

OPEN

Genomic scan of selective sweeps in Djallonké (West African Dwarf) sheep shed light on adaptation to harsh environments

Isabel Álvarez¹, Iván Fernández¹, Amadou Traoré², Lucía Pérez-Pardal³, Nuria A. Menéndez-Arias¹ & Félix Goyache^{1*} 

The Djallonké (West African Dwarf) sheep is a small-sized haired sheep resulting from a costly evolutionary process of natural adaptation to the harsh environment of West Africa including trypanosome challenge. However, genomic studies carried out in this sheep are scant. In this research, genomic data of 184 Djallonké sheep (and 12 Burkina-Sahel sheep as an outgroup) generated using medium-density SNP Chips were analyzed. Three different statistics (iHS, XP-EHH and nSL) were applied to identify candidate selection sweep regions spanning genes putatively associated with adaptation of sheep to the West African environment. A total of 207 candidate selection sweep regions were defined. Gene-annotation enrichment and functional annotation analyses allowed to identify three statistically significant functional clusters involving 12 candidate genes. Genes included in Functional Clusters associated to selection signatures were mainly related to metabolic response to stress, including regulation of oxidative and metabolic stress and thermotolerance. The bovine chromosomal areas carrying QTLs for cattle trypanotolerance were compared with the regions on which the orthologous functional candidate cattle genes were located. The importance of cattle BTA4 for trypanotolerant response might have been conserved between species. The current research provides new insights on the genomic basis for adaptation and highlights the importance of obtaining information from non-cosmopolite livestock populations managed in harsh environments.

Drift and selection may cause selective sweep signatures on the genome, characterized by reduced diversity and extensive linkage disequilibrium. Favorable alleles increase in frequency more rapidly than expected for a neutral genetic scenario and further alter the allele frequencies of genomic variants near the beneficial mutation. The identification of signatures of selection has been carried out on several sheep breeds using genome-wide single nucleotide polymorphism (SNP) data¹. Most analyses allowed to identify signatures of positive selection for economically important traits such as parasite resistance², dairy production^{3–5}, meat yield and growth^{4,6,7}, fat metabolism and deposition^{8–10} or reproduction^{11,12}. Other traits affecting breeds' definition such as morphology, horn shape or coat color have caused selective sweeps on the sheep genome as well^{6,12,13}.

However, there is an increasing interest in identifying selection signatures caused by human-mediated selection pressure for adaptation to different environmental conditions contrasting in climate (mainly precipitations and temperature), vegetation and altitude^{10,14–17}. This approach allows to identify sheep populations (or breeds) more robustly suited to future climate changes and to obtain new insights on the history of the species after domestication.

Domestication and dissemination of sheep is a complex process, initiated about 11,000 years before present (yBP), which included different waves, specialization for different production goals and adaptation to different environmental conditions¹⁸. When compared with Northern Africa, the initial appearance of pastoralist communities based on small stock in Sub-Saharan West Africa may have been delayed by thousand years¹⁹. Reviewing the literature, Muiga and Hanotte²⁰ suggested that domestic sheep entered into central Nile valley and central Sahara around 6,000 yBP reaching West Africa by 3,700 yBP. This lag in successful introduction of small ruminant

¹Servicio Regional de Investigación y Desarrollo Agroalimentario, E-33394, Gijón, Spain. ²Institut de l'Environnement et des Recherches Agricoles (INERA), Ouagadougou, 04 BP 8645, Burkina Faso. ³CIBIO - InBIO, Universidade do Porto, 4485-661, Vairão, Portugal. *email: fgoyache@serida.org

livestock into West Africa stems from new animal diseases encountered by pastoral colonists entering biogeographic zones south of the Sahel, namely trypanosomiasis¹⁹. Native West African domestic ruminants (cattle, sheep and goat) share distinctive traits, namely trypanotolerance and small size. These species were subject to similar processes of adaptation to an extremely difficult environmental scenario and to trypanosome challenge²⁰.

Djallonké (West African Dwarf) sheep^{21,22} are a small sized haired sheep population spread in the humid West and Central African regions. Djallonké sheep have the ability to maintain undepressed production and reproduction performance under parasite challenge even in the presence of a persistent parasitaemia. Therefore, they are considered resilient to trypanosome challenge²³.

Most studies aiming at the identification of selective sweeps on the sheep genome have been carried out using highly structured populations including different cosmopolitan (commercial) and local sheep breeds^{4,6,24}. Although there exist methodologies accounting for population structuring, such as hapFLK²⁵, genetic heterogeneity amongst breeds carrying distinct mutations with similar phenotypic effects may lead selective sweeps to not reach statistical significance, particularly when sample size is moderate⁴.

In this research, genomic data generated using medium-density Chips on a sample of Burkina Faso Djallonké sheep were assessed to identify candidate selection sweep regions spanning genes associated with different biological pathways putatively involved in the adaptation of sheep to the environmental conditions of the humid areas of West Africa. The genomic profile of this population, compared with that of a small sample of Burkina-Sahel sheep when necessary, was explored for signatures of selection using various Extended Haplotype Homozygosity (EHH-)based statistics²⁶. The corresponding cattle orthologous genomic regions were compared with the bovine chromosomal areas in which Quantitative Traits Loci (QTL) for trypanotolerance-related traits have been reported²⁷.

Methods

Sample used. A total of 184 DNA samples of Djallonké lambs (64 males and 120 females) previously analyzed using microsatellites²⁸ were available. Up to 166 individuals were sampled in the surroundings of Mangodara (latitude 9°53'59.99"N; longitude 4°20'59.99"W; Comoé province), southern Burkina Faso. The other 18 individuals (9 males and 9 females) were sampled in Dédougou (latitude 12°27'48.17"N; longitude 3°27'38.7"W; Mouhoun province), southwestern Burkina Faso. Both sampling areas belong to the environmental Sudan-Guinea Savannah humid region (tse-tse challenged) of Burkina Faso^{21,22}. The Sudan-Guinea Savannah environmental region has annual rainfall higher than 900 millimeters (with precipitations in Mangodara and Dédougou varying from 1,000 to 1,200 millimeters per year), predominance of woodlands and savannahs, and temperatures varying from 19 °C to 36 °C²⁹. Additionally, 12 individuals belonging to the Burkina-Sahel sheep breed were sampled in the surroundings of Dori (latitude 14°02'7.44"N; longitude 0°02'4.20"E; Séno Province), located on the arid environmental Sahel region of Northern Burkina Faso, to be used as an outgroup. The Sahel environmental region has less than 600 millimeters of rainfall per year, temperatures varying from 15 °C to 43 °C and bushy vegetation²⁹. Morphological and genetic description of both the Djallonké and the Burkina-Sahel sheep populations of Burkina Faso is available in the literature^{21,22,30}. The management of the Djallonké sheep of Burkina Faso has been described as well³¹. Sheep breeding in Burkina Faso is based on small-holders and carried out in extensive conditions with no supplementation. In the Sudan-Sahel environmental area, animals perform communal grazing of native pasture with no restrictions during the dry season and about 12 hours a day during the rainy season. In the Sahel area, management system is traditional and extensive with little or no management inputs: shelter or minerals. Sheep are managed by their owners (Fulani ethnic group). Communal grazing is less frequent. In all cases, use of fodder crops and conserved forages is negligible. Planned matings are not usual.

SNP genotyping and quality control. The whole dataset was typed using the Ovine 50 K SNP BeadChip following standard protocols (<http://www.illumina.com>). The software GenomeStudio (Illumina Inc., San Diego, CA) was used to generate standard.ped and map files. Sample and marker-based quality control measures were performed using the program PLINK V 1.09³². GenCall score cutoff of 0.15 and average sample call rate of 99% were considered. All unmapped SNPs, those mapping to sexual chromosomes, SNPs with a genotyping rate lower than 90% or those below a Minor Allele Frequency threshold of 0.05 were removed. To avoid departures from Hardy-Weinberg proportions due to genotyping errors, SNPs that did not pass Hardy-Weinberg test for $P \leq 0.001$ were removed as well. A total of 46,977 SNPs located on the 26 ovine autosomes passed the quality control for the whole sample analyzed.

Diversity and population structuring analyses. Genetic diversity of the available population was assessed using the software VCFtools³³. Within subpopulation (Djallonké and Burkina-Sahel) inbreeding and between subpopulations F_{ST} were computed.

