which genes have been or are being shaped by natural selection. Spatial analysis with geographical information systems (GIS) and environmental variables, along with molecular data, were included in the landscape genomic analysis to uncover the genomic regions under selection and identify candidate environmental factors that cause this selection. Multiple univariate logistic regressions were carried out to test for association between allele frequencies at marker loci and environmental variables. The analyses considered variables that can be summarized by differences in temperature and precipitation, which could be indirectly correlated to water availability. The Samβada analyses showed that more significant associations and more selected SNPs were associated with differences in temperature than with differences in precipitation. Therefore, temperature appears to have played a bigger role in the adaptation processes of goat breeds.

In addition, our analyses examined the signals that were detected in the genome after a detailed modeling of genetic divergence and population structure. Complex admixture and strong population bottlenecks can mimic the effect of selection by reducing local variability considerably. Therefore, to detect these signals, we applied the FLK and hapFLK approaches to individuals with a unique genetic background, without recent admixture or population bottlenecks. Finally, signatures of selection left by recent human selection on goat populations that are linked to production purposes or phenotypic standardization were analyzed.

Taken together, these analyses contributed to provide a genome-wide picture of the genes and genomic regions that have been subject to selection and led to signatures of selection that concern different biological pathways or group of genes, which were detected by one or more types of analyses.

#### Genes related to milk production

Among all the genes that can contribute to milk yield, quality and processing, the casein genes are one of the most important gene families in ruminants, since they have a major role in the production of cheese products from milk. In the goat genome, this gene cluster is located on chromosome 6 between 85.95 and 86.25 Mb. Analysis of the ROH scores of the breeds selected for milking purposes showed that the longest region with loss of variability coincides with the casein region on chromosome 6 between 85.9 and 86.2 Mb. The goat breeds used for milk

and cheese production belong mainly to the European breeds, particularly from Italy and France, which suggests that the signals identified are related to a common direction of selection within these breeds, i.e. that they are used mainly to produce whole milk (20% of production) and cheeses (80% of production) [42]. Many dairy goat breeds of Europe have been developed through breeding with the two most specialized dairy breeds, i.e. Saanen and Alpine from Switzerland [21]. Due to its popularity as a dairy breed, Saanen goats are distributed across more than 80 countries worldwide [5]. The genomic region that contains the cluster of casein genes was also detected by FLK analysis, which identified a region containing this gene cluster in the group of Alpine breeds and a region close to this cluster in the group of Eastern African populations. The different analyses and particularly the FLK and hapFLK analyses also detected other regions related to milk production. Chromosome 6 carries several genes that are related to milk production and found in several ruminant species. Another region on chromosome 6 that contains the *ABCG2* gene was detected in the Egyptian, southwestern and southeastern European groups, which was further confirmed by the iHS signal detected for the Murciano-Granadina and Argentata dell'Etna breeds. The *ABCG2* gene matched with a previously identified signature of selection in sheep [22] and polymorphisms in this gene have been associated with milk yield and composition in cattle and sheep [43, 44]. The ROH/Fst/XP\_EHH analyses detected another consensus region on chromosome 11 that contains the *EFEMP1* gene, which is associated with conjugated linoleic lipid contents in the meat of Wagyu Angus cattle breed [45]. A region on chromosome 1 between 110 and 130 Mb that was detected in the Alpine and southwestern European groups is homologous to a genomic region in dairy cattle for which selective pressure was previously reported [46]. A separate signature of selection approximately 5 Mb away from this region was also found in the Italian local breed Ciociara Grigia, which is used for milk production. Brito et al. [13] recently identified other genomic regions that may be associated with milk production, using a partially overlapping subset of breeds [13]. This difference may be due to a different breed composition of the dataset used and underlines the fact that there may be some breed-specific regions related to milk composition. Therefore, more tailored experiments are needed to decipher the genetics behind milk production in goat.

#### Genes related to fiber production

For the group of “fiber-producing” goat breeds, the strongest signature of selection in all three types of analyses (ROH, XP-EHH and  $F_{ST}$ ) was detected in the genomic regions containing the *CUX1* and *PLOD3* genes. The



detected regions in Angora and Ankara were confirmed by the FLK and hapFLK analyses. These genes are associated with hair development: *CUX1* is associated with wavy hairs and curly whiskers in mouse [47] and *PLOD3* may play a role in the formation of hairs or in their texture [48]. The roles of these genes are consistent with the presence of a signature of selection only in the two Angora breeds (Angora and Ankara), which have curly hairs, and not in the Cashmere breed, which has long and wavy hairs. For the Cashmere breed, Brito et al. [13] proposed several alternative genomic regions. The  $F_{ST}$  analyses comparing the Cashmere vs. the Angora and Ankara breeds that revealed a signature of selection on chromosome 25 did not show any overlapping regions with those reported in [13], probably because the approach used in [13] highlighted other genomic regions. The ROH analyses carried out on the data for the Cashmere breed revealed a region on chromosome 22 that overlapped with one of the regions reported by Brito et al. [13] and was identified by a  $F_{ST}$  analysis of Cashmere against several other breeds. Thus, this may be an interesting region to investigate further for fiber production in the Cashmere breed. In addition, it suggests that different genomic regions are involved with fiber production in Cashmere and Angora/Ankara breeds.

#### Genes related to meat production

In the analyses comparing breeds for meat production, some of the detected regions contain genes associated with muscle formation. On chromosome 3, we identified a region that includes the *AMPDI* and *NRAS* genes, in agreement with previous findings in cattle and mice. *AMPDI* is involved in the deamination of AMP in skeletal muscles [49] and its disruption was reported to influence the expression of neighboring genes, such as *NRAS*, in mice [50]. In addition, allelic variants of the *AMPDI* gene are associated with traits such as heart girth and body weight in Chinese beef cattle [51]. Another gene within the same region, *TSHB* may also be associated with muscular functionality. *TSHB* encodes a subunit of the thyroid stimulating hormone (TSH), which plays a pivotal role in regulating thyroid activity, stimulating colloid reabsorption and in the release of the  $T_4$  hormone [52]. The role of thyroid function in growth is known, and different levels of TSH are associated with hyperthyroidism or hypothyroidism [53]. In fact, high levels of TSH are commonly found in obese human children and adolescents [54]. In goat, this gene has been characterized in relation with reproductive seasonality [55].

The main breed of the “meat-producing” group is the Boer breed, which is selected for meat production and for which several breeding programs have been developed. In spite of this, no evidence of strong selection

in the Boer breed for genes related to meat production was detected in our analyses. A possibility is that variation in meat production traits in this population may have a highly polygenic basis (many alleles of small effects determine the trait). Because of the size of the samples and the availability of phenotype records, we used approaches that are not suited to detect such selection events. Other targeted studies on the Boer breed are needed to evaluate the possible genomic impact of breeding programs in this population.

