




OPEN

Characterization of heterozygosity-rich regions in Italian and worldwide goat breeds

Giorgio Chessari^{1,7}, Andrea Criscione^{1,7}, Donata Marletta¹, Paola Crepaldi², Baldassare Portolano³, Arianna Manunza⁴, Alberto Cesarani^{5,6}, Filippo Biscarini⁴ & Salvatore Mastrangelo³

Heterozygosity-rich regions (HRR) are genomic regions of high heterozygosity, which may harbor loci related to key functional traits such as immune response, survival rate, fertility, and other fitness traits. This study considered 30 Italian and 19 worldwide goat breeds genotyped with the Illumina GoatSNP50k BeadChip. The aim of the work was to study inter-breed relationships and HRR patterns using Sliding Window (SW) and Consecutive Runs (CR) detection methods. Genetic relationships highlighted a clear separation between non-European and European breeds, as well as the north–south geographic cline within the latter. The Pearson correlation coefficients between the descriptive HRR parameters obtained with the SW and CR methods were higher than 0.9. A total of 166 HRR islands were detected. CHI1, CHI11, CHI12 and CHI18 were the chromosomes harboring the highest number of HRR islands. The genes annotated in the islands were linked to various factors such as productive, reproductive, immune, and environmental adaptation mechanisms. Notably, the Montecristo feral goat showed the highest number of HRR islands despite the high level of inbreeding, underlining potential balancing selection events characterizing its evolutionary history. Identifying a species-specific HRR pattern could provide a clearer view of the mechanisms regulating the genome modelling following anthropogenic selection combined with environmental interaction.

In diploid organisms, single nucleotide differences observed between paternal and maternal chromosomes are called heterozygous sites¹. With the continued development and cost reduction of high-throughput DNA sequencing and genotyping technologies, researchers have powerful tools for studying animal genomes through whole-genome molecular markers. However, little has been done to analyze genomic aspects of heterozygosity in livestock populations², and the limited evidence suggests that increased heterozygosity over time may be attributed to selection³. Heterozygosity-rich regions (HRR), also known as runs of heterozygosity (ROHet), are a recently emerged analytical concept and refer to regions of consecutive heterozygous sites detected between paternal and maternal chromosomes in diploid organisms⁴. The analysis of HRR aims to identify genomic regions with high genetic variability, to provide information about the populations' genetic diversity and evolutionary history⁵, as well as to identify specific segments in the genome where increased genetic diversity could be beneficial^{6,7}. HRR may harbor loci associated to key functional traits such as immune response, survival rate, fertility and other fitness traits⁸ and avoid the deleterious effects of harmful homozygous genotypes^{9–11}. Williams et al.⁴ first introduced the concept of HRR in livestock and suggested that some of these regions were under balancing selection and contained recessive lethal mutations in cattle. Along this line, further studies were subsequently carried out in cattle^{6,7}, horses³, sheep^{8,12} and pigs^{9,13,14}, but less commonly in goat species¹⁵.

¹Dipartimento Agricoltura, Alimentazione e Ambiente, University of Catania, Via Santa Sofia 100, 95123 Catania, Italy. ²Dipartimento Scienze Agrarie e Ambientali, Produzione, Territorio, Agroenergia, University of Milan, Via Giovanni Celoria 2, 20133 Milan, Italy. ³Dipartimento Scienze Agrarie, Alimentari e Forestali, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy. ⁴CNR, Institute of Agricultural Biology and Biotechnology (IBBA), Via Bassini 15, 20133 Milan, Italy. ⁵Dipartimento di Agraria, University of Sassari, Viale Italia 39, 07100 Sassari, Italy. ⁶Animal and Dairy Science Department, University of Georgia, 425 River Road, 30602 Athens, GA, USA. ⁷These authors contributed equally: Giorgio Chessari and Andrea Criscione. ✉email: andrea.criscione@unict.it

Goats (*Capra hircus*) were domesticated around 10,500 years ago in the Fertile Crescent¹⁶. After dispersion from its center of domestication, this species has undergone an intense adaptation process and occupied different agroecological areas worldwide¹⁷. Different breeds and goat strains have been selected for milk, meat and fiber production, playing an important role in the livestock sector around the world. Given the wide range of genetic variability and its ability to populate very different geographical areas and climates and produce in conditions of low anthropogenic input, the goat species is the best for studying genetic diversity and adaptation^{18,19}. Thanks to these features, this species provides one of the most suitable models for understanding the patterns and distribution of HRR. Furthermore, Italy is the European country with the highest number of goat breeds, and it provides a precious reservoir of genetic diversity shaped by its varied history, environment, climate, and farming traditions²⁰. Therefore, in this study, we investigated the genomes of Italian and worldwide goat breeds, to detect and characterize their HRR patterns and reveal regions of heterozygosity (HRR islands) which contain candidate genes related to specific traits.

Results

Genetic diversity indices

The genetic diversity indices were calculated per breed and then averaged per geographical group (see Table 1 for details on the dataset). The summary statistics showed a high variability all over the breeds (see Supplementary Table S1). On average, Asian breeds revealed the lowest values of expected ($H_E = 0.340$) and observed ($H_O = 0.332$) heterozygosity, and the highest average level of inbreeding (0.243). Conversely, the Turkish breeds (KIL and KLS) deviated from the Asian trend and showed values in contrast (H_O and H_E higher than 0.396 and inbreeding lower than 0.091). Similarly, African breeds showed a mean H_O of 0.371 and a mean H_E of 0.366, but a notable lower F_{IS} equal to 0.152. All European breed groups (Europe, Alpine arch, and Italy) reported average heterozygosity indices higher than 0.386 and average inbreeding coefficient lower than 0.121. Values of the Italy group of breeds spanned from ARG ($H_O = 0.416$, $H_E = 0.412$ and $F_{IS} = 0.051$) to MNT_I ($H_O = 0.271$, $H_E = 0.263$ and $F_{IS} = 0.381$) that also represented the range of the whole dataset. ARG, MLG, and JON were the breeds with the highest H_O (0.416, 0.415, and 0.415, respectively), while RCC, ARG, and M×S showed the highest H_E (0.414, 0.412, and 0.412, respectively). The lowest F_{IS} was estimated in ARG (0.051) and the highest, excluding the island-isolated MNT_I, was reported in GIR (0.174). The *MAF* averaged per geographical group followed the trend of

CODE	Breed	Name	n_{raw}	n_{final}	GEO	CODE	Breed	Name	n_{raw}	n_{final}	GEO
Italy	ARG	Argentata dell'Etna*	48	30	Italy	Alpine arch	ALP	Camosciata delle Alpi*	143	30	Italy
Italy	ASP	Capra dell'Aspromonte*	24	24	Italy	Alpine arch	BIO	Bionda dell'Adamello*	24	24	Italy
Italy	BIA	Bianca Monticellana*	24	24	Italy	Alpine arch	LIV	Capra di Livo-Lariana*	24	23	Italy
Italy	CAP	Capestrina*	24	24	Italy	Alpine arch	NVE	Nera di Verzasca*	19	19	Italy
Italy	FAC	Facciuta della Valnerina*	24	24	Italy	Alpine arch	ORO	Orobica*	23	23	Italy
Italy	FUL	Fulva del Lazio*	22	22	Italy	Alpine arch	RCC	Roccoverano*	28	28	Italy
Italy	GAR	Garganica*	40	30	Italy	Alpine arch	SAA	Saanen*	44	30	Italy
Italy	GCI	Grigia Ciociara*	43	30	Italy	Alpine arch	VAL	Valdostana*	24	24	Italy
Italy	GIR	Girgentana*	59	30	Italy	Alpine arch	VLS	Vallesana*	24	24	Italy
Italy	GRF	Garfagnana*	28	27	Italy	Alpine arch	VPS	Capra della Val Passiria*	24	24	Italy
Italy	JON	Jonica*	16	16	Italy						
Italy	MAL	Maltese*	31	30	Italy	Africa	ABR	Abergelle**	53	30	Ethiopia
Italy	MES	Messinese*	24	24	Italy	Africa	BRK	Barki**	153	30	Egypt
Italy	MNT_I	Montecristo (island)*	24	24	Italy	Africa	GUE	Guera**	25	25	Mali
Italy	MON	Capra di Montefalcone*	24	24	Italy	Africa	GUM	Gumez**	41	30	Ethiopia
Italy	M×S	Maltese x Sarda*	36	30	Italy	Africa	NBN	Nubian**	84	30	Egypt
Italy	NIC	Nicastrese*	24	24	Italy						
Italy	RME	Rossa Mediterranea*	78	30	Italy	Asia	BEZ	Bezoar**	7	7	Wild
Italy	SAR	Sarda*	33	30	Italy	Asia	JAT	Jattan**	24	24	Pakistan
Italy	TER	Capra di Teramo*	43	30	Italy	Asia	KAC	Kachan**	24	24	Pakistan
						Asia	KIL	Kil**	25	25	Turkey
Europe	CRS	Corse**	30	30	Corsica	Asia	KLS	Kilis**	40	30	Turkey
Europe	FSS	Fosses**	26	26	France	Asia	PAT	Pateri**	37	30	Pakistan
Europe	LNR	Landrace**	120	30	Denmark	Asia	TAP	Tapri**	24	24	Pakistan
Europe	MLG	Malaguena**	42	30	Spain						
Europe	PTV	Poitevine**	29	29	France	Brazil	CAN	Caninde**	31	30	Brazil
Europe	PYR	Pyrenean**	27	27	France						

Table 1. Dataset composition. *CODE* geographic group, *Breed* breed's acronym, *Name* breed's full name, n_{raw} the raw breed's size, n_{final} reduced size after data management and quality control, *GEO* geographic area of breeding. *Reference: Cortellari et al.²⁰. **Reference: Stella et al.¹⁸.

H_E , showing the highest value in Italy group (0.308) and the lowest in the Asia group (0.261). RCC highlighted the highest breed value (0.328), while CAN the lowest (0.229).

Genetic relationships

A representation of genetic relationships within- and between-breeds comes from multidimensional scaling analysis. Two different graphs of the same MDS analysis were generated: the first (Fig. 1) shows the breeds grouped according to their geographic breeding area as reported in Table 1, and the second (see Supplementary Fig. S1) represents all breeds separately. In Fig. 1, the first component (C1 = 29.88%) clearly separated the European and Italian breeds from the rest of the dataset: African and Asian breeds clustered according to their different breeding location, while European, Alpine, and Italian breeds partially overlapped and showed a north–south gradient of variation. Focusing on this cluster, the spatial breeds grouping highlighted the effective geographic distribution within each country and between countries, with the Southern, Center, and Northern (Alpine cluster) Italian breeds extending along a line. The second component (C2 = 11.25%) reported a partial overlapping between Spanish and Egyptian breeds, as well as between Turkish, Bezoar and the European macro cluster (Alpine, Italian, and French breeds). Considering both C1 and C2 components, the Spanish MLG breed and PYR from France slightly detached from the European macro cluster (see Supplementary Fig. S1). The Neighbor-Net based on pairwise Reynolds' genetic distances also shows distinct clustering according to the different geographic distribution of breeds (see Supplementary Fig. S2). In particular, a phylogenetic divergence between European and non-European breeds was highlighted, with the MAL and MLG breeds at the basis of the separation. The Italian breeds highlighted a complex interweaving of nodes and positioned at the center of the Net. Alpine and French breeds clustered showing a certain degree of distinction from other European breeds. The Brazilian breed (CAN), MNT_I and breeds from Asia showed the highest degree of divergence with respect to the analyzed goat breeds.