A clustering analysis was carried out using the program Admixture v1.23^{34,35} which calculates maximum likelihood estimates of individual ancestries based on data provided by multiple loci. Analyses were conducted for $1 \leq K \leq 8$ being K the number of clusters given the data. The optimal number of clusters was determined via cross-validation as the value of K exhibiting the lower cross-validation error compared to other K values. Data set was divided into 5 folders for each K. Folders were sequentially used as test sets while the other four were used for training.

The program PLINK V 1.09³² was used to compute Principal Component Analysis (PCA). Eigenvectors computed for each individual were used to construct dispersion plots accounting for the Admixture Cluster to which individuals were assigned. The 75% confidence regions of the relationships among individuals within cluster were illustrated using contour plots. Plots were constructed using the library ggplot2 of R (<http://CRAN.R-project.org/>).

Identification of selective sweeps. Three complementary EHH-based statistics, integrated Haplotype Score³⁶ (iHS), XP-EHH³⁷ and nSL³⁸, were used to assess genome-wide signatures of selection in Djallonké sheep. These three approaches were proposed to overcome the main concern affecting the original EHH test²⁶ i.e. the yield of a high number of false positives due to the strong influence on results of the demographic history of the population under studied.

The iHS test compares the EHH between derived and ancestral alleles within a population searching for loci where the derived allele resides on a longer haplotype than the ancestral allele³⁶. In this way, the iHS values are less affected by the demographic history of the population and are more suitable for identifying incomplete sweeps, where the selected allele is not fixed in the sample.

The XP-EHH test compares the frequencies of the selected haplotypes between two populations to detect ongoing or nearly fixed selection signatures³⁷. Therefore, the XP-EHH test is more robust than the EHH statistics in scenarios of complete selective sweeps. This test has an increased power to detect selection signatures using small sample sizes and allows to use groups of genetically similar populations.

In contrast to iHS and XP-EHH, which measures the length of haplotypes in terms of genetic distance and thereby uses recombination, the nSL statistics³⁸ measures haplotype lengths in terms of the number of segregating sites in the sample, making it more robust to recombination rate variations. In addition, although both iHS and nSL target at the identification of incomplete sweeps, the nSL statistic has improved power in detecting soft sweeps³⁸.

Both large positive and negative iHS, XP-EHH and nSL scores were considered potentially informative. Although negative iHS scores would mean that a derived allele would have swept up in frequency and, therefore, could be considered of more interest, large positive values can indicate that ancestral alleles hitchhike with the selected site³⁹. Furthermore, scenarios in which selection switch to favor an ancestral allele cannot be discarded. Similar rationale can be applied to the recombination-free counterpart of iHS, the nSL test. In the case of XP-EHH scores large positive and negative values would identify selection events in the studied population and in the reference population, respectively. Both cases are informative in scenarios of poor differentiation.

All estimates were carried out using the program *selscan* v.1.0.4⁴⁰, freely available at <http://github.com/szpiech/selscan>, fitting the parameters recommended by the authors: maximum EHH extension in bp (“–max-extend” option) 1000000, maximum gap allowed between two SNPs in bp (“–max-gap” option) 200000, EHH decay cutoff (“–cutoff” option) 0.05. Whatever the statistics computed, the output results for each SNP were frequency-normalized over all chromosomes using the program *norm*, provided with *selscan*. This normalization was carried out using default parameters as well: number of frequency bins (“–bins” option) 100. As implemented by default in the program *norm*, normalized values higher than |2| were used to identify SNPs under selection.

The program *selscan* is ‘dumb’ with respect ancestral/derived alleles, simply assigning the “reference” allele depending on previous arbitrary allele coding⁴⁰. Unstandardized iHS scores are then computed as $\ln(iHH1/iHH0)$ ⁴⁰, being iHH the integrated haplotype homozygosity for the ancestral (0) and derived (1) haplotypes. There is empirical evidence suggesting that iHS values obtained from a given subset of SNPs with or without ancestral allele information (i.e. random assignment) are highly consistent⁴¹. Therefore, this has been proved to be a reliable strategy.

Previous studies aiming at the ascertainment of the genomic diversity in Djallonké sheep used as reference population Sahel sheep sampled in the same geographical area⁴². Therefore, Burkina-Sahel individuals were used for the XP-EHH analyses.

Candidate selection sweep regions were defined starting from: a) individual SNPs identified as being under selection by at least two of the statistics applied; and b) adjacent SNPs identified as being under selection by at least two different statistics. Two SNPs were considered adjacent if their 75 kb up- or down-stream regions overlapped. All other SNPs identified as being under selection by any statistics with their 75 kb up- or down-stream regions sequentially intersecting with the 75 kb regions surrounding the SNPs selected using the strategies a) and b) described above were assigned to the corresponding selection sweep region.

Functional characterization of the candidate regions. Candidate genes were considered if their boundaries fell within 75 kb up- or down-stream the selection sweep regions defined. Protein-coding genes found within the candidate regions were retrieved from the Ensembl Genes 91 database, based on the Oar v3.1 ovine reference genome (<http://www.livestockgenomics.csiro.au/sheep/oar3.1.php>) using the BioMart tool⁴³. The Ensembl Genes 91 database and the BioMart tool were also used to identify the corresponding orthologous cattle genes, based on the UMD3.1 bovine reference genome. All the identified genes were processed using the functional annotation tool implemented in DAVID Bioinformatics resources 6.8⁴⁴ to determine enriched functional terms. An enrichment score of 1.3, which is equivalent to the Fisher exact test P-value of 0.05, was used as a threshold to define the significantly enriched functional terms in comparison to the whole bovine reference genome background. Relationships among genomic features in different chromosome positions were represented using the R software package shinyCircos⁴⁵.

Overlap with cattle trypanotolerance-related QTLs. The bovine QTLs for trypanotolerance-related traits²⁷ which were mapped on the bovine Btau 4.0 reference genome assembly were downloaded from the cattle QTL database (<http://www.animalgenome.org/cgi-bin/QTLdb/BT/index>). The QTL genome coordinates were then re-mapped on the bovine UMD 3.1 reference genome assembly using the NCBI genome remapping online service (<https://www.ncbi.nlm.nih.gov/genome/tools/remap>). The intersectBed function of the BedTools software⁴⁶ was used to overlap these trypanotolerance-related QTLs with the bovine regions on which the orthologous functional ovine candidate genes identified were located.

