The genome-wide structure of two economically important indigenous Sicilian cattle breeds[1]

**S. Mastrangelo,*[2] M. Saura,† M. Tolone,* J. Salces-Ortiz,†
R. Di Gerlando,* F. Bertolini,‡ L. Fontanesi,‡ M. T. Sardina,* M. Serrano,†[3] and B. Portolano*[3]**

*Dipartimento di Scienze Agrarie e Forestali, University of Palermo, Viale delle Scienze, 90128 Palermo, Italy;
†Departamento de Mejora Genética Animal, INIA, Carretera de la Coruña Km 7.5, 28040 Madrid, Spain; and ‡Department of Agricultural and Food Sciences, Division of Animal Sciences, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy

ABSTRACT: Genomic technologies, such as high-throughput genotyping based on SNP arrays, provided background information concerning genome structure in domestic animals. The aim of this work was to investigate the genetic structure, the genome-wide estimates of inbreeding, coancestry, effective population size (N_e), and the patterns of linkage disequilibrium (LD) in 2 economically important Sicilian local cattle breeds, Cinisara (CIN) and Modicana (MOD), using the Illumina Bovine SNP50K v2 BeadChip. To understand the genetic relationship and to place both Sicilian breeds in a global context, genotypes from 134 other domesticated bovid breeds were used. Principal component analysis showed that the Sicilian cattle breeds were closer to individuals of *Bos taurus taurus* from Eurasia and formed non-overlapping clusters with other breeds. Between the Sicilian cattle breeds, MOD was the most differentiated, whereas the animals belonging to the CIN breed showed a lower value of assignment, the presence of substructure, and genetic links with the MOD breed. The average molecular inbreeding and coancestry coefficients were moderately high, and the current estimates of N_e were low in both breeds. These values indicated a low genetic variability. Considering levels of LD between adjacent markers, the average r^2 in the MOD breed was comparable to those reported for others cattle breeds, whereas CIN showed a lower value. Therefore, these results support the need of more dense SNP arrays for a high-power association mapping and genomic selection efficiency, particularly for the CIN cattle breed. Controlling molecular inbreeding and coancestry would restrict inbreeding depression, the probability of losing beneficial rare alleles, and therefore the risk of extinction. The results generated from this study have important implications for the development of conservation and/or selection breeding programs in these 2 local cattle breeds.

Key Words: genetic diversity, genetic structure, Sicilian cattle breeds, single nucleotide polymorphisms

INTRODUCTION

The global decline in livestock genetic diversity is mainly the result of the massive use of a small number of selected breeds. However, in recent years, an increasing interest in recovering and preserving local breeds has taken place (Sechi et al., 2007). An interesting situation is represented by 2 Sicilian cattle breeds, Cinisara (**CIN**) and Modicana (**MOD**). These 2 breeds are well adapted to the harshness of marginal mountain areas of Sicily because of their good grazing characteristics and resistance to environmental conditions. Their economic importance lies in the traditional production systems of 2 typical pasta filata cheeses: Caciocavallo Palermitano and Ragusano P.D.O. (protected designation of origin; Guastella et al., 2011). Therefore, the socioeconomic and ecological values and the historical, cultural, and genetic heritage of these 2 breeds are unquestionable.

Maintaining high levels of genetic diversity and limiting the increase of inbreeding are the premise of any conservation program (Frankham et al., 2002). Single nucleotide polymorphism genotyping allows

high-throughput interrogation of hundreds of thousands loci with high precision at an affordable cost that enables large-scale studies (Matukumalli et al., 2009). Moreover, with the application of genome-wide SNP genotyping, studying the extent of linkage disequilibrium (**LD**; the nonrandom association of alleles at different loci) in livestock populations has gained more attention. The need to explore the genetic characteristics of these 2 Sicilian cattle breeds arises to determine the genetic status of these populations and their risk of damage and/or extinction.

The aim of this study was to characterize Cinisara and Modicana breeds from a genetic perspective because no previous studies on these breeds have been reported. In particular, we aimed to investigate the patterns of genetic structure, genetic variability, and LD.

MATERIALS AND METHODS

Description of Sicilian Cattle Breeds

The CIN breed is characterized by a uniform black coat and, less frequently, by a color-sided (lineback) coat. This breed is mainly reared in a restricted area of Sicily (northwest), and currently, about 5,000 animals are enrolled in the herd book, with an average herd size of 25 animals (Associazione Italiana Allevatori [**AIA**], 2014). The MOD breed is characterized by a solid red coat and is reared mainly in the eastern part of the island. The number of animals enrolled in the herd book is about 4,700, with an average herd size of 20 animals (AIA, 2014). Neither breed is subject to breeding programs, and in the last 50 yr both breeds have undergone a progressive reduction in size, mainly because of the mechanization of agriculture and the introduction of more specialized and productive cosmopolitan breeds.