Heterozygosity-rich regions

The HRR analysis was performed using two methods of investigation (CR and SW) and reported a total of 13,612 HRR for CR approach and 13,558 HRR for SW approach. Individuals reporting no HRR belonged to Asian breeds (3 animals for KAC, 1 animal for PAT, and 1 animal for TAP). Table 2 shows the descriptive statistics of HRR per breed and geographic cluster. Pearson correlation coefficients were calculated between average values (over the 49 breeds) of the four parameters (N_{HRR} , L_{HRR} , S_{HRR} and D_{HRR}) obtained with the two methods (CR and SW), reporting an $r > 0.999$ for N_{HRR} , S_{HRR} and D_{HRR} and an $r = 0.947$ for L_{HRR} . Based on the high overlapping of the HRR results and high correlation coefficients, we here referred only to the results of the CR approach that reported the highest total number of HRR. Figure 2 graphically presented the results for each parameter per breed and geographical group. The Alpine group had the highest average values within the dataset ($AN_{HRR} = 12.66$, $AL_{HRR} = 0.51$, $AS_{HRR} = 6.38$ and $AD_{HRR} = 0.0026$), with ALP and SAA showing high values above the average both within the group and in the entire dataset. Similar averages were exhibited by the breeds of the Europe group, with MLG breed showing the highest values. The Italy group had mean values of the HRR parameters close to those of Alpine and Europe and showed MON and M×S with the highest values, while MNT_I and GIR were the ones with the lowest. The rest of the goat breeds analyzed showed a fair level of rich heterozygosity and values not far from the group means ($AN_{HRR} = 11.59$, $AL_{HRR} = 0.50$, $AS_{HRR} = 5.80$ and $AD_{HRR} = 0.0024$). Both African and Asian breeds reported low values for all parameters. Within the first group, breeds showed comparable average values, while in the second group, KIL and KLS breeds from Turkey differentiated and showed values above the average.

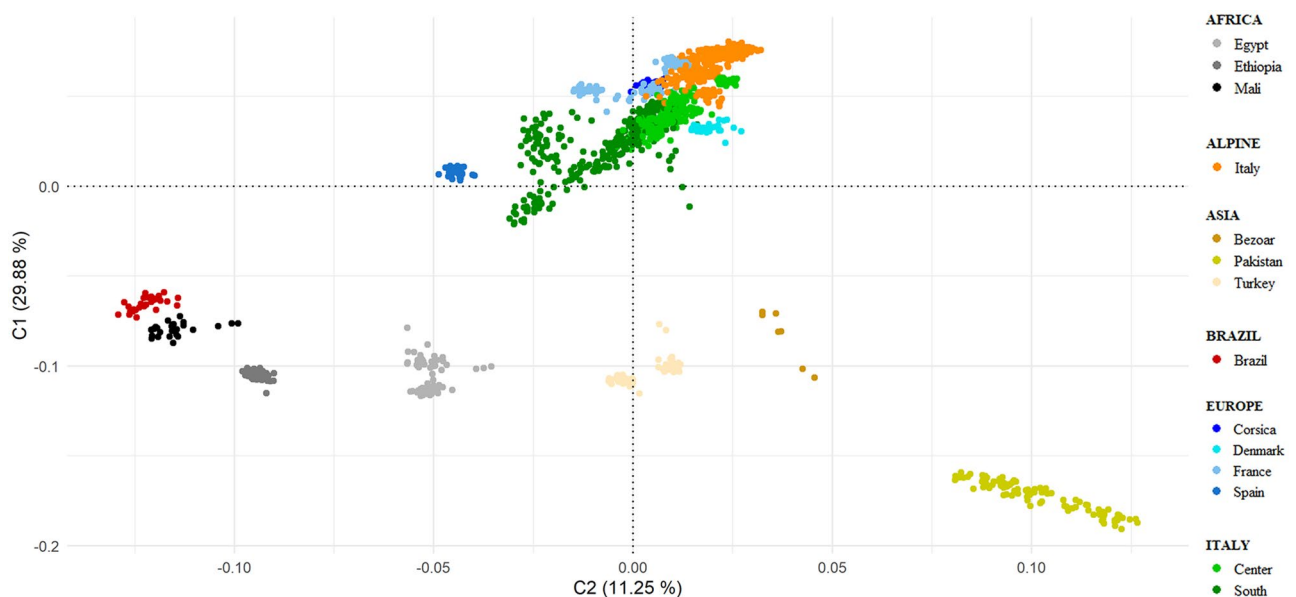


Figure 1. Multidimensional scaling (MDS) plot according to C1 (29.88%) and C2 (11.25%) components. The geographical clusters are represented by different shades of a color. See Table 1 for a full definition of the dataset.

CODE	Breed	CR		SW		Breed	Variable	CR		SW		Breed	Variable	CR		SW	
		Mean ± s.d	Mean ± s.d	Mean ± s.d	Mean ± s.d			Mean ± s.d	Mean ± s.d	Mean ± s.d	Mean ± s.d			Mean ± s.d	Mean ± s.d		
ARG		N_{HRR}	11.37 ± 2.93	11.33 ± 2.87	N_{HRR}			11.07 ± 3.61	11.07 ± 3.61	N_{HRR}			N_{HRR}	13.58 ± 4.86	13.58 ± 4.86	N_{HRR}	13.58 ± 4.86
		L_{HRR}	0.50 ± 0.03	0.50 ± 0.03	L_{HRR}			0.50 ± 0.03	0.50 ± 0.03	L_{HRR}			L_{HRR}	0.50 ± 0.03	0.50 ± 0.03	L_{HRR}	0.50 ± 0.03
		S_{HRR}	5.71 ± 1.66	5.66 ± 1.55	S_{HRR}	GCI		5.51 ± 1.80	5.51 ± 1.80	S_{HRR}		MON	S_{HRR}	6.82 ± 2.58	6.82 ± 2.58	S_{HRR}	6.82 ± 2.58
		D_{HRR}	0.0023 ± 0.0007	0.0023 ± 0.0006	D_{HRR}			0.0022 ± 0.0007	0.0022 ± 0.0007	D_{HRR}			D_{HRR}	0.0028 ± 0.0011	0.0028 ± 0.0011	D_{HRR}	0.0028 ± 0.0011
ASP		N_{HRR}	10.92 ± 2.76	10.88 ± 2.85	N_{HRR}			8.73 ± 3.12	8.70 ± 3.11	N_{HRR}			N_{HRR}	13.57 ± 3.21	13.43 ± 3.28	N_{HRR}	13.43 ± 3.28
		L_{HRR}	0.50 ± 0.04	0.50 ± 0.04	L_{HRR}			0.49 ± 0.05	0.49 ± 0.05	L_{HRR}			L_{HRR}	0.51 ± 0.05	0.50 ± 0.03	L_{HRR}	0.50 ± 0.03
		S_{HRR}	5.44 ± 1.45	5.42 ± 1.47	S_{HRR}	GIR		4.30 ± 1.57	4.28 ± 1.57	S_{HRR}		M×S	S_{HRR}	6.84 ± 1.57	6.65 ± 1.60	S_{HRR}	6.65 ± 1.60
		D_{HRR}	0.0022 ± 0.0006	0.0022 ± 0.0006	D_{HRR}			0.0017 ± 0.0006	0.0017 ± 0.0006	D_{HRR}			D_{HRR}	0.0028 ± 0.0006	0.0027 ± 0.0007	D_{HRR}	0.0027 ± 0.0007
BIA		N_{HRR}	10.88 ± 3.80	10.83 ± 3.80	N_{HRR}			12.07 ± 3.34	12.04 ± 3.32	N_{HRR}			N_{HRR}	10.42 ± 3.76	10.42 ± 3.76	N_{HRR}	10.42 ± 3.76
		L_{HRR}	0.49 ± 0.03	0.49 ± 0.03	L_{HRR}			0.51 ± 0.03	0.50 ± 0.03	L_{HRR}			L_{HRR}	0.50 ± 0.05	0.50 ± 0.05	L_{HRR}	0.50 ± 0.05
		S_{HRR}	5.35 ± 1.87	5.28 ± 1.84	S_{HRR}	GRF		6.12 ± 1.83	6.06 ± 1.79	S_{HRR}		NIC	S_{HRR}	5.25 ± 1.91	5.25 ± 1.91	S_{HRR}	5.25 ± 1.91
		D_{HRR}	0.0022 ± 0.0008	0.0021 ± 0.0008	D_{HRR}			0.0025 ± 0.0007	0.0025 ± 0.0007	D_{HRR}			D_{HRR}	0.0021 ± 0.0008	0.0021 ± 0.0008	D_{HRR}	0.0021 ± 0.0008
CAP		N_{HRR}	12.25 ± 3.34	12.21 ± 3.30	N_{HRR}			11.13 ± 4.32	11.00 ± 4.41	N_{HRR}			N_{HRR}	11.47 ± 2.80	11.40 ± 2.76	N_{HRR}	11.40 ± 2.76
		L_{HRR}	0.50 ± 0.04	0.49 ± 0.03	L_{HRR}			0.51 ± 0.05	0.49 ± 0.04	L_{HRR}			L_{HRR}	0.51 ± 0.04	0.51 ± 0.04	L_{HRR}	0.51 ± 0.04
		S_{HRR}	6.05 ± 1.58	6.00 ± 1.51	S_{HRR}	JON		5.61 ± 2.12	5.43 ± 2.21	S_{HRR}		RME	S_{HRR}	5.89 ± 1.55	5.80 ± 1.49	S_{HRR}	5.80 ± 1.49
		D_{HRR}	0.0025 ± 0.0006	0.0024 ± 0.0006	D_{HRR}			0.0023 ± 0.0009	0.0022 ± 0.0009	D_{HRR}			D_{HRR}	0.0024 ± 0.0006	0.0024 ± 0.0006	D_{HRR}	0.0024 ± 0.0006
FAC		N_{HRR}	12.54 ± 3.97	12.50 ± 3.95	N_{HRR}			12.77 ± 4.05	12.60 ± 4.12	N_{HRR}			N_{HRR}	12.10 ± 2.89	12.03 ± 2.89	N_{HRR}	12.03 ± 2.89
		L_{HRR}	0.50 ± 0.04	0.50 ± 0.04	L_{HRR}			0.50 ± 0.04	0.49 ± 0.03	L_{HRR}			L_{HRR}	0.50 ± 0.04	0.50 ± 0.03	L_{HRR}	0.50 ± 0.03
		S_{HRR}	6.25 ± 2.12	6.20 ± 2.08	S_{HRR}	MAL		6.38 ± 2.08	6.190 ± 2.147	S_{HRR}		SAR	S_{HRR}	6.10 ± 1.56	6.02 ± 1.53	S_{HRR}	6.02 ± 1.53
		D_{HRR}	0.0025 ± 0.0009	0.0025 ± 0.0009	D_{HRR}			0.0026 ± 0.0008	0.0025 ± 0.0009	D_{HRR}			D_{HRR}	0.0025 ± 0.0006	0.0024 ± 0.0006	D_{HRR}	0.0024 ± 0.0006
FUL		N_{HRR}	12.64 ± 3.99	12.64 ± 3.99	N_{HRR}			11.46 ± 2.52	11.46 ± 2.52	N_{HRR}			N_{HRR}	11.30 ± 2.29	11.27 ± 2.29	N_{HRR}	11.27 ± 2.29
		L_{HRR}	0.50 ± 0.04	0.50 ± 0.04	L_{HRR}			0.49 ± 0.04	0.49 ± 0.04	L_{HRR}			L_{HRR}	0.51 ± 0.05	0.51 ± 0.05	L_{HRR}	0.51 ± 0.05
		S_{HRR}	6.28 ± 1.83	6.28 ± 1.83	S_{HRR}	MES		5.62 ± 1.44	5.62 ± 1.44	S_{HRR}		TER	S_{HRR}	5.85 ± 1.31	5.83 ± 1.31	S_{HRR}	5.83 ± 1.31
		D_{HRR}	0.0026 ± 0.0007	0.0026 ± 0.0007	D_{HRR}			0.0023 ± 0.0006	0.0023 ± 0.0006	D_{HRR}			D_{HRR}	0.0024 ± 0.0005	0.0024 ± 0.0005	D_{HRR}	0.0024 ± 0.0005
GAR		N_{HRR}	12.97 ± 4.75	12.97 ± 4.75	N_{HRR}			8.54 ± 2.55	8.54 ± 2.55	N_{HRR}			N_{HRR}	11.59 ± 1.35	11.55 ± 1.34	N_{HRR}	11.55 ± 1.34
		L_{HRR}	0.50 ± 0.03	0.50 ± 0.03	L_{HRR}			0.50 ± 0.03	0.50 ± 0.03	L_{HRR}			L_{HRR}	0.50 ± 0.01	0.50 ± 0.01	L_{HRR}	0.50 ± 0.01
		S_{HRR}	6.41 ± 2.29	6.41 ± 2.29	S_{HRR}	MNT_1		4.27 ± 1.28	4.27 ± 1.28	S_{HRR}		Average	S_{HRR}	5.80 ± 0.69	5.75 ± 0.67	S_{HRR}	5.75 ± 0.67
		D_{HRR}	0.0026 ± 0.0009	0.0026 ± 0.0009	D_{HRR}			0.0017 ± 0.0005	0.0017 ± 0.0005	D_{HRR}			D_{HRR}	0.0024 ± 0.0003	0.0023 ± 0.0003	D_{HRR}	0.0023 ± 0.0003