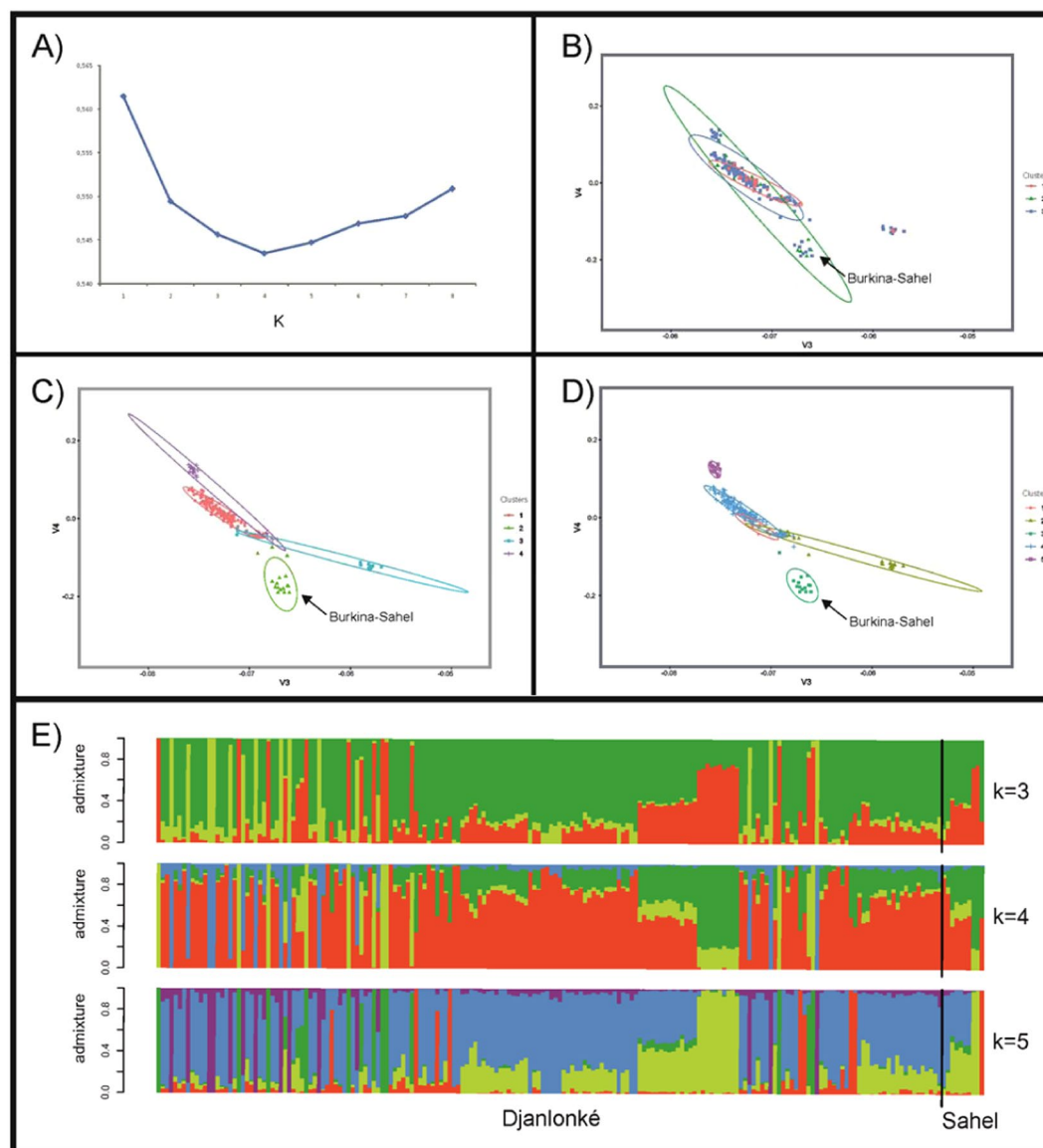


Figure 1. Plots showing cross-validation errors for each K tested using the program Admixture v1.23, with the lowest error for K = 4 (Plot A), and dispersion of the Djallonké and Burkina-Sahel individuals according the two first factors computed via PCA (Plots B, C and D). Factors 1 and 2 explained 3.2% and 2.4% of the total variance, respectively. Contour plots illustrate the 75% confidence region of the relationships among the individuals assigned to each Cluster identified using Admixture v1.23 for K = 3 (Plot B), K = 4 (Plot C) and K = 5 (Plot D). The location of most Burkina-Sahel individuals is highlighted. Furthermore, barplots of individual ancestries estimated using Admixture v1.23 for K = 3, K = 4 and K = 5 are shown (Plot E).

Results

Genetic diversity and structuring of data. Mean inbreeding (and standard deviation) in Djallonké sheep was of 0.058 ± 0.070 . This was higher than the value computed for the Burkina-Sahel sample (0.014 ± 0.020). Differentiation between Djallonké and Burkina-Sahel sheep, assessed computing weighted Weir and Cockerham's F_{ST} , was 0.066.

The results of the admixture analysis informed that the lowest cross-validation error was at K = 4 (Fig. 1). However, differences between cross-validation errors for K = 3, K = 4 and K = 5 were lower than 0.005. Most Djallonké individuals (144 for K = 3, 145 for K = 4 and 140 for K = 5) grouped into a single Cluster. However, assignment of the Burkina-Sahel individuals into a given Cluster was not clear. No differentiation was found between Djallonké sheep sampled either in Mangodara or Dédougou. Most Dédougou samples (from 75% for K = 5 to 100% for K = 3) clustered within the main group formed by Mangodara samples.

Furthermore, most Burkina-Sahel individuals (10) clustered with Djallonké individuals for K = 3 (Fig. 1B). Although Burkina-Sahel individuals tended to form particular Clusters for K = 4 and K = 5 (Figs. 1C,D), they shared these Clusters with a number of Djallonké individuals for K = 4 (4) and for K = 5 (7).

PCA identified 19 factors with eigenvalue higher than 1 which accounted for a total of 25.1% of the genetic variance. Figure 1 shows the dispersion of the individuals and the 75% confidence region of the relationships between individuals assigned to each Cluster identified using Admixture v1.23. Independently of the number of K considered, PCA did not allow to assess a clear separation of the individuals. This was particularly remarkable for K = 3 (Fig. 1B) in which the 75% confidence region of dispersion of the individuals assigned to each of the three clusters overlapped. Although this did not happen for K = 4 and K = 5, PCA illustrated a weak structure in the population analyzed: a significant number of individuals assigned to different clusters were highly intermingled in the middle of the plots.

Candidate regions under selection. Figure 2 shows the Manhattan plots illustrating the SNPs identified as being under selection pressure on all ovine autosomes according to the three tests assayed. After normalization 974 SNPs (Supplementary Table S1), 1,633 (Supplementary Table S2) and 110 (Supplementary Table S3) SNPs were considered to be under selection using the iHS, XP-EHH and nSL tests, respectively.

After crossing the output of the three tests applied, a total of 207 candidate selection sweep regions on all autosomes, except for OAR20 and OAR21, were defined (Table 1; Supplementary Table S4). These regions comprised a total of 34,957,703 bp and 555 different SNPs. Up to 153 SNPs were identified as being under selection by two different statistics. Four chromosomes (OAR1, 2, 3 and 23) gathered 46% of the candidate selection sweep regions.

Identification of functional candidate genes linked to selection sweeps. Gene-annotation enrichment analysis allowed to identify a total of 491 potential candidate ovine genes in the 75 kb up- and down-stream regions surrounding the selection sweep regions defined previously. A full description of these 491 ovine genes, including their identification, description and location is given in Supplementary Table S5. Details on the corresponding orthologous cattle genes are given in Supplementary Table S5 as well.

Functional annotation conducted on these genes allowed to identify 22 different functional term clusters (Supplementary Table S6). However, only three functional term clusters were significantly enriched (enrichment score higher than 1.3; Table 2; Fig. 3). Functional Cluster 1 (enrichment score = 3.24) included four genes (*ALB*, *GC*, *AFP* and *AFM*) involved in transport functions of the plasma membrane and homeostasis. Functional Cluster 2 (enrichment score = 1.39) included four genes (*MC2R*, *CIB1*, *MC5R* and *CAV1*) involved in angiogenesis, answer to ischemia and hypoxia. Functional Cluster 3 (enrichment score = 1.31) included four genes (*LDLRAD4*, *LRP11*, *CFI* and *VLDLR*) involved in lipid metabolism and response to temperature stress.

A description of the genes involved in the definition of these three functional clusters is given in Table 3. The four genes forming Functional Cluster 1 were located on OAR6 and in close vicinity spanning a chromosomal area of 87,639 bp (from position 88,136,611 to position 88,224,250). Functional Cluster 3 included gene *CFI* which was located on OAR6 as well, although on a distant area. Functional Clusters 2 and 3 had neighboring genes located on OAR23 (genes *MC2R* and *MC5R* for Cluster 2 and gene *LDLRAD4* for Cluster 3) on an area spanning 247,960 bp.

Correspondence with cattle trypanotolerance-related QTLs. The relationship between the orthologous cattle genes listed in Table 3 and the cattle QTLs reported for trypanotolerance-related traits²⁷ are given in Supplementary Table S7. Only the cattle gene *CAV1*, assigned to Functional Cluster 2, intersected with a QTL (QTL ID no. 10515) reported on BTA4 for trypanosomes' load.

Discussion

As expected^{22,47}, the population analyzed is poorly structured. This is true even if the Burkina-Sahel individuals are considered (Fig. 1). This situation is consistent with the more common West African livestock scenario in which no artificial selection programmes exist and long-distance livestock trading is intense.

Although Sahelian and Djallonké sheep show large morphological differences and are expected to have different ancestral origins^{21,22}, the Burkina-Sahel sheep do not show a clear separation from the Djallonké individuals (Fig. 1). At the neutral loci level, it is assumed that genetic differentiation in West African livestock ruminant populations is basically due to geographic distance^{47,48}. Morphology, production goals or differences in origin do not result in high genetic differentiation. The poor genetic structuring identified may be due to different unobserved founder events underlying data as a consequence of a very traditional management system in which unplanned natural matings are the rule²⁸.