DNA Sampling and Genotyping

A total of 144 animals from 14 farms were sampled. Samples consisted of 72 CIN (68 cows and 4 bulls) and 72 MOD (69 cows and 3 bulls) animals born between 1999 and 2010 and chosen on the basis of their phenotypic profiles (morphological traits such as coat color) and information supplied by farmers (year of birth). The number of animals sampled per farm ranged from 8 to 12 individuals. For these cattle breeds pedigree data were not available. About 10 mL of blood were collected from the caudal vein using tubes with EDTA as an anticoagulant. Genomic DNA was extracted from buffy coats of nucleated cells using the salting-out method (Miller et al., 1988). The concentration of extracted DNA was assessed with the NanoDrop ND-1000 spectrophotometer (NanoDrop Technologies, Wilmington, DE).

CIN and MOD animals were genotyped for 54,609 SNP using the Bovine SNP50K v2 BeadChip (Illumina Inc., San Diego, CA) according to the standard operating procedures recommended by the manufacturer. Only the SNP located in the autosomes were considered in further analyses. Therefore, unmapped SNP were discarded. Markers were filtered according to quality criteria that included 1) call frequency (proportion of samples with the genotype at each locus; ≥ 0.98), 2) minor allele frequency (**MAF**; ≥ 0.05), and 3) Hardy-Weinberg equilibrium (**HWE**; P-value = 0.001). Single nucleotide polymorphisms that did not satisfy these quality criteria were excluded.

Genetic Relationship and Population Structure between Sicilian and Other Cattle Breeds

To understand the genetic relationship within and among breeds, genotypes from 1,543 other animals belonging to 134 domesticated bovid breeds were used (Decker et al., 2013). These breeds arose from 3 domesticated (sub)species: *Bos javanicus, B. taurus indicus,* and *B. t. taurus*. Data editing was performed using SAS (SAS Inst. Inc., Cary, NC) and PLINK (Purcell et al., 2007). Single nucleotide polymorphisms that did not satisfy the following quality criteria were discarded: 1) SNP located on autosomes and common to all breeds, 2) MAF ≥ 0.0005, and 3) SNP with genotyping rate ≥ 0.90 according to Decker et al. (2014).

An identity-by-state (**IBS**) matrix of genetic distances (D) containing each pairwise combination of all individuals was generated using PLINK (Purcell et al., 2007). Classical MDS analysis was applied to explore the similarities in the matrix. The cluster and mds plot options of PLINK (Purcell et al., 2007) were used. It should be noted that when MDS is applied to D, it is numerically identical to principal component analysis (**PCA**; Purcell et al., 2007). The model-based clustering algorithm implemented in Admixture software (Alexander and Lange, 2011) was used to examine patterns of ancestry and admixture in our data set. Unlinked SNP were selected using the indep-pairwise option of PLINK (Purcell et al., 2007) to reduce the impact of the SNP ascertainment bias phenomenon, with the following parameters: 50 SNP per window, a shift of 5 SNP between windows, and r^2 threshold of 0.2. We used the default (5-fold) Admixture cross-validation procedure for values of K from 2 through 20, from which estimated prediction errors are obtained for each K value by adopting a kind of leave-one-out approach. The K value that minimizes this estimated prediction errors is then assumed to be the most suitable. The graphical representation was generated using the statistical R software (R Development Core Team, 2011).

Genetic Structure of Sicilian Cattle Breeds

Structure version 2.3.1 (Pritchard et al., 2000) was used to analyze the genetic structure, identify the true number of populations (clusters), and assign the individuals to each cluster and population. Genotypes from 96 animals of the Italian Holstein (**HOL**) cattle breed were included in the analysis. The software estimates the natural logarithm of the probability (ln Pr) that a given genotype (G) is part of a given population (K). The following quality control was applied: 1) SNP located on autosomes and in common between the 3 breeds, 2) MAF $\geq$ 0.05, 3) SNP with genotyping rate $\geq$ 0.95, and 4) HWE (P-value = 0.001). Unlinked SNP were selected as described above, and from these markers, a random set of 10,000 SNP was used. Analysis was performed considering both the admixture model and the correlated allele frequencies between populations (Pritchard et al., 2000). Lengths of the burn-in and Monte Carlo Markov chain were 50,000 steps and 100,000 iterations, respectively. To estimate the most likely number of clusters in the data set, ln Pr($G|K$) was calculated for each K ranging from 1 through 6 without prior information on breed and farm origins, running 50 independent replicates for each K. The estimates of genetic distances of each pairwise combination of individuals and MDS analysis were performed as reported above (Purcell et al., 2007). The graphical representation was depicted using the statistical R software (http://www.R-project.org/). The same software was used to visualize the IBS matrix using the Heatmap function.