Italy

Continued

CODE	Breed	CR		SW		Breed	CR		SW		Breed	CR		SW	
		Variable	Mean ± s.d	Variable	Mean ± s.d		Variable	Mean ± s.d	Variable	Mean ± s.d		Variable	Mean ± s.d	Variable	Mean ± s.d
ALP		N_{HRR}	17.33 ± 3.88	N_{HRR}	17.27 ± 3.92		N_{HRR}	10.96 ± 3.42	N_{HRR}	10.96 ± 3.42		N_{HRR}	10.42 ± 3.08	N_{HRR}	10.42 ± 3.08
		L_{HRR}	0.51 ± 0.03	L_{HRR}	0.51 ± 0.03	ORO	L_{HRR}	0.50 ± 0.03	L_{HRR}	0.50 ± 0.03	VAL	L_{HRR}	0.50 ± 0.04	L_{HRR}	0.50 ± 0.04
		S_{HRR}	8.83 ± 1.97	S_{HRR}	8.76 ± 2.04		S_{HRR}	5.46 ± 1.77	S_{HRR}	5.46 ± 1.77		S_{HRR}	5.19 ± 1.49	S_{HRR}	5.19 ± 1.49
		D_{HRR}	0.0036 ± 0.0008	D_{HRR}	0.0036 ± 0.0008		D_{HRR}	0.0022 ± 0.0007	D_{HRR}	0.0022 ± 0.0007		D_{HRR}	0.0021 ± 0.0006	D_{HRR}	0.0021 ± 0.0006
BIO		N_{HRR}	10.71 ± 2.51	N_{HRR}	10.63 ± 2.45		N_{HRR}	13.79 ± 3.89	N_{HRR}	13.75 ± 3.86		N_{HRR}	10.50 ± 3.39	N_{HRR}	10.38 ± 3.50
		L_{HRR}	0.50 ± 0.04	L_{HRR}	0.49 ± 0.04	RCC	L_{HRR}	0.52 ± 0.03	L_{HRR}	0.52 ± 0.04	VLS	L_{HRR}	0.50 ± 0.07	L_{HRR}	0.48 ± 0.03
		S_{HRR}	5.36 ± 1.45	S_{HRR}	5.23 ± 1.31		S_{HRR}	7.13 ± 1.99	S_{HRR}	7.07 ± 1.94		S_{HRR}	5.14 ± 1.57	S_{HRR}	4.95 ± 1.66
		D_{HRR}	0.0022 ± 0.0006	D_{HRR}	0.0021 ± 0.0005		D_{HRR}	0.0029 ± 0.0008	D_{HRR}	0.0029 ± 0.0008		D_{HRR}	0.0021 ± 0.0006	D_{HRR}	0.0029 ± 0.0007
LIV		N_{HRR}	11.87 ± 3.90	N_{HRR}	11.83 ± 3.92		N_{HRR}	18.13 ± 4.96	N_{HRR}	18.00 ± 5.05		N_{HRR}	11.38 ± 3.49	N_{HRR}	11.38 ± 3.49
		L_{HRR}	0.51 ± 0.06	L_{HRR}	0.51 ± 0.04	SAA	L_{HRR}	0.50 ± 0.03	L_{HRR}	0.50 ± 0.03	VPS	L_{HRR}	0.51 ± 0.04	L_{HRR}	0.51 ± 0.04
		S_{HRR}	6.05 ± 2.00	S_{HRR}	5.98 ± 1.97		S_{HRR}	9.12 ± 2.56	S_{HRR}	8.98 ± 2.62		S_{HRR}	5.83 ± 1.89	S_{HRR}	5.83 ± 1.89
		D_{HRR}	0.0025 ± 0.0008	D_{HRR}	0.0024 ± 0.0008		D_{HRR}	0.0037 ± 0.0010	D_{HRR}	0.0037 ± 0.0011		D_{HRR}	0.0024 ± 0.0008	D_{HRR}	0.0024 ± 0.0008
NVE		N_{HRR}	11.53 ± 2.84	N_{HRR}	11.47 ± 2.93							N_{HRR}	12.66 ± 2.85	N_{HRR}	12.61 ± 2.83
		L_{HRR}	0.50 ± 0.04	L_{HRR}	0.49 ± 0.03						Average	L_{HRR}	0.51 ± 0.01	L_{HRR}	0.50 ± 0.01
		S_{HRR}	5.73 ± 1.30	S_{HRR}	5.65 ± 1.43							S_{HRR}	6.38 ± 1.48	S_{HRR}	6.31 ± 1.47
		D_{HRR}	0.0023 ± 0.0005	D_{HRR}	0.0023 ± 0.0006							D_{HRR}	0.0026 ± 0.0006	D_{HRR}	0.0026 ± 0.0006
CRS		N_{HRR}	11.97 ± 3.10	N_{HRR}	11.93 ± 3.03							N_{HRR}	11.41 ± 3.55	N_{HRR}	11.26 ± 3.50
		L_{HRR}	0.49 ± 0.03	L_{HRR}	0.49 ± 0.03	MLG	L_{HRR}	0.51 ± 0.04	L_{HRR}	0.50 ± 0.03	PYR	L_{HRR}	0.51 ± 0.04	L_{HRR}	0.50 ± 0.04
		S_{HRR}	5.88 ± 1.54	S_{HRR}	5.83 ± 1.41		S_{HRR}	6.55 ± 1.88	S_{HRR}	6.45 ± 1.85		S_{HRR}	5.81 ± 1.98	S_{HRR}	5.68 ± 1.92
		D_{HRR}	0.0024 ± 0.0006	D_{HRR}	0.0024 ± 0.0006		D_{HRR}	0.0027 ± 0.0008	D_{HRR}	0.0028 ± 0.0008		D_{HRR}	0.0024 ± 0.0008	D_{HRR}	0.0028 ± 0.0008
FSS		N_{HRR}	13.88 ± 3.76	N_{HRR}	13.81 ± 3.78							N_{HRR}	12.26 ± 0.95	N_{HRR}	12.19 ± 0.96
		L_{HRR}	0.50 ± 0.03	L_{HRR}	0.50 ± 0.03							L_{HRR}	0.51 ± 0.01	L_{HRR}	0.50 ± 0.01
		S_{HRR}	6.89 ± 1.89	S_{HRR}	6.84 ± 1.91	PTV	S_{HRR}	6.04 ± 1.54	S_{HRR}	6.04 ± 1.54	Average	S_{HRR}	6.18 ± 0.44	S_{HRR}	6.12 ± 0.44
		D_{HRR}	0.0028 ± 0.0008	D_{HRR}	0.0028 ± 0.0008		D_{HRR}	0.0025 ± 0.0006	D_{HRR}	0.0025 ± 0.0006		D_{HRR}	0.0025 ± 0.0002	D_{HRR}	0.0025 ± 0.0002
LNR		N_{HRR}	11.50 ± 4.10	N_{HRR}	11.43 ± 4.13										
		L_{HRR}	0.52 ± 0.04	L_{HRR}	0.51 ± 0.04										
		S_{HRR}	5.94 ± 2.18	S_{HRR}	5.86 ± 2.21										
		D_{HRR}	0.0024 ± 0.0009	D_{HRR}	0.0024 ± 0.0009										
ABR		N_{HRR}	6.83 ± 2.25	N_{HRR}	6.80 ± 2.20							N_{HRR}	8.20 ± 3.49	N_{HRR}	8.16 ± 3.40
		L_{HRR}	0.52 ± 0.06	L_{HRR}	0.51 ± 0.05							L_{HRR}	0.50 ± 0.03	L_{HRR}	0.50 ± 0.04
		S_{HRR}	3.56 ± 1.32	S_{HRR}	3.50 ± 1.22	GUE	S_{HRR}	4.07 ± 1.71	S_{HRR}	4.05 ± 1.67	NBN	S_{HRR}	3.12 ± 1.59	S_{HRR}	3.12 ± 1.59
		D_{HRR}	0.0015 ± 0.0005	D_{HRR}	0.0014 ± 0.0005		D_{HRR}	0.0017 ± 0.0007	D_{HRR}	0.0017 ± 0.0007		D_{HRR}	0.0013 ± 0.0007	D_{HRR}	0.0013 ± 0.0007
BRK		N_{HRR}	7.37 ± 3.88	N_{HRR}	7.33 ± 3.84							N_{HRR}	7.17 ± 2.79	N_{HRR}	7.13 ± 2.79
		L_{HRR}	0.48 ± 0.05	L_{HRR}	0.48 ± 0.05							L_{HRR}	0.50 ± 0.04	L_{HRR}	0.50 ± 0.01
		S_{HRR}	3.59 ± 1.97	S_{HRR}	3.56 ± 1.94	GUM	S_{HRR}	3.60 ± 1.44	S_{HRR}	3.56 ± 1.43	Average	S_{HRR}	3.59 ± 0.34	S_{HRR}	3.56 ± 0.33
		D_{HRR}	0.0015 ± 0.0008	D_{HRR}	0.0014 ± 0.0008		D_{HRR}	0.0015 ± 0.0006	D_{HRR}	0.0015 ± 0.0009		D_{HRR}	0.0015 ± 0.0001	D_{HRR}	0.0015 ± 0.0002