#### Genes related to coat color

ROH analysis of the data for the three groups of solid coat colors detected signatures of selection near at least five genes that are known to be involved in the color and pattern definition of the coat: *ADAMTS20*, *MC1R*, *ASIP*, *SOX18* and *TIMP3*. Two of these genes (*MC1R* and *SOX18*) were specifically detected in the comparison between solid Black and solid White individuals, whereas the remaining three were identified in both solid Black vs. solid White and solid Black and Red vs. solid White groups.

The well-known *MC1R* gene is involved in the genetic determinism of color [56] and located on chromosome 18 in goats. It plays a major role in controlling the switch from eumelanin (black and brown) to pheomelanin (yellow to red; [57]) produced by melanocytes. This gene has already been studied in several species [58], such as cat [59], horse [60], dog [61], pig [62], cattle [63], sheep [64] and even in goat [65–68]. Although described in other species, no direct association between *MC1R* and red coat color is known in goat.

*ASIP* is a competitive antagonist of *MSH* (*melanocyte stimulating hormone*) for binding *MC1R*, enabling the switch from eumelanin to pheomelanin [69]. Similarly to what we observed for SNPs within the region surrounding *MC1R*, SNPs within the region around *ASIP* seemed to be able to distinguish the solid Red from the solid Black and White groups. The presence of an association with the Red/Black group in the region surrounding this gene seems to confirm its role also in goats, as hypothesized in several studies [11, 70]. Moreover, the signal on this gene was also found with the differentiation tests (FLK, hapFLK) in Alpine, Poitevine and Valdostana goats with black-and-tan pigmentation pattern. This further confirms the potential role of this gene in modulating the melanocyte activity in goats, in particular its involvement in the switch from black to red pigments, whereas the white color may to be caused by other gene(s).

The third candidate gene identified is *SOX18*, which encodes a transcription factor that plays an important role in hair follicles and blood vessel development during



embryogenesis in disheveled mice, a semi-dominant mutation characterized by coat sparseness [71, 72]. In addition, alleles of this gene seem to be responsible for a dark phenotype in mice [57, 73].

*ADAMTS20* located on chromosome 5 encodes a highly conserved metalloprotease. This gene is required for melanoblast survival and mediating Kit signaling in skin colonization such as the belted white locus in inbred mouse colonies [74–76]. It is also important in multiple biological functions, such as delayed palate closure [77] and soft tissue syndactyly [78]. Another gene detected with these approaches, *TIMP3*, is known to inhibit the activity of metalloproteinases, and shows high specificity and selectivity for the ADAM and ADAMTS families [79] that are involved in melanoma cancerous cell development [80]. A signature of selection that included this gene was also found in the Southwestern European goat group and in particular in the Pyrenean goat population, which is characterized by a piebald black and white pattern, in the Argentata dell'Etna breed, which is characterized by a silver coat colour, and in the group of breeds from Northwestern Africa where Sahel, Peulh and Thari breeds are characterized by white or white spotted coat colours. However, for the last three breeds, no reliable pictures were available to evaluate the coat color pattern.

The signatures of selection near the *KIT* gene observed in the Kacchan and Abergelle breeds could be related to their spotted phenotypes (in this case, this phenotype is localized mainly on the anterior parts of the animals) as already observed in other species. It should be noted that the Argentata dell'Etna breed has a silver phenotype and that, in the fox, an analogous phenotype is caused by a mutation on an autosomal copy of the *KIT* gene [81].

#### Genes related to other traits

In this work, a strong association was found between the results of the landscape genomics analyses and those obtained using independent analyses based on other types of more categorical classifications, such as the Köppen classification. Several signals were associated with environmental parameters and these loci behave atypically in comparison with the theoretical distribution for neutral loci. Among the 13 common SNPs that differentiated the Dry and Temperate/Continental groups, i.e. between hot-dry and temperate-cold areas, all showed a different major allele, and this may be the sign of natural selection that is driving alleles in opposite directions for the adaptation of the breeds to different environments. These SNPs were also confirmed by landscape genomics analyses, i.e. although they are not included in the top 57 SNPs, they displayed a high significance for one or more bioclimatic variables. Several of the genes around these 13 SNPs are associated with production traits in

cattle. *LYPLA1* encodes a hormone that acts as a ghrelin inhibitor and is, therefore, involved in the regulation of the appetite, as shown in rats [82]. This gene is also associated with feed intake and weight gain in cattle [83]. Other genes, such as *RGS20* and *SOX17* may play a role in pubertal development, as shown in Brahman cattle [84]. Regarding *CLYBL*, it was shown to be differentially expressed in cows with different milk citrate contents [85]. A region detected by FLK and landscape genomics and which is unique in the Continental group through  $F_{ST}$  analysis contains the *CAPN10* and *RNPEPL1* genes. Calpains are calcium-regulated proteases involved in cellular functions that include muscle proteolysis both ante- and postmortem. In livestock, they play important roles in muscle growth and development, myoblast fusion, and differentiation [86]. The early post-mortem cleavage of these proteins leads to the tenderization of meat and, thus, *calpain* genes have been associated with meat tenderness [87, 88]. The *RNPEPL1* gene is differentially expressed in pigs with different muscularity traits [89]. FLK analyses detected the *RXFP2* gene in the Southwestern European group and the central Asian group, particularly in the Thari and Blanca de Rasquera breeds. This well-known gene is responsible for the polled condition in most domestic sheep populations and associated with horn phenotypes in the Soay breed, a feral archaic sheep from Scottish islands, but it has not been shown to be involved in the development of horns in goats [22]. This signature of selection suggests that the genetic determinism of horns in these two goat populations may be specific compared to other goat breeds.

The *LCORL* gene was present in a signature of selection detected in the Egyptian group, particularly in the Nubian breed. It has been associated with signatures of selection linked to animal size in cattle [90, 91] and in European commercial and local pig breeds [92].

#### Insulin, glucose pathway and oxidative stress response genes

Temperature variation detected by the Samβada analyses can operate as a primary environmental stressor. A major cellular effect caused by multiple environmental stressors is the generation of reactive oxygen species (ROS) that leads to oxidative stress. This was reported in livestock species such as cattle and chickens. In goats, it could be supported by genes that are involved in pathways linked to temperature changes, such as insulin or glucose pathways and oxidative stress response. The *GPR37L1* gene is primarily expressed in neuronal cells and has a role in protecting primary astrocytes against oxidative stress [93]. Another gene related to oxidative stress is the *INS* gene, its level in pancreatic cells changing in case of oxidative stress [94], which stimulates the *HSP* (*heat shock*

*protein*) gene in cardiac tissue [95]. The *INS* gene is also the target of other important pathways related to insulin production, response of the organism to insulin and glucose metabolism. These biological processes are the most predominant in the landscape analyses. Heat stress can also affect the response of an organism to insulin stimulus, which in turn affects lipid and carbohydrate pathways [96, 97]. Several of the genes involved in pathways related to temperature changes were close to SNPs identified in our analyses and are associated with production traits in several species. For example, *IGF2* that is present in pathways related to insulin, glucose, glycogen, lipid, and carbohydrate processes is one of the most important genes for meat production and heat stress response in several species [98, 99].