Inbreeding and Coancestry Molecular Coefficients, Rates of Inbreeding and Coancestry, and Effective Population Size in Sicilian Cattle Breeds

Following Caballero and Toro (2002), the molecular coancestry (f) coefficient between pairs of individuals is defined as the probability that 2 alleles taken at random, 1 from each individual, are identical in state. Similarly, the molecular inbreeding (F) coefficient for each individual is defined as the probability that the 2 alleles of a locus are identical in state. The molecular f coefficient between individuals i and j was calculated as:

$$f_{ij} = (1/N) \sum_{n=1}^{N} \left[\left(\sum_{k=1}^{2} \sum_{m=1}^{2} I_{nk(i)m(j)} \right) / 4 \right],$$

where N is the number of markers and $I_{nk(i)m(j)}$ is the identity of the kth allele from individual i with the mth allele from the animal j at locus n, which takes a value of 1 if alleles are identical and 0 if they are not. The molecular F coefficient for individual i was calculated as $F = 2f_{ii} - 1$ (i.e., f_{ii} is the molecular self-coancestry). Thus, F was estimated as the proportion of homozygous genotypes. The rate of molecular inbreeding per year

Table 1. Estimates of genetic diversity for Cinisara and Modicana cattle breeds: Inbreeding (F) and coancestry (f) coefficients, rate of inbreeding (ΔF) and coancestry (Δf) per year, SD, and effective population size estimated from the rate of inbreeding ($N_e F$) and the rate of coancestry ($N_e f$)

Breed	$F \pm$ SD	$f \pm$ SD	ΔF	Δf	$N_e F$	$N_e f$
Cinisara	0.68 ± 0.024	0.67 ± 0.030	0.004	0.022	19.38	3.77
Modicana	0.69 ± 0.020	0.70 ± 0.030	0.007	0.010	11.90	7.94

(ΔF) was computed by regressing the natural logarithm of $(1 - F)$ on year of birth. The rate of molecular coancestry per year (Δf) was estimated in the same way, that is, by regressing the natural logarithm of $(1 - f)$ for each pair of individuals on year of birth. Ne was estimated from the rates of inbreeding ($N_e F = 1/2L\Delta F$) and coancestry ($N_e f = 1/2L\Delta f$) per generation, assuming a generation interval L of 6 yr.

Linkage Disequilibrium

A standard descriptive LD parameter, the squared correlation coefficient of allele frequencies at a pair of loci (r^2), was obtained using PLINK (Purcell et al., 2007). Pairwise r^2 between adjacent SNP was calculated on each chromosome. Moreover, r^2 was estimated for all pairwise combinations of SNP using Haploview version 4.2 (Barrett et al., 2005). For each chromosome, the pairwise r^2 was calculated for SNP between 5 kb and 50 Mb apart. To visualize the LD pattern per chromosome, r^2 values were stacked and plotted as a function of intermarker distance categories.

RESULTS AND DISCUSSION

A descriptive summary of chromosomes and SNP is given in Supplemental Data Table 1. The number of SNP per chromosome (**BTA**) after all filtering ranged from 2,635 (BTA1) to 746 (BTA27) in CIN and from 2,470 (BTA1) to 718 (BTA27) in MOD. The average density of SNP per Mb was 15 for CIN and 14 for MOD. The overall mean MAF was 0.279 ± 0.004 for CIN and 0.276 ± 0.007 for MOD; these values are in agreement with those reported by Matukumalli et al. (2009) in a study on development and characterization of a high-density SNP genotyping assay for several cattle breeds. Among the 54,609 SNP included in the array, only 52,886 mapping to bovine autosomes were considered for further analysis. After screening, the final number of samples and SNP were 71 and 40,461 for CIN and 69 and 37,968 for MOD cattle breeds, respectively. The major point for SNP exclusion was MAF (11,090 SNP rejected for CIN and 13,829 SNP for MOD).
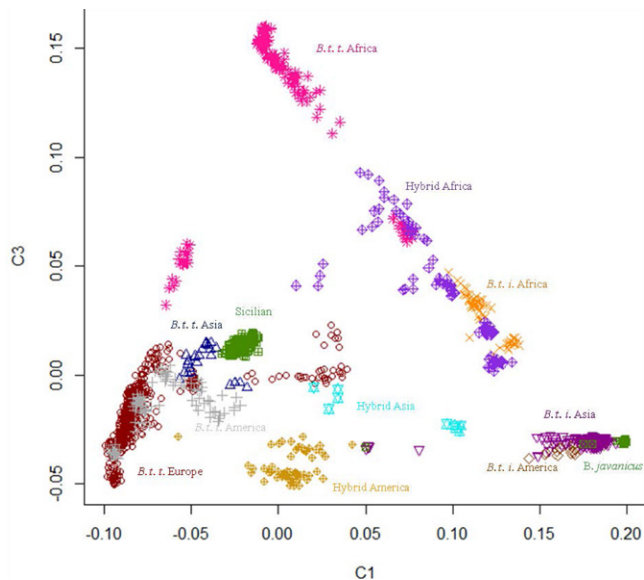
Figure 1. Principal component analysis of 1,683 animals plotted for the first and third components. Points are colored according to breed geographical origin. See online version for figure in color.