Continued

CODE	Breed	CR		SW		Breed	Variable	CR		SW		Breed	Variable	CR		SW	
		Mean ± s.d	Mean ± s.d	Mean ± s.d	Mean ± s.d			Mean ± s.d	Mean ± s.d	Mean ± s.d	Mean ± s.d			Mean ± s.d	Mean ± s.d		
Asia	BEZ	N_{HRR}	1.86 ± 0.90	1.86 ± 0.90	9.96 ± 3.81	9.96 ± 3.81		N_{HRR}	9.96 ± 3.81	9.96 ± 3.81			N_{HRR}	4.20 ± 2.12	4.20 ± 2.12	4.20 ± 2.12	4.20 ± 2.12
		L_{HRR}	0.44 ± 0.03	0.44 ± 0.03	0.48 ± 0.02	0.48 ± 0.02	KIL	L_{HRR}	0.48 ± 0.02	0.48 ± 0.02	PAT	L_{HRR}	0.56 ± 0.17	0.56 ± 0.17	0.56 ± 0.17	0.56 ± 0.17	
		S_{HRR}	0.83 ± 0.42	0.83 ± 0.42	4.77 ± 1.86	4.77 ± 1.86		S_{HRR}	4.77 ± 1.86	4.77 ± 1.86		S_{HRR}	2.30 ± 1.05	2.30 ± 1.05	2.30 ± 1.05	2.30 ± 1.05	
		D_{HRR}	0.0003 ± 0.0002	0.0003 ± 0.0002	0.0019 ± 0.0007	0.0019 ± 0.0008		D_{HRR}	0.0019 ± 0.0007	0.0019 ± 0.0008		D_{HRR}	0.0009 ± 0.0004	0.0009 ± 0.0004	0.0009 ± 0.0004	0.0009 ± 0.0004	
	JAT	N_{HRR}	3.46 ± 1.64	3.46 ± 1.64	8.50 ± 2.43	8.50 ± 2.43		N_{HRR}	8.50 ± 2.43	8.50 ± 2.43			N_{HRR}	3.96 ± 2.05	3.96 ± 2.05	3.96 ± 2.05	3.96 ± 2.05
		L_{HRR}	0.50 ± 0.05	0.50 ± 0.05	0.51 ± 0.03	0.51 ± 0.03	KLS	L_{HRR}	0.51 ± 0.03	0.51 ± 0.03	TAP	L_{HRR}	0.50 ± 0.13	0.50 ± 0.13	0.50 ± 0.13	0.50 ± 0.13	
		S_{HRR}	1.77 ± 0.91	1.77 ± 0.91	4.31 ± 1.22	4.25 ± 1.22		S_{HRR}	4.31 ± 1.22	4.25 ± 1.22		S_{HRR}	2.03 ± 1.04	2.03 ± 1.04	2.03 ± 1.04	2.03 ± 1.04	
		D_{HRR}	0.0007 ± 0.0004	0.0007 ± 0.0004	0.0018 ± 0.0005	0.0017 ± 0.0005		D_{HRR}	0.0018 ± 0.0005	0.0017 ± 0.0005		D_{HRR}	0.0008 ± 0.0004	0.0008 ± 0.0004	0.0008 ± 0.0004	0.0008 ± 0.0004	
	KAC	N_{HRR}	2.21 ± 1.67	2.21 ± 1.67				N_{HRR}						N_{HRR}	4.88 ± 3.12	4.88 ± 3.12	4.87 ± 3.11
		L_{HRR}	0.45 ± 0.18	0.45 ± 0.18				L_{HRR}						L_{HRR}	0.49 ± 0.04	0.49 ± 0.04	0.49 ± 0.04
		S_{HRR}	1.15 ± 0.90	1.15 ± 0.90				S_{HRR}						S_{HRR}	2.45 ± 1.52	2.45 ± 1.52	2.44 ± 1.51
		D_{HRR}	0.0008 ± 0.0004	0.0005 ± 0.0004				D_{HRR}						D_{HRR}	0.0010 ± 0.0006	0.0010 ± 0.0006	0.0010 ± 0.0006
Brazil	CAN	N_{HRR}	8.60 ± 2.97	8.60 ± 2.97				N_{HRR}									
		L_{HRR}	0.51 ± 0.05	0.51 ± 0.05				L_{HRR}									
		S_{HRR}	4.39 ± 1.53	4.39 ± 1.53				S_{HRR}									
		D_{HRR}	0.0018 ± 0.0006	0.0018 ± 0.0006				D_{HRR}									

Table 2. Descriptive statistics for Heterozygosity-rich regions (HRR) by consecutive runs (CR) and sliding window (SW) methods. Averages per geographic group are also tabulated (AN_{HRR} , AL_{HRR} , AS_{HRR} and AD_{HRR}). CODE geographical cluster, Breed breed's acronym, N_{HRR} average HRR number, L_{HRR} average length of HRR (in Mbp), S_{HRR} mean genome length covered by HRR segments (in Mbp), D_{HRR} mean coefficient of diversity, s.d. standard deviation.

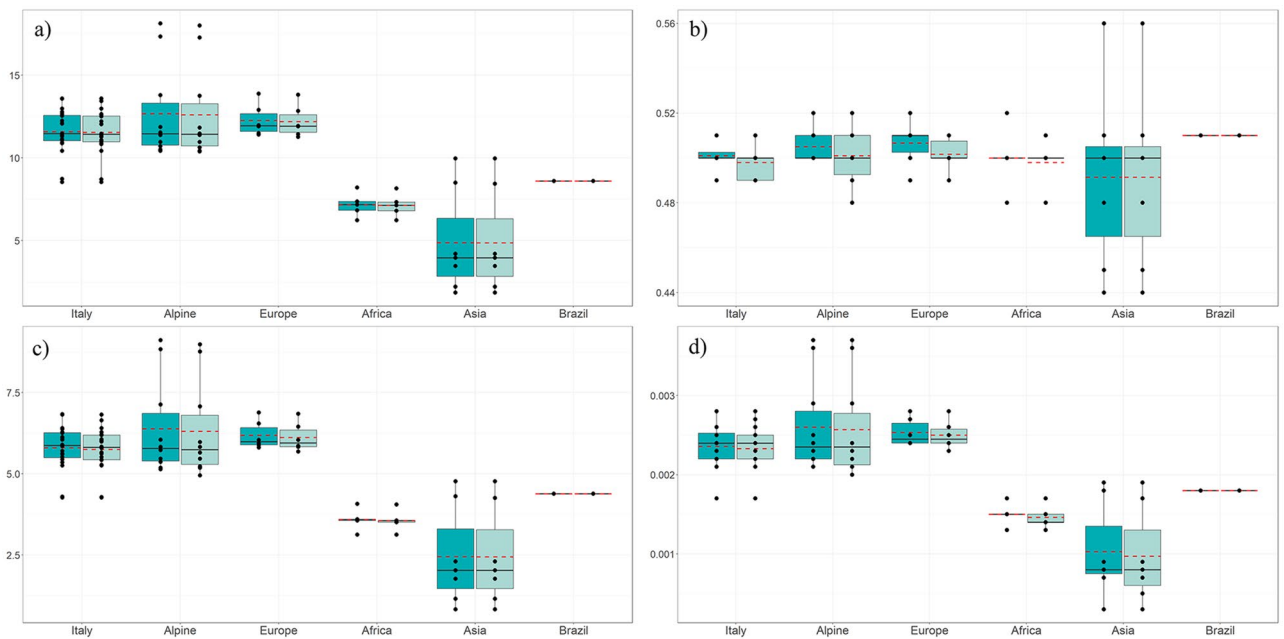


Figure 2. Boxplots of the Heterozygosity-rich regions (HRR) indices aggregated by geographic area. **(a)** N_{HRR} = average number of HRR per breed, **(b)** L_{HRR} = average length of HRR per breed (in Mbp), **(c)** S_{HRR} = mean genome length covered by HRR segments (in Mbp), **(d)** D_{HRR} = average coefficient of diversity. The different colors indicate the two approaches, CR (blue) and SW (light blue). Each boxplot extends from the 25th to the 75th percentile and shows the average (red dashed line) and the median (black horizontal line) group value. The black dots represent the single-breed results.

Heterozygosity islands and gene annotation

According to the filter applied for the identification of HRR islands, 1,665 SNPs were considered to form a total of 166 HRR islands (see Supplementary Table S2). The average number of islands on the overall dataset was 3.39 per breed, and the highest number of HRR and islands was found in MNT_I (103 and 16, respectively), while no HRR islands were detected for the BRK breed. To explain the workflow in HRR detection, we report the results of SAA. The 0.1 percentile threshold for cutting off the SNP-in-HRR p -values allowed the detection of 164 SNPs for both CR and SW. These markers identified 89 in-breed HRR, but only 21 exceeded the frequency (20% within-breed) and length (minimum 4 SNPs) thresholds, for a final detection of 3 HRR islands. Table 3 reports details of each HRR island including genomic coordinates, breeds, and annotated genes. CHI1, CHI11, CHI12 and CHI18 were the chromosomes harboring the highest number of HRR islands (sum of breeds' HRR islands), reporting a total of 33 (3 unique), 16 (4 unique), 25 (4 unique) and 36 (4 unique) islands, respectively. Several breeds had HRR islands in common. Notably, 31 (including 22 Italian breeds) out of 49 breeds shared an island in CHI1 (CHI1-A from 131.88 to 132.54 Mb). Similarly, a portion of the genome ranging from 36.36 to 40.01 Mb on CHI18, including two HRR islands, was largely shared by several European breeds: 19 breeds had the CHI18-A island in common, and 13 breeds the CHI18-B. Finally, 15 breeds with different geographic distributions reported a common HRR island in CHI12 (CHI12-A from 49.80 to 51.28 Mb). All HRR islands are plotted in Fig. 3, highlighting the chromosome and breeds involved. The gene enrichment performed on genes annotated within HRR islands revealed a total of 74 different processes, mainly biological processes involved into regulation of metabolic processes (CHI12-A), sensor organ development (CHI11-C) and activation of immune response (CHI8-C, CHI16-D) (see Supplementary Table S3). No results were found for CHI1-A shared island.

Discussion

In livestock, selective breeding tends to progressively decrease the diversity and resilience of target breeds²¹. Locally reared breeds still seem to maintain their rustic characteristics thanks to the lower anthropic impact²². Goats are well known for being the most suitable livestock species for adapting to harsh environments^{18,19}. In particular, local goat breeds represent a way to the sustainability of animal production in marginal areas in both developed and developing countries^{18,23}. The domestication process has triggered different evolutionary trends, which over the centuries have led to the development of several breeds with different productive aptitudes. This study aimed to provide a general overview of the relationship among the studied goat breeds and to highlight their HRR patterns as repositories of advantageous gene diversity related to adaptation and resilience processes^{6,7}. Identifying a species-specific HRR pattern could provide a clearer view of the mechanisms regulating the genome modelling following anthropogenic selection combined with environmental interaction.

Genetic diversity indices

The high degree of variability in the dataset reflected the expected biodiversity^{19,23,24} among different breeds. The choice of the dataset, composed of goat breeds claiming a general dairy productive orientation, led to an