In scenarios of poor differentiation amongst individuals, the use of EHH-based tests to identify genomic areas under selection is preferable to F_{ST} -based methods which need at least two different populations in dataset¹. We further tested this running the software BayeScan v. 2.1⁴⁹ on our Djallonké genotypes. Only four SNPs (OAR1_22330907.1, OAR12_17378202.1, s56240.1 on OAR14 and OAR16_10423797.1) were identified as being under diversifying selection (positive α values). The inclusion of the Burkina-Sahel individuals in the analyses allowed to identify ten SNPs under diversifying selection (Supplementary Table S8). However, the set of SNPs listed above and their 75 kb up- or down-stream regions did not overlap with the genomic regions surrounding the SNPs identified using EHH-based tests.

The 207 selection sweep regions identified in Djallonké sheep spanned about 35 Mbp which is about 1.3% of the roughly 2.65 Gbp covered by SNPs typed. These figures are consistent with those reported in previous works using similar conservative criteria in the definition of selection signatures. The concordant genomic regions identified using two different EHH-based statistics on three different sheep breeds (Sunite, German Mutton and Dorper) typed with the same SNP Chip used here spanned chromosomal regions summing up from 1,049,753 bp (Dorper) to 1,507,244 bp (Sunite)²⁴. The analysis of the genomic profile, using three different methods (F_{ST} , iSH and RsB), of three Brazilian sheep breeds gave only 5 (out of 246) coincident selection signatures¹². When only two tests were considered the number of coincident selection signatures was 37¹².

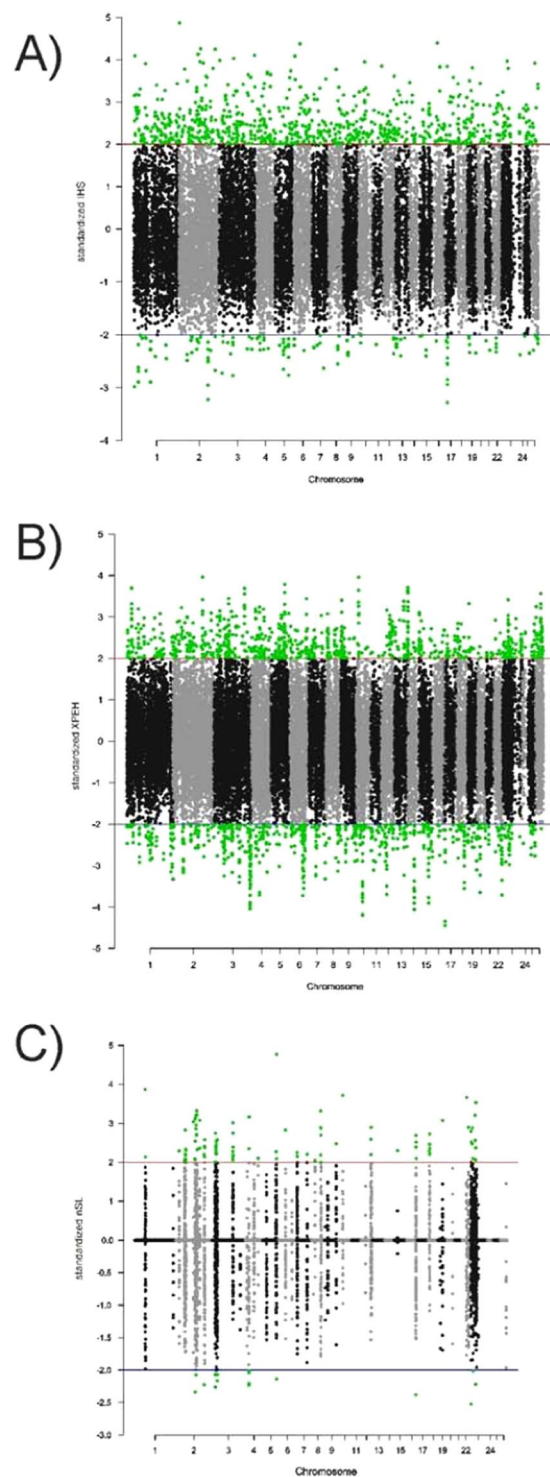


Figure 2. Manhattan plots summarizing results corresponding to iHS (Plot **A**), XP-EHH (Plot **B**) and nSL (Plot **C**) analyses. SNPs with normalized values higher than $|2|$ are highlighted in green.

Biological importance of the functional clusters identified. Gene-annotation enrichment analysis carried out allowed to identify various functional term clusters involved in signaling pathways associated directly or indirectly with environmental adaptation, such as control of metabolic stress, homeostasis, modulation of immune and inflammatory responses, cell proliferation and migration (Supplementary Table S6). The statistically significant Functional Clusters identified depicted particular genetic aspects of adaptation to harsh environments. Although some of the genes included in such Functional Clusters are involved in immune response (e.g. *GC* and *CFI* genes), they are mainly related to metabolic response to stress.

OAR	N	Length	n
1	18	3549100	67 (10) ^a
2	34	6581006	44 (34)
3	27	5892878	105 (19)
4	16	2450404	45 (12)
5	8	1845327	27 (10)
6	10	2590704	43 (5)
7	4	521393	11 (1)
8	9	1097043	17 (5)
9	5	983744	17 (2)
10	4	1077755	16 (9)
11	2	116626	2 (2)
12	7	822135	16 (4)
13	7	960058	18 (3)
14	1	57444	2 (0)
15	6	914153	21 (3)
16	6	181219	7 (3)
17	4	578619	9 (4)
18	10	1669082	27 (7)
19	2	153579	3 (1)
20			
22	3	486195	7 (2)
23	17	1407368	34 (11)
24	2	286693	3 (2)
25	1	107671	2 (2)
26	4	627507	12 (2)
Totals	207	34957703	555 (153)

Table 1. Summary of the candidate selection sweep regions identified per ovine chromosome (OAR). The number of selection sweeps identified (N), the sum of their lengths (in bp) and the sum of the SNPs defining them (n) are given per chromosome. ^asum of the SNPs identified by two different EHH-based statistics.

Functional Cluster 1 included four genes (*ALB*, *GC*, *AFP* and *AFM*) belonging to the family of albumin genes. Because of their location at the same chromosome locus and in the same transcriptional orientation across species, they are proposed to originate from common predecessor and to be cooperatively regulated⁵⁰. Albumin (*ALB*) is involved on transport of metals, fatty acids, cholesterol, bile pigments, and drugs. It is a key element in the regulation of osmotic pressure and distribution of fluid between different compartments. In general, albumin represents the major and predominant antioxidant in plasma, a body compartment known to be exposed to continuous oxidative stress⁵¹. Variants of the albumin genes family, including the fetal counterpart of serum albumin (*AFP*) and afamin/alpha-albumin (*AFM*) genes, are associated with the development of the metabolic syndrome in adult humans^{52,53}. In turn, the *GC* gene is part of the complement system, a key component of innate immunity and susceptibility to diseases with a major role in tissue homeostasis, degeneration, and regeneration⁵⁴. The *GC* gene has major functions including the modulation of immune and inflammatory responses via regulation of chemotaxis and macrophage activation, transport by binding of fatty acids in collaboration with albumin and control of bone development⁵⁵.

Functional Cluster 2 included two genes (*MC2R* and *MC5R*) encoding proteins belonging to the G protein-coupled receptors family involved in the response to stress via binding plasma adrenocorticotropin hormone (ACTH). While the *MC2R* gene acts on adrenal steroidogenesis and regulation of the glucocorticoid axis, the *MC5R* gene regulates the function of sebaceous glands with effect on water repulsion and thermoregulation⁵⁶. In turn, the *CIB1* gene encodes a protein involved in cell proliferation, particularly proplatelets, therefore acting on thrombopoiesis⁵⁷, while the *CAV1* gene has a critical role in signal transduction and trafficking for its interplay with steroid receptors and is associated with the metabolic syndrome in humans⁵⁸.

Genes assigned to Functional Cluster 3 have a major role on response to stress and immunity. The *LDLRAD4* gene encodes a protein involved in the regulation of the transforming growth factor- β (TGF- β) signaling⁵⁹ which plays a pivotal role in cell differentiation, apoptosis, cell migration, production of matrix proteins, angiogenesis, and anti-proliferative responsiveness. The *VLDLR* gene is also involved in the regulation of cell proliferation and migration⁶⁰. Another low density lipoprotein receptor gene belonging to this Cluster (*LPR11*) is associated with chronic stress caused by food or water deprivation or elevated or cold temperature⁶¹. Finally, the complement factor I (*CFI*) gene encodes a serine protease forming part of the complement system, which is involved in innate and adaptive immune responses for prevention and control of diseases⁶².

