scientific reports

Check for updates

OPEN

Local adaptations of Mediterranean sheep and goats through an integrative approach

Bruno Serranito[1,2,12], Marco Cavalazzi[3,12], Pablo Vidal[4], Dominique Taurisson-Mouret[3,5], Elena Ciani[6], Marie Bal[3], Eric Rouvellac[3], Bertrand Servin[7], Carole Moreno-Romieux[7], Gwenola Tosser-Klopp[7], Stephen J. G. Hall[8], Johannes A. Lenstra[9], François Pompanon[10], Badr Benjelloun[10,11] & Anne Da Silva[1✉]

Small ruminants are suited to a wide variety of habitats and thus represent promising study models for identifying genes underlying adaptations. Here, we considered local Mediterranean breeds of goats (n = 17) and sheep (n = 25) from Italy, France and Spain. Based on historical archives, we selected the breeds potentially most linked to a territory and defined their original cradle (i.e., the geographical area in which the breed has emerged), including transhumant pastoral areas. We then used the programs PCAdapt and LFMM to identify signatures of artificial and environmental selection. Considering cradles instead of current GPS coordinates resulted in a greater number of signatures identified by the LFMM analysis. The results, combined with a systematic literature review, revealed a set of genes with potentially key adaptive roles in relation to the gradient of aridity and altitude. Some of these genes have been previously implicated in lipid metabolism (SUCLG2, BMP2), hypoxia stress/lung function (BMPR2), seasonal patterns (SOX2, DPH6) or neuronal function (TRPC4, TRPC6). Selection signatures involving the PCDH9 and KLH1 genes, as well as NBEA/NBEAL1, were identified in both species and thus could play an important adaptive role.

Sheep and goats were among the first mammalian livestock species to be domesticated, in a process that started at ca. 10,000–9500 B.P. in Mesopotamia[1]. The subsequent expansion across the Mediterranean Basin via maritime transport included different waves of colonization. The remains testifying to these founding periods along the northern coasts of the western Mediterranean are dated from 8100 to 7700 B.P.[2].

Waves of dispersal, which have dispatched small groups of individuals (founder effect) to very diverse habitats, are the starting point of the process that led to the emergence of breeds[3]. Traditional pastoralism, characterized by limited human intervention (i.e., "soft" artificial selection[4]), and in conjunction with a strong influence of natural selection, led over the millennia to populations genetically differentiated and adapted to various agro-climatic conditions[5]. Sheep and goats provide an interesting model for deepening our knowledge of adaptive mechanisms under extreme environmental conditions.

In industrialized countries, artificial breeding has intensified over the past 200 years[4], focusing mainly on agronomic traits related to milk, meat production and fiber quality. Many breeds have experienced negative impacts on their ability to adapt, their level of hardiness and sometimes their fitness by crossbreeding and/or intensive artificial selection[4,6]. In contrast, the so-called local breeds have remained at least partially, connected to their original environments. We suggest it is among these local breeds that selection signatures related to environmental adaptation will be most detectable. For example, studies on local sheep in Ethiopia[7] and Tibet[8,9] have identified

several genes involved in adaptation to thermal stress and hypoxia. Kim et al.[10], considering both local goats and sheep of hot and arid environment, have highlighted genes implicated in the adaptation to thermal stress.

This study aimed at identifying the genomic bases of adaptation in local breeds of goats and sheep distributed along the Mediterranean arc, in Italy, the south of France and Spain. The area under consideration harbours contrasting environments, such as high altitude areas (Alps, Pyrenees, Corsica), coastal areas, wetlands and arid areas, with closely related sheep and goat breeds[11–13]. All these features provide an ideal situation for the identification of adaptation signatures.

We hypothesize that selection signatures relevant to environmental adaptation, will be more reliably identified by reference to the environmental conditions of the breed's cradle of origin, rather than to those of the locations where the animals were actually sampled. Indeed, breeds of domestic animals are at the interface between anthropized and natural environments; their evolutionary trajectory is thus determined by both the environment and human societies. First, (i) sheep and goat breeds have originally emerged within socially structured human groups, sometimes referred to as clans or tribes, occupying a well-defined territory, and within which herds were transmitted from generation to generation[14, 15]. In industrialized countries, the traditional links between territory, social group and breeds has become distorted. Thus, herds of a given breed can be found outside of its original cradle. Hence, the search for a selection signature may be optimized by considering the environmental conditions of the breed's cradle, where the genomes have been shaped over time, rather than the current location of herds. Moreover, (ii) we should consider that traditional pastoralism was largely based on transhumance. During these periods, the animals are kept outdoors, most often in mountain altitude summer pastures, such that the strength of natural selection is particularly pronounced. Even if transhumance practices are increasingly abandoned[16, 17], they have fundamentally shaped the genetic adaptations of sheep and goats.

The objective of this study was to investigate the regions of the genome underlying environmental adaptations, i.e. selection signatures linked to environmental variables. Selection signatures are identified as genomic areas comprising a number of neighboring markers under selection. We used the PCAdapt[18] and LFMM[19] methods. PCAdapt uses principal component analysis to describe population structure and identifies candidate markers as outliers in terms of inferred population structure. The LFMM approach searches for significant associations with environmental factors while controlling for neutral population structure. These approaches were coupled with a thorough analysis of the literature to identify the signatures including genes most likely related to environment adaptations. Moreover, we used datasets that have been previously investigated, in order to assess if the use of historical and anthropological dimensions of breeds improves the detection of selection signatures. We reasoned that (i) the inclusion of authentic local breeds in the analysis will limit the background noise caused by breeds for which the link to the environment is weak or has eroded, for instance by crossbreeding or migration; and (ii) targeting the geographical area in which populations have evolved rather than the present distribution will optimize the identification of adaptive genomic regions. Next, the proposed approach was implemented to search for selection signatures underlying local adaptation. We attached particular importance to the selection signatures involving homologous genomic regions in sheep and goats.

Results

Breed analyses. The dataset includes 17 goat breeds and 25 sheep breeds, selected on the basis of historical and anthropological documentation (Table 1, see Supplementary Table 1 for the initial list of 83 breeds and "Material and methods" section for an explanation of the choice of breeds for analysis). Figures 1 and 2 show the geographical cradles of the different breeds and provide details on the phenotypes of the animals and the environmental conditions at the cradle level.

For 15 of the 42 selected sheep and goat breeds considered, we found documented evidence targeting ancestral populations at their origin, dating back several millennia (see Supplementary Table 2 for historical references). The remaining breeds have a history of at least several centuries, with the exception of 4 Italian goat breeds (DIT, CCG, GAR, ARG) and 3 Italian sheep breeds (DEL, BIE, VAL), which were recognized about a hundred years ago under their current names, but which have an ancient history characterized by mixtures of populations native to the territories considered. Seventeen breeds were classified as "critical" or "endangered" according to FAO[20]. Considering jointly the goat and sheep breeds, 30 breeds have been described as showing adaptations to stressful environments and 29 are, or were, transhumant. For most transhumant breeds the transhumance is, or was, vertical, so we included their summer mountain pasture in the geographical definition of the cradle (see Supplementary Table 2).

The Admixture analysis is detailed in Supplementary text 1. For both goats and sheep (Figure 3 in the Supplementary text 1), the cross-entropy curve decreases as K increases, which is characteristic of a distance isolation pattern[21], moreover, the Mantel tests postulated distance isolation patterns (p value = 0.049 for goats and 0.001 for sheep). In this case, to correct for the limited genetic structure, we used K = 1 and K = 2, in the LFMM analyses, as recommended by Dalongeville et al.[22], De Kort et al.[23] and Capblancq et al.[24].

Comparison of approaches. We performed the LFMM analyses on the goat dataset using three approaches for characterising the environment of each breed (see "Identification and analysis of genomic selection signatures" and "Experimental test of the method" sections in "Material and methods" section and Fig. 3). The poor results obtained with the Annual Mean Precipitation are not shown (i.e., the different methods could only reveal two signatures each, with NBEA identified in all cases). Notably, for the Altitude, the individual GPS method did not detect any signatures; the GPS area method identified only two signatures, whereas 10 signatures were recorded with the cradle method. For the Mean Annual Temperature, five signatures were observed with the individual GPS method, 10 with the GPS area method, and 15 by the cradle method. The process of clustering the environ-

Breed name	Code	Country	Cradle location
Goat dataset			
Bionda dell Adamello	BIO (n = 24)	Italy	The Alpine massif
Valdostana	VAL (n = 24)	Italy	The Alpine massif
Orobica	ORO (n = 24)	Italy	The Alpine massif
Val Passiria	VSS (n = 24)	Italy	The Alpine massif
Di Teramo	DIT (n = 24)	Italy	Abruzzo massif
Ciociara Grigia	CCG (n = 19)	Italy	Aurunci Mountains
Garganica	GAR (n = 20)	Italy	The Gargano promontory
Argentata Dell'Etna	ARG (n = 25)	Italy	Mount Etna
Nicastrese	NIC (n = 25)	Italy	The province of Catanzaro
Girgentana	GGT (n = 30)	Italy	The Agrigento coast in Sicily
Corse	CRS (n = 29)	France	Corsican massifs
Provençale	PVC (n = 18)	France	The "Provence of hills"
Pyrénéenne	PYR (n = 25)	France	Pyrenees
Bermeya	BEY (n = 24)	Spain	Asturias
Blanqua De Rasquera	RAS (n = 20)	Spain	Southern Catalonia
Malaguena	MLG (n = 41)	Spain	Malaga
Mallorquina	MAL (n = 20)	Spain	Island of Mallorca
Sheep dataset			
Manech Tête Rousse	MTR (n = 25)	France	Basque hillsides
Tarasconnaise	TAR (n = 15)	France	Central Pyrenees
Causse Du Lot	CDL (n = 20)	France	Plateaus of the Lot
Limousine	LIM (n = 18)	France	Millevaches plateau
Rava	RAV (n = 20)	France	The Massif Central
Blanche Du Massif	BMC (n = 20)	France	The Massif Central
Noire Du Velay	NVE (n = 19)	France	The Massif Central
Préalpes Du Sud	PAS (n = 17)	France	The Southern Alps
Mourerous	MOUR (n = 16)	France	The Maritime Alps
Corse	COR (n = 16)	France	The Corsican massifs
Delle Langhe	DEL (n = 24)	Italy	The Alpine massif
Biellese	BIE (n = 21)	Italy	The Alpine massif
Bergamasca	BER (n = 24)	Italy	The Alpine massif
Alpagota	ALP (n = 24)	Italy	The Alpine massif
Massese	MAS (n = 24)	Italy	The Alpi Apuane mountains
Appenninica	APP (n = 24)	Italy	The Apennine mountains
Laticauda	LAT (n = 24)	Italy	Campania and Calabria
Bagnolese	BAG (n = 23)	Italy	Campania
Valle Del Bellice	VAL (n = 24)	Italy	Sicily
Sardinian White	SAR (n = 24)	Italy	Sardinia
Latxa	LATX (n = 24)	Spain	Pyrenees
Gallega	GAL (n = 27)	Spain	Gallicia
Churra	CHU (n = 30)	Spain	Castile and León
Ojalada	OJA (n = 24)	Spain	The province of Soria
Roja Mallorquina	(MAL, n = 28)	Spain	Island of Mallorca

Table 1. Description of datasets. Breeds, number of individuals, cradle location. n: number of individuals.

ments allowed the detection of 13 signatures via the GPS areas and 19 (maximum value) via the cradles (see details in Supplementary Figures 3 and 4).