Genetic Relationship and Population Structure between Sicilian and Other Cattle Breeds

Principal component analysis and ancestry models were used to cluster animals, to explore the relationships among and within breeds, and to place the Sicilian breeds in a global context. As demonstrated in previous studies in human (Tishkoff et al., 2009) and livestock species such as cattle (Bovine HapMap Consortium, 2009; Gautier et al., 2010) and sheep breeds (Kijas et al., 2012), the combination of principal component 1 (PC1), principal component 2 (PC2), and principal component 3 (PC3) allowed the separation of individuals according to their geographic origin. A total of 40,958 SNP shared among 136 breeds (134 from DRYAD and 2 Sicilian breeds) and 1,683 individuals were analyzed.

The Sicilian breeds showed outlier behavior for the values of PC2, with all individuals spread over the entire range of variability pertaining to the second component (Supplemental Data Fig. 1). Therefore, the observation that all breeds except the Sicilian ones had a small range for PC2 (from -0.005 to 0.058) required plotting PC1 and PC3. An analogous situation was reported by Kijas et al. (2009) in a study on the genetic structure of sheep breeds. Despite the different numbers of SNP (40,958 vs. 43,043) and the different components (PC1 and PC3 vs. PC1 and PC2) used in our study, Fig. 1 shows the same results as those reported by Decker et al. (2014), such as the separation between *B. t. taurus* and *B. t. indicus* breeds and the divergences between African and Eurasian taurines. Principal component 1 separated the individuals into several nonoverlapping clusters that corresponded with (sub)species and the geographical origin of each breeds. Using PC3, the Sicilian cattle breeds were closer

to individuals of *B. t. taurus* from Europe (that included Italian cattle breeds), according to their geographic distribution, and to individuals of *B. t. taurus* from Asia. In fact, as reported by Decker et al. (2014), European cattle breeds were exported to Asia and admixed with Asian taurines. Moreover, the Sicilian breeds formed nonoverlapping clusters with other breeds. Kijas et al. (2012), in a genome-wide analysis of the world's sheep breeds, reported a similar result for the Sicilian breeds that formed a separate clusters with respect to the others. Ancestry models with 4 ancestral populations implemented in Admixture (Fig. 2) provided the same results as those reported by Decker et al. (2014) that used an ancestry model with 3 ancestral populations. In fact, the 3 major groups of Asian indicine, Eurasian taurine, and African taurine were consistently observed (Fig. 2). Admixture corroborated the findings based on the PCA using the first 2 components; that is, the Sicilian breeds formed a separate cluster (Supplemental Data Fig. 1). These results reflected the differences between breeds resulting from separate domestic events, geographic dispersal and isolation, and breed formation (Decker et al., 2014).