### **HOXC genes family and related genes**

Three SNPs located on chromosome 5 were detected by  $F_{ST}$  analyses in the Tropical group. SNPs with the highest  $F_{ST}$  value were included in the *HOXC* cluster, and in the landscape genomics analysis they were associated with the bioclimatic variables defined as “isothermality” and “temperature seasonality”. A genome scan between geographical groups to detect signals of early adaptation identified the same region. The *HOX* gene family (*HOXA*, *HOXB*, *HOXC* and *HOXD* clusters) controls the body plan of the embryo along the craniocaudal (head–tail) axis. The *HOX* gene family was recently associated with mating/sexual behavior in *Drosophila*, *C. elegans* and mice [100, 101]. Selective pressure on some *HOX* genes was also detected in bats and aquatic mammals [102, 103]. Recently, a signature of selection near the *HOXA* and *HOXC* gene clusters was also reported in sheep [22]. One of the strongest associations in the Samβada analysis was detected close to the *PBX1* gene, which is involved in the morphological development of several parts of the body and tissues. *PBX1* interacts with *HOX* gene clusters, which includes the *HOXC* gene [104, 105]. It is interesting to note that most of the breeds assigned to the Tropical group were characterized by a smaller size (e.g. Cameroon goat, Naine, Small east Africa, and African Dwarf). It has been reported that the small body size of Zebu cattle in tropical areas can contribute to heat tolerance, since small animals have a higher surface to body mass ratio, together with the fact that small size correlates also with low nutritional requirements [106]. These observations may apply also to goats and such genes can determine an adaptation towards a smaller size and therefore better heat management in areas where hot temperatures are constant.

### **Genes linked to circadian clock**

The genome scan for signatures of selection between geographical groups and the landscape genomics analysis

detected several genes that are linked to circadian clock rhythm suggesting a possible association with adaptation at different latitudes. These included *SOX14*, *NOCT*, *RAI1* and *TH*. *SOX14* is required to drive the development of a functional network supporting light-circadian behaviors [107]. *NOCT* encodes a circadian deadenylase expressed at high levels during the night in several mammalian tissues and was recently implicated in circadian control of metabolism [108]. *RAI1* has an important function in the maintenance of circadian rhythmicity. In humans, it is crucial for normal regulation of circadian rhythms, lipid metabolism and neurotransmitter function [109]. This gene disrupts the transcription of the *CLOCK* gene, a key component of the mammalian circadian oscillator that regulates many other critical genes involved in metabolism, physiology, and behavior. A chromosomal deletion of *RAI1* in mice is associated with a short circadian period, whereas in humans it is associated with the Smith-Magenis syndrome, a pathology that is characterized by an inverted melatonin rhythm, sleep disturbances, abnormal feeding, and cognitive disturbance [110, 111]. Finally, *TH* is involved in the production of tyrosine hydroxylase, which is important for the nervous system to function normally. Tyrosine hydroxylase is the rate-limiting enzyme in the synthesis of catecholamine, and hence plays a key role in the physiology of adrenergic neurons and receptors. It is involved in the production of neurotransmitters that control physical movements and involuntary body processes, such as the regulation of blood pressure and body temperature. Moreover, in the response to photoperiod, it is well known that tyrosine hydroxylase neurons have estrogen receptors that are involved in the physiological cyclic activity induced by photoperiod [112].

Another gene that could be linked to the reproductive seasonality is *TSHB*, which has been reported to be associated with meat production and reproductive seasonality in goats [55].

### **Conclusions**

The goat is one of the most adaptable livestock species distributed worldwide across a large variety of climatic and geographical areas and used by humans for different purposes. Natural and artificial pressures have led to different genomic signatures of selection across the genomes of many goat breeds. By applying different kinds of analyses, we were able to detect (1) different allelic distributions worldwide that are associated with bioclimatic variables and groups, (2) signatures of selection that differentiate breeds raised for different economical purposes, and (3) population structure clustering that represent sub-continental groups. We were also able to identify several genomic regions that contain genes related to processes such as milk-, meat- or fiber-related

production, coat color, glucose pathway, oxidative stress response, size and the internal circadian clock. These results provide a first comprehensive picture into the global domestication process and adaptation of the goat and highlights several genes that have contributed to the differentiation of this species worldwide.

## Additional files

**Additional file 1.** Treemix plots. Each page shows the output of a continental/sub-continental group

**Additional file 2: Figure S1.** MDS plot of the breeds, grouped by production purpose: milk, meat and fiber groups. Group colors: milk = green, meat = red, fiber = blue; MDS and box plots of the first two components pre-filtering (upper) and MDS plots after filtering. **Figure S2.** Manhattan plot of the FLK results for the sub-continental group after filtering steps. Sub-geographical group names are given on top of each plot. Chromosomes are alternately red and black. **Figure S3.** Manhattan plot of the hapFLK results for the sub-continental group after filtering steps. Sub-geographical group names are given on top of each plot. Chromosomes are alternately red and black. **Figure S4.** Genomic distribution of FLK and hapFLK signals across population groups. For each chromosome and each sub-continental group, the chromosomal position detected with at least one of the two approaches is indicated. **Figure S5.** Signatures of selection on chromosome 5 for the North western Africa, South eastern Africa and South western Africa groups. North western Africa (NWA): red; South eastern Europe (SEE): green; South western Europe (SWE): blue. The table (bottom-right) reports the genes within the region in which a signature was detected. **Figure S6.** Signatures of selection on chromosome 6 for the Central Asia, East Africa and South east Europe groups. Central Asia (CA): red; East Africa (EA): green; South east Europe (SEE): blue. The table (upper-right) reports the genes within the region in which a signature was detected. **Figure S7.** Signatures of selection on chromosome 13 for the Alpines and Central Asia groups. Alpines (Alps): red; Central Asia (CA): blue. The table (right) reports the genes within the region in which a signature was detected. **Figure S8.** Signatures of selection on chromosome 1 for the Alpines and South Western Europe groups. Alpines (Alps): red; South Western Europe (SWE): blue. The table (bottom-right) reports the genes within the region in which a signature was detected. **Figure S9.** FLK signals on chromosome 6 around the *casein* gene cluster. The cluster of unannotated genes between *YTHDC1* and *SULT1B1* consists of genes coding for glucuronosyltransferase enzymes. Alps: red; East Africa (EA): blue. The table (right) reports the genes within the casein cluster region. **Figure S10.** Signatures of selection on chromosome 6 for the Egypt, South Eastern Europe and South Western Europe groups. Egypt (Egypt): red; South Eastern Europe (SEE): green; South Western Europe (SWE): blue. The table (bottom-right) reports the genes within the region in which a signature was detected. **Figure S11.** Signatures of selection on chromosome 12 for the Central Asia and South Western Europe groups. Central Asia (CA): red; South Western Europe (SWE): blue. The table (right) reports the genes within the region in which a signature was detected. **Figure S12.** ROH,  $F_{ST}$  and XP-EHH results for the group of "meat-producing" goat breeds. Types of analysis are indicated with different plot colors, within the most external squared-based circle, where each color represent a chromosome (chromosome number outside the squares): green (external) = ROH; blue (middle) =  $F_{ST}$ ; violet (internal): XP-EHH. For the three analyses, the regions above the threshold are marked in red. **Figure S13.** ROH,  $F_{ST}$  and XP-EHH results for the group of "milk-producing" goat breeds. Types of analysis are indicated with different plot colors, within the most external squared-based circle, where each color represent a chromosome (chromosome number outside the squares): green (external) = ROH; blue (middle) =  $F_{ST}$ ; violet (internal): XP-EHH. For the three analyses, the regions above the threshold are marked in red. **Figure S14.** CDA for the region on chromosome 25 detected for the group of "fiber-producing" goat breeds. (a): LONG and (b) SHORT: (b). Left: CDA plot. Right: Correlation value of the