Goat dataset: track for selection signatures. From the 37 signatures, 21 were identified by PCAdap, 19 by LFMM; 3 signatures were supported by both methods (Table 2). Seven signatures were observed on chromosome 6, and four on chromosome 12. Genes associated with these signatures have been documented in the literature (Supplementary Table 3). Of the 21 signatures highlighted by PCAdapt, 15 were found to be involved in adaptive processes and of the 19 signatures highlighted by LFMM, 15 were linked to adaptive processes according to the literature.

The LFMM analysis was based on the PCA (Principal Component Analysis)/HCPC (Hierarchical Clustering on Principal Components) ranking in which the cradles were clustered according to the similarity of their
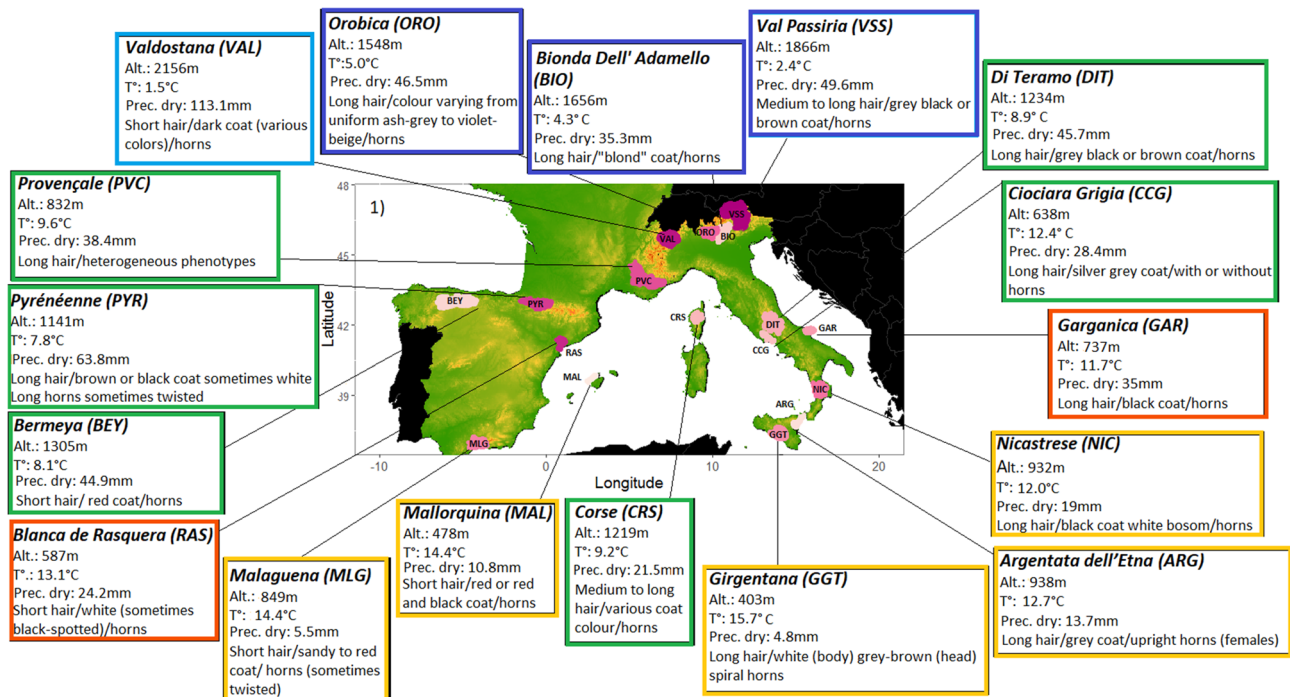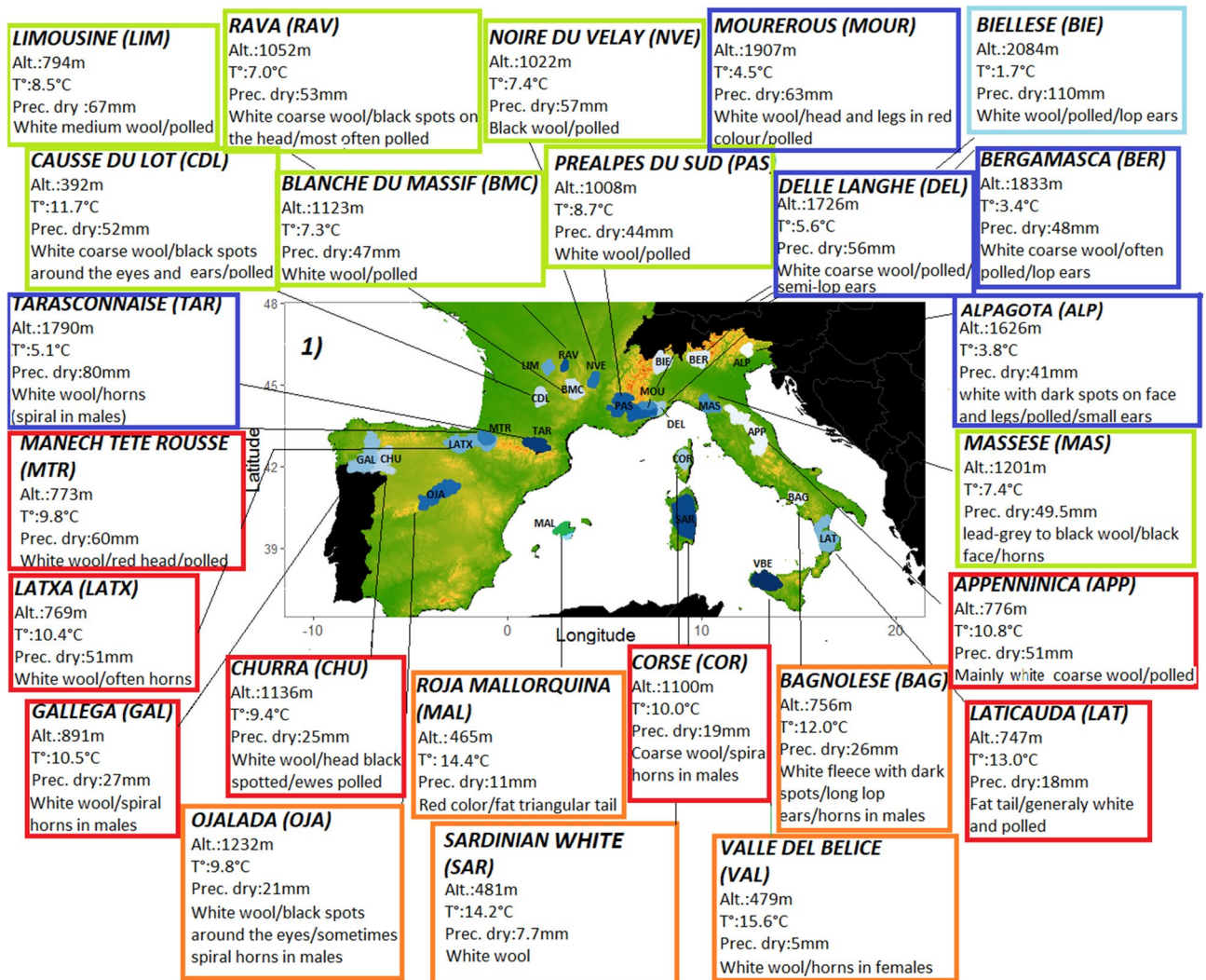
Figure 1. Description of goat cradles as a function of environmental variables: synthetic descriptions of cradles and some characteristic breed phenotypic traits. The breed labels are color coded according to the clustering ranking defined by the PCA/HCPC analysis shown in Fig. 4.

environments (Fig. 4). The PCA analysis was largely driven by the first component (76.9% of variance explained by PC1). The cradles appeared to be distributed along a South–North axis, following a decreasing temperature and aridity gradient towards the negative values of PC1, and an increasing altitudinal gradient towards the positive ones. HCPC clustering, indicating K = 5 as the optimal number of clusters, was as follows: the first group (1) clustered breeds from southern Italy, Sicily, southern Spain and the island of Majorca, i.e. regions with high annual mean temperatures and very low rainfall during the dry season. Group (2) clustered the Spanish (Blanca de Rasquera, RAS) and Italian (Garganica, GAR) breeds, raised under low rainfall also during the wet season. The third group (3) clustered Spanish, French and Italian goat breeds, in an intermediate position with regard to the aridity gradient during the dry season. Groups (4) and (5) corresponded to Italian breeds in regions of high altitude, low average temperatures and high precipitation (Fig. 4). PCA/HCPC processed with environmental data recorded at the GPS area level (see Supplementary Figure 4) also showed five clusters that followed a temperature and precipitation gradient, but without an altitude gradient (Supplementary Figure 4.3), resulting in a different ranking clustering than the one obtained by the cradle method (Fig. 4) and a lower number of signatures highlighted by LFMM (Fig. 3).

The LFMM analysis, based on the cluster ranking obtained by the cradle method (Fig. 5), showed several strong signals near genes that have been implicated in adaptation: SOX2, ADAMTS20, UBE2R2/UBAP2, DPH6, NBEA, TRPC4, TRPC6, and UHRF1BP1/C23H6ORF106. Strong signals detected by PCAdapt analysis and also found in the literature as implicated in adaptive processes corresponded to the genes SIL1/CTNNA1, UBE2R2/UBAP2, SHQ1, SUCLG2 as well as a large area on chromosome 12 containing RNF17, ZMYM2, PARP4 and XPO4 and the vast intergenic area between KLH1 and PCDH9. A particularly strong signal was found on chromosome 18 at locus LOC108637979. The detailed results can be viewed in Supplementary Tables 3, 5 and 6.

Sheep dataset: track for selection signatures. From the 26 selection signatures, 8 were identified by PCAdapt, 23 by LFMM; 5 signatures were detected by both methods (Table 2). Genes associated with these signatures have been documented in the literature (Supplementary Table 4), all PCAdapt signatures and 18 out of 23 LFMM signatures could be linked to environmental adaptation.