General discussion on adaptation to West African environment. The genomic regions harboring candidate genes of functional importance identified in the current study are, in general, different to those

Functional Cluster (enrichment score)	Category	Term and Description	Genes	P-Value	Fold Enrichment
Cluster 1 (3.24)	INTERPRO	IPR014760:Serum albumin, N-terminal	ENSOARG00000013782 (ALB), ENSOARG00000012835 (GC), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	8.61E-05	38.40
	INTERPRO	IPR000264:ALB/AFP/VDB	ENSOARG00000013782 (ALB), ENSOARG00000012835 (GC), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	8.61E-05	38.40
	INTERPRO	IPR020857:Serum albumin, conserved site	ENSOARG00000013782 (ALB), ENSOARG00000012835 (GC), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	8.61E-05	38.40
	SMART	SM00103:ALBUMIN	ENSOARG00000013782 (ALB), ENSOARG00000012835 (GC), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	8.81E-05	37.89
	INTERPRO	IPR020858:Serum albumin-like	ENSOARG00000013782 (ALB), ENSOARG00000012835 (GC), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	1.70E-04	32.00
Cluster 2 (1.39)	PIR_SUPERFAMILY	PIRSF002520:serum albumin	ENSOARG00000013782 (ALB), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	1.19E-03	47.29
	INTERPRO	IPR021177:Serum albumin/Alpha-fetoprotein	ENSOARG00000013782 (ALB), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	2.51E-03	36.00
	GOTERM_BP_DIRECT	GO:0006810~transport	ENSOARG00000013782 (ALB), ENSOARG00000013966 (AFP), ENSOARG00000014129 (AFM)	4.33E-01	2.04
	INTERPRO	IPR023415:Low-density lipoprotein (LDL) receptor class A, conserved site	ENSOARG00000003950 (MC2R), ENSOARG00000002239 (MC5R), ENSOARG00000001337 (CAVI)	1.70E-02	12.47
	SMART	SM00192:LDLa	ENSOARG00000003950 (MC2R), ENSOARG00000002239 (MC5R), ENSOARG00000001337 (CAVI)	5.20E-02	8.08
Cluster 3 (1.31)	INTERPRO	IPR002172:Low-density lipoprotein (LDL) receptor class A repeat	ENSOARG00000003950 (MC2R), ENSOARG00000002128 (CIB1), ENSOARG00000002239 (MC5R), ENSOARG00000001337 (CAVI)	7.59E-02	4.04
	INTERPRO	IPR023415:Low-density lipoprotein (LDL) receptor class A, conserved site	ENSOARG00000002144 (LDLRAD4), ENSOARG00000002898 (LRP11), ENSOARG00000005291 (CFI), ENSOARG000000013303 (VLDLR)	3.83E-02	5.33
	SMART	SM00192:LDLa	ENSOARG00000002144 (LDLRAD4), ENSOARG00000002898 (LRP11), ENSOARG00000005291 (CFI), ENSOARG000000013303 (VLDLR)	5.10E-02	4.74
	INTERPRO	IPR002172:Low-density lipoprotein (LDL) receptor class A repeat	ENSOARG00000002144 (LDLRAD4), ENSOARG00000002898 (LRP11), ENSOARG00000005291 (CFI), ENSOARG000000013303 (VLDLR)	5.96E-02	4.46
	INTERPRO	IPR023415:Low-density lipoprotein (LDL) receptor class A, conserved site	ENSOARG00000002144 (LDLRAD4), ENSOARG00000002898 (LRP11), ENSOARG00000005291 (CFI), ENSOARG000000013303 (VLDLR)	5.96E-02	4.46

Table 2. Significantly enriched functional term clusters (enrichment score higher than 1.3) for genes identified within the 75 kb regions flanking the candidate selection sweeps identified using genome-wide iHS, XP-EHH and nSL statistics following DAVID analysis.

previously reported in the literature. This is not surprising in the case of works aiming at the characterization of the effect on the sheep genome of directional selection for dairy^{3,5} or meat traits^{6,7}. However, interestingly enough, it also occurred when reports aiming at the identification of genomic signals of adaptation to extreme environments were considered. It is worth pointing out that most studies in this field aimed at the identification of genomic areas related to adaptation to high altitude^{10,17} or to high temperature in basically arid environments^{15,63}. Such works involved several sheep populations bred in different geographical areas highly contrasting in environments, production systems and goals^{14,15,63}. Therefore, the ability to identify selection signatures related to thermotolerance or hypoxia may be affected by the influence of others more likely linked to production or reproduction performance.

The Djallonké population analyzed here is a homogeneous population managed in a harsh, hot and humid environment subject to different disease challenges, namely trypanosomosis²³. Although it is assumed that genes related to immune response and thermotolerance are basic for environmental adaptation, the lack of concordance between the genomic signals identified here and those previously reported would suggest that the genomic areas involved in resistance to hot climate stress may be different if they are identified in populations bred either in humid or arid areas. In this respect, the genomic areas conserved between sheep and goat indigenous to the hot

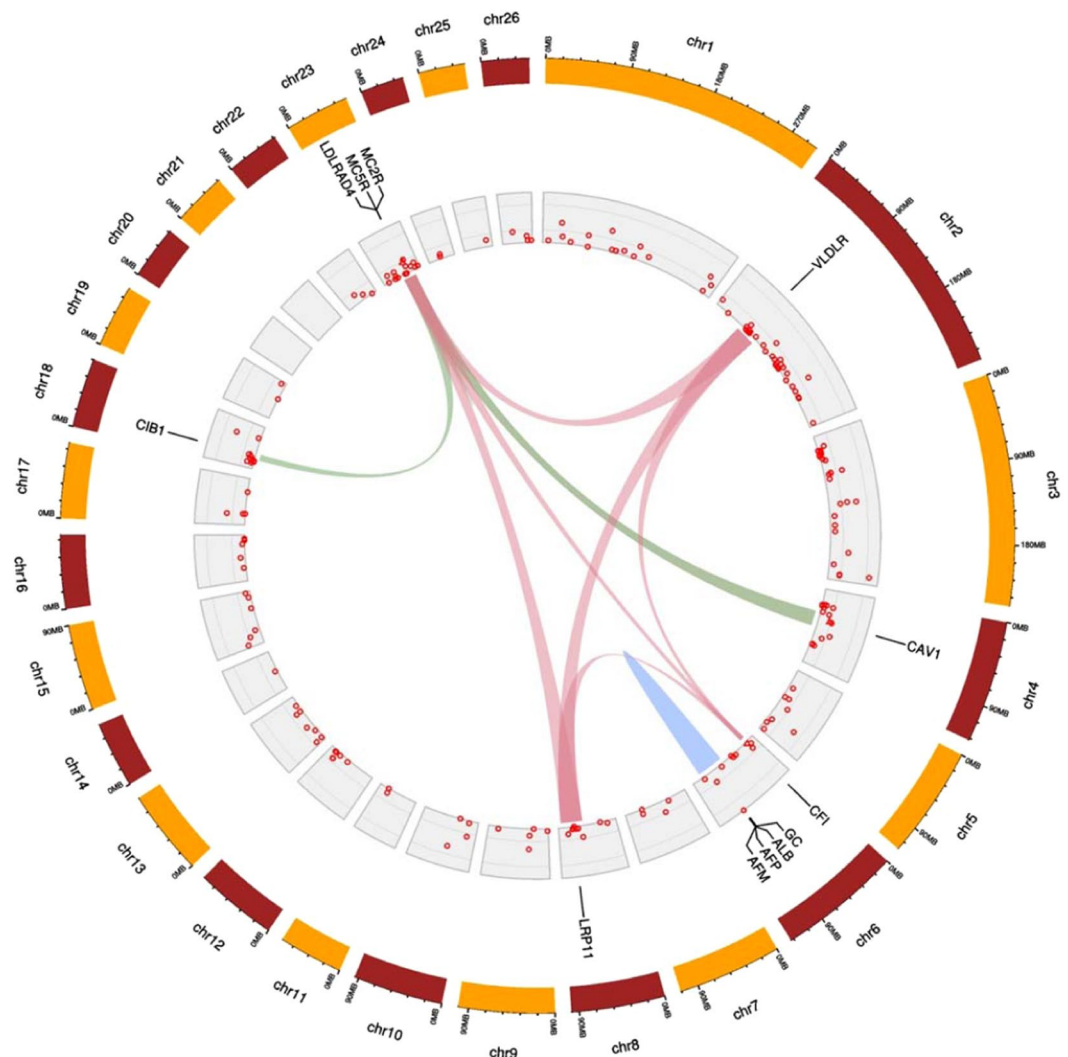


Figure 3. Circular map, created using the ShinyCircos⁴⁵ software, summarizing information on Functional Clusters identified using selection sweeps in Djallonké sheep. Chromosomes are shown in the outermost circle. The innermost circles show the distribution of selection sweeps within a chromosome. Differences in height among selection sweeps inform on the number of SNPs defining them. Candidate genes forming the significantly enriched functional term clusters are indicated besides their genomic localization. At the center of the map, links among candidate genes belonging to the same functional cluster are illustrated using the same color.