SNPs used for the analyses for CAN1 and CAN2. **Figure S15.** CDA for the region on chromosome 25 detected for the group of "meat-producing" goat breeds. (a): LONG and (b) SHORT: (b). Left: CDA plot. Right: Correlation value of the SNPs used for the analyses for CAN1 and CAN2. **Figure S16.** CDA for the region on chromosome 25 detected for the group of "milk-producing" goat breeds. (a): LONG and (b) SHORT: (b). Left: CDA plot. Right: Correlation value of the SNPs used for the analyses for CAN1 and CAN2. **Figure S17.** MDS plot of breeds considered for the panel of coat colors. Breed codes and subdivisions based on the coat color pattern are indicated in the right part of the plot. **Figure S18.** CDA for the region on chromosome 18 near the *MC1R* gene detected for the group of coat color breeds. (a): LONG and (b) SHORT: (b). Left: CDA plot. Right: Correlation value of the SNPs used for the analyses for CAN1 and CAN2. **Figure S19.** CDA for the region on chromosome 13 near the *ASIP* gene detected for the group of coat color breeds. (a): LONG and (b) SHORT: (b). Left: CDA plot. Right: Correlation value of the SNPs used for the analyses for CAN1 and CAN2. **Figure S20.** CDA for the region on chromosome 5 near the *ADAMTS20* gene detected for the group of coat color breeds. (a): LONG and (b) SHORT: (b). Left: CDA plot. Right: Correlation value of the SNPs used for the analyses for CAN1 and CAN2. **Figure S21.** MDS plot of the filtered dataset considering components 1 and 2. Animals are color-coded based on the Köppen classification of groups: Tropical (green), Dry (red), Temperate (orange), Continental (blue). **Figure S22.**  $F_{ST}$  plot of the comparison of the Dry group vs. the other groups. The threshold line in red represents the 0.995 of the percentile distribution ( $F_{ST} = 0.398$ ). **Figure S23.**  $F_{ST}$  plot of the comparison of the Temperate group vs. the other groups. The threshold line in red represents the 0.995 of the percentile distribution ( $F_{ST} = 0.320$ ). **Figure S24.**  $F_{ST}$  plot of the comparison of the Continental group vs. the other groups. The threshold line in red represents the 0.995 of the percentile distribution ( $F_{ST} = 0.507$ ).

**Additional file 3: Table S1.** Environmental variables considered for the landscape genomic analysis. **Table S2.** FLK and hapFLK windows for the sub-continental groups after filtering steps. Overlapping or partially overlapping regions across the different geographical subdivisions are reported with the same letter as used for the population group (from <sup>a</sup> to <sup>m</sup>). **Table S3.** Selective sweeps of early adaptation. The bold values indicate overlap or partial overlap between comparisons. **Table S4** Common regions between ROH and XP-EHH and/or  $F_{ST}$  analyses for the group of "fiber-producing" goat breeds and genes within these regions. **Table S5.** Common regions between ROH and XP-EHH and/or  $F_{ST}$  analyses for the group of "meat-producing" goat breeds and genes within these regions. **Table S6.** Common regions between ROH and XP-EHH and/or  $F_{ST}$  analyses for the group of "milk-producing" goat breeds and genes within these regions. **Table S7.** Top four regions detected for the coat color groups (Black, White and Red), indicated comparisons between the groups and genes located nearby these regions. **Table S8.** Results of the most significant associations involving the 57 filtered SNPs. For each SNP: probe, genotype, genomic coordinate, associated environmental variable ("Env") and scores and statistical test output values ("Gscore", "WaldScore", "Efron", "AIC", "Beta\_0" and "Beta\_1") are reported. **Table S9.** List of genes located nearby ( $\pm 100$  kb) the 57 filtered SNPs obtained by the landscape genomics analysis. **Table S10.** GO term biological processes. **Table S11.** MAF (minor allele frequency) and major allele of the 13 SNPs shared by Dry and Temperate/Continental groups. **Table S12.** SamBada significant results for the SNPs of the  $F_{ST}$  analyses. For each SNP probe: chromosome and position are reported, as well as Environment, G score, Wald Score, AIC, Abs\_Beta1 and the Köppen group in which the SNP was detected by the  $F_{ST}$  analyses. **Table S13.** List of genes located nearby ( $\pm 100$  kb) of the 65 SNPs selected with the  $F_{ST}$  approach and confirmed by SamBada. The \* symbol indicates the SNPs with the highest value ( $> 0.999$ ) also in landscape genomic analysis

**Additional file 4.** Graphical representation of ROH and iHS results at the breed level of the signals FLK and hapFLK in the sub continental groups. Left panel: FLK (points) and hapFLK (line) signatures, with as header the chromosomal region investigated. Middle panels: ROH signatures. Right panels: iHS signatures [see Additional file 2]

**Additional file 5.** Genotype distribution of the SNPs detected in the landscape genomics analysis based on the GPS coordinates. Each page

represent a SNP, with its coordinates reported on the upper right part of each sheet, with environmental variable associated and statistics (G score, Beta 1 and AIC). The three possible genotypes are labelled with three different colors: AA = red; AB = green; BB = blue

#### Authors' contributions

PC, FB, BS, and SJ conceived the study, LC, MM, AT, FB, BS, ER, PC, and EN curated the data, BS, FB, AT, ER, AC, GC, RC, GM, ESK, TS, and SJ contributed to data analysis, BS, AT, FB, ER, AC, GC, and RS contributed to the graphical representations, AdaptMap consortium contributed the resources, FB, PC, BS, AT, and ER helped to write the paper, MFR, MA, EN, BDR, CVT, BG, GTK, AS, and MM helped to edit the versions of manuscript, PC, AS, MA, EN, BDR, CVT, BG, GTK, and AS were in charge of the project administration, PC and FB supervised the work, AdaptMap and MFR were in charge of the funding acquisition. All authors read and approved the final manuscript.