Figure 6 shows the PCA/HCPC results used for the LFMM analysis. As in Fig. 4 for goats, PC1 accounted for a large part of the variation (76.0%) and showed similar South-North gradients of temperature and aridity (Fig. 6). The optimal K = 5 generated the following clusters: group (1) concerned regions with highest aridity and temperature levels, including the Sicilian, Southern Italian, the island of Majorca and Central Spanish breeds. Groups (2) and (3) were in an intermediate position, in term of dryness and wetness. Group (2) corresponded to regions showing slightly higher average temperatures than group (3) and high levels of precipitation during the wet season (especially for Gallega GAL, Manech Tête Rousse MTR and Latxa LATX), while group (3) included regions of higher altitudes (except for Causse du Lot CDL). Groups (4) and (5), including the Pyrenean and Alpine breeds of the three countries, corresponded to areas of high altitudes and significant precipitation levels (Fig. 6). LFMM on the basis of this cluster ranking obtained by the cradle method (Fig. 7) showed several

Figure 2. Description of sheep cradles as a function of environmental variables: synthetic descriptions of cradles and some characteristic breed phenotypic traits. The breed labels are color coded according to the clustering ranking defined by the PCA/HCPC analysis shown in Fig. 6.

strong signals near genes that previously have been implicated in environmental adaptation: NBEAL1, MSRB3, RXFP2, BMP2. Several other regions identified by LFMM included genes not previously been implicated in environmental adaptation (Supplementary Table 4).

The strongest PCAdapt signals corresponded to: PAK1 and the KLHL1/PCDH9 intergenic region, both described previously as being involved in environmental adaptation. A broad signal was on chromosome 6, near the genes HERC6, SPP1, LAP3, also previously involved in adaptation. The detailed results can be viewed in Supplementary Tables 4, 5 and 7.

Comparison of sheep and goat signatures. The genes NBEA (NBEA in goat and NBEAL1 in sheep), SHQ1 and VPS13 genes (goat VPS13B and sheep VPS13D), the regions containing HERC6 and ABCG2 as well as the KLHL1/PCDH9 intergenic region appeared to be under selection in both goat and sheep (Table 2). Most of the selection signatures (74%), or genes included in the signatures, were found in the literature on mammals and birds as being related to environmental adaptation (Supplementary Tables 3 and 4). Of these, the most noteworthy include the following: 16 goat and 14 sheep selection signatures previously identified in Chinese sheep adapted to extreme environments[9], four signatures in goats and three in sheep found in Ethiopian sheep kept at high-altitude[7] and three goat and six sheep signatures detected in Egyptian sheep and goats from arid environments[10].
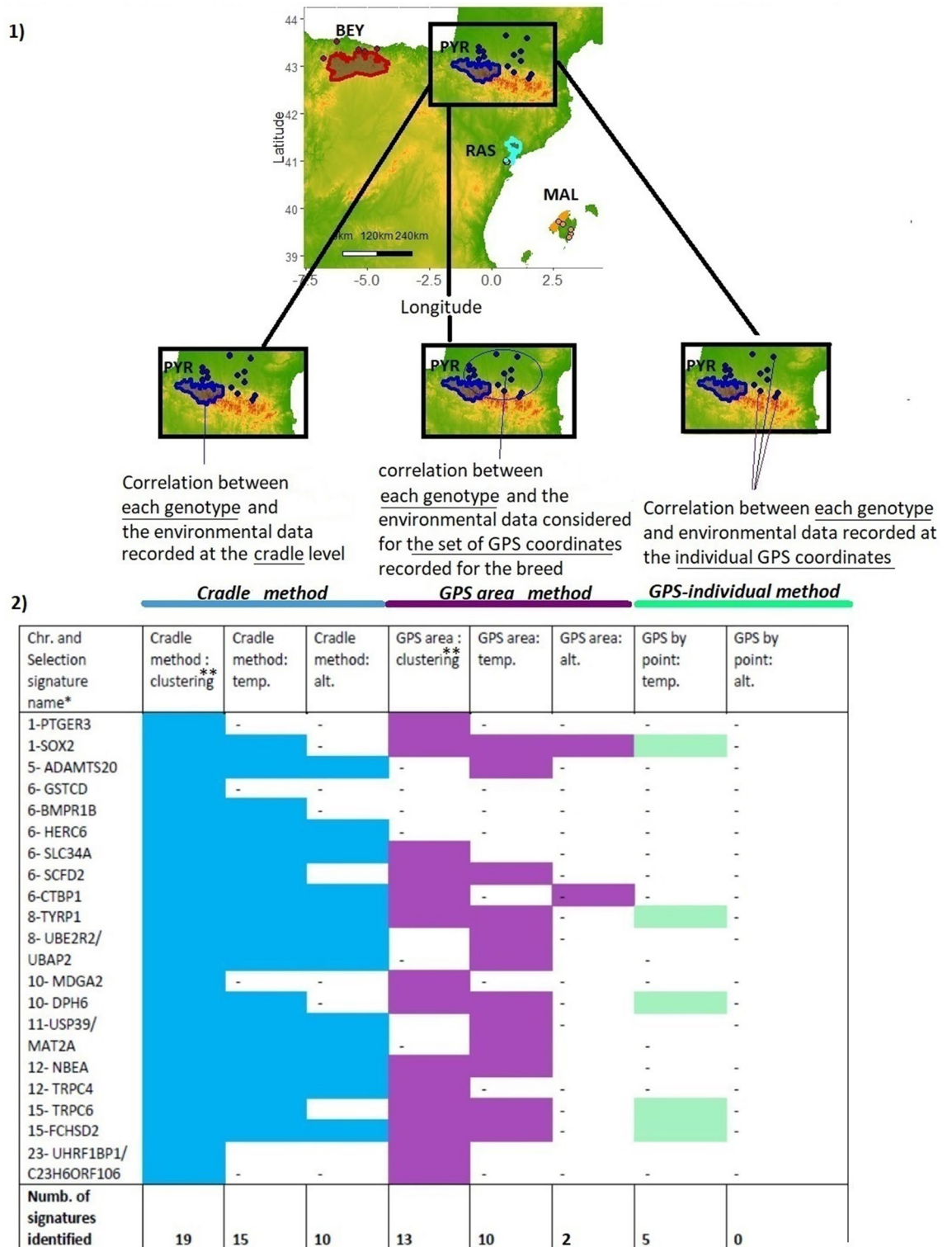
Figure 3. Experimental test of the method. The signature search was done via LFMM. The environment was characterized either at the level of the GPS coordinates (individual GPS method) or considering all the GPS points recorded for a given breed (GPS area method), or at the level of the cradle (cradle method). The environmental variables Altitude and Mean Annual Temperature were considered individually. Chr.: chromosome number. *Selection signature are named according to the gene(s) most likely targeted by the selection process, taking into account the distance of the SNPs from the genes, the associated *p* values and the number of SNPs near or in the gene(s); it is possible to know in detail the genes associated with each selection signature (see Supplementary Table 3). **Clustering: the environment was characterized considering all the variables at the same time and according to the PCA/HCPC process.
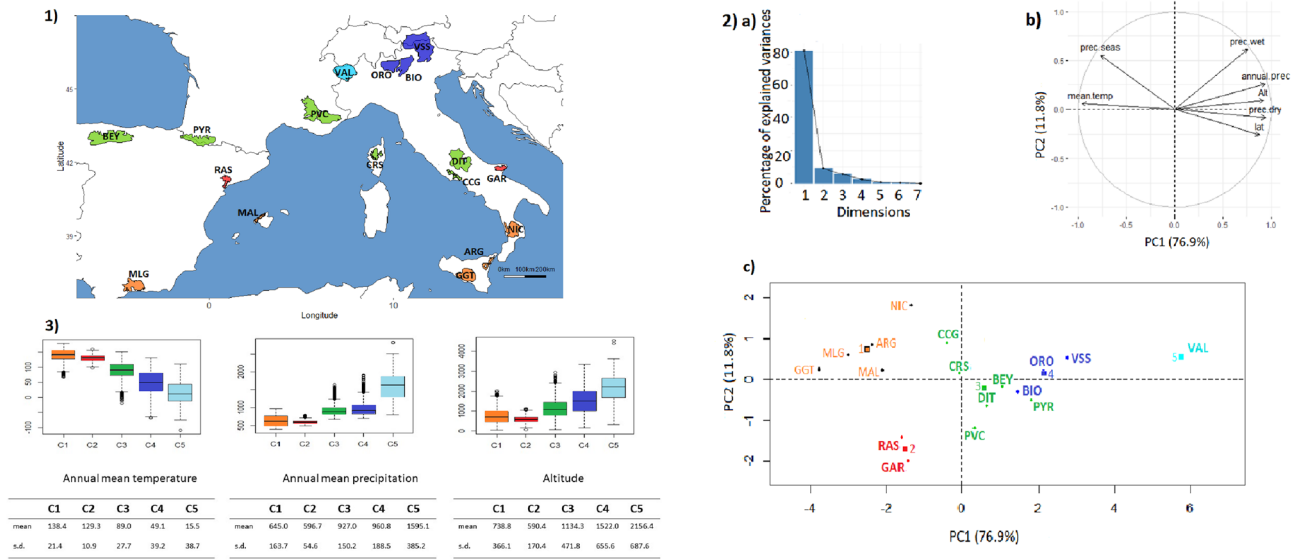
Goat

Chr	Signature range	PCAdapt	LFMM	Gene(s) most likely under selection
1	46047752-46411087		●	PTGER3
	84242981-84773966		●	SOX2
5	36576428-37074733		●	ADAMTS20
6	19677067-20300302		●	GSTCD/TET2
	30139296-30874402		●	BMPR1B
	36835876-36949805	●	●	**HERC6**
	45415286-45625340		●	SLC34A2
	69082499-69758377		●	SCFD2
	85991683-86155374		●	CSN1S1/CSN2
	116501734-117144177		●	CTBP1
7	59624626-60848321	●		SIL1/CTNNA1
8	74894754-75321840	●	●	UBE2R2/UBAP2
	31846675-32220312		●	TYRP1
10	49806948-50111884	●		TCF12
	61579137-63362001		●	MDGA2
	71545588-71700461		●	DPH6
11	37793580-38186897	●		LOC106502619
	48845878-49036557		●	USP39/MAT2A
	61247655-61456836	●		WDPCP
12	43598089-44471339	●		**KLHL1/PCDH9**
	48839256-51252236	●		RNF17/ZMYM2/PARP4/XPO4
	60637258-60950668	●	●	**NBEA**
	62109741-62503627		●	TRPC4
13	46458560-46473488	●		PRNP
14	16942897-17652266	●		**VPS13**B
15	6473101-6606528		●	TRPC6
	30213116-30556088		●	FCHSD2
18	36054355-37092141	●		NFATC3
	39636636-39980690	●		LOC108637979
19	32969117-33274858	●		PIGL/NCOR1
20	13992024-14204949	●		ADAMTS6
22	28867395-29079702	●		**SHQ1**
	33594470-34062919	●		SUCLG2
23	36461204-36962562	●		BTBD9
	39855572-40059589		●	UHRF1BP1/C23H6ORF106
25	30481819-30744262	●		AUTS2
26	28596734-29498263	●		FBXW4/BTRC/GBF1