arid environment of Egypt⁶³ are completely different than those identified as functionally important in Djallonké sheep. In any case, it cannot be discarded that the history of the population is affecting results. The only coincidence between our study and others in the literature is a selection signature on OAR23 carrying out genes *MC5R* and *MC2R* in Ethiopian Menz and Red Maasai breeds (East African sheep)¹³. While the Ethiopian Menz is a fat-tailed coarse woolled sheep probably descending from very ancient importations from Arabia (<http://eth.dagris.info/node/2448>), the Red Maasai is a hair sheep like Djallonké, although medium to large body sized, with well known resistance to gastrointestinal parasites⁶⁴.

Trypanosome challenge has been hypothesized to be a major historical force influencing the formation of native West African domestic ruminant populations including Djallonké sheep¹⁹. In consequence, we could expect that orthologous cattle chromosomal areas in which some QTLs for trypanotolerance related traits were previously identified²⁷ would coincide with the genomic areas harboring genes of functional importance for adaptation in Djallonké sheep. However, only one orthologous candidate gene located on BTA4 (*CAV1*) had a clear relationship with those QTLs.

QTL information on cattle trypanotolerance was refined experimental herds⁶⁵ and outbred populations^{66,67}. Such analyses suggested that QTLs located on BTA2, BTA4 and BTA7 (with genes *ARHGAP15*⁶⁵, *TICAM1*⁶⁵, *CXCR4*⁶⁶ and *INHBA*⁶⁷ being the stronger candidates to underlie these QTLs) could have a major role on cattle trypanotolerance response. Although recent analyses failed in identifying causal mutations on the cattle *ARHGAP15*, *TICAM1*, *CXCR4* and *INHBA* genes^{68–70} the importance of BTA2, BTA4 and BTA7 for

Functional	Gene		Sheep gene information				Orthologous cattle gene information			
Cluster	name	description	EnsemblID	OAR	start	end	Cattle EnsemblID	BTA	start	end
Cluster 1	<i>ALB</i>	albumin [Source:NCBI gene;Acc:443393]	ENSOARG00000013782	6	88136611	88159187	ENSBTAG00000017121	6	90232762	90251126
	<i>GC</i>	GC, vitamin D binding protein [Source:HGNC Symbol;Acc:HGNC:4187]	ENSOARG00000012835	6	86619919	86657661	ENSBTAG00000013718	6	88695940	88739180
	<i>AFP</i>	alpha fetoprotein [Source:HGNC Symbol;Acc:HGNC:317]	ENSOARG00000013966	6	88166794	88190190	ENSBTAG00000017131	6	90258976	90280522
	<i>AFM</i>	afamin [Source:HGNC Symbol;Acc:HGNC:316]	ENSOARG00000014129	6	88198267	88224250	ENSBTAG00000047833	6	90291078	90312046
Cluster 2	<i>MC2R</i>	melanocortin 2 receptor [Source:NCBI gene;Acc:443482]	ENSOARG00000003950	23	43893855	43894748	ENSBTAG00000011038	24	44006264	44024553
	<i>CIB1</i>	calcium and integrin binding 1 [Source:NCBI gene;Acc:100141300]	ENSOARG00000012128	18	20723490	20727112	ENSBTAG000000021275	21	22043354	22046525
	<i>MC5R</i>	Ovis aries melanocortin 5 receptor (MC5R), mRNA. [Source:RefSeq mRNA;Acc:NM_001078656]	ENSOARG00000002239	23	43867867	43870160	ENSBTAG00000009143	24	43983880	43984857
	<i>CAV1</i>	caveolin 1 [Source:NCBI gene;Acc:445468]	ENSOARG00000001337	4	51786039	51820791	ENSBTAG00000017869	4	52173110	52208687
Cluster 3	<i>LDLRAD4</i>	low density lipoprotein receptor class A domain containing 4 [Source:HGNC Symbol;Acc:HGNC:1224]	ENSOARG00000002144	23	43646788	43801055	ENSBTAG00000009139	24	43861727	43919241
	<i>LRP11</i>	LDL receptor related protein 11 [Source:VGNC Symbol;Acc:VGNC:30988]	ENSOARG00000002898	8	73814860	73836141	ENSBTAG00000032007	9	88115145	88183314
	<i>CFI</i>	complement factor I [Source:NCBI gene;Acc:101119307]	ENSOARG000000005291	6	15708801	15753015	ENSBTAG000000034501	6	16765034	16815352
	<i>VLDLR</i>	very low density lipoprotein receptor [Source:HGNC Symbol;Acc:HGNC:12698]	ENSOARG00000013303	2	70468444	70486178	ENSBTAG00000018517	8	42109679	42141155

Table 3. Description of the sheep genes included in the three functional clusters displaying enrichment score higher than 1.3. Besides the gene name and description, the following information is provided: the identification of the gene retrieved from the Ensembl Genes 91 database, the ovine chromosome (OAR) on which the gene is located and the positions (in bp) of start and end of the gene within the chromosome. The same information is given for the orthologous cattle genes.

trypanotolerance cannot be neglected. The case of BTA4 would have been confirmed by the current analyses. Moreover, the sheep chromosomal areas orthologous to the QTLs reported on BTA4 (OAR4) might have been conserved between species to play a role for adaptation to the same harsh environment and trypanosome challenge.

In summary, the genome of the Djallonké sheep provided new insights on the genomic basis for adaptation. This sheep population is subject to the particularly harsh environment of the hot-humid, trypanosome challenged, West Africa. The genomic areas identified are associated with innate immunity and thermotolerance. Results further suggest that genes involved in the regulation of oxidative and metabolic stress can be target of research for the ascertainment of the genetic basis of adaptation. Moreover, our findings suggest that the functional importance of cattle BTA4 for trypanotolerant response might have been conserved between species. Furthermore, our findings highlight the importance of obtaining information from non-cosmopolite livestock populations managed in particularly harsh environments.

Ethics statement. Blood and hair root samples used here were collected by veterinary practitioners with the permission and in presence of the owners. For this reason, permission from the Ethics Committee for Health Research in Burkina Faso (Joint Order 2004-147/MS/MESSE of May 11, 2004) was not required. In all instances, veterinarians followed standard procedures and relevant national guidelines to ensure appropriate animal care.

Data availability

The dataset used and analyzed during the current study is available from the corresponding author on reasonable request.