#### Author details

<sup>1</sup> Department of Animal Science, Iowa State University, Ames, IA 50011, USA. <sup>2</sup> National Institute of Aquatic Resources, Technical University of Denmark (DTU), 2800 Lyngby, Denmark. <sup>3</sup> GenPhySE, INRA, Université de Toulouse, INPT, ENVT, 31326 Castanet Tolosan, France. <sup>4</sup> Dipartimento di Medicina Veterinaria, Università degli Studi di Milano, 20133 Milan, Italy. <sup>5</sup> Laboratory of Geographic Information Systems (LASIG), School of Architecture, Civil and Environmental Engineering (ENAC), Ecole Polytechnique Fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland. <sup>6</sup> Recombinetics Inc, St Paul 55104, MN, USA. <sup>7</sup> Consiglio per la ricerca in agricoltura e l'analisi dell'economia agraria (CREA) - Research Centre for Animal Production and Acquaculture, 00015 Monterotondo, Roma, Italy. <sup>8</sup> Centre for Research in Agricultural Genomics (CRAG), CSIC-IRTA-UAB-UB, Campus Universitat Autònoma de Barcelona, Bellaterra, 08193 Barcelona, Spain. <sup>9</sup> DIANA Dipartimento di Scienze Animali, della Nutrizione e degli Alimenti, Università Cattolica del S. Cuore, 29100 Piacenza, Italy. <sup>10</sup> BioDNA Centro di Ricerca sulla Biodiversità e sul DNA Antico, Università Cattolica del S. Cuore, 29100 Piacenza, Italy. <sup>11</sup> Fondazione Parco Tecnologico Padano (PTP), 26900 Lodi, Italy. <sup>12</sup> Department of Support, Production and Animal Health, School of Veterinary Medicine, São Paulo State University (UNESP), Araçatuba, Brazil. <sup>13</sup> Animal Genomics and Improvement Laboratory, ARS USDA, Beltsville, MD 20705, USA. <sup>14</sup> Center for Quantitative Genetics and Genomics, Aarhus University, 8830 Tjele, Denmark.

#### Acknowledgements

Authors are grateful to all breeders and AdaptMap members who provided data to the initiative (<http://www.goatadaptmap.org/>) that are described in Stella et al. [13]. The authors from the University of Milan are grateful to Stefano Frattini who was supported by the GenHome project, for supporting coordinating meetings. Support for Francesca Bertolini was provided by State of Iowa, Hatch and Ensminger funds. Financial support provided in part by the Illumina Greater Good grant is appreciated.

#### Competing interests

The authors declare that they have no competing interests.

#### Availability of data and materials

Genotyping data collected by AdaptMap (<http://www.goatadaptmap.org/>) will be shared on Dryad (<https://doi.org/10.5061/dryad.v8g21pt>).

#### Consent for publication

Not applicable.

#### Ethical statement

All authors declare that animal samples were obtained in compliance with local/national laws in force at the time of sampling. Data exchange was in accordance with national and international regulations, and approved by the owners.

#### Funding

Funding for FB was provided by the Ensminger Endowment, Hatch and State of Iowa. Funding. AT was supported by the PhD fellowship of the University of Milan.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 10 November 2017 Accepted: 15 October 2018

Published online: 19 November 2018

#### References

- Naderi S, Rezaei HR, Pompanon F, Blum MGB, Negrini R, Naghash HR, et al. The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc Natl Acad Sci USA*. 2008;105:17659–64.
- <http://www.fao.org/faostat/en/> Accessed 10 February 2017.
- Skapetas B, Bampidis V. Goat production in the World: present situation and trends. *Livest Res Rural Dev*. 2016;28:200.
- Nicoloso L, Bomba L, Colli L, Negrini R, Milanese M, Mazza R, et al. Genetic diversity of Italian goat breeds assessed with a medium-density SNP chip. *Genet Sel Evol*. 2015;47:62.
- FAO. The second report on the state of the world's animal genetic resources for food and agriculture. BD Scherf, D. Pilling editors. FAO Commission on Genetic Resources for Food and Agriculture Assessments. Rome. 2015.
- Qanbari S, Simianer H. Mapping signatures of positive selection in the genome of livestock. *Livest Sci*. 2014;166:133–43.
- Lenstra JA, Groeneveld LF, Eding H, Kantanen J, Williams JL, Taberlet P, et al. Molecular tools and analytical approaches for the characterization of farm animal genetic diversity. *Anim Genet*. 2012;43:483–502.
- Tosser-Klopp G, Bardou P, Bouchez O, Cabau C, Crooijmans R, Dong Y, et al. Design and characterization of a 52 K SNP chip for goats. *PLoS One*. 2014;9:e86227.
- Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). *Nat Biotechnol*. 2013;31:135–41.
- Bickhart DM, Rosen BD, Koren S, Sayre BL, Hastie AR, Chan S, et al. Single-molecule sequencing and chromatin conformation capture enable de novo reference assembly of the domestic goat genome. *Nat Genet*. 2017;49:643–50.
- Burren A, Neuditschko M, Signer-Hasler H, Frischknecht M, Reber I, Menzi F, et al. Genetic diversity analyses reveal first insights into breed-specific selection signatures within Swiss goat breeds. *Anim Genet*. 2016;47:727–39.
- Talenti A, Bertolini F, Pagnacco G, Pilla F, Ajmone-Marsan P, Rothschild MF, et al. The Valdostana goat: a genome-wide investigation of the distinctiveness of its selective sweep regions. *Mamm Genome*. 2017;28:114–28.
- Brito LF, Kijas JW, Ventura RV, Sargolzaei M, Porto-Neto LR, Cánovas A, et al. Genetic diversity and signatures of selection in various goat breeds revealed by genome-wide SNP markers. *BMC Genomics*. 2017;18:229.
- Kim ES, Elbeltagy AR, Aboul-Naga AM, Rischkowsky B, Sayre B, Mwacharo JM, et al. Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* (Edinb). 2016;116:255–64.
- Benjelloun B, Alberto FJ, Streeter I, Boyer F, Coissac E, Stucki S, et al. Characterizing neutral genomic diversity and selection signatures in indigenous populations of Moroccan goats (*Capra hircus*) using WGS data. *Front Genet*. 2015;6:107.
- Lai FN, Zhai HL, Cheng M, Ma JY, Cheng SF, Ge W, et al. Whole-genome scanning for the litter size trait associated genes and SNPs under selection in dairy goat (*Capra hircus*). *Sci Rep*. 2016;6:38096.
- Guan D, Luo N, Tan X, Zhao Z, Huang Y, Na R, et al. Scanning of selection signature provides a glimpse into important economic traits in goats (*Capra hircus*). *Sci Rep*. 2016;6:36372.
- Stucki S, Orozco-terWengel P, Colli L, Kabi F, Masembe C, Negrini R, et al. Sambada in Uganda: landscape genomics study of traditional cattle breeds with a large SNP dataset. In Proceedings of the IALE 2013 European Congress: 9-12 September 2013; Manchester. 2013.