Sheep

Chr	Signature range	PCAdapt	LFMM	Gene(s) most likely under Selection
1	117132460-117619369		●	POU2F1
	170496673-170755505		●	BBX
	198199858-198342295		●	MASP1
2	74396276-74749751		●	KDM4C
	203734535-204610983		●	**NBEAL1**
3	153576602-154585886	●		MSRB3
4	24727308-25337674		●	ISPD
	43272227-43454930		●	MAGI2
	47614996-48783775		●	BCAP29/SLC26A4
5	51252017-51859391		●	ARHGAP26
6	36121839-37756801* 37536930-38567860**	●	●	LCORL/**HERC6**
8	62665338-63271621		●	ARFGEF3/PERP
9	34730111-35192293		●	ATP6V1H
10	29413536-29776019* 29479711-29940445**	●	●	RXFP2
	38196759-39403332* 37941892-39043416**	●	●	**PCDH9**
	42194236-43935671	●		**PCDH9 /KLHL1**
12	38573067-39683281		●	**VPS13**D
13	48552093-49208171	●		BMP2
	55025081-56301025		●	CDH26
	57302923-57809890		●	RAB22A
15	59813250-60219075		●	DCDC5
17	55110665-55329873		●	CIT
18	66257896-68025345		●	TRAF3/CDC42BPB
19	28133753-28962852		●	PDZRN3/**SHQ1**
21	17885334-18117515	●		PAK1
25	19073058-19380300	●		JMJD1C

Table 2. Selection signatures identified in goats and sheep by LFMM and PCAdapt using the cradle method. Each signature was named according to the gene(s) most likely targeted by the selection process, taking into account the distance of the SNPs from the genes, the associated p values and the number of SNPs near or in the gene(s). All annotated genes associated with each selection signature are highlighted in Supplementary Tables 3 and 4. *PCAdapt signature; **LFMM signature; Chr.: chromosome number, in bold genes identified in both datasets.

Discussion

This study proposed a new genomic landscape approach for the identification of selection signatures in livestock breeds. The key aspects of this approach are (i) filtering the breeds to be included in the study and (ii) matching the genotypes to the environmental conditions recorded not at the current coordinates of the individuals, but at the level of the areas identified as their cradle of origin.

Performance of cradle versus GPS method. We targeted French, Spanish and Italian breeds with strong historical links to their territory of origin, in order to build datasets for analysis. For most of the breeds, this link was present for several centuries. The majority of the 63 signatures identified in the current study had not been detected in the initial studies of these datasets[24–30]. A high proportion (74%) of the highlighted selection signatures were previously identified by studies on environmental genetic adaptation in birds and mammals from different parts of the world, providing support for our approach. Further corroboration came from the many signatures identified in our study that have been previously reported in sheep or goats from extreme environments in different parts of the world ([7, 9, 10, 31, 32], see Supplementary Tables 3 and 4 for a complete list), and also from the several signatures found in both sheep and goats. We applied particularly stringent screens and reinforced our results with a thorough analysis of the literature in order to limit false positives as much as possible. Although our results suggest that the conducted approach can complement classical approaches by

Figure 4. Goat cradles analyses as a function of environmental variables. (**1**) Mapping display of the cradles coloured according to the result of the clustering. (**2**) HCPC analysis. (a) Scree plot. (b) PCA correlation circle. (c) PCA score plot of goat cradles. Colours represent the main clusters obtained by HCPC (K = 5). PC 1 and PC 2 represent respectively principal components 1 and 2; numbers in brackets show the variance explained by each PC. (**3**) Boxplot representation for the variables Mean Annual Temperature, Mean Annual Precipitation and Altitude for the different HCPC groups and associated numerical values. Alt.: Altitude in meters, T°: Temperature in °C × 10, prec.: precipitation in millimeters, min.: minimal, max.: maximal, s.d.: standard deviation.

Figure 5. Genome scan for selection signatures in local goat breeds obtained by the LFMM approach and via the cradle method. Each signature was named according to the gene(s) most likely targeted by the selection process, taking into account the distance of the SNPs from the genes, the associated *p* values and the number of SNPs near or in the gene(s).

providing new results, formal mathematical modeling as well as the identification of candidate mutations and the appropriate functional analyses are required to validate the results.

Results from our approach highlighted a few points. First, the vast majority of small ruminant breeds have been shaped by traditional transhumant breeding practices. However, the individual GPS method will be able to consider summer grazing areas (i) only for breeds and herds always managed according to these practices while transhumance practices are declining and (ii) only if the sampling is carried out during the summer period. Consistent with this, we observed that GPS methods revealed very few signatures with the Altitude variable, whereas it was possible to detect 53% of the total signatures identified with the cradle approach by the consideration of this single variable. Secondly, considering the Mean Annual Temperature variable, (i) the difference between the number of signatures detected by the GPS methods and the cradle approach was clear but less pronounced than
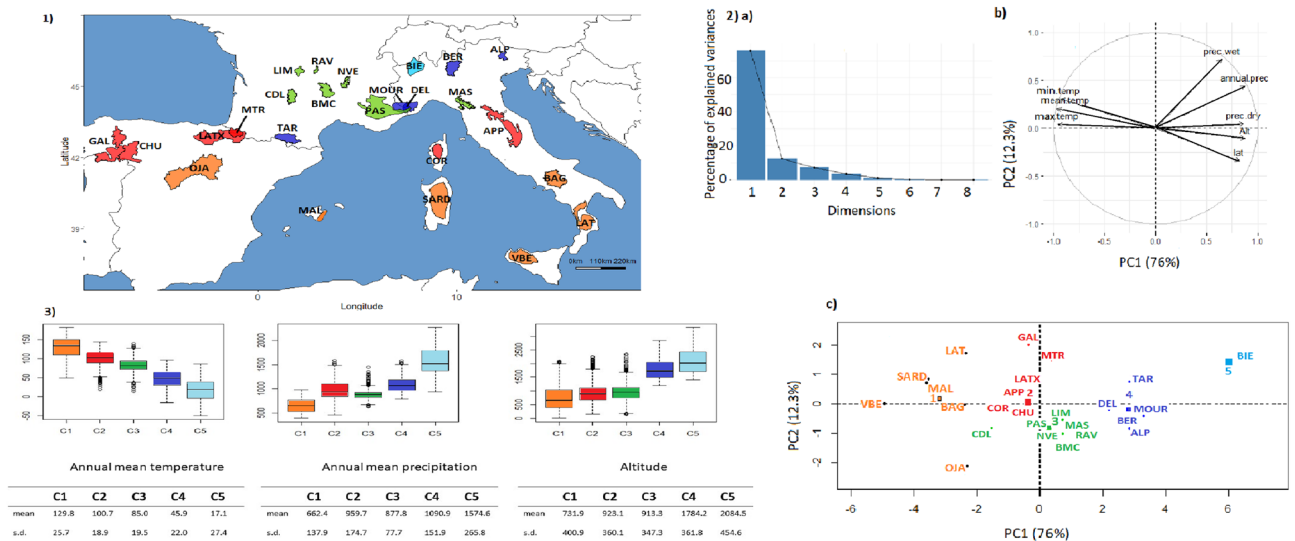
Figure 6. Sheep cradles analyses as a function of environmental variables. (**1**) Mapping display of the cradles coloured according to the result of the clustering. (**2**) HCPC analysis. (a) Scree plot. (b) PCA correlation circle. (c) PCA score plot of goat cradles. Colours represent the main clusters obtained by HCPC (K = 5). PC 1 and PC 2 represent respectively principal components 1 and 2; numbers in brackets show the variance explained by each PC. (**3**) Boxplot representation for the variables Mean Annual Temperature, Mean Annual Precipitation and Altitude for the different HCPC groups and associated numerical values. Alt.: Altitude in meters, T°: Temperature in °C × 10, prec.: precipitation in millimeters, min.: minimal, max.: maximal, s.d.: standard deviation.

Figure 7. Genome scan for selection signature in local sheep breeds obtained by the LFMM approach and via the cradle method. Each signature was named according to the gene(s) most likely targeted by the selection process, taking into account the distance of the SNPs from the genes, the associated *p* values and the number of SNPs near or in the gene(s).

for the Altitude variable and (ii) the correlation of genotypes with environmental variables recorded at the level of the area defined by the set of the breed's GPS coordinates (GPS area method) detected more signatures than considering the environmental variable at individual GPS coordinates (individual GPS method). This suggested that there are benefits of smoothing environmental data at the area level rather than considering it at individual points. Finally, the greatest number of selection signatures was identified by considering all the environmental variables at the same time and using the cradle method including transhumant pasture areas.

Thus, there are several strengths of our novel cradle method. First, taking into account the history of the breeds avoids including breeds that could interfere with the signal (i.e. breeds which do have a strong link with a given environment or whose link has been distorted). Secondly, in industrialized countries, agricultural intensification in recent decades has led to major shifts in practices, particularly the abandonment of transhumance, and also significantly changed the distribution areas of herds. Thus, the environmental characteristics recorded

at the current GPS points are different from those of the original cradle (see Supplementary Text 2 for details). Considering the cradle of the breed rather than the current distribution thus links with more accuracy the genetic determinants of adaptation with the environment that has mainly shaped the breed's genome. Thirdly, we combined the PCAdapt algorithm, which is particularly suited to the detection of selection signatures in heterogeneous populations[18], with LFMM, which identifies the links between genomic and environmental variation. On the basis of the scientific literature we eliminated PCAdapt signatures that appeared to be related to agronomic selection and researched the links established in the literature between other signatures and environmental adaptation. It can be hypothesized that PCAdapt signals not detected by LFMM reflect local adaptation to patchy environmental conditions unrelated to the environmental gradient. The two methods thus provide complementary information. It can be noted that the cradles were organized in both datasets according to a gradient of temperature and altitude, as shown by the PCA analyses; LFMM thus highlighted areas of the genome whose variation was linked to this gradient. The correlation between temperature/humidity variables and altitude is a classic issue in ecology that can be addressed by including microclimatic variables[33] and/or considering datasets designed in such a way as to allow the evaluation of temperature variations while altitude remains unchanged, or vice versa. These considerations provide an opening for further research that would improve our ability to link signatures more specifically to a specific factor.