Chromosome	Code	From	To	Breed	Genes
1	CHI1-A	131,887,965	132,542,315	ALP, ARG, LNR, GUE, KIL, BIO, BIA, MLG, GUM, KLS, LIV, CAB, PYR, NBN, TAP, VAL, FAC, VPS, FUL, GAR, GCI, GIR, JON, MAL, MES, MON, M×S, NIC, RME, SAR, TER	SLC35G2, STAG1, PCCB
	CHI1-B	119,276,440	119,721,981	MNT_I	
	CHI1-C	29,543,132	30,017,015	CAN	
2	CHI2-A	61,002,073	61,568,008	MNT_I	
3	CHI3-A	24,833,445	25,571,553	MNT_I, LNR, GUE, JAT, GUM, KLS, TAP	DMRTA2, FAF1, CDKN2C
	CHI3-B	68,974,808	69,377,131	VLS	TGFBR3, BRDT, EPHX4, LOC102178967
	CHI3-C	61,360,209	61,753,348	MNT_I	BTBD8, SAMD13, LOC102168228, DNASE2B, RPF1, GNG5, SPATA1, CTBS, SSX2IP
	CHI3-D	104,971,744	105,344,042	MNT_I	ASH1L, DAP3, MSTO1, GON4L, YY1AP1, SYT11, RIT1, KHDC4, RXFP4, ARHGEF2
4	CHI4-A	78,787,857	79,098,602	FSS	
5	CHI5-A	69,680,879	70,284,360	ABR	ASCL4, RTCB, BPIFC, FBXO7, SYN3, TIMP3
	CHI5-B	73,464,107	73,868,942	CAN	MYH9, TXN2, FOXRED2, EIF3D, CACNG2, IFT27
6	CHI6-A	23,810,805	24,438,981	MNT_I	PPP3CA
7	CHI7-A	59,924,105	60,515,404	CAP, TER	SIL1, CTNNA1, LRRTM2
	CHI7-B	1,891,473	2,346,215	VLS	PJA2
	CHI7-C	71,887,410	72,271,623	VLS	NSD1, FGFR4, ZNF346, UIMC1, HK3, UNC5A
8	CHI8-A	74,954,737	75,321,840	ALP, VPS	AQP3, NOL6, UBE2R2, UBAP2
	CHI8-B	38,476,907	38,695,165	MNT_I	RANBP6, KIAA2026, MLANA
	CHI8-C	39,057,865	39,289,111	PAT	CD274, PLGRKT, JAK2, INSL6
9	CHI9-A	50,569,758	51,104,960	MNT_I	SYNCRIP
10	CHI10-A	51,511,315	51,976,359	FAC, M×S, LNR	ADAM10, LOC102184991, MINDY2, SLTM, RNF111, CCNB2, MYO1E
	CHI10-B	49,756,418	50,111,884	ORO	TCF12
11	CHI11-A	37,760,059	38,565,140	ALP, ARG, LNR, BEZ, LIV, CAP, ORO, GRF, SAA, NIC, TER	CFAP36, PPP4R3B, PNPT1, EFEMP1, CCDC85A
	CHI11-B	93,996,800	94,603,621	FAC, JON, GUE	CRB2, DENND1A
	CHI11-C	60,993,493	61,515,733	PTV	EHBPI, OTX1, WDPCP
	CHI11-D	71,126,715	71,509,659	MLG	BABAM2
12	CHI12-A	49,805,838	51,283,345	BIO, GRF, CRS, ABR, JAT, ORO, RME, GUM, KAC, RCC, SAR, NBN, PAT, PTV, TAP	ATP12A, RNF17, CENPJ, PARP4, MPHOSPH8, PSPC1, ZMYM5, GJA3, GJB2, GJB6, CRYL1, IFT88, IL17D, EEF1AKMT1, XPO4, LATS2, SAP18, SKA3, MRPL57, ZDHHC20, MICU2
	CHI12-B	43,916,255	44,471,339	VPS, FUL, MLG, MAL, MNT_I, M×S, RME, SAR	
	CHI12-C	44,789,048	45,251,150	VAL	
	CHI12-D	36,642,968	37,043,615	MNT_I	
13	CHI13-A	22,047,464	22,437,842	ORO, CRS, LNR	SKIDA1, MLLT10, DNAJC1
	CHI13-B	46,470,353	46,869,126	ORO, VLS	PRNP, PRND, RASSF2, SLC23A2, TMEM230, PCNA
15	CHI15-A	65,986,490	66,544,220	MNT_I	GUCY1A2, CWF19L2, C11orf97, FUT4, PIWIL4
	CHI15-B	72,589,246	73,074,663	MNT_I	CNTN5
	CHI15-C	44,309,482	44,822,467	LNR	PDE3B, CYP2R1, LOC102189885, CRSP-1, CRSP-2, LOC102188877
16	CHI16-A	33,556,420	33,900,422	FUL	WDR64, CHML, KMO, FH
	CHI16-B	54,180,217	54,526,196	MAL	RABGAP1L
	CHI16-C	43,284,994	43,460,646	TER	RERE
	CHI16-D	35,427,296	35,987,762	CAN	NME7, LOC108637771, CCDC181, SLC19A2, F5, SELP, SELL, SELE
17	CHI17-A	54,979,564	55,407,415	ORO	INPP4B
	CHI17-B	3,201,875	3,848,371	CAN	TTC28, PITPNB
	CHI17-C	60,310,137	61,213,780	PAT	TMEM184C, PRMT9, ARHGAP10, NR3C2
Continued					

Chromosome	Code	From	To	Breed	Genes
18	CHI18-A	36,360,862	37,201,445	ALP, ASP, CRS, GUE, KIL, VAL, BIA, MLG, VLS, CAP, VPS, FUL, FSS, GCI, PTV, GIR, GRE, MES, MON, M×S, SAR	<i>RIPOR1, CTFC, CARMIL2, ACD, PARD6A, C16orf86, ENKD1, GFOD2, RANBP10, CENPT, TSNAXIP1, THAP11, NUTF2, EDC4, NRN1L, PSKH1, PSMB10, LCAT, SLC12A4, DPEP3, DPEP2, DDX28, DUS2, ESRP2, NFATC3, PLA2G15, SLC7A6, SLC7A6OS, PRMT7, SMPD3, ZFP90</i>
	CHI18-B	39,567,796	40,014,717	ALP, CAP, BIO, FAC, LIV, GAR, NVE, GCI, ORO, MON, SAA, VPS, PTV	<i>ZFHX3</i>
	CHI18-C	40,300,941	40,654,294	LNR	<i>PKD1L3, IST1, ZNF821, ATXN1L, AP1G1, PHLPP2, MARVELD3</i>
	CHI18-D	3,788,363	4,113,124	CAN	<i>LOC102173875, LOC102174148, LOC102176156, BCAR1, BCNT, P97BCNT, TMEM170A, LOC102176721, TMEM231, GABARAPL2</i>
20	CHI20-A	23,340,984	23,958,224	MNT_I	<i>IL31RA, DDX4, SLC38A9, PLPP1, MTREX, LOC102190074, DHX29, CCNO, MCIDAS</i>
21	CHI21-A	51,297,564	51,616,572	BIA	
	CHI21-B	46,698,076	47,218,999	MNT_I	<i>MIPOL1, FOXA1</i>
	CHI21-C	64,171,824	64,509,595	MNT_I	<i>EML1, EVL, DEGS2, YY1, SLC25A29, SLC25A47</i>
22	CHI22-A	2,644,329	3,178,952	MNT_I	<i>CMC1, AZI2, ZCWPW2</i>
23	CHI23-A	36,431,298	36,745,815	RCC, VAL, BIA	<i>ZFAND3</i>
24	CHI24-A	43,465,500	44,003,472	SAA	<i>PTPN2, SEH1L, LDLRAD4, FAM210A, RNMT, MC5R, MC2R</i>
	CHI24-B	20,161,766	20,444,877	VAL	<i>KIAA1328</i>
25	CHI25-A	41,639,292	42,026,061	GUE	<i>CHST12, EIF3B, SNX8, MRM2, NUDT1, MAD1L1</i>
27	CHI27-A	36,831,781	37,182,122	BIA	<i>AGA, NEIL3</i>
	CHI27-B	12,114,641	12,380,291	MON	<i>FGFR1, LETM2, NSD3, PLPP5, DDHD2, BAG4, LOC106503669, LSM1, STAR, ASH2L</i>

Table 3. Distribution of HRR islands per chromosome. For each HRR island, the table reports the island code, spanning range in base pairs (from-to), the list of breeds carrying the island (Breed) and the NCBI name of the annotated genes.

evident imbalance between the European panorama and the rest of the geographical groups. Therefore, the average diversity values per group need to be considered a function of the notable numerical disparity between European goats and breeds from other parts of the world. Colli et al.¹⁹ showed a pattern of observed heterozygosity in worldwide goat breeds that matched our results. They highlighted a clear association of this parameter with geography reflecting climatic conditions and breeding management. Therefore, the higher values are probably caused by admixture events between breeds reared in areas with extensive practices (i.e., transhumance in South Italy). The two Sicilian breeds—Argentata dell’Etna and Messinese—represent a known case of occasional admixture linked to the shared breeding area, which impacts on high values of both H_O and H_E ²⁵. Despite their status as cosmopolitan breeds and notoriously more subject to selection, the Saanen and the Camosciata delle Alpi showed relatively high heterozygosity and low molecular inbreeding, proving the proper management of their selective plan. Notably, the feral Montecristo goat resulted in the lowest heterozygosity of the whole dataset, likely reflecting a strong inbreeding due to its geographic isolation. Somenzi et al.²⁶ have retraced the evolutionary history of this feral population, suggesting repeated bottleneck events and founder effects that characterized the demographic history of the insular goat.

Genetic relationships

Genetic relationships analyses pointed out the differences between European and non-European goats, likely as a consequence of gene pool divergence between domesticated breeds and ancestral populations from the regions of the Fertile Crescent¹⁹. We can also assume that physical barriers, such as long distances and mountain ranges, have marked this divergence. Moreover, bias related to the design of the Illumina chip panel must be considered to explain the clear divergence of the two macro goat clusters^{24,27}. The Multidimensional-scaling analysis highlighted the high cohesiveness of the European breeds in their totality (Mediterranean, Central-Europe and Alpine goats) when compared with the rest of the data set. The north–south geographic cline and the greater homogeneity of some breeds such as the Malaguena and the Pyrenean, were particularly evident. More significant variability among the different breeds of the other geographic groups was found (e.g., Pakistani vs Turkish goats, Ethiopian vs Egyptian goats vs Malian goats), probably due to several factors involving the diverse histories of local communities, the different levels of gene flow affecting breeds from different geographic areas (e.g. Pakistani vs Turkish goats)^{19,28}, the eventual influence of the European genome²⁹, as well as the original complexity of the African and Asian goat stock origin^{30,31}. The phylogenetic nodes of the Neighbor-Net based on pairwise Reynolds’ genetic distances better represented the genetic relationships among breeds, confirming the genetic closeness of the Europe, Alpine and Italy geographic groups and their divergence from the African and Asiatic strains. The Montecristo goat showed a close relationship to the other insular goat (Corse) and genetic proximity to the Italy group possibly because of the recent inputs of domestic stocks (twentieth century) as already reported in a previous study²⁶.

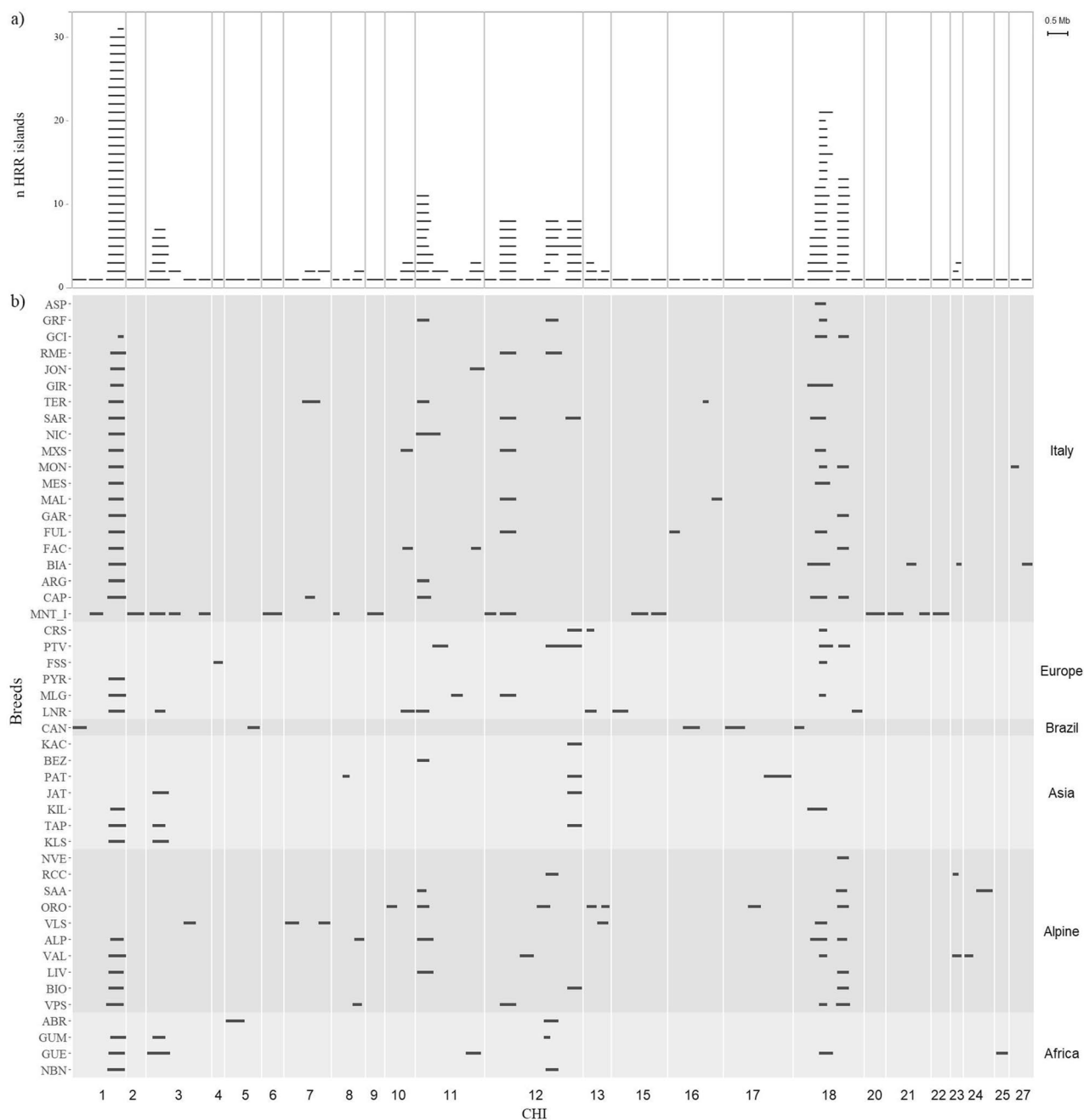


Figure 3. Graphical distribution of Heterozygosity-Rich Regions (HRR) islands per chromosome. Detected HRR in the meta-population (a) and per breed (b). The gap between consecutive islands within a chromosome and between consecutive chromosomes do not correspond to the real distance in bps.