19. Pariset L, Joost S, Marsan PA, Valentini A, Econogene Consortium (EC). Landscape genomics and biased FST approaches reveal single nucleotide polymorphisms under selection in goat breeds of North-East Mediterranean. *BMC Genet*. 2009;10:7.
20. Colli L, Joost S, Negrini R, Nicoloso L, Crepaldi P, Ajmone-Marsan P, et al. Assessing the spatial Dependence of adaptive loci in 43 European and Western Asian goat breeds using AFLP markers. *PLoS One*. 2014;9:e86668.
21. Colli L, Milanese M, Talenti A, Bertolini F, Chen M, Crisà A, et al. Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet Sel Evol*. <https://doi.org/10.1186/s12711-018-0422-x>.
22. Fariello MI, Servin B, Tosser-Klopp G, Rupp R, Moreno C, International Sheep Genomics Consortium, et al. Selection signatures in worldwide sheep populations. *PLoS One*. 2014;9:e103813.
23. Liu Z, Ji Z, Wang G, Chao T, Hou L, Wang J. Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *BMC Genomics*. 2016;17:863.
24. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26:841–2.
25. Bonhomme M, Chevalet C, Servin B, Boitard S, Abdallah J, Blott S, et al. Detecting selection in population trees: the Lewontin and Krakauer test extended. *Genetics*. 2010;186:241–62.
26. Fariello MI, Boitard S, Naya H, SanCristobal M, Servin B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics*. 2013;193:929–41.
27. Pickrell JK, Pritchard JK. Inference of population splits and mixtures from genome-wide allele frequency data. *PLoS Genet*. 2012;8:e1002967.
28. Scheet P, Stephens M. A fast and flexible statistical model for large-scale population genotype data: applications to inferring missing genotypes and haplotypic phase. *Am J Hum Genet*. 2006;78:629–44.
29. Storey JD, Tibshirani R. Statistical significance for genomewide studies. *Proc Natl Acad Sci USA*. 2003;100:9440–5.
30. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol*. 2006;4:e72.
31. Marras G, Rossoni A, Schwarzenbacher H, Biffani S, Biscarini F, Nicolazzi EL, zanardi: an open-source pipeline for multiple-species genomic analysis of SNP array data. *Anim Genet*. 2017;48:121.
32. Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10:5–6.
33. Gautier M, Klassmann A, Vitalis R. rehh 2.0: a reimplementation of the R package rehh to detect positive selection from haplotype structure. *Mol Ecol Resour*. 2017;17:78–90.
34. Venables WN, Ripley BD. MASS: modern applied statistics with S. New York: Springer; 2002.
35. Szpiech ZA, Hernandez RD. Selscan: an efficient multithreaded program to perform EHH-based scans for positive selection. *Mol Biol Evol*. 2014;31:2824–7.
36. Stucki S, Orozco-terWengel P, Forester BR, Duruz S, Colli L, Masembe C, et al. High performance computation of landscape genomic models including local indicators of spatial association. *Mol Ecol Resour*. 2017;17:1072–89.
37. Bartholomé E, Belward AS. GLC2000: a new approach to global land cover mapping from Earth observation data. *Int J Remote Sens*. 2005;26:1959–77.
38. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. 2013;14:128.
39. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res*. 2016;44:W90–7.
40. Peel MC, Finlayson BL, McMahon TA. Updated world map of the Köppen-Geiger climate classification. *Hydrol Earth Syst Sci*. 2007;11:1633–44.
41. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience*. 2015;4:7.
42. Escareño L, Salinas-Gonzalez H, Wurzinger M, Iñiguez L, Sölkner J, Meza-Herrera C. Dairy goat production systems: status quo, perspectives and challenges. *Trop Anim Health Prod*. 2012;45:17–34.
43. Cohen-Zinder M, Seroussi E, Larkin DM, Loor JJ, Everts-van der Wind A, Lee JH, et al. Identification of a missense mutation in the bovine ABCG2 gene with a major effect on the QTL on chromosome 6 affecting milk yield and composition in Holstein cattle. *Genome Res*. 2005;15:936–44.
44. Áryasi M, Komlósi I, Kent MP, Czeglédi L, Gulyás G, Jávora A. Investigation of polymorphisms and association of the ABCG2 gene with milk production traits in sheep. *Livest Sci*. 2013;154:64–8.
45. Zhang L, Michal JJ, O'Fallon JV, Pan Z, Gaskins CT, Reeves JJ, et al. Quantitative genomics of 30 complex phenotypes in Wagyu x Angus F1 progeny. *Int J Biol Sci*. 2012;8:838–58.
46. Stella A, Ajmone-Marsan P, Lazzari B, Boettcher P. Identification of selection signatures in cattle breeds selected for dairy production. *Genetics*. 2010;185:1451–61.
47. Sansregret L, Nepveu A. The multiple roles of CUX1: insights from mouse models and cell-based assays. *Gene*. 2008;412:84–94.
48. Salo AM, Cox H, Farndon P, Moss C, Grindulis H, Risteli M, et al. A connective tissue disorder caused by mutations of the lysyl hydroxylase 3 gene. *Am J Hum Genet*. 2008;83:495–503.
49. Coley W, Rayavarapu S, Pandey GS, Sabina RL, Van der Meulen JH, Ampong B, et al. The molecular basis of skeletal muscle weakness in a mouse model of inflammatory myopathy. *Arthritis Rheum*. 2012;64:3750–9.
50. Pan Y, Zhang L, Liu Q, Li Y, Guo H, Peng Y, et al. Insertion of a knockout-first cassette in Ampd1 gene leads to neonatal death by disruption of neighboring genes expression. *Sci Rep*. 2016;6:35970.
51. Wei CB, Wang JQ, Chen FY, Niu H, Li K. DNA sequence polymorphism within the bovine *adenosine monophosphate deaminase 1 (AMPD1)* gene is associated with production traits in Chinese cattle. *Genet Mol Res*. 2015;14:1025–33.
52. Ortiga-Carvalho T, Chiamolera M, Pazos-Moura C, Wondisford FE. Hypothalamus-pituitary-thyroid axis. *Compr Physiol*. 2016;6:1387–428.
53. Medrano RF, Hua HJ. Advances in thyroid hormones function relate to animal nutrition. *Ann Thyroid Res*. 2016;2:45–52.
54. Reinehr T. Thyroid function in the nutritionally obese child and adolescent. *Curr Opin Pediatr*. 2011;23:415–20.
55. Huang DW, Wang JX, Liu QY, Chu MX, Di R, He JN, et al. Analysis on DNA sequence of TSHB gene and its association with reproductive seasonality in goats. *Mol Biol Rep*. 2013;40:1893–904.
56. Valverde P, Healy E, Jackson I, Rees JL, Thody AJ. Variants of the melanocyte-stimulating hormone receptor gene are associated with red hair and fair skin in humans. *Nat Genet*. 1995;11:328–30.
57. Bennett DC, Lamoreux ML. Pigment gene focus the color loci of mice – A genetic century. *Pigment Cell Res*. 2003;16:333–44.
58. Switonski M, Mankowska M, Salamon S. Family of melanocortin receptor (MCR) genes in mammals' mutations, polymorphisms and phenotypic effects. *J Appl Genet*. 2013;54:461–72.
59. Gustafson NA, Gandolfi B, Lyons LA. Not another type of potato: mC1R and the russet coloration of Burmese cats. *Anim Genet*. 2017;48:116–20.
60. Marklund L, Moller MJ, Sandberg K, Andersson L. A missense mutation in the gene for melanocyte-stimulating hormone receptor (*MC1R*) is associated with the chestnut coat color in horses. *Mamm Genome*. 1996;7:895–9.
61. Dreger DL, Schmutz SM. A new mutation in MC1R explains a coat color phenotype in 2 "Old" breeds: Saluki and Afghan hound. *J Hered*. 2010;101:644–9.
62. Li J, Yang H, Li J, Li H, Ning T, Pan XR, et al. Artificial selection of the melanocortin receptor 1 gene in Chinese domestic pigs during domestication. *Heredity (Edinb)*. 2010;105:274–81.
63. Klungland H, Vage DI, Gomez-Raya L, Adalsteinsson S, Lien S. The role of melanocyte-stimulating hormone (MSH) receptor in bovine coat color determination. *Mamm Genome*. 1995;6:636–9.
64. Yang GL, Fu DL, Lang X, Wang YT, Cheng SR, Fang SL, et al. Mutations in *MC1R* gene determine black coat color phenotype in Chinese sheep. *Sci World J*. 2013;2013:675382.
65. Fontanesi L, Beretti F, Riggio V, Dall'Olio S, González EG, Finocchiaro R, et al. 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 Genet*. 2009;10:47.