It should be noted that the breeds, considered today, come from ancestral populations whose histories are largely unknown. Indeed, the various traces and historical documents indicating the existence of breeds on territories for several centuries, or even millennia, refer to the ancestral populations from which they originated. Hence, selection signatures identified may predate the formation of the breed or alternatively, reflect rapid recent changes. In addition, environmental features of a cradle may not be unique and may also have changed over time. The environmental characterization was inferred from climatic data recorded during the twentieth century. This constitutes a bias, which in the context of current climate change, must be carefully considered. We expect that our work will contribute to the synthesis of landscape genomics studies (e.g.[34]) by focusing on the particularities of livestock at the interface of the natural and anthropized environment.

Selection signatures identified in this study. The highlighted selection signatures can be categorized according to physiology and phenotypes. In Supplementary Text 3, we discussed the genes that seem to be involved (i) in lipid storage, (ii) in seasonal patterns and circadian behaviors, (iii) in coat colour and horns, (iv) in hypoxia and/or heat stress response, (v) in immunity, (vi) in lung function, and (vii) in neuronal function. We also discussed selection signatures that seemed difficult to classify because of their length and the diversity of the genes within them. Here we only focus on the most robust signals, considering both the *p* values and the number of SNPs involved in the selection signatures, and for which other studies have previously shown a link between the genes involved in the signatures and environmental adaptation.

Goat selection signatures. We identified SUCLG2 in goat by the PCAdapt approach. It was found under selection in sheep in arid Egyptian environments[10] and in the Chinese desert[9], but also in yaks adapted to extreme altitudes[35]. The gene is involved in the propanoate metabolic pathway and was significantly associated with growth in pigs[36]. Moreover, the latter study suggested that SUCLG2 plays a key role in the regulation of POU1F1, which is well known to be involved in growth function, and belongs to the same family as POU2F1, detected in sheep by our study. The study by Schmidt et al.[37] in the blind subterranean mole rat, known to show remarkable tolerance to hypoxia and cancer resistance, also highlighted SUCLG2 and POU1F1 in the liver transcriptome. It was hypothesized that the energy-saving responses triggered in hepatic metabolic pathways may be crucial adaptations to low oxygen levels. Finally, the study of Tian et al.[38], assessing energy metabolism-related genes in hypoxia-tolerant mammals, identified SUCLG2 as candidate gene of interest in the liver of cetaceans.

SOX2 and DPH6 were found correlated with the environmental gradient in goats by LFMM. In mice, SOX2 is expressed in adult SCN neurons and positively regulates transcription of the core clock gene, Period2, implicated in behavioral rhythms linked to environmental light cycles[39]. A link between SOX2 and cold adaptation has been found in marmots[40]. Interestingly, BBX, another Sox protein that belongs to the HMG box superfamily of DNA-binding proteins, was also identified in sheep in our study. DPH6 belongs to the circadian rhythm-related GO categories. Moreover it was found under selection in northern European cattle[41] and yaks[42], and associated with plateau conditions in Chinese sheep[9].

It is interesting to note that the TRPC4 gene, on chromosome12, and the TRPC6 gene, on chromosome 15, are among the strongest signatures highlighted by the LFMM approach. Transient receptor potential channel (TRPC) proteins have been characterized as molecular substrates mediating receptor-activated cation influx. TRPC4 and TRPC6, in particular, have been shown to strongly contribute to synaptic information transfer in neuronal dendrites via the Ca^{2+}-dependent release of neurotransmitter[43]. The increase in dendritic γ-aminobutyric acid (GABA) release from thalamic interneurons appears critically dependent on these TRP proteins[44]. Hence, these genes may mediate hot and cold sensation and affect endothelial-dependent regulation of vascular tone[45]. Moreover, there is strong evidence for an important function of TRPC6 in pulmonary vasculature[46] and a functional role in hypoxia induction, potentially via the metabolism regulation of HIF-1α[31, 32, 47, 48]. In addition, TRPC6 was associated with high altitude adaptation in Tibetan highlanders[49] and TRPC4 was implicated in body temperature regulation in cattle[50].

Sheep selection signatures. Both PCAdapt and LFMM approaches identified the gene MSRB3. This gene has been subject to high selection pressure in sheep[51] and has been suggested as a functional candidate for associations with ear morphology in numerous studies. It has been reported in large-eared sheep[52]. Its link with ears was also highlighted in pigs[53, 54] and dogs[55]. Our study showed a correlation between SNPs near this gene

and the environmental gradient. Interestingly, a recent study[56] associated genetic variation in this gene and fat deposition in sheep. Furthermore, Webster et al.[55] revealed that this gene was in genetic linkage with variants in HMGA2, a neighbouring gene, which influences body mass in dogs. Moreover, HMGA2 is known to be involved in adipose tissue, development and obesity in mouse[57]. This selection signature, for which the role of MSRB3 and HGMA2 is not yet elucidated, appears to be of central importance in terms of adaptation, because it was identified under selection in highland Ethiopian sheep[7], in Tibetan sheep[8], in Tibetan Yaks[35] and in Tibetan dogs[58].

Like MSRB3, both methods (PCAdapt and LFMM) implicated BMP2, which initiates osteoblast and adipocyte differentiation. It has previously been found in association with dry conditions in Chinese[9] and Egyptian sheep[10]. It was also suggested to be under selection at high altitude, in Tibetan sheep[8] and in yaks[35]. Adaptive mechanisms governed by BMP2 could be linked to lipid storage capacities, particularly at the level of the tail[56, 59, 60].

Finally, RXFP2, well known to be under strong selection due to its role in horn development[26, 51, 61, 62], was identified by both PCAdapt and LFMM approaches (see details in Supplementary Text 3).

Selection signatures highlighted in both species. An interesting result concerns the area between the PCDH9 and KLHL1 genes that was identified as a selection signature in both sheep and goats. This region appears to be a genetic desert that extends over a syntenic segment conserved on bovine chromosome 12, goat chromosome 12 and sheep chromosome 10. Kim et al.[10] also found this genomic region to show evidence of selection in both sheep and goats that are adapted to arid environments, and suggested a major role for this region. PCDH9 has been found to be implicated in autism disorders[63]. Protocadherins are thought to be implicated in various aspects of neuronal development and functions. PCDH9 in particular is involved in synaptic cell adhesion[64, 65]. KLHL1 was found related to neuron motion and neuromuscular process[66] and it may play a role in organizing the actin cytoskeleton of the brain cells.

Another selection signature common to both species and most interesting is the neurobeachin signal (NBEA in goat, and NBEAL1 in sheep). In the present study, LFMM revealed a correlation of NBEA and the environmental gradient, in both goats and sheep; in goat it was also identified by PCAdapt. NBEA was previously reported to be associated with high altitude in Ethiopian sheep[7], Chinese sheep[9], cattle[67], and yaks[35]. Furthermore, it was previously associated with body temperature regulation in cattle[50] and was suggested to be under selection in Ugandan and Moroccan goats[68, 69]. NBEA is a brain specific A-kinase anchor protein (AKAP), which is required for synaptic surface expression of glutamate and GABA receptors. Neurons lacking the BEACH (beige-Chediak/Higashi) domain protein Neurobeachin (NBEA) show strongly reduced synaptic responses caused by a reduction in surface levels of glutamate and GABAA receptors. Hence, NBEA plays an essential role in thermal adaptation through the regulation of synaptic transmission[70, 71].

Interestingly, Alberto et al.[72] identified NBEA, and also, HERC6 and SLC34A2, all detected in the current study, as involved in selective sweeps that differentiate domestic from wild sheep and goat populations, possibly indicating predomestic selection on these genomic areas.

Finally, the signature near NBEAL1 is bordered in sheep by BMPR2, for which mutations have been associated with high Altitude Pulmonary Hypertension (APH) in Kyrgyz Highlanders[73] and in cattle[74]. Moreover, this gene was found to be associated with desert and plateau conditions in Chinese sheep[9]. BMP2, which belongs to the TGF-β superfamily and is able to activate BMPR2, was observed as a major signal in sheep in this study (see above). Hypoxia may act on BMPR2 activity[75] and on BMP2 signaling in the pulmonary vasculature[76].

Conclusion

Our study provides an approach for integrating ecological, historical and cultural approaches in the search for selection signatures in small ruminants, which as domestic animals are at the very interface of natural and anthropized environments. One of the key step is the characterization of the environment at the cradle level, thus targeting the geographic area, including transhumant summer pastures, where the breed has evolved by traditional practices. Local breeds are derived from populations selected for centuries and sometimes millennia and have thus developed long-standing links between genome and environment. In current times, local breeds are increasingly neglected and have become endangered, or in some cases have even disappeared. For those that remain, traditional practices, including the transhumance that has been one of the main driving forces in their development, are being abandoned[16, 17]. This study emphasizes that local breeds are invaluable resources of identifying environmental adaptation in the context of climate changes.

Material and methods

Methodology. A summary diagram of the main steps of the proposed approach is shown in Supplementary Figure 1.

Selection of breeds. The aim was to keep breeds with the highest probability of having retained strong adaptations to the local environment, i.e. breeds least affected by the disruptions inherent to the intensification of agricultural practices. For the study, we have targeted the Mediterranean arc in France, Italy and Spain. This region has the advantage of harboring closely related sheep and goat breeds despite large environmental contrasts[11–13]. We initially, considered 32 breeds of goats and 51 breeds of sheep (Supplementary Table 1). Subsequently, we selected 17 goat breeds and 25 sheep breeds (Figs. 1 and 2, and Table 1) on the basis of historical and anthropological documentation and according to the following criteria: (i) the breeds are local (i.e. cosmopolitan breeds were excluded); (ii) the breeds have featured in the local tradition for at least one century; (iii) any recorded introgression of exotic breeds was limited and occurred more than 100 years ago. If available, recent evaluations of admixture were also used to select breeds to include (see[25, 26–29, 77]). Outline descriptions of the selected breeds are in Supplementary Table 2.

Geographic definition of cradle of origin. We determined cradles of origin on the basis of historical and anthropological documentation. Particular attention has been paid to transhumance practices (see Figs. 1 and 2 and Supplementary Table 2 for details). We used ArcGIS[78] to map the cradles in such a way as to include areas of summer pastures when so-called vertical transhumance was recorded. The horizontal transhumance, involving movements covering long distances without significant altitude differences[79], cannot be directly taken into account because a large horizontal range would complicate the definition of the cradles.