Received: 12 February 2019; Accepted: 11 December 2019;

Published online: 18 February 2020

References

- Paim, T., Ianella, P., Paiva, S. R., Caetano, A. & McManus, C. M. Detection and evaluation of selection signatures in sheep. *Pesq. Agropec. Bras.* **53**, 527–539, <https://doi.org/10.1590/s0100-204x2018000500001> (2018).
- McRae, K. M., McEwan, J. C., Dodds, K. G. & Gemmell, N. J. Signatures of selection in sheep bred for resistance or susceptibility to gastrointestinal nematodes. *BMC Genom.* **15**, 637, <https://doi.org/10.1186/1471-2164-15-637> (2014).
- Gutiérrez-Gil, B. *et al.* Application of selection mapping to identify genomic regions associated with dairy production in sheep. *PLoS One* **9**, e94623, <https://doi.org/10.1371/journal.pone.0094623> (2014).
- Manunza, A. *et al.* Population structure of eleven Spanish ovine breeds and detection of selective sweeps with BayeScan and hapFLK. *Sci. Rep.* **6**, 27296, <https://doi.org/10.1038/srep27296> (2016).
- Moioli, B., Scatà, M. C., Steri, R., Napolitano, F. & Catillo, G. Signatures of selection identify loci associated with milk yield in sheep. *BMC Genet.* **14**, 76, <https://doi.org/10.1186/1471-2156-14-76> (2013).
- Purfield, D. C., McParland, S., Wall, E. & Berry, D. P. The distribution of runs of homozygosity and selection signatures in six commercial meat sheep breeds. *PLoS One* **12**, e0176780, <https://doi.org/10.1371/journal.pone.0176780> (2017).
- Wang, H. *et al.* Genome-wide specific selection in three domestic sheep breeds. *PLoS One* **10**, e0128688, <https://doi.org/10.1371/journal.pone.0128688> (2015).
- Moradi, M. H., Nejati-Javaremi, A., Moradi-Shahrbabak, M., Dodds, K. G. & McEwan, J. C. Genomic scan of selective sweeps in thin and fat tail sheep breeds for identifying of candidate regions associated with fat deposition. *BMC Genet.* **13**, 10, <https://doi.org/10.1186/1471-2156-13-10> (2012).
- Yuan, Z. *et al.* Selection signature analysis reveals genes associated with tail type in Chinese indigenous sheep. *Anim. Genet.* **48**, 55–66, <https://doi.org/10.1111/age.12477> (2017).
- Wei, C. *et al.* Genome-wide analysis reveals population structure and selection in Chinese indigenous sheep breeds. *BMC Genomics* **16**, 194, <https://doi.org/10.1186/s12864-015-1384-9> (2015).
- Liu, Z. Z. *et al.* Genome-wide analysis reveals signatures of selection for important traits in domestic sheep from different ecoregions. *BMC Genomics* **17**, 863, <https://doi.org/10.1186/s12864-016-3212-2> (2016).
- Gouveia, J. J. *et al.* Genome-wide search for signatures of selection in three major Brazilian locally adapted sheep breeds. *Livest. Sci.* **197**, 36–45, <https://doi.org/10.1016/j.livsci.2017.01.006> (2017).
- Fariello, M.-I. *et al.* Selection signatures in worldwide sheep populations. *PLoS One* **9**, e103813, <https://doi.org/10.1371/journal.pone.0103813> (2014).
- Ly, F.-H. *et al.* Adaptations to climate-mediated selective pressures in sheep. *Mol. Biol. Evol.* **31**, 3324–3343, <https://doi.org/10.1093/molbev/msu264> (2014).
- Mwacharo, J. M. *et al.* Genomic footprints of dryland stress adaptation in Egyptian fat-tail sheep and their divergence from East African and western Asia cohorts. *Sci. Rep.* **7**, 17647, <https://doi.org/10.1038/s41598-017-17775-3> (2017).
- Yang, J. *et al.* Whole-genome sequencing of native sheep provides insights into rapid adaptations to extreme environments. *Mol. Biol. Evol.* **33**, 2576–2592, <https://doi.org/10.1093/molbev/msw129> (2016).
- Wei, C. *et al.* Genome-wide analysis reveals adaptation to high altitudes in Tibetan sheep. *Sci. Rep.* **6**, 26770, <https://doi.org/10.1038/srep26770> (2016).
- Chessa, B. *et al.* Revealing the history of sheep domestication using retrovirus integrations. *Science* **324**, 532–536, <https://doi.org/10.1126/science.1170587> (2009).
- Gifford-Gonzalez, D. Animal Disease Challenges to the Emergence of Pastoralism in Sub-Saharan Africa. *Afr. Archaeol. Review* **17**, 95–139, <https://doi.org/10.1023/A:1006601020217> (2000).
- Muigai, A. W. T. & Hanotte, O. The origin of African sheep: archaeological and genetic perspectives. *Afr. Archaeol. Rev.* **30**, 39–50, <https://doi.org/10.1007/s10437-013-9129-0> (2013).
- Traoré, A. *et al.* Multivariate characterization of morphological traits in Burkina Faso sheep. *Small Rum. Res.* **80**, 62–67, <https://doi.org/10.1016/j.smallrumres.2008.09.011> (2008).
- Álvarez, I. *et al.* Microsatellite analysis characterizes Burkina Faso as a genetic contact zone between Sahelian and Djallonké sheep. *Anim. Biotech.* **20**, 47–57, <https://doi.org/10.1080/10495390902786926> (2009).
- Geerts, S., Osaer, S., Goossens, B. & Faye, D. Trypanotolerance in small ruminants of sub-Saharan Africa. *Trends Parasitol.* **25**, 132–138, <https://doi.org/10.1016/j.pt.2008.12.004> (2009).
- Zhao, F.-P. *et al.* A genome scan of recent positive selection signatures in three sheep populations. *J. Int. Agr.* **15**, 162–164, [https://doi.org/10.1016/S2095-3119\(15\)61080-2](https://doi.org/10.1016/S2095-3119(15)61080-2) (2016).
- Fariello, M.-I., Boitard, S., Naya, H., Sancristobal, M. & Servin, B. Detecting signatures of selection through haplotype differentiation among hierarchically structured populations. *Genetics* **193**, 929–941, <https://doi.org/10.1534/Genetics.112.147231> (2013).
- Sabeti, P. C. *et al.* Detecting recent positive selection in the human genome from haplotype structure. *Nature* **419**, 832–837, <https://doi.org/10.1038/nature01140> (2002).
- Hanotte, O. *et al.* Mapping of quantitative trait loci controlling trypanotolerance in a cross of tolerant West African N'Dama and susceptible East African Boran cattle. *Proc. Nat. Acad. Sci. USA* **100**, 7443–7448, <https://doi.org/10.1073/pnas.1232392100> (2003).
- Álvarez, I. *et al.* Usefulness of running animal models in absence of pedigrees: estimation of genetic parameters for gastrointestinal parasite resistance traits in Djallonké sheep of Burkina Faso. *Small Rum. Res.* **161**, 81–88, <https://doi.org/10.1016/j.smallrumres.2018.01.020> (2018).
- Ouadba, J. M. Development of national monograph on the biological diversity of Burkina Faso: Data gathering, ecological considerations (in French) pp 45 (Minist. Envir. et de l'Eau, Ouagadougou, 1997).
- Álvarez, I. *et al.* Genetic relationships of the Cuban hair sheep inferred from microsatellite polymorphism. *Small Rum. Res.* **104**, 89–93, <https://doi.org/10.1016/j.smallrumres.2011.10.025> (2012).
- Traoré, A. *et al.* Resistance to gastrointestinal parasite infection in Djallonké sheep. *Animal* **11**, 1354–1362, <https://doi.org/10.1017/S1751731116002640> (2017).
- Purcell, S. *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am. J. Hum. Genet.* **81**, 559–75, <https://doi.org/10.1086/519795> (2007).
- Danecek, P. *et al.* The variant call format and VCFtools. *Bioinformatics* **27**, 2156–2158, <https://doi.org/10.1093/bioinformatics/btr330> (2011).
- Alexander, D. H. & Lange, K. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. *BMC Bioinform.* **12**, 246, <https://doi.org/10.1186/1471-2105-12-246> (2011).
- Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664, <https://doi.org/10.1101/gr.094052.109> (2009).
- Voight, B. F., Kudaravalli, S., Wen, X. & Pritchard, J. K. A map of recent positive selection in the human genome. *PLoS Biol.* **4**, e72, <https://doi.org/10.1371/journal.pbio.0040072> (2006).
- Sabeti, P. C. *et al.* Genome-wide detection and characterization of positive selection in human populations. *Nature* **449**, 913–918, <https://doi.org/10.1126/science.1124309> (2007).
- Ferrer-Admetlla, A., M. Liang, M., Korneliussen, T. & Nielsen, R. On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol. Biol. Evol.* **31**, 1275–1291, <https://doi.org/10.1093/molbev/msu077> (2014).
- Schlamp, F. *et al.* Evaluating the performance of selection scans to detect selective sweeps in domestic dogs. *Mol. Ecol.* **25**, 342–356, <https://doi.org/10.1111/mec.13485> (2016).