Heterozygosity-rich regions

HRR are heterozygous genomic regions which are potentially associated with disease resistance, immunity, and adaptation processes. The higher the level of diversity, the better the biological response of species to environmental changes or new diseases. Consecutive Runs and Sliding Window are the two approaches currently used to detect stretches of consecutive homozygosity (ROH) and HRR in livestock species e.g.^{3,7,8,32,33}. Although only few studies have focused on identifying HRR islands with both methods of detection, it seems that the use of the Consecutive Runs approach is preferred^{3,7,15}. In this study, we investigated both methods, highlighting the overlapping results for HRR identification. Strong correlation between CR and SW was estimated for all the parameters, indicating nearly identical results for all breeds. In a similar study on HRR in pigs, Bordonaro et al.¹⁴ found a Pearson correlation coefficient higher than 0.96 for all four parameters (N_{ROH} , L_{ROH} , S_{ROH} , F_{ROH}), confirming the obtaining of overlapping results between the two approaches. As mentioned by Biscarini et al.⁶, the choice of the parameter values is of fundamental importance in detecting ROH/HRR because of its effect on results. The information present in the scientific literature regarding the setting of the detection parameters is not

very detailed. According to Mulim et al.⁷, the allowed number of opposite and missing SNP and the gap between two consecutive markers are the factors that mainly affect the detection of runs. In our preliminary screening, we evaluated the number of HRR found by testing several combinations of parameters (the minimum number of SNPs in a run, the minimum length, and the number of missing/opposite markers). The minimum number of SNPs in a HRR (*minSNP*) and the maximum number of SNP with opposite genotype in a window (*maxOppWindow*) were the parameters that most influenced the number of HRR found in SW method, while for the CR method it was the number of opposite genotypes allowed in a HRR (*maxOppRun*)³⁴. Selli et al.⁸ conducted a methodological study on several combinations of detection parameters in Sliding Window approach to investigate HRR. They found that a reduction in the minimum number of SNPs in a run (*minSNP*) and in a window (*windowSize*), as well as the number of homozygous and missing SNPs allowed (*maxOppRun* and *maxMissRun*), caused a decrease in HRR found, whereas a reduction in the minimum length of the run (*minLengthBps*) caused an increase in the number of HRR. Similarly, *minSNP* and *windowSize* positively affect the HRR length. In this scenario, it is unclear how different combinations might affect the results, and further analyses are needed⁶. In general, the within-breed occurrence of HRR is lower than that of ROH^{2,3,6}. Indeed, we found an average number of HRR per breed lower than 18, while Cortellari et al.²³ found an average number of ROH of 31.5 in the same Italian breeds data. Other studies on HRR reported an average of 9.5 in semi-feral Chillingham cattle population known to be characterized by a strong inbreeding⁴, less than 9 in the Italian Maremmana cattle breed⁶, 121.5 in Pinzgauer cattle breed², 57.8 in commercial turkeys³⁵, 52.2 in a local horse breed³, ~40 in Duroc pig⁹ and 139.6 in sheep⁸. These notable differences may be attributed to the species of interest, to the breeds analysed, and to the different SNP arrays and parameters used for the HRR detection. All the cited studies also found that HRR were much shorter than ROH, confirming our general mean of ~500 kb in length and reflecting low coverage of the entire genome (maximum of 6.38 Mb for the Alpine breed group). These results might be related to the absence of missing and opposite SNPs in the detection approach, that usually reduces the size of detected ROH⁶ and, similarly, of HRR. We highlighted a notable difference between the results obtained for European (Italy, Alpine, Europe groups) and non-European clusters (Africa, Asia, Brazil). We assume that this result, also reported by Li et al.¹⁵ in Chinese goats compared to worldwide breeds, could come from ascertainment bias related to the SNP array design based on European goat breeds^{24,27}. Worth of note are the outcomes of Maltese × Sarda (M × S) goats. This population represents the only crossbreed sample included in the data set and the high values for all the indices were expected and probably related to the combination of relatively different genomes.

Similarly, Mulim et al.⁷ reported the borderline case of the Montana cattle that combines different genomes due to the crossbreeding between *Bos taurus indicus* and *Bos taurus taurus* and, consequently, showed the highest amount of HRR in their dataset.

Heterozygosity islands and gene annotation

The identification of recurrent HRR for each breed highlighted wide regions of shared heterozygosity in the chromosomes CHI1, CHI11, CHI12, and CHI18. Williams et al.⁴ observed that heterozygosity-rich regions in the Chillingham cattle genome appear to be not randomly distributed but rather grouped in particular chromosomal locations. Moreover, some of the detected heterozygosity islands had already been recognized as homozygosity islands in previous studies^{15,36}. Classifying the same chromosomal regions as both homozygosity-rich and heterozygosity-rich emphasizes the diversity among breeds within the goat species. The different breeding strategies and the different production orientations¹⁵ rather than the adaptation to different environments have differentially shaped the genome of breeds¹⁴. CHI1-A (131.88–132.54 Mb) was the most shared island within the whole dataset. *STAG1* and *PCCB* genes, mapped within CHI1-A, are linked to reproduction-related traits and were reported to experience balancing selection in goat breeds, playing a significant role in the domestication of the species¹⁵. Notably, this HRR island was not present in either Bezoar or Montecristo, the two wild goat populations in the dataset. The second most common HRR island was CHI18-A (36.36–37.20 Mb). It was detected mainly in Italian and Alpine breeds, matched a ROH island found in goats from Europe, Africa and Asia by Bertolini et al.³⁶, and it was reported as a possible signature of selection for fiber production with a low number of SNPs supporting this finding³⁷. This region harbors genes involved in productive mechanisms in other species, such as the *RIPOR1* gene for feeding intake control in the Nellore cattle breed³⁸ or the *PSKH1* gene listed as a candidate gene for mean staple length, live weight, greasy fleece weight, and mean fiber diameter in Merino sheep³⁹. The *EDC4* gene was reported to mediate the post-transcriptional regulation of IL-6 cytokine playing an important role in immune response⁴⁰. The nearby island CHI18-B (39.56–40.01 Mb) was detected in 7 Alpine breeds out of 10 and in 5 breeds of the Italy group, allowing us to hypothesize a possible correlation to a well-defined geographic area. This short region harbors the *ZFH3* gene, also known as *ATBFI*, which has a well-known function in humans as a transcription factor that regulates myogenic and neuronal differentiation, with similar effects on Chinese native goat breeds where its polymorphisms influence animal growth rate⁴¹. As far as found in literature, a study on 3 Chinese local goat breeds confirmed the association of this gene with development traits⁴². In our results, we did not find this HRR island in any of the six Asian breeds. Bertolini et al.³⁶ found two ROH regions in goat chromosome 12 (from 43 to 44 Mb in European breeds and from 50 to 51 Mb in worldwide goats) which overlapped two of our HRR hotspots CHI12-B (43.91–44.47 Mb) and CHI12-A (49.80–51.28 Mb), harboring genes that are related to ectodermal, nervous system, and hearing function (*GJB2* and *GJB6*)^{43,44} as well as gonad development (*SAP18*)⁴⁵. CHI12-A was also reported as a ROH hotspot by Li et al.¹⁵ in Chinese and other worldwide goat breeds; the authors underlined this region in common with a ROH hotspot in Chinese sheep breeds⁴⁶ and possibly under parallel selection between goats and sheep species before the domestication due to the presence of a series of genes (*GJA3*, *GJB2* and *GJB6*) associated with perception senses, such as sight and hearing essential for surviving in harsh environments (lack of food and presence of wild enemies). These genes were also identified under positive selection for growth (body size, skeletal and embryonic development)

in Boar goat⁴⁷ and in Barki goat and sheep⁴⁸. Signatures of selection localized in CHI12-A (49.80–51.28 Mb) were reported in goats by Serranito et al.²² which listed a series of annotated genes involved in the adaptation of small ruminants to arid environments (*RNF17*, *ATP12A*, *GJB2*) and high altitude (*ATP12A*, *PARP4*, *ZMYM5*, *PSPCI*, *CENPJ*, *GJB2*). Moreover, the interleukin17D (*IL17D*) gene, spanning from ~ 50.90 to ~ 50.93 Mb and belonging to the *IL17* family of cytokines, was closely associated to host defense and immune response in humans^{49,50}. This region perfectly matched the one found in 5 commercial and local goat breeds by Biscarini et al.⁵¹. CHI11-A (37.76–38.56 Mb) was a shared island among 11 breeds, mainly from the Alpine and Italy groups. This region overlapped the ROH island found by Bertolini et al.³⁶, which detected a signature of selection in a previous study³⁷. In our dataset, the two cosmopolitan breeds (Saanen and Camosciata delle Alpi) reported this HRR island in which several mapped genes have a role in livestock production. In particular, the *PPP4R3B* gene was associated to thermotolerance in the African N'Dama cattle breed⁵²; Berihulay et al.⁵³ found this gene involved in gluconeogenesis and lipidic metabolism in Abergelle goats. The *PNPT1* gene was reported implicated into RNA transport in pigs⁵³ and Bandur sheep⁵⁴. Previous studies linked the *EFEMP1* gene with traits of interest in livestock species. Zhang et al.⁵⁵ identified this gene as a significant influencer of the oleic acid content in the meat of the Wagyu × Angus cattle breed. Or else, *EFEMP1* was previously identified as a differentiated expressed gene in muscle for residual feed intake in Nelore steers and seems to be regulated by a transcriptional factor (*TCF4*)⁵⁶. The worth of notice are the islands in CHI8, even though they are present only in few breeds: CHI8-A (74.95–75.32 Mb) and CHI8-C (39.05–39.29 Mb) seem to have a strong correlation with immune response. In particular, the *AQP3* gene in the first island was found in a ROH hotspot in Ganxi and Guangfeng goats by Li et al.¹⁵ who stated its involvement in immunoreactions. The *CD274* and *JAK2* genes in CHI8-C were reported in several biological processes for immune response in our gene enrichment results.