Environmental characterization. Bioclimatic data were obtained from the WorldClim database (v.1.4[80]) covering the period from 1960 to 1990, with a spatial resolution of 30 arc-seconds in the WGS84 datum. We considered the bioclimatic variables most relevant to highlight the major contrasts in temperature and humidity between the cradles, i.e.: BIO_1 = Annual Mean Temperature, BIO_5 = Maximum Temperature of Warmest Month, BIO_6 = Minimum Temperature of Coldest Month, BIO_{12} = Annual Precipitation, BIO_{13} = Precipitation of Wettest Month, BIO_{14} = Precipitation of Driest Month, BIO_{15} = Precipitation Seasonality (Coefficient of Variation). Altitude information was collected from the SRTM 90 m Digital Elevation Database (v.4.1)[81]. Finally, latitude was taken as a proxy for luminosity and seasonality. This procedure was performed in R 3.5.2[82] using the R package RSAGA[83].

Because our objective was not to estimate allele-environment correlations between each SNP and each environmental variable at a time, but to capture the relation between the genome and the environment considered as a whole, the cradles were clustered according to their climatic profiles, following this procedure:

For each cradle, the environmental variables of 10,000 random location points were extracted, using R packages sp[84, 85] and rgdal[86].

In order to identify groups of breeds with relatively similar environmental conditions in their cradle area, principal component analysis (PCA) was performed on the environmental variables. Prior to the analysis, Spearman correlation coefficient identified highly correlated variables (i.e. correlation coefficient r ≥ 0.9), in which case the biologically less relevant variable was removed to facilitate graphical presentation. The PCA analysis was followed by hierarchical clustering on principal components (HCPC), using Euclidian distances and Ward's method. These multivariate analyses were performed with R software, using the FactoMineR package[87].

The LFMM[19] analyses were conducted following the cluster ranking of the cradles obtained through the PCA/HCPC procedure. Indeed, this procedure, allows classifying the cradles by considering their climatic resemblance defined by the set of all the recorded climatic variables. This ranking is then used to establish the link between genome and environment by LFMM, in order to optimize the statistical power of the analysis.

Genotyping. For goats, we used the AdaptMap dataset, including breeds genotyped with the Caprine SNP50 BeadChip (available via Dryad: https://doi.org/10.5061/dryad.v8g21pt), which contains genotypes for 53,547 SNPs. For sheep, we merged a dataset of French breeds obtained with the Illumina Ovine HD SNP Beadchip (Zenodo repository https://doi.org/10.5281/zenodo.237116) with a dataset of Italian breeds ([77], provided by the authors) and a dataset of Spanish breeds ([27], provided by the authors), obtained with the Illumina Ovine SNP50 BeadChip. SNP data from French breeds were extracted from the 600 K variation using the Ovine SNP50 BeadChip coordinates of SNPs on the OAR v3.1 reference genome assembly using Vcftools[88]. Merging with the SNP50 data resulted in 40,455 genotypes. SNPs and animals were pruned with PLINK v1.07[89] using the following filtering thresholds: (i) SNP call rate ≤ 97%; (ii) SNP minor allele frequency (MAF) ≤ 1%; (iii) animals displaying ≥ 10% of missing genotypes. After filtration of the merged datasets we retained 50,329 genotypes for 416 goats and 32,168 genotypes for 555 sheep (initial datasets included 420 and 576 individuals, respectively).

Identification and analysis of genomic selection signatures. We identified loci that may contribute to local adaptation by two methods: (i) the individual-based multivariate approach PCAdapt[18], via the pcadapt R package, and (ii) the latent factor mixed models LFMM, as implemented in the R package LEA[90], for testing genotype-environment association.

PCAdapt scans the genome to detect outliers with respect to the population structure. Unlike population-based approaches, the package can incorporate admixed individuals and does not assume prior knowledge of population structure. As recommended by Luu et al.[18], the optimal number of principal components (i.e., K, the optimal number of genetic groups) was determined using the graphical PCAdapt function, by varying K from 1 to 35, and following "Cattell's rule", i.e., keeping PCs that correspond to eigenvalues to the left of the lower straight line in the screeplot[91]. This led us to choose K = 15 for goats and K = 20 for sheep (see Supplementary Figure 6). Candidate SNPs were identified by calculating the False Discovery Rate (FDR, α = 0.05) of the *p* values associated with Mahalanobis distance estimated by PCAdapt, using the R package q value[92].

LFMM software assesses associations between genetic variation (response variable) and environmental factors (explanatory variables) using a linear mixed model and controlling for neutral genetic structure, such as population history and isolation-by-distance, with (random) latent factors via an MCMC algorithm. The number of latent factors was chosen in accordance with the Admixture[93] analyses. We ran LFMM using 1000 sweeps for burn-in and 9000 additional sweeps. Since LFMM uses a stochastic algorithm, ten runs with different seeds were performed. We combined the Z-scores from the ten repetitions and computed the median. We then recalibrated Z-scores by dividing by the inflation factor before computing the *p* values. We then chose significant associations based on FDR (α = 0.05) using the R package qvalue. LFMM was used to test for the correlation between the genetic variations and the environment taken as a whole (i.e. the variables were not considered one by one, except to test the method), but the analysis was performed once, analyzing the correlation between genomic variations and the ranking obtained by the PCA/HPCP clustering of the cradles environments.

For both PCAdapt and LFMM, we applied a stringent screen to identify selection signatures: candidate-selected regions were required to have at least 3 SNPs ≤ 500 kb apart, previously identified based on FDR and showing p value ≤ 10^{-9}. The window was chosen on the basis of previous evidence that LD in sheep and goats does not persist beyond 500 kb (see Supplementary Figure 2). It should be noted that the size of the window considered is widely used and that the selection criteria retained are among the most stringent (see[53, 54, 94–98]). For each analysis, genes within a region spanning 100 kb upstream and downstream of the candidate selection regions, were annotated in such a way that genes including one or more SNPs identified as selected were specifically studied, but genes bordered by SNPs identified as selected were also considered. The chromosomal regions under selection pressure were inspected using the NCBI Genome Data Viewer (https://www.ncbi.nlm.nih.gov/genome/gdv/), Oar_v3.1 for sheep and ARS1 for goats. Biological functions of the genes identified were inferred from the literature. Studies focusing on environmental adaptation signatures involving the same genes as those identified here were compiled. As top candidate genes for adaptation, we considered genes identified by our study and by at least one other study focusing on environmental adaptation.

Experimental test of the method. We compared the cradle-based method with the use of the individual GPS coordinates. This comparison was carried out only for the goat dataset, for which the AdaptMap Consortium provided us with the geographical locations for all breeds considered.

For each GPS coordinate, the environmental variables were extracted for the circular region of radius of 5 km centered on the GPS point. We considered the most discriminating variables, Annual Mean Temperature, Annual Mean Precipitation and Altitude, in order to compare the average information obtained by breed from the GPS coordinates with that obtained from the cradles.

The LFMM analysis proceeded as follows (see Fig. 3):

- The individual GPS method: we assessed the LFMM correlation between individual genotypes and the variables Annual Mean Temperature, Annual Mean Precipitation and Altitude, considered separately and extracted for each individual GPS coordinate.
- The GPS area method: we considered environmental data for the set of all individual GPS coordinates recorded for the breed. (i) The LFMM correlation was calculated between each genotype and the average value obtained for the different variables considered separately (Annual Mean Temperature, Annual Mean Precipitation and Altitude). (ii) As an alternative approach, in order to group the different GPS areas according to their similarity, we used all the environmental variables recorded at the set of the breed GPS coordinates and followed the PCA/HCPC process as explained in "Environmental characterization" section. Then we tested LFMM correlation between the genotypes and the resulting cluster ranking.
- The novel cradle method: For each breed, we considered environmental data at the cradle level. The LFMM correlation was tested in two steps, as for the GPS area method.

Complementary analyses. *LD decay.* To provide an insight into the overall levels of LD in the different breeds, genome-wide pairwise r^2 values of SNPs separated by a maximum distance of 10 Mb, were calculated with PLINK software.

Admixture analyses. To assess population genetic structure we used ADMIXTURE software[99]. Prior to these analyses, SNP pruning was used to select a subset of SNPs with minimal linkage disequilibrium via the—indep option of PLINK with the following parameters: 50 SNPs per window, a shift of five SNPs between windows, and a Variation Inflation Factor threshold of two (corresponding to $r^2 > 0.5$). SNPs identified by the study as potentially selected were removed. For K = 2 to K = 25, ten independent runs were performed. The entropy criterion was calculated via the sNMF function implemented in the R package LEA to assess the number of ancestral populations that best explains the genotypic data[100, 101]. We used the program CLUMPAK ([102], http://clumpak.tau.ac.il) to analyze the multiple independent runs at a single K and visualize the results.

Mantel test. Finally, we assessed the correlation between genetic distances and geographic distances using the Mantel test via the mantel.randtest function, implemented in the R adegenet package[103].

Data availability
For goats, we used the AdaptMap dataset, shared on Dryad: https://doi.org/10.5061/dryad.v8g21pt. For sheep, we used the French dataset online at the Zenodo repository https://doi.org/10.5281/zenodo.237116, and the Italian dataset partially available in the Figshare repository via https://doi.org/10.23644/uu.8947346 (for the few missing breeds please contact Elena Ciani: elena.ciani@uniba.it and Johannes A. Lenstra: j.a.lenstra@uu.nl).

References
1. Vigne, J.-D. Early domestication and farming: What should we know or do for a better understanding?. *Anthropozoologica* **50**(2), 123–150. https://doi.org/10.5252/az2015n2a5 (2015).
2. Zeder, M. A. Animal domestication in the Zagros: An update and directions for future research. *MOM Édit.* **49**(1), 243–277 (2008).

3. Sponenberg, D. P. & Bixby, D. E. *Managing Breeds for a Secure Future: Strategies for Breeders and Breed Associations* (ALBC, 2007).

4. Taberlet, P. *et al.* Are cattle, sheep, and goats endangered species?. *Mol. Ecol.* **17**(1), 275–284. https://doi.org/10.1111/j.1365-294X.2007.03475.x (2008).

5. Berihulay, H., Abied, A., He, X., Jiang, L. & Ma, Y. Adaptation mechanisms of small ruminants to environmental heat stress. *Anim. Open Access J. MDPI* **9**(3), 75. https://doi.org/10.3390/ani9030075 (2019).

6. Leroy, G., Baumung, R., Boettcher, P., Scherf, B. & Hoffmann, I. Review: Sustainability of crossbreeding in developing countries; definitely not like crossing a meadow.... *Animal* **10**(2), 262–273. https://doi.org/10.1017/S175173111500213X (2016).

7. Edea, Z., Dadi, H., Dessie, T. & Kim, K.-S. Genomic signatures of high-altitude adaptation in Ethiopian sheep populations. *Genes Genomics* **41**(8), 973–981. https://doi.org/10.1007/s13258-019-00820-y (2019).

8. Wei, C. *et al.* Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. *Sci. Rep.* **6**(1), 26770. https://doi.org/10.1038/srep26770 (2016).