66. Nicoloso L, Negrini R, Ajmone-Marsan P, Crepaldi P. On the way to functional agro biodiversity: coat colour gene variability in goats. *Animal*. 2012;6:41–9.
67. Badaoui B, Manunza A, Castelló A, D'Andrea M, Pilla F, Capote J, et al. Technical note: advantages and limitations of authenticating Palmeta goat dairy products by pyrosequencing the melanocortin 1 receptor (*MC1R*) gene. *J Dairy Sci*. 2014;97:7293–7.
68. Kırıkçı K, Noce A, Zidi A, Serradilla JM, Carrizosa J, Urrutia B, et al. Analysing the diversity of the caprine melanocortin 1 receptor (*MC1R*) in goats with distinct geographic origins. *Small Rumin Res*. 2016;145:7–11.
69. Suzuki I, Tada A, Ollmann MM, Barsh GS, Im S, Lamoreux ML, et al. Agouti signaling protein inhibits melanogenesis and the response of human melanocytes to alpha-melanotropin. *J Invest Dermatol*. 1997;108:838–42.
70. Badaoui B, D'Andrea M, Pilla F, Capote J, Zidi A, Jordana J, et al. Polymorphism of the goat Agouti signaling protein gene and its relationship with coat color in Italian and Spanish breeds. *Biochem Genet*. 2011;49:523–32.
71. Hosking BM, Wang SCM, Chen SL, Penning S, Koopman P, Muscat GEO. *SOX18* directly interacts with *MEF2C* in endothelial cells. *Biochem Biophys Res Commun*. 2001;287:493–500.
72. James K, Hosking B, Gardner J, Muscat GEO, Koopman P. *Sox18* mutations in the ragged mouse alleles jagged-like and opossum. *Genesis*. 2003;36:1–6.
73. Fitch KR, McGowan KA, Raamsdonk CD Van, Fuchs H, Lee D, Puech A, H, et al. Genetics of dark skin in mice. *Genes Dev*. 2003;17:214–28.
74. Rao C, Foerzler D, Loftus SK, Liu S, McPherson JD, Jungers KA, et al. A defect in a novel *ADAMTS* family member is the cause of the belted white-spotting mutation. *Development*. 2003;130:4665–72.
75. Silver DL, Hou L, Somerville R, Young ME, Apte SS, Pavan WJ. The secreted metalloprotease *ADAMTS20* is required for melanoblast survival. *PLoS Genet*. 2008;4:e1000003.
76. Dubail J, Apte SS. Insights on *ADAMTS* proteases and *ADAMTS*-like proteins from mammalian genetics. *Matrix Biol*. 2015;44–46:24–37.
77. Enomoto H, Nelson CM, Somerville RPT, Mielke K, Dixon LJ, Powell K, et al. Cooperation of two *ADAMTS* metalloproteases in closure of the mouse palate identifies a requirement for versican proteolysis in regulating palatal mesenchyme proliferation. *Development*. 2010;137:4029–38.
78. McCulloch DR, Nelson CM, Dixon LJ, Silver DL, Wylie JD, Lindner V, et al. *ADAMTS* metalloproteases generate active versican fragments that regulate interdigital web regression. *Dev Cell*. 2009;17:687–98.
79. Jackson HW, Defamie V, Waterhouse P, Khokha R. *TIMPs*: versatile extracellular regulators in cancer. *Nat Rev Cancer*. 2016;17:38–53.
80. Cruz-Muñoz W, Kim I, Khokha R. *TIMP-3* deficiency in the host, but not in the tumor, enhances tumor growth and angiogenesis. *Oncogene*. 2006;25:650–5.
81. Johnson JL, Kozysa A, Kharlamova AV, Gulevich RG, Perelman PL, Fong HWF, et al. Platinum coat color in red fox (*Vulpes vulpes*) is caused by a mutation in an autosomal copy of *KIT*. *Anim Genet*. 2015;46:190–9.
82. Shanado Y, Kometani M, Uchiyama H, Koizumi S, Teno N. Lysophospholipase I identified as a ghrelin deacylation enzyme in rat stomach. *Biochem Biophys Res Commun*. 2004;325:1487–94.
83. Lindholm-Perry AK, Kuehn LA, Smith TPL, Ferrell CL, Jenkins TG, Freetly HC, et al. A region on *BTA14* that includes the positional candidate genes *LYPLA1*, *XKR4* and *TMEM68* is associated with feed intake and growth phenotypes in cattle. *Anim Genet*. 2012;43:216–9.
84. Fortes MRS, Lehnert SA, Bolormaa S, Reich C, Fordyce G, Corbet NJ, et al. Finding genes for economically important traits: brahman cattle puberty. *Anim Prod Sci*. 2012;52:143–50.
85. Cánovas A, Rincón G, Islas-Trejo A, Jimenez-Flores R, Laubscher A, Medrano JF. RNA sequencing to study gene expression and single nucleotide polymorphism variation associated with citrate content in cow milk. *J Dairy Sci*. 2013;96:2637–48.
86. Fox JEB, Goll DE, Reynolds CC, Phillips DR. Identification of two proteins (actin-binding protein and P235) that are hydrolyzed by endogenous  $Ca^{2+}$ -dependent protease during platelet aggregation. *J Biol Chem*. 1985;260:1060–6.
87. Ouali A. Proteolytic and physicochemical mechanisms involved in meat texture development. *Biochimie*. 1992;74:251–65.
88. Koochmaria M. Muscle proteinases and meat aging. *Meat Sci*. 1994;36:93–104.
89. Ropka-Molik K, Żukowski K, Eckert R, Piórkowska K, Oczkowicz M, Gurgul A, et al. Whole transcriptome analysis of the porcine muscle tissue of breeds differing in muscularity and meat quality traits. *Livest Sci*. 2015;182:93–100.