Heterozygosity islands in feral Montecristo goat

The evolutionary history of the feral Montecristo goat, geographically isolated in the homonymous island, subject to repeated bottleneck phenomena, and in the absence of anthropic management, albeit with a partial introduction of domestic germplasm²⁶, represents an outlier case in the analyzed dataset. Moreover, notwithstanding the highest inbreeding coefficient (F_{IS}) and the lowest heterozygosities (H_O and H_E) of the whole dataset, Montecristo goat reported the highest number of HRR. Events of balancing selection characterizing this feral goat population might be assumed during the shape of its genomic structure. Despite a high inbreeding, these events may have led to an increase in short segments of heterozygosity, that are strongly associated with fitness and survival traits^{2,12}. Indeed, *FAF1* gene in CHI3-A (24.83–25.57 Mb) was highly associated with tolerance to *Theileria* infection in African cattle⁵⁷. On chromosome 3, HRR islands (CHI3-C and CHI3-D) harbored two genes (*SPATA1* and *ASHIL*) connected to reproduction traits in livestock species^{58–60}. Moreover, *SYT11* may play an important role in adaptation to different environmental conditions in Kenyan goat breeds⁶¹, while *ARHGEF2* gene's function in livestock is still unclear, but may be involved in epithelial barrier permeability affecting host-microbial interactions in the rumen⁶². The HRR island in CHI6-A (23.81–24.43 Mb) was also found as a HRR region in MNT_I by Somenzi et al.²⁶, which attributed a reproduction function due to the presence of *PPP3CA* gene associated to fecundity traits and litter size in small ruminants. The HRR islands in CHI8-B (38.47–38.69 Mb), in CHI15-A (65.98–66.54 Mb), and CHI15-B (72.58–73.07 Mb) showed other candidate genes related with immune response traits such as *RANBP6*⁶³, *GUCY1A2*⁶⁴, *PIWIL4*⁶⁵, and *CNTN5*⁶⁶. The *IL31RA* gene in CHI20-A (23.34–23.95 Mb) is a cytokine receptors known to be involved in human inflammation and allergic diseases⁶⁷. Moreover, *SLC38A9* gene in the same HRR island was identified as an indicator of heat stress in bovine⁶⁸. Finally, the last gene worth of interest is *EML1* in CHI21-C (64.17–64.51), which plays a role in processes of exocytosis (the process of releasing vesicle content to the extracellular environment) in the molecular pathway of Ab production by B lymphocytes⁶⁹. However, we cannot exclude that the high number of HRR islands found in the feral goat Montecristo derives from the bias linked to the chip design based on the domesticated goat genome.

Conclusions

This work studied the genetic relationships among goat breeds with dairy aptitude and investigated their continuous heterozygosity patterns. We confirmed a notable divergence between the European lineage and the Asian and African breeds. European goats have shown a clear north–south geographic cline and genetic interconnections probably due to gene flow between geographically close breeds. The investigation of heterozygosity-rich regions highlighted specific portions of the goat genome associated with different functional factors. The distribution of the HRR islands in the goat genome seems to be mainly related to the breeds' geography. A species-specific HRR pattern, possibly shaped by adaptation mechanisms, might provide a clearer view of the mechanisms modelling the genome as a function of anthropic selection. Interestingly, some heterozygosity hotspots showed overlap with ROH islands reported in the literature. This possibly indicates genomic areas targeted by selective breeding which has shaped the goat genome differently in different production contexts. The methodological part of the manuscript gave new insight into the standardization of HRR detection, highlighting the point of overlap between two different detection methods. Further investigation is needed to improve the parameter settings of HRR detection, possibly involving different species and breeds genotyped with different densities of SNP arrays.

Methods

Dataset and quality control

A dataset including worldwide goat breeds was generated by merging the Italian Goat Consortium2 (IGC2) and the ADAPTmap repositories described in Cortellari et al.²⁰ and Stella et al.¹⁸, respectively. The breeds involved in the analyses were selected for their milk-production aptitude, excluding those for meat or fiber, and grouped according to their geographic breeding area, resulting in a reduced dataset comprehensive of 1,289 individuals

belonging to 49 goat breeds (Table 1). Italian breeds were divided into two different groups (Southern and Center breeds labelled as Italy, while Northern breeds as Alpine arch) due to their isolation and divergent morpho-aptitude factors. The random sampling selection procedure implemented in the `BITE R` package⁷⁰ was used to select a maximum of 30 representative individuals per breed. All animals were genotyped with the Illumina GoatSNP50k BeadChip, including a total of 53,347 SNPs²⁴. Chromosomal coordinates of SNPs were referred to the *ARSI* genome assembly (Assembly GCA_001704415.1) using the commands `--update-map` and `--update-chr`. The software PLINK ver. 1.9⁷¹ was used to perform filtering and quality control. After removing the unmapped SNPs and markers located on sexual chromosomes, the additional quality parameters were set as follows: minor allele frequency ≥ 0.05 (`--maf 0.05`), SNP genotype call rate ≥ 0.95 (`--geno 0.05`) and individual genotype call rate ≥ 0.80 (`--mind 0.20`), resulting in 48,544 SNPs and 1287 goats.

Genetic diversity indices

The software PLINK ver. 1.9⁷¹ was used to estimate the observed and expected heterozygosity (H_O and H_E), the molecular inbreeding coefficient (F_{IS}) and calculate the minor allele frequencies (*MAF*) for each breed, using the commands `--hardy`, `--het` and `--freq` respectively. Genetic diversity indices values were averaged per geographical group.

Genetic relationships

To explore the genetic relationships among breeds, the multidimensional scaling (MDS) analysis was performed based on pairwise identity-by-state (IBS) distances between individuals (`--read-genome --cluster --ppc 0.0001 --mds-plot 30 eigvals`) using PLINK ver. 1.9⁷¹. In addition, the Neighbor-Net based on pairwise Reynolds' genetic distances was obtained using software ARLEQUIN ver. 3.5.2.2⁷² and plotted using SPLITSTREE4 ver. 4.14.8⁷³.

Heterozygosity-rich regions detection

Heterozygosity-rich regions (HRR) were investigated using the R package *detectRUNS* ver. 0.9.6³³, by two different methods of detection: Sliding Window (SW)⁷⁴ and Consecutive Runs (CR)⁷⁵. The absence of clear guidelines from literature led us to test range of values for the detection parameters³⁴. Screening involved several combinations, varying the minimum number of SNPs in a HRR (10–25), the minimum length (10^3 bp to 10^6 bp), and the number of missing/opposing markers (0–3). The evaluation criteria to choose the parameters setting were based on (i) avoiding missing or opposite genotype per run; (ii) obtaining a comparable number of HRR between the two methods (SW and CR). The following parameters were finally set: (i) the minimum number of SNPs included in the HRR was 10; (ii) the number of missing or opposite genotypes were set to zero; (iii) the maximum gap between consecutive SNPs was set to 1 Mb; (iv) the minimum HRR length was set to 250 kb. Additional parameters were set for the SW method: v) sliding window of 10 SNPs for HRR; vi) no missing or opposite genotypes were allowed in the window; (vii) the minimum density of one SNP every 100 kb; (viii) the threshold to call a SNP within a HRR was set to 0.05. The HRR pattern per breed was investigated through the average HRR number (N_{HRR}), the average length of HRR in Mbp (L_{HRR}) and the mean genome length covered by HRR segments in Mbp (S_{HRR}). Finally, the total length of the genome covered by HRR per breed was divided by the total autosomal genome length covered by the SNP array (~ 2.4 Gb) to evaluate the coefficient of diversity (D_{HRR})^{9,14}. In addition, the four descriptive breed-statistics were aggregated by geographic area (AN_{HRR} , AL_{HRR} , AS_{HRR} and AD_{HRR}).

Heterozygosity islands detection and gene annotation

HRR islands per breed were identified comparing the results of SW and CR methods of detection. Based on the standard normal z-score, obtained from all the SNPs-within-HRR incidence per breed and method, *p*-values were calculated and the top 0.1% of SNPs were selected to constitute an island^{14,76,77}. Only those markers identified simultaneously with both methods were considered. HRR islands in a breed were considered those regions with frequency $\geq 20\%$ and a minimum of 4 SNPs. The genomic coordinates of HRR islands were examined using the Ensemble browser for the goat genome, according to the assembly *ARSI* (Assembly GCA_001704415.1) to retrieve annotated gene lists. Gene Ontology (GO) and the enrichment analysis of annotated genes were conducted for each island using the open-source Database for Annotation, Visualization, and Integrated Discovery ver. 2021 package⁷⁸. For the GO terms and Kyoto Encyclopaedia of Genes and Genomes (KEGG) pathway analysis, the level of significance for the enriched biological processes was set as *p*-value < 0.05 . Corrections for multiple testing were made by applying the Bonferroni test.

Data availability

The datasets generated and analyzed during the current study are available in the Mendeley Data repository, <https://data.mendeley.com/datasets/hnd59x6gmg/120> and Dryad repository, <https://datadryad.org/stash/dataset/doi:10.5061/dryad.v8g21pt18>.

Received: 19 September 2023; Accepted: 4 December 2023

Published online: 02 January 2024

References

1. Renaud, G., Hanghoj, K., Korneliussen, T. S., Willerslev, E. & Orlando, L. Joint estimates of heterozygosity and runs of homozygosity for modern and ancient samples. *Genetics* **212**, 587–614 (2019).
2. Ferenčaković, M. *et al.* Mapping of heterozygosity rich regions in Austrian pinzgauer cattle. *Acta Agric. Slov.* **5**, S41–44 (2016).