Genetic Structure of Sicilian Cattle Breeds

Principal component analysis was also used to cluster animals and explore in detail the relationship within and among Sicilian cattle breeds. As mentioned before, a sample of HOL was included in this analysis. A total of 38,925 SNP and 236 individuals were analyzed. The PCA of the first 2 components showed that animals from the 3 cattle breeds formed nonoverlapping clusters and are clearly separated populations. In addition, a clear genetic division between the Holstein and Sicilian cattle breeds was observed (Fig. 3). The Heatmap of genetic similarity corroborated the findings obtained with the PCA and showed that individuals from HOL and MOD were closer to individuals from their same population than to individuals of the CIN breed that showed the presence of substructure (Supplemental Data Fig. 2). Genotype data were also analyzed using a Bayesian clustering algorithm to search for admixture between populations and to infer population structure. Results from analysis of admixture considering a range from 1 through 6 potential clusters (K) pointed out that the greatest average likelihood value with the smallest variance among replicates was obtained for $K = 4$. The value of ln Pr($G|K$) increased from $K = 1$ to $K = 4$, reaching a plateau at $K = 4$, did not show a significant fluctuation from $K = 4$ to $K = 5$, and then decreased for $K = 6$ (Supplemental Data Fig. 3). For $K = 2$ Sicilian breeds clustered together and separated from HOL. For $K = 3$, although HOL belonged to a differentiated cluster, CIN and MOD presented some admixture between them. In addition, a low degree of introgression
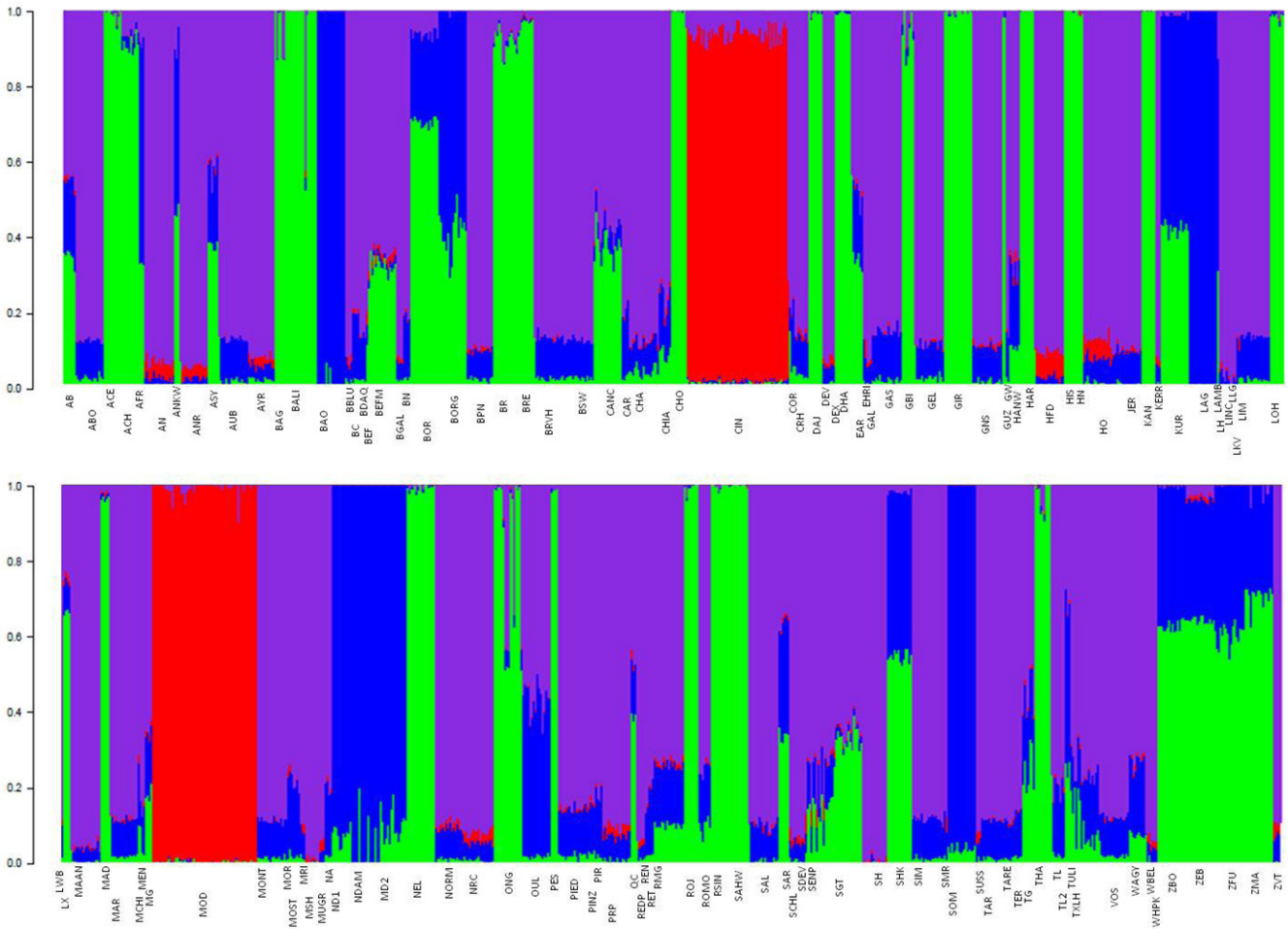
Figure 2. Ancestry models with 4 ancestral populations (*K* = 4). Violet represents *Bos taurus taurus,* green represents *B. javanicus* and *B. t. indicus,* blue represents African *B. t. taurus*, and red represents Sicilian breeds. See online version for figure in color.