9. Yang, J. *et al.* Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.* **33**(10), 2576–2592. https://doi.org/10.1093/molbev/msw129 (2016).

10. Kim, E. S. *et al.* Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* **116**(3), 255–264. https://doi.org/10.1038/hdy.2015.94 (2016).

11. Ciani, E. *et al.* On the origin of European sheep as revealed by the diversity of the Balkan breeds and by optimizing population-genetic analysis tools. *Genet. Sel. Evol. GSE* **52**, 1–14. https://doi.org/10.1186/s12711-020-00545-7 (2020).

12. Colli, L. *et al.* Genome-wide SNP profiling of worldwide goat populations reveals strong partitioning of diversity and highlights post-domestication migration routes. *Genet. Sel. Evol. GSE* **50**, 1–20. https://doi.org/10.1186/s12711-018-0422-x (2018).

13. Kijas, J. W. *et al.* A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. *PLoS ONE* **4**(3), e4668. https://doi.org/10.1371/journal.pone.0004668 (2009).

14. Brisebarre, A. *Races ovines, systèmes d'élevage et représentations des éleveurs.* in *Développement rural, environnement et enjeux territoriaux. Regards croisés Oriental marocain et Sud-Est tunisien* (dir. Bonte, P., Elloumi, M., Guillaume, H. & Mahdi, M.) 63–78 (Cérès Ed., 2009).

15. Hall, S. J. G. Livestock biodiversity as interface between people, landscapes and nature. *People Nat.* **1**(3), 284–290. https://doi.org/10.1002/pan3.23 (2019).

16. Caballero, R. *et al. Grazing Systems and Biodiversity in Mediterranean Areas: Spain, Italy and Greece* (Pastos, 2011).

17. Collantes, F. The demise of European Mountain Pastoralism: Spain 1500–2000. *Nomadic People* **13**(2), 124–145 (2009).

18. Luu, K., Bazin, E. & Blum, M. G. B. pcadapt: An R package to perform genome scans for selection based on principal component analysis. *Mol. Ecol. Resour.* **17**(1), 67–77. https://doi.org/10.1111/1755-0998.12592 (2017).

19. Frichot, E., Schoville, S. D., Bouchard, G. & François, O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol. Biol. Evol.* **30**(7), 1687–1699. https://doi.org/10.1093/molbev/mst063 (2013).

20. FAO. The State of the World's Animal Genetic Resources for Food and Agriculture, edited by B. Rischkowsky & D. Pilling. Rome. (2007).

21. François, O. *Running Structure-Like Population Genetic Analyses with R. R Tutorials in Population Genetics* 1–9 (U. Grenoble-Alpes, 2016).

22. Dalongeville, A., Benestan, L., Mouillot, D., Lobreaux, S. & Manel, S. Combining six genome scan methods to detect candidate genes to salinity in the Mediterranean striped red mullet (*Mullus surmuletus*). *BMC Genomics* **19**, 1–13. https://doi.org/10.1186/s12864-018-4579-z (2018).

23. De Kort, H., Vandepitte, K., Mergeay, J., Mijnsbrugge, K. V. & Honnay, O. The population genomic signature of environmental selection in the widespread insect-pollinated tree species *Frangula alnus* at different geographical scales. *Heredity* **115**(5), 415–425. https://doi.org/10.1038/hdy.2015.41 (2015).

24. Capblancq, T., Luu, K., Blum, M. G. B. & Bazin, E. Evaluation of redundancy analysis to identify signatures of local adaptation. *Mol. Ecol. Resour.* **18**(6), 1223–1233. https://doi.org/10.1111/1755-0998.12906 (2018).

25. Bertolini, F. *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* **50**(1), 57. https://doi.org/10.1186/s12711-018-0421-y (2018).

26. Fariello, M.-I. *et al.* Selection signatures in worldwide sheep populations. *PLoS ONE* **9**(8), e103813. https://doi.org/10.1371/journal.pone.0103813 (2014).

27. Manunza, A. *et al.* Population structure of eleven Spanish ovine breeds and detection of selective sweeps with BayeScan and hapFLK. *Sci. Rep.* **6**(1), 1–10. https://doi.org/10.1038/srep27296 (2016).

28. Oget, C., Servin, B. & Palhière, I. Genetic diversity analysis of French goat populations reveals selective sweeps involved in their differentiation. *Anim. Genet.* **50**(1), 54–63. https://doi.org/10.1111/age.12752 (2019).

29. Rochus, C. M. *et al.* Revealing the selection history of adaptive loci using genome-wide scans for selection: An example from domestic sheep. *BMC Genomics* **19**(1), 71. https://doi.org/10.1186/s12864-018-4447-x (2018).

30. Ruiz-Larrañaga, O. *et al.* Genomic selection signatures in sheep from the Western Pyrenees. *Genet. Sel. Evol. GSE* **50**, 1–12. https://doi.org/10.1186/s12711-018-0378-x (2018).

31. Wang, Q., Wang, D., Yan, G., Sun, L. & Tang, C. TRPC6 is required for hypoxia-induced basal intracellular calcium concentration elevation, and for the proliferation and migration of rat distal pulmonary venous smooth muscle cells. *Mol. Med. Rep.* **13**(2), 1577–1585. https://doi.org/10.3892/mmr.2015.4750 (2016).

32. Wang, X. *et al.* Whole-genome sequencing of eight goat populations for the detection of selection signatures underlying production and adaptive traits. *Sci. Rep.* **6**, 38932. https://doi.org/10.1038/srep38932 (2016).

33. Graae, B. *et al.* On the use of weather data in ecological studies along altitudinal and latitudinal gradients. *Oikos* **121**, 3–19. https://doi.org/10.1111/j.1600-0706.2011.19694.x (2011).

34. Rellstab, C., Gugerli, F., Eckert, A. J., Hancock, A. M. & Holderegger, R. A practical guide to environmental association analysis in landscape genomics. *Mol. Ecol.* **24**(17), 4348–4370. https://doi.org/10.1111/mec.13322 (2015).

35. Qi, X. *et al.* The transcriptomic landscape of yaks reveals molecular pathways for high altitude adaptation. *Genome Biol. Evol.* **11**(1), 72–85. https://doi.org/10.1093/gbe/evy264 (2019).

36. Yang, F., Wang, Q., Wang, M., He, K. & Pan, Y. Associations between gene polymorphisms in two crucial metabolic pathways and growth traits in pigs. *Chin. Sci. Bull.* **57**(21), 2733–2740. https://doi.org/10.1007/s11434-012-5328-3 (2012).

37. Schmidt, H. *et al.* Hypoxia tolerance, longevity and cancer-resistance in the mole rat Spalax—A liver transcriptomics approach. *Sci. Rep.* https://doi.org/10.1038/s41598-017-13905-z (2017).

38. Tian, R. *et al.* Adaptive evolution of energy metabolism-related genes in hypoxia-tolerant mammals. *Front. Genet.* **8**, 205. https://doi.org/10.3389/fgene.2017.00205 (2017).

39. Cheng, A. H. *et al.* SOX2-dependent transcription in clock neurons promotes the robustness of the central circadian pacemaker. *Cell Rep.* **26**(12), 3191-3202.e8. https://doi.org/10.1016/j.celrep.2019.02.068 (2019).

40. Bai, L. *et al.* Hypoxic and cold adaptation insights from the Himalayan Marmot Genome. *IScience* **11**, 519–530. https://doi.org/10.1016/j.isci.2018.11.034 (2019).

41. Stronen, A. V., Pertoldi, C., Iacolina, L., Kadarmideen, H. N. & Kristensen, T. N. Genomic analyses suggest adaptive differentiation of northern European native cattle breeds. *Evol. Appl.* https://doi.org/10.1111/eva.12783 (2019).

42. Lan, D. *et al.* Genetic diversity, molecular phylogeny, and selection evidence of Jinchuan Yak revealed by whole-genome rese-quencing. *G3 (Bethesda, Md.)* **8**(3), 945–952. https://doi.org/10.1534/g3.118.300572 (2018).

43. Chen, J. *et al.* Deletion of TRPC6 attenuates NMDA receptor-mediated Ca^{2+} entry and Ca^{2+}-induced neurotoxicity following cerebral ischemia and oxygen-glucose deprivation. *Front. Neurosci.* **11**, 138. https://doi.org/10.3389/fnins.2017.00138 (2017).

44. Munsch, T., Freichel, M., Flockerzi, V. & Pape, H.-C. Contribution of transient receptor potential channels to the control of GABA release from dendrites. *Proc. Natl. Acad. Sci. U. S. A.* **100**(26), 16065–16070. https://doi.org/10.1073/pnas.2535311100 (2003).

45. Duan, J. *et al.* Structure of the mouse TRPC4 ion channel. *Nat. Commun.* **9**, 1–10. https://doi.org/10.1101/282715 (2018).

46. Malczyk, M. *et al.* The role of transient receptor potential channel 6 channels in the pulmonary vasculature. *Front. Immunol.* **8**, 707. https://doi.org/10.3389/fimmu.2017.00707 (2017).

47. Li, S. *et al.* Crucial role of TRPC6 in maintaining the stability of HIF-1α in glioma cells under hypoxia. *J. Cell Sci.* **128**(17), 3317–3329. https://doi.org/10.1242/jcs.173161 (2015).

48. Xu, L. *et al.* Chronic hypoxia increases TRPC6 expression and basal intracellular Ca^{2+} concentration in rat distal pulmonary venous smooth muscle. *PLoS ONE* **9**(11), e112007. https://doi.org/10.1371/journal.pone.0112007 (2014).

49. Deng, L. *et al.* Prioritizing natural-selection signals from the deep-sequencing genomic data suggests multi-variant adaptation in Tibetan highlanders. *Natl. Sci. Rev.* **6**(6), 1201–1222. https://doi.org/10.1093/nsr/nwz108 (2019).

50. Howard, J. T. *et al.* Beef cattle body temperature during climatic stress: A genome-wide association study. *Int. J. Biometeorol.* **58**, 1665–1672. https://doi.org/10.1007/s00484-013-0773-5 (2013).

51. Kijas, J. W. *et al.* Genome-wide analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. *PLoS Biol.* **10**(2), e1001258. https://doi.org/10.1371/journal.pbio.1001258 (2012).

52. Wei, C. *et al.* Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* **16**(1), 1–12. https://doi.org/10.1186/s12864-015-1384-9 (2015).

53. Chen, M. *et al.* Genome-wide detection of selection signatures in Chinese indigenous Laiwu pigs revealed candidate genes regulating fat deposition in muscle. *BMC Genet.* **19**, 1–9. https://doi.org/10.1186/s12863-018-0622-y (2018).