3. Santos, W. B. *et al.* Fine-scale estimation of inbreeding rates, runs of homozygosity and genome-wide heterozygosity levels in the Mangalarga Marchador horse breed. *J. Anim. Breed. Genet.* **138**, 161–173 (2021).
4. Williams, J. L. *et al.* Inbreeding and purging at the genomic level: the Chillingham cattle reveal extensive, non-random SNP heterozygosity. *Anim. Genet.* **47**, 19–27 (2016).
5. Samuels, D. C. *et al.* Heterozygosity ratio, a robust global genomic measure of autozygosity and its association with height and disease risk. *Genetics* **204**, 893–904 (2016).
6. Biscarini, F., Mastrangelo, S., Catillo, G., Senczuk, G. & Ciampolini, R. Insights into genetic diversity, runs of homozygosity and heterozygosity-rich regions in Maremmana semi-feral cattle using pedigree and genomic data. *Animals (Basel)* **10**, 2285 (2020).
7. Mulim, H. A. *et al.* Characterization of runs of homozygosity, heterozygosity-enriched regions, and population structure in cattle populations selected for different breeding goals. *BMC Genomics* **23**, 209 (2022).
8. Selli, A. *et al.* Detection and visualization of heterozygosity-rich regions and runs of homozygosity in worldwide sheep populations. *Animals (Basel)* **11**, 2696 (2021).
9. Ruan, D. *et al.* Assessment of heterozygosity and genome-wide analysis of heterozygosity regions in two Duroc pig populations. *Front. Genet.* **12**, 812456 (2022).
10. VanRaden, P. M., Olson, K. M., Null, D. J. & Hutchison, J. L. Harmful recessive effects on fertility detected by absence of homozygous haplotypes. *J. Dairy Sci.* **94**, 6153–6161 (2011).
11. Biscarini, F. *et al.* Use of SNP genotypes to identify carriers of harmful recessive mutations in cattle populations. *BMC Genomics* **17**, 857 (2016).
12. Tsartsianidou, V. *et al.* A comprehensive genome-wide scan detects genomic regions related to local adaptation and climate resilience in Mediterranean domestic sheep. *Genet. Sel. Evol.* **53**, 90 (2021).
13. Chen, Z. *et al.* Heterozygosity and homozygosity regions affect reproductive success and the loss of reproduction: A case study with litter traits in pigs. *Comput. Struct. Biotechnol. J.* **20**, 4060–4071 (2022).
14. Bordonaro, S. *et al.* Genome-wide population structure, homozygosity, and heterozygosity patterns of Nero Siciliano pig in the framework of Italian and cosmopolitan breeds. *Anim. Genet.* **00**, 1–15 (2023).
15. Li, G. *et al.* Genome-wide estimates of runs of homozygosity, heterozygosity, and genetic load in two Chinese indigenous goat breeds. *Front. Genet.* **13**, 774196 (2022).
16. Naderi, S. *et al.* The goat domestication process inferred from large-scale mitochondrial DNA analysis of wild and domestic individuals. *Proc. Natl. Acad. Sci. USA* **105**, 17659–17664 (2008).
17. Denoyelle, L. *et al.* VarGoats project: A dataset of 1159 whole-genome sequences to dissect *Capra hircus* global diversity. *Genet. Sel. Evol.* **53**, 86 (2021).
18. Stella, A. *et al.* AdaptMap: Exploring goat diversity and adaptation. *Genet. Sel. Evol.* **50**, 61 (2018).
19. 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.* **50**, 58 (2018).
20. Cortellari, M. *et al.* The climatic and genetic heritage of Italian goat breeds with genomic SNP data. *Sci. Rep.* **11**, 10986 (2021).
21. Miller, M. A. & Zachery, J. F. *Pathologic basis of veterinary disease*. 6th edn St Louis (Elsevier, 2017).
22. Serrano, B. *et al.* Local adaptations of Mediterranean sheep and goats through an integrative approach. *Sci. Rep.* **11**, 21363 (2021).
23. Cortellari, M. *et al.* Runs of homozygosity in the Italian goat breeds: Impact of management practices in low-input systems. *Genet. Sel. Evol.* **53**, 92 (2021).
24. Tosser-Klopp, G. *et al.* Design and characterization of a 52K SNP chip for goats. *PLoS One* **9**, e86227 (2014).
25. Mastrangelo, S. *et al.* Genome-wide patterns of homozygosity reveal the conservation status in five Italian goat populations. *Animals (Basel)* **11**, 1510 (2021).
26. Somenzi, E. *et al.* The SNP-based profiling of Montecristo feral goat populations reveals a history of isolation, bottlenecks, and the effects of management. *Genes (Basel)* **13**, 213 (2022).
27. Peñaloza, C. *et al.* Development and testing of a combined species SNP array for the European seabass (*Dicentrarchus labrax*) and gilthead seabream (*Sparus aurata*). *Genomics* **113**, 2096–2107 (2021).
28. Pereira, F. *et al.* Tracing the history of goat pastoralism: new clues from mitochondrial and Y chromosome DNA in North Africa. *Mol. Biol. Evol.* **26**, 2765–2773 (2009).
29. Missohou, A., Talaki, E. & Laminou, I. M. Diversity and genetic relationships among seven West African goat breeds. *Asian-Australas. J. Anim. Sci.* **19**, 1245–1251 (2006).
30. Gifford-Gonzalez, D. & Hanotte, O. Domesticating animals in Africa: Implications of genetic and archaeological findings. *J. World Prehist.* **24**, 1–23 (2011).
31. Naderi, S. *et al.* Large-scale mitochondrial DNA analysis of the domestic goat reveals six haplogroups with high diversity. *PLoS One* **2**, e1012 (2007).
32. Dixit, S. P. *et al.* Genome-wide runs of homozygosity revealed selection signatures in *Bos indicus*. *Front. Genet.* **11**, 92 (2020).
33. Biscarini, F., Cozzi, P., Gaspa, G. & Marras, G. detectRUNS: An R package to detect runs of homozygosity and heterozygosity in diploid genomes. <https://orca.cardiff.ac.uk/108906/>. Accessed 28 August 2023 (2018).
34. Biscarini, F., Cozzi, P., Ramirez-Diaz, J., Stella, A. & Manunza, A. in *Proceedings of 25th Congress on Animal Production Science: innovations and sustainability for future generations (ASPA)* (2023).
35. Marras, G. *et al.* in *Proceedings of the 11th World Congress of Genetics Applied to Livestock Production (WCGALP)* (2018).
36. Bertolini, F. *et al.* Genome-wide patterns of homozygosity provide clues about the population history and adaptation of goats. *Genet. Sel. Evol.* **50**, 59 (2018).
37. Bertolini, F. *et al.* Signatures of selection and environmental adaptation across the goat genome post-domestication. *Genet. Sel. Evol.* **50**, 57 (2018).
38. Mota, L. F. M. *et al.* Meta-analysis across Nellore cattle populations identifies common metabolic mechanisms that regulate feed efficiency-related traits. *BMC Genomics* **23**, 424 (2022).
39. Zhao, B. *et al.* Integration of a single-step genome-wide association study with a multi-tissue transcriptome analysis provides novel insights into the genetic basis of wool and weight traits in sheep. *Genet. Sel. Evol.* **53**, 56 (2021).
40. Seto, E., Yoshida-Sugitani, R., Kobayashi, T. & Toyama-Sorimachi, N. The assembly of EDC4 and Dcp1a into processing bodies is critical for the translational regulation of IL-6. *PLoS One* **10**, e0123223 (2015).
41. Zhang, X. *et al.* Novel nucleotide variations, haplotypes structure and associations with growth related traits of goat AT Motif-Binding Factor (ATBF1) gene. *Asian-Australas. J. Anim. Sci.* **28**, 1394–1406 (2015).
42. Wei, Z. *et al.* Detection of insertion/deletions (indels) of the ATBF1 gene and their effects on growth-related traits in three indigenous goat breeds. *Arch. Anim. Breed.* **61**, 311–319 (2018).
43. Lamartine, L. *et al.* Mutations in GJB6 cause hidrotic ectodermal dysplasia. *Nat. Genet.* **26**, 142–144 (2000).
44. Pandya, A. *et al.* Frequency and distribution of GJB2 (connexin 26) and GJB6 (connexin 30) mutations in a large North American repository of deaf probands. *Genet. Med.* **5**, 295–303 (2003).
45. Feng, Y., Peng, X., Li, S. & Gong, Y. Isolation and characterization of sexual dimorphism genes expressed in chicken embryonic gonads. *Acta Biochim. Biophys. Sin. (Shanghai)* **41**, 285–294 (2009).
46. Abied, A. *et al.* Genome-wide analysis revealed homozygosity and demographic history of five Chinese sheep breeds adapted to different environments. *Genes (Basel)* **11**, 1480 (2020).

47. Onzima, R. B. *et al.* Genome-wide characterization of selection signatures and runs of homozygosity in Ugandan goat breeds. *Front. Genet.* **9**, 318 (2018).
48. Kim, E. S. *et al.* Multiple genomic signatures of selection in goats and sheep indigenous to a hot arid environment. *Heredity (Edinb)* **116**, 255–264 (2016).
49. Lee, Y., Clinton, J., Yao, C. & Chang, S. H. Interleukin-17D promotes pathogenicity during infection by suppressing CD8 T cell activity. *Front. Immunol.* **10**, 1172 (2019).
50. Huang, J. *et al.* Interleukin-17D regulates group 3 innate lymphoid cell function through its receptor CD93. *Immunity* **54**, 673–686 e674 (2021).
51. Biscarini, F., Manunza, A., Cozzi, P. & Stella, A. in *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)* (2022).
52. Taye, M. *et al.* Exploring evidence of positive selection signatures in cattle breeds selected for different traits. *Mamm. Genome* **28**, 528–541 (2017).
53. Berihulay, H. *et al.* Whole genome resequencing reveals selection signatures associated with important traits in Ethiopian indigenous goat populations. *Front. Genet.* **10**, 1190 (2019).
54. Arora, R. *et al.* Transcriptome profiling of longissimus thoracis muscles identifies highly connected differentially expressed genes in meat type sheep of India. *PLoS One* **14**, e0217461 (2019).
55. Zhang, L. *et al.* Quantitative genomics of 30 complex phenotypes in Wagyu x Angus F1 progeny. *Int. J. Biol. Sci.* **8**, 838–858 (2012).
56. de Lima, A. O. *et al.* Potential biomarkers for feed efficiency-related traits in Nelore cattle identified by co-expression network and integrative genomics analyses. *Front. Genet.* **11**, 189 (2020).
57. Wragg, D. *et al.* A locus conferring tolerance to Theileria infection in African cattle. *PLoS Genet.* **18**, e1010099 (2022).
58. Wang, J. J. *et al.* Genome-wide detection of selective signals for fecundity traits in goats (*Capra hircus*). *Gene* **818**, 146221 (2022).
59. Cui, L. X. *et al.* Knockdown of ASH1L methyltransferase induced apoptosis inhibiting proliferation and H3K36 methylation in bovine cumulus cells. *Theriogenology* **161**, 65–73 (2021).
60. Giesecke, K. *et al.* Evaluation of SPATA1-associated markers for stallion fertility. *Anim. Genet.* **40**, 359–365 (2009).
61. Waineina, R. W., Okeno, T. O., Ilatsia, E. D. & Ngeno, K. In *Proceedings of 12th World Congress on Genetics Applied to Livestock Production (WCGALP)* (2022).
62. Zang, X. W. *et al.* Heritable and nonheritable rumen bacteria are associated with different characters of lactation performance of dairy cows. *mSystems* **7**, e00422 (2022).
63. Liu, L. *et al.* Study of the integrated immune response induced by an inactivated EV71 vaccine. *PLoS One* **8**, e54451 (2013).
64. Berton, M. P. *et al.* Genomic regions and pathways associated with gastrointestinal parasites resistance in Santa Ines breed adapted to tropical climate. *J. Anim. Sci. Biotechnol.* **8**, 73 (2017).
65. Mousel, M. R. *et al.* Genes involved in immune, gene translation and chromatin organization pathways associated with Mycoplasma ovipneumoniae presence in nasal secretions of domestic sheep. *PLoS One* **16**, e0247209 (2021).
66. Li, G. S. *et al.* Genome-wide association study of bone quality and feed efficiency-related traits in Pekin ducks. *Genomics* **112**, 5021–5028 (2020).
67. He, Y. *et al.* Interleukin-31 receptor alpha is required for basal-like breast cancer progression. *Front. Oncol.* **10**, 816 (2020).
68. Fu, L. *et al.* Effect of heat stress on bovine mammary cellular metabolites and gene transcription related to amino acid metabolism, amino acid transportation and mammalian target of rapamycin (mTOR) signaling. *Animals (Basel)* **11**, 3153 (2021).
69. de Klerk, B. *et al.* A genome-wide association study for natural antibodies measured in blood of Canadian Holstein cows. *BMC Genomics* **19**, 694 (2018).
70. Milanese, M. *et al.* BITE: An R package for biodiversity analyses. <https://doi.org/10.1101/181610>. Accessed 28 August 2023. (2017).
71. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7 (2015).
72. Excoffier, L. & Lischer, H. E. Arlequin suite ver 3.5: A new series of programs to perform population genetics analyses under Linux and Windows. *Mol. Ecol. Resour.* **10**, 564–567 (2010).
73. Huson, D. H. & Bryant, D. Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254–267 (2006).
74. Bjelland, D. W., Weigel, K. A., Vukasinovic, N. & Nkrumah, J. D. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. *J. Dairy Sci.* **96**, 4697–4706 (2013).
75. Marras, G. *et al.* Analysis of runs of homozygosity and their relationship with inbreeding in five cattle breeds farmed in Italy. *Anim. Genet.* **46**, 110–121 (2015).
76. Criscione, A. *et al.* Genome-wide survey on three local horse populations with a focus on runs of homozygosity pattern. *J. Anim. Breed. Genet.* **139**, 540–555 (2022).
77. Gorssen, W., Meyermans, R., Janssens, S. & Buys, N. A publicly available repository of ROH islands reveals signatures of selection in different livestock and pet species. *Genet. Sel. Evol.* **53**, 2 (2021).
78. Huang, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* **4**, 44–57 (2008).

Author contributions

G.C., A.Cr., and S.M. conceived the study. D.M., P.C., B.P., A.M., A.Ce., and F.B. organized the database. G.C., A.Cr., F.B. and S.M. performed the statistical analysis. D.M., P.C. and B.P. were involved in the supervision of the work. G.C., A.Cr., and S.M. were involved in the interpretation of the data and wrote the first draft of the manuscript. All authors contributed to the writing of the manuscript and approved the submitted version.

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/s41598-023-49125-x>.

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

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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024