with HOL was observed in CIN. A graphic representation of the estimated membership coefficients to the 4 clusters for each breed is shown in Fig. 4. HOL was the most differentiated breed, with 99.8% of the individuals assigned to cluster 1, whereas CIN animals showed a lower value of assignment, with a proportion of 46.9% of the individuals assigned to cluster 3 and 41.6% assigned to cluster 2. Cluster 4 included MOD individuals with 79.7% of the assignment value. Furthermore, 17.2% of MOD individuals were assigned to cluster 3, and 10.8% of CIN individuals were assigned to cluster 4 (Supplemental Data Table 2). Again, the results evidenced that MOD represented a well-supported group, with some degree of introgression with CIN genes, whereas CIN was split in 2 groups. Therefore, model-based clustering suggested that admixture has occurred and genetic links exist between the CIN and MOD breeds. These results may be explained by considering that CIN and MOD are 2 ancient cattle breeds reared in the same area, with a possible (natural or artificial) gene flow between them and with a common history and breeding practices. The CIN herd book was established in 1996, and before that date, the individuals of the CIN breed were
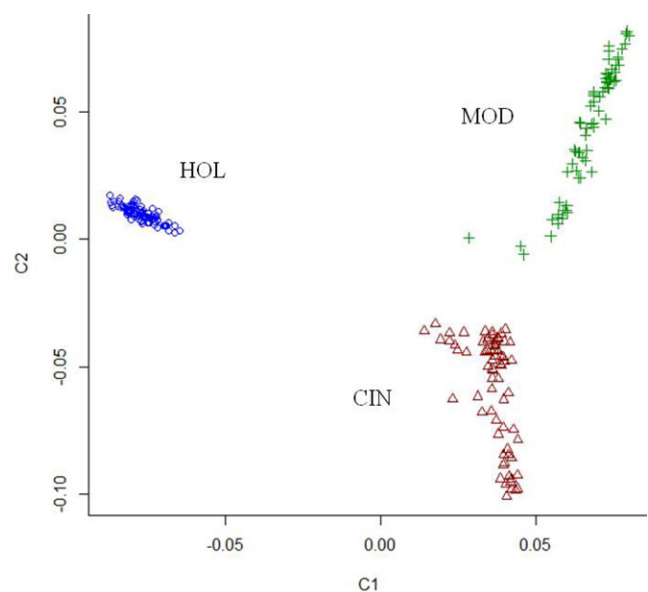
Figure 3. Principal components analysis among Holstein (HOL), Cinisara (CIN), and Modicana (MOD) cattle breeds. See online version for figure in color.
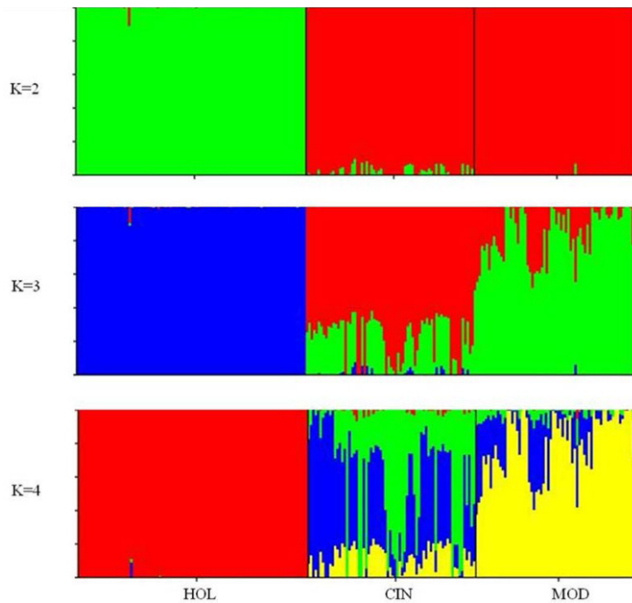
Figure 4. Bar plots obtained from Structure software, analyzing the population structure of Holstein (HOL), Cinisara (CIN), and Modicana (MOD) cattle breeds for *K* ranging from 2 to 4. See online version for figure in color.

registered in the MOD herd book, which was established in 1975. Furthermore, for CIN, the clustering analysis suggested that the breed did not constitute a homogenous population and had the presence of substructure. Moreover, this substructure in the CIN cattle breed was confirmed by the PCA analysis performed with only the 2 Sicilian cattle breeds (Supplemental Data Fig. 4). However, according to the herd book and morphological traits recorded, it did not seem there were crossbred individuals among the samples used in this work; hence, the genetic structure detected for the CIN breed could be due to the geographical isolation and proximity of some farms and the sampling of CIN animals from independent farms. In the Sicilian farming system, particularly for local breeds, natural mating is the common practice, and the exchange of males among herds is quite unusual (Mastrangelo et al., 2012). This leads to an increase of inbreeding within the population and would generate a population subdivision as consequence of genetic drift (e.g., the Wahlund effect). A PCA conducted considering only CIN animals (Supplemental Data Fig. 5) showed that individuals that clustered together belonged to farms located in the same geographic area. In fact, as suggested by Pritchard et al. (2000), the inferred clusters are not necessarily the corresponding real ancestral populations, and they can be determined by sampling schemes. Moreover, the presence of a substructure could evoke concerns about the generation of false-positive results when using LD mapping as the only means to locate genes underlying complex traits (Beghain et al., 2013). Therefore, the observed genetic structure was in agreement with the demographic history that occurred during the formation of the 2 Sicilian cattle breeds.