54. Chen, C. *et al.* Copy number variation in the MSRB3 gene enlarges porcine ear size through a mechanism involving miR-584-5p. *Genet. Sel. Evol. GSE* **50**, 1–18. https://doi.org/10.1186/s12711-018-0442-6 (2018).

55. Webster, M. T. *et al.* Linked genetic variants on chromosome 10 control ear morphology and body mass among dog breeds. *BMC Genomics* **16**, 474. https://doi.org/10.1186/s12864-015-1702-2 (2015).

56. Mastrangelo, S. *et al.* Novel and known signals of selection for fat deposition in domestic sheep breeds from Africa and Eurasia. *PLoS ONE* **14**(6), e0209632. https://doi.org/10.1371/journal.pone.0209632 (2019).

57. Xi, Y. *et al.* HMGA2 promotes adipogenesis by activating C/EBPβ-mediated expression of PPARγ. *Biochem. Biophys. Res. Commun.* **472**(4), 617–623. https://doi.org/10.1016/j.bbrc.2016.03.015 (2016).

58. Gou, X. *et al.* Whole-genome sequencing of six dog breeds from continuous altitudes reveals adaptation to high-altitude hypoxia. *Genome Res.* **24**(8), 1308–1315. https://doi.org/10.1101/gr.171876.113 (2014).

59. Yuan, Z. *et al.* Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Anim. Genet.* **48**(1), 55–66. https://doi.org/10.1111/age.12477 (2017).

60. Zhu, C. *et al.* *Identification of Genes Associated with Sheep Tail Fat Deposition*. Retrieved from https://www.preprints.org/manuscript/201906.0093/v1 (2019).

61. Allais-Bonnet, A. *et al.* Novel insights into the bovine polled phenotype and horn ontogenesis in Bovidae. *PLoS ONE* **8**(5), e63512. https://doi.org/10.1371/journal.pone.0063512 (2013).

62. Johnston, S. E. *et al.* Genome-wide association mapping identifies the genetic basis of discrete and quantitative variation in sexual weaponry in a wild sheep population. *Mol. Ecol.* **20**(12), 2555–2566. https://doi.org/10.1111/j.1365-294X.2011.05076.x (2011).

63. Oksenberg, N., Stevison, L., Wall, J. D. & Ahituv, N. Function and regulation of AUTS2, a gene implicated in autism and human evolution. *PLoS Genet.* **9**(1), e1003221. https://doi.org/10.1371/journal.pgen.1003221 (2013).

64. Hayashi, S. & Takeichi, M. Emerging roles of protocadherins: From self-avoidance to enhancement of motility. *J. Cell Sci.* **128**(8), 1455–1464. https://doi.org/10.1242/jcs.166306 (2015).

65. Seong, E., Yuan, L. & Arikkath, J. Cadherins and catenins in dendrite and synapse morphogenesis. *Cell Adhes. Migr.* **9**(3), 202–213. https://doi.org/10.4161/19336918.2014.994919 (2015).

66. Shin, D.-H. *et al.* Deleted copy number variation of Hanwoo and Holstein using next generation sequencing at the population level. *BMC Genomics* **15**(1), 240. https://doi.org/10.1186/1471-2164-15-240 (2014).

67. Zeng, X. *Angus Cattle at High Altitude: Pulmonary Arterial Pressure, Estimated Breeding Value and Genome-Wide Association Study* (PhD thesis). (Colorado State University, 2017).

68. Benjelloun, B. *Diversité des génomes et adaptation locale des petits ruminants d'un pays méditerranéen : le Maroc* (PhD thesis) (Université Grenoble Alpes, France, 2015).

69. Onzima, R. B. *et al.* Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan Goat Breeds. *Front. Genet.* **9**, 318. https://doi.org/10.3389/fgene.2018.00318 (2018).

70. Farzana, F. *et al.* Neurobeachin regulates glutamate- and GABA-receptor targeting to synapses via distinct pathways. *Mol. Neurobiol.* **53**(4), 2112–2123. https://doi.org/10.1007/s12035-015-9164-8 (2016).

71. Nair, R. *et al.* Neurobeachin regulates neurotransmitter receptor trafficking to synapses. *J. Cell Biol.* **200**(1), 61–80. https://doi.org/10.1083/jcb.201207113 (2013).

72. Alberto, F. J. *et al.* Convergent genomic signatures of domestication in sheep and goats. *Nat. Commun.* **9**, 1–9. https://doi.org/10.1038/s41467-018-03206-y (2018).

73. Iranmehr, A. *et al.* Novel insight into the genetic basis of high-altitude pulmonary hypertension in Kyrgyz highlanders. *Eur. J. Hum. Genet. EJHG* **27**(1), 150–159. https://doi.org/10.1038/s41431-018-0270-8 (2019).

74. Newman, J. H. *et al.* High-altitude pulmonary hypertension in cattle (Brisket disease): Candidate genes and gene expression profiling of peripheral blood mononuclear cells. *Pulmon. Circ.* **1**(4), 462–469. https://doi.org/10.4103/2045-8932.93545 (2011).

75. Yang, X., Kong, Q., Zhao, C., Cai, Z., & Wang, M. New pathogenic variant of BMPR2 in pulmonary arterial hypertension. *Cardiology in the Young*, **29**(4), 462–466. https://doi.org/10.1017/S1047951119000015 (2019).

76. Anderson, L. *et al.* Bmp2 and Bmp4 exert opposing effects in hypoxic pulmonary hypertension. *Am. J. Physiol. Regul. Integr. Comp. Physiol.* **298**(3), R833–R842. https://doi.org/10.1152/ajpregu.00534.2009 (2009).

77. Ciani, E. *et al.* Genome-wide analysis of Italian sheep diversity reveals a strong geographic pattern and cryptic relationships between breeds. *Anim. Genet.* **45**(2), 256–266. https://doi.org/10.1111/age.12106 (2014).

78. ESRI. *ArcGIS Desktop: Release 10* (Environmental Systems Research Institute, 2011).

79. Ruiz, M. & Ruiz, J. P. Ecological history of transhumance in Spain. *Biol. Conserv.* **37**, 73–86 (1986).

80. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).

81. Jarvis, A., Reuter, H. I., Nelson, A. & Guevara, E. Hole-filled seamless SRTM dataV4, International Centre for Tropical Agriculture (CIAT). Available from https://srtm.csi.cgiar.org (2008).

82. R Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2018). https://www.R-project.org.

83. Brenning, A. Statistical geocomputing combining R and SAGA: The example of landslide susceptibility analysis with generalized additive models. In *Hamburger Beitraege zur Physischen Geographie und Landschaftsoekologie* (eds Böhner, J. *et al.*) 23–32 (SAGA, 2008).

84. Bivand, R. S., Pebesma, E. & Gomez-Rubio, V. *Applied Spatial Data Analysis with R* 2nd edn (Springer, 2013). http://www.asdar-book.org/.

85. Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. *R News* **5**(2), 9–13. https://CRAN.R-project.org/doc/Rnews/ (2005).

86. Keitt, T. H., Bivand, R., Pebesma, E. & Rowlingson, B. rgdal: Bindings for the geospatial data abstraction library. Copy at http://www.tinyurl.com/h8w8n29 (2010).

87. Le, S., Josse, J. & Husson, F. FactoMineR: An R package for multivariate analysis. *J. Stat. Softw* **25**(1), 1–18. https://doi.org/10.18637/jss.v025.i01 (2008).

88. Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**(15), 2156–2158. https://doi.org/10.1093/bioinformatics/btr330 (2011).

89. Purcell, S. *et al.* PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**(3), 559–575. https://doi.org/10.1086/519795 (2007).

90. Frichot, E. & François, O. LEA: An R package for landscape and ecological association studies. *Methods Ecol. Evol.* **6**(8), 925–929. https://doi.org/10.1111/2041-210X.12382 (2015).

91. Cattell, R. B. The Scree plot test for the number of factors. *Multivar. Behav. Res.* **1**, 140–161 (1966).

92. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* **100**, 9440–9445. https://doi.org/10.1073/pnas.1530509100 (2003).

93. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**(6755), 788–791 (1999).

94. Ablondi, M., Viklund, Å., Lindgren, G., Eriksson, S. & Mikko, S. Signatures of selection in the genome of Swedish warmblood horses selected for sport performance. *BMC Genomics* **20**(1), 717. https://doi.org/10.1186/s12864-019-6079-1 (2019).

95. Avila, F., Mickelson, J. R., Schaefer, R. J. & McCue, M. E. Genome-wide signatures of selection reveal genes associated with performance in American Quarter Horse subpopulations. *Front. Genet.* **9**, 249. https://doi.org/10.3389/fgene.2018.00249 (2018).

96. Chen, M. *et al.* Identification of selective sweeps reveals divergent selection between Chinese Holstein and Simmental cattle populations. *Genet. Sel. Evol.* **48**(1), 76. https://doi.org/10.1186/s12711-016-0254-5 (2016).

97. Cheruiyot, E. K. *et al.* Signatures of selection in admixed dairy cattle in Tanzania. *Front. Genet.* **9**, 607. https://doi.org/10.3389/fgene.2018.00607 (2018).

98. López, M. E. *et al.* Multiple selection signatures in farmed Atlantic Salmon adapted to different environments across hemispheres. *Front. Genet.* **10**, 901. https://doi.org/10.3389/fgene.2019.00901 (2019).

99. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**(9), 1655–1664. https://doi.org/10.1101/gr.094052.109 (2009).

100. Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**, 246 (2011).

101. Frichot, E., Mathieu, F., Trouillon, T., Bouchard, G. & Francois, O. Fast and efficient estimation of individual ancestry coefficients. *Genetics* **196**, 973–983 (2014).

102. Kopelman, N. M., Mayzel, J., Jakobsson, M., Rosenberg, N. A. & Mayrose, I. Clumpak: A program for identifying clustering modes and packaging population structure inferences across K. *Mol. Ecol. Resour.* **15**(5), 1179–1191. https://doi.org/10.1111/1755-0998.12387 (2015).

103. Jombart, T. Adegenet: A R package for the multivariate analysis of genetic markers. *Bioinformatics* **24**, 1403–1405 (2008).

Acknowledgements

Author contributions

B.S. and M.C.: contributed to data analyses; P.V.: performed anthropological analyses; D.T.-M.: performed historical analyses; E.C., B.S., C.M.-R. and G.T.-K.: acquisition of data; M.B. and E.R.: contributed to the conception of the study; J.A.L.: acquisition of data/review and editing; S.J.G.H.: review and editing; F.P.: contributed to the conception of the study/review and editing; B.B.: contributed to data analyses/review and editing; A.D.S.: designed research/analyzed data/wrote the paper; all authors contributed to the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-021-00682-z.

Correspondence and requests for materials should be addressed to A.S.

Reprints and permissions information is available at www.nature.com/reprints.

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