90. Xu L, Bickhart DM, Cole JB, Schroeder SG, Song J, Van Tassel CP, et al. Genomic signatures reveal new evidences for selection of important traits in domestic cattle. *Mol Biol Evol*. 2015;32:711–25.
91. Randhawa IAS, Khatkar MS, Thomson PC, Raadsma HW. Composite selection signals for complex traits exemplified through bovine stature using multibreed cohorts of European and African *Bos taurus*. *G3 (Bethesda)*. 2015;5:1391–401.
92. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci USA*. 2012;109:19529–36.
93. Meyer RC, Giddens MM, Schaefer SA, Hall RA. *GPR37* and *GPR37L1* are receptors for the neuroprotective and glioprotective factors prosaptide and prosaposin. *Proc Natl Acad Sci USA*. 2013;110:9529–34.
94. Kajimoto Y, Kaneto H. Role of oxidative stress in pancreatic beta-cell dysfunction. *Ann NY Acad Sci*. 2004;1011:168–76.
95. Li G, Alii S, Currie RW. Insulin induces myocardial protection and *Hsp70* localization to plasma membranes in rat hearts. *Am J Physiol Heart Circ Physiol*. 2006;291:H1709–21.
96. Victoria Sanz Fernandez M, Johnson JS, Abuajamieh M, Stoakes SK, Seibert JT, Cox L, et al. Effects of heat stress on carbohydrate and lipid metabolism in growing pigs. *Physiol Rep*. 2015;3:e12315.
97. Sanz Fernandez MV, Stoakes SK, Abuajamieh M, Seibert JT, Johnson JS, Horst EA, et al. Heat stress increases insulin sensitivity in pigs. *Physiol Rep*. 2015;3:e12478.
98. Satrapa RA, Razza EM, Castilho AC, Simões RA, Silva CF, Nabhan T, et al. Differential expression of IGF family members in heat-stressed embryos produced in vitro from OPU-derived oocytes of Nelore (*Bos indicus*) and Holstein (*Bos taurus*) cows. *Reprod Domest Anim*. 2013;48:1043–8.
99. Nezer C, Moreau L, Brouwers B, Coppiepers W, Detilleux J, Hanset R, et al. An imprinted QTL with major effect on muscle mass and fat deposition maps to the *IGF2* locus in pigs. *Nat Genet*. 1999;21:155–6.
100. Emmons SW, Lipton J. Genetic basis of male sexual behavior. *J Neurobiol*. 2003;54:93–110.
101. Zakany J, Duboule D. A genetic basis for altered sexual behavior in mutant female mice. *Curr Biol*. 2012;22:1676–80.
102. Liang L, Shen YY, Pan XW, Zhou TC, Yang C, Irwin DM, et al. Adaptive evolution of the *Hox* gene family for development in bats and dolphins. *PLoS One*. 2013;8:e65944.
103. Nery MF, Borges B, Dragalzew AC, Kohlsdorf T. Selection on different genes with equivalent functions: the convergence story told by *Hox* genes along the evolution of aquatic mammalian lineages. *BMC Evol Biol*. 2016;16:113.
104. Lu Q, Knoepfler PS, Scheele J, Wright DD, Kamps MP. Both *Pbx1* and *E2A-Pbx1* bind the DNA motif ATCAATCAA cooperatively with the products of multiple murine *Hox* genes, some of which are themselves oncogenes. *Mol Cell Biol*. 1995;15:3786–95.
105. Ravasi T, Suzuki H, Cannistraci CV, Katayama S, Bajic VB, Tan K, et al. An atlas of combinatorial transcriptional regulation in mouse and man. *Cell*. 2010;140:744–52.
106. FAO. Animal genetic resources: strategies for improved use and conservation. Proceedings of the 2nd Meeting of the FAO/UNEP Expert Panel: June 1986; Warsaw; 1987.
107. Delogu A, Sellers K, Zagoraiou L, Bocianowska-Zbrog A, Mandal S, Guimera J, et al. Subcortical visual shell nuclei targeted by ipRGCs develop from a *Sox14*+ GABAergic progenitor and require *Sox14* to regulate daily activity rhythms. *Neuron*. 2012;75:648–62.
108. Douris N, Green CB. NOC out the fat: a short review of the circadian deadenylyase Nocturnin. *Ann Med*. 2008;40:622–6.
109. Elsea SH, Williams SR. Smith-Magenis syndrome: haploinsufficiency of *RAI1* results in altered gene regulation in neurological and metabolic pathways. *Expert Rev Mol Med*. 2011;13:e14.
110. Williams SR, Zies D, Mullegama SV, Grotewiel MS, Elsea SH. Smith-Magenis syndrome results in disruption of *CLOCK* gene transcription and reveals an integral role for *RAI1* in the maintenance of circadian rhythmicity. *Am J Hum Genet*. 2012;90:941–9.

111. De Leersnyder H. Smith-Magenis syndrome. *Handb Clin Neurol*. 2013;111:295–6.
112. Lehman MN, Karsch FJ. Do gonadotropin-releasing hormone, tyrosine hydroxylase-, and beta-endorphin-immunoreactive neurons contain estrogen receptors? A double-label immunocytochemical study in the Suffolk ewe. *Endocrinology*. 1993;133:887–95.
113. Stella A, Nicolazzi EL, Van Tassell CP, Rothschild M, Colli L, Rosen BD, et al. AdaptMap: Exploring goat diversity and adaptation. *Genet Sel Evol*. 2018. <https://doi.org/10.1186/s12711-018-0427-5>.

**Ready to submit your research? Choose BMC and benefit from:**

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

**At BMC, research is always in progress.**

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