Table 2. Average distance (bp), linkage disequilibrium (r^2), and SD between adjacent SNP on each chromosome (BTA) in Cinisara (CIN) and Modicana (MOD) breeds

| BTA | CIN | | MOD | |
	Average spacing adjacent SNP (bp)	$r^2 \pm$ SD	Average spacing adjacent SNP (bp)	$r^2 \pm$ SD
1	60,021	0.181 ± 0.238	58,098	0.214 ± 0.255
2	63,612	0.179 ± 0.232	61,716	0.209 ± 0.256
3	60,938	0.177 ± 0.232	60,062	0.210 ± 0.256
4	61,549	0.162 ± 0.221	60,710	0.234 ± 0.274
5	72,329	0.167 ± 0.226	71,942	0.220 ± 0.267
6	58,056	0.180 ± 0.241	58,351	0.220 ± 0.266
7	64,388	0.184 ± 0.240	62,042	0.218 ± 0.263
8	60,443	0.171 ± 0.228	58,098	0.210 ± 0.260
9	66,665	0.175 ± 0.233	65,751	0.223 ± 0.268
10	61,063	0.169 ± 0.237	58,692	0.215 ± 0.264
11	60,839	0.168 ± 0.235	58,802	0.196 ± 0.253
12	69,317	0.159 ± 0.218	67,516	0.191 ± 0.239
13	59,774	0.153 ± 0.216	56,511	0.188 ± 0.242
14	57,228	0.174 ± 0.231	57,189	0.205 ± 0.257
15	63,277	0.158 ± 0.211	61,566	0.197 ± 0.240
16	63,377	0.171 ± 0.236	61,228	0.197 ± 0.238
17	60,442	0.161 ± 0.224	59,340	0.195 ± 0.246
18	61,410	0.160 ± 0.220	61,525	0.198 ± 0.244
19	58,028	0.152 ± 0.216	56,431	0.183 ± 0.227
20	57,784	0.150 ± 0.214	55,933	0.202 ± 0.242
21	65,287	0.168 ± 0.228	62,641	0.225 ± 0.262
22	61,524	0.156 ± 0.221	58,749	0.182 ± 0.237
23	60,931	0.138 ± 0.201	59,823	0.168 ± 0.212
24	60,866	0.161 ± 0.214	59,955	0.199 ± 0.242
25	55,880	0.162 ± 0.210	52,970	0.193 ± 0.245
26	60,401	0.142 ± 0.200	57,705	0.180 ± 0.238
27	60,767	0.140 ± 0.203	57,675	0.170 ± 0.226
28	61,577	0.141 ± 0.202	60,056	0.171 ± 0.223
29	59,659	0.146 ± 0.203	58,993	0.176 ± 0.226
Mean	61,636	0.162 ± 0.222	63,734	0.200 ± 0.247

Inbreeding and Coancestry Molecular Coefficients, Rates of Inbreeding and Coancestry, and Effective Population Size in Sicilian Cattle Breeds

Traditionally, *F* and *f* are estimated on the basis of pedigree information (Wright, 1922), but in most cases this information is unavailable or inaccurate in local breeds. Currently, with the availability of high-density SNP arrays, these coefficients can be estimated accurately in the absence of pedigree information (Allendorf et al., 2010; Li et al., 2011). The average molecular *F* and *f* coefficients were 0.68 ± 0.024 and 0.67 ± 0.03 in CIN and 0.69 ± 0.020 and 0.70 ± 0.03 in MOD, respectively (Table 1). Similar results were reported by Saura et al. (2013) in a study on genome-wide estimates of *F* and *f* coefficients in an endangered strain of Iberian pigs. High values of *F* and *f* coefficients in local breeds with low population size, such as CIN and MOD, can compromise the viability
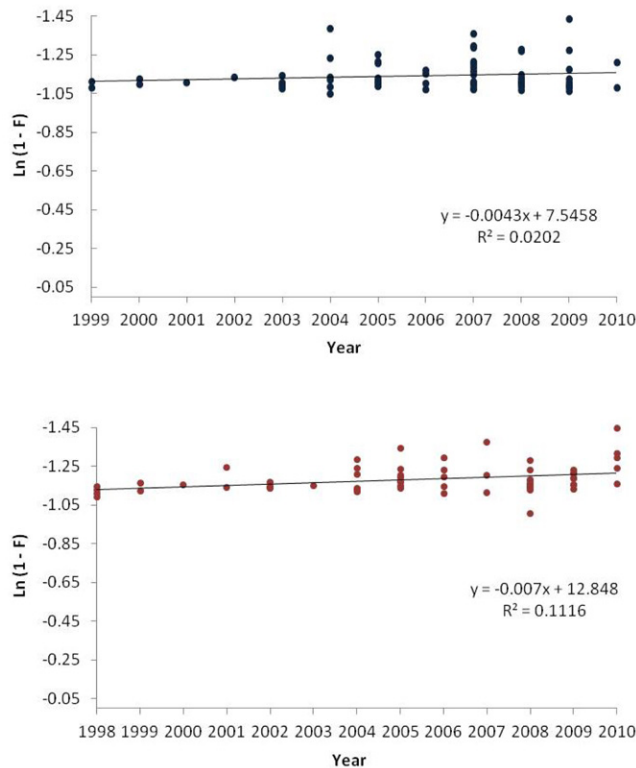
Figure 5. Rate of inbreeding per year in (top) Cinisara and (bottom) Modicana cattle breeds. See online version for figure in color.