40. Szpiech, Z. A. & Hernandez, R. D. *selscan*: an efficient multithreaded program to perform EHH based scans for positive selection. *Mol. Biol. Evol.* **31**, 2824–2827, <https://doi.org/10.1093/molbev/msu211> (2014).
41. Cardoso, D. F. *et al.* Genome-wide scan reveals population stratification and footprints of recent selection in Nelore cattle. *Genet. Sel. Evol.* **50**, 22, <https://doi.org/10.1186/s12711-018-0381-2> (2018).
42. Yaro, M. *et al.* Analysis of pooled genome sequences from Djallonké and Sahelian sheep of Ghana reveals co-localisation of regions of reduced heterozygosity with candidate genes for disease resistance and adaptation to a tropical environment. *BMC Genomics* **20**, 816, <https://doi.org/10.1186/s12864-019-6198-8> (2019).
43. Kinsella, R. J. *et al.* Ensembl BioMart: a hub for data retrieval across taxonomic space. Database bar030, pmid:21785142, <https://doi.org/10.1093/database/bar030> (2011).
44. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57, <https://doi.org/10.1038/nprot.2008.211> (2009).
45. Yu, Y., Ouyang, Y. & Yao, W. shinyCircos: an R/Shiny application for interactive creation of Circos plot. *Bioinformatics* **34**, 1229–1231, <https://doi.org/10.1093/bioinformatics/btx763> (2018).
46. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842, <https://doi.org/10.1093/bioinformatics/btq033> (2010).
47. Traoré, A. *et al.* Ascertaining gene flow patterns in livestock populations of developing countries: a case study in Burkina Faso goat. *BMC Genet.* **13**, 35, <https://doi.org/10.1186/1471-2156-13-35> (2012).
48. Missohou, A. *et al.* Genetic diversity and differentiation in nine West African local goat breeds assessed via microsatellite polymorphism. *Small Rum. Res.* **99**, 20–24, <https://doi.org/10.1016/j.smallrumres.2011.04.005> (2011).
49. Foll, M. & Gaggiotti, O. E. A genome scan method to identify selected loci appropriate for both dominant and codominant markers: A Bayesian perspective. *Genetics* **180**, 977–993, <https://doi.org/10.1534/genetics.108.092221> (2008).
50. Lazarevich, N. L. Molecular mechanisms of alpha-fetoprotein gene expression. *Biochemistry* **65**, 117–133 (2000).
51. Roche, M., P. Rondeau, P., Singh, N. R., Tarnus, E. & Bourdon, E. The antioxidant properties of serum albumin. *FEBS Lett.* **582**, 1783–1787, <https://doi.org/10.1016/j.febslet.2008.04.057> (2008).
52. Chen, Y. *et al.* Association between alpha-fetoprotein and metabolic syndrome in a Chinese asymptomatic population: a cross-sectional study. *Lipids Health Dis.* **15**, 85, <https://doi.org/10.1186/s12944-016-0256-x> (2016).
53. Dieplinger, H. & Dieplinger, B. Afamin — A pleiotropic glycoprotein involved in various disease states. *Clin. Chim. Acta* **446**, 105–110, <https://doi.org/10.1016/j.cca.2015.04.010> (2015).
54. Phielers, J., García-Martin, R., Lambris, J. D. & Chavakis, T. The role of the complement system in metabolic organs and metabolic diseases. *Semin. Immunol.* **25**, 47–53, <https://doi.org/10.1016/j.smim.2013.04.003> (2013).
55. Speckaert, M. M., Speckaert, R., van Geel, N. & Delanghe, J. R. Vitamin D Binding Protein: a multifunctional protein of clinical importance. *Adv. Clin. Chem.* **63**, 1–57, <https://doi.org/10.1016/B978-0-12-800094-6.00001-7> (2014).
56. Ramachandrapa, S., Gorrigan, R. J., Clark, A. J. L. & Chan, L. F. The melanocortin receptors and their accessory proteins. *Front. Endocrinol.* **4**, 9, <https://doi.org/10.3389/fendo.2013.00009> (2013).
57. Kostyak, J. C., Naik, M. U. & Naik, U. P. 2012. Calcium- and integrin-binding protein 1 regulates megakaryocyte ploidy, adhesion, and migration. *Blood* **119**, 838–846, <https://doi.org/10.1182/blood-2011-04-346098> (2012).
58. Baudrand, R. *et al.* A prevalent caveolin-1 gene variant is associated with the metabolic syndrome in Caucasians and Hispanics. *Metabolism* **64**, 1674–1681, <https://doi.org/10.1016/j.metabol.2015.09.005> (2015).
59. Itoh, S. & Itoh, F. TMEPAI family: involvement in regulation of multiple signaling pathways. *J. Biochem.* **164**, 195–204, <https://doi.org/10.1093/jb/mvy059> (2018).
60. Di, Y. *et al.* TFPI or uPA-PAI-1 complex affect cell function through expression variation of type II very low density lipoprotein receptor. *FEBS Lett.* **584**, 3469–3473, <https://doi.org/10.1016/j.febslet.2010.07.005> (2010).
61. Xu, J. *et al.* Genetic regulatory network analysis reveals that low density lipoprotein receptor-related protein 11 is involved in stress responses in mice. *Psychiatry Res.* **220**, 1131–1137, <https://doi.org/10.1016/j.psychres.2014.09.002> (2014).
62. Roversi, P. *et al.* Structural basis for complement factor I control and its disease-associated sequence polymorphisms. *Proc. Natl. Acad. Sci. USA* **108**, 12839–12844, <https://doi.org/10.1073/pnas.1102167108> (2011).
63. Kim, E.-S. B. *et al.* Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity* **116**, 255–264, <https://doi.org/10.1038/hdy.2015.94> (2016).
64. Benavides, M. V. *et al.* Identification of novel loci associated with gastrointestinal parasite resistance in a red Maasai x Dorper backcross population. *PLoS One* **10**, e0122797, <https://doi.org/10.1371/journal.pone.0122797> (2015).
65. Noyes, H. *et al.* Genetic and expression analysis of cattle identifies candidate genes in pathways responding to *Trypanosoma congolense* infection. *Proc. Natl. Acad. Sci. USA* **108**, 9304–9309, <https://doi.org/10.1073/pnas.1013486108> (2011).
66. Dayo, G. K. *et al.* Detection of selection signatures within candidate regions underlying trypanotolerance in outbred cattle populations. *Mol. Ecol.* **18**, 1801–1813, <https://doi.org/10.1111/j.1365-294X.2009.04141.x> (2009).
67. Dayo, G. K. *et al.* Association studies in QTL regions linked to bovine trypanotolerance in a West African crossbred population. *Anim. Genet.* **43**, 123–132, <https://doi.org/10.1111/j.1365-2052.2011.02227.x> (2012).
68. Álvarez, I., Pérez-Pardal, L., Traoré, A., Fernández, I. & Goyache, F. African cattle do not carry unique mutations on the exon 9 of the ARHGAP15 gene. *Anim. Biotech.* **27**, 9–12, <https://doi.org/10.1080/10495398.2015.1053606> (2016).
69. Álvarez, I., Pérez-Pardal, L., Traoré, A., Fernández, I. & Goyache, F. Lack of haplotype structuring for two candidate genes for trypanotolerance in cattle. *J. Anim. Breed. Genet.* **133**, 105–114, <https://doi.org/10.1111/jbg.12181> (2016).
70. Álvarez, I., Pérez-Pardal, L., Traoré, A., Fernández, I. & Goyache, F. Lack of specific alleles for the bovine Chemokine (C-X-C) receptor type 4 (CXCR4) gene in West African cattle questions its role as a candidate for trypanotolerance. *Infect. Genet. Evol.* **42**, 30–33, <https://doi.org/10.1016/j.meegid.2016.04.029> (2016).

Acknowledgements

This work was partially supported by the grant MICIIN-FEDER AGL2016-77813-R. CompGen. LP-P is supported by grant SFRH/BPD/94518/2013.

Author contributions

F.G. and I.A. conceived and planned the project; F.G., I.F. and L.P.-P. did the data analyses; F.G., I.A. and I.F. wrote the paper; A.T. undertook sampling and discussed and interpreted genetic data in light of breeding evidence; I.A., and N.M.-A. obtained samples and did laboratory work; L.P.-P. and N.M.-A. discussed and interpreted genetic data in light of the statistical evidence. All authors gave final approval for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-59839-x>.

Correspondence and requests for materials should be addressed to F.G.

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.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020