Figure 6. Rate of coancestry per year in (top) Cinisara and (bottom) Modicana cattle breeds. See online version for figure in color.

of the populations. In fact, in terms of genetic variability, the endangered populations are less diverse, probably because of the reduced number of animals. The estimate of the molecular f coefficient was slightly greater than those reported by other authors for local cattle breeds and populations characterized by a reduction in their population sizes (Ginja et al., 2010; Maretto et al., 2012). Bozzi et al. (2012), in a study for conservation of Tuscan cattle breeds using microsatellite markers, obtained a maximum value for the f coefficient of 0.48. The results for CIN and MOD were expected considering the reduced number of reared animals and a farming system where mating with close relatives can be quite frequent. Rates of molecular inbreeding (ΔF) and coancestry (Δf) per year were 0.004 and 0.022 for CIN and 0.007 and 0.010 for MOD, respectively (Table 1, Fig. 5 and 6). Estimates of N_e from ΔF and Δf are shown in Table 1. The N_e values estimated from ΔF were about 19 animals in CIN and 12 animals in MOD, and those calculated from Δf were about 4 and 8 animals in the CIN and MOD cattle breeds, respectively. The current estimates of N_e were low, probably because of the intrapopulation genetic structuring. In animal breeding, the recommendation is to maintain N_e of at least 50 to 100 individuals (Frankham, 1995). Monitoring F and f coefficients is crucial to implement conservation programs because it is a strategy to avoid inbreeding depression and, in the extreme the risk of extinction (Frankham et al., 2002)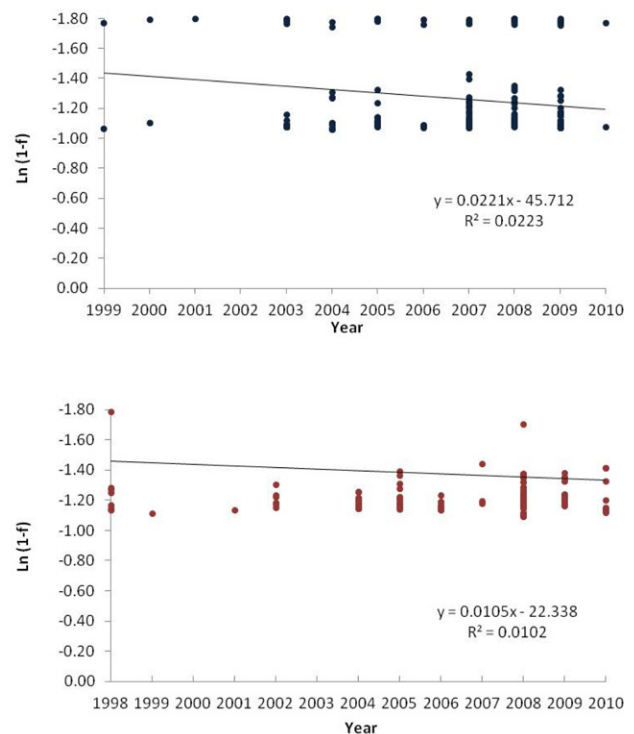. Moreover, managing the rates of inbreeding (F) and coancestry (f) provides a general framework to control the loss of variability by avoiding or alleviating the reductions in viability and fertility (Villanueva et al., 2004).

Linkage Disequilibrium

The extent of LD is often used to determine the optimal number of markers required for fine mapping of QTL (Meuwissen and Goddard, 2000), genomic selection (Meuwissen et al., 2001), increasing the understanding of genomic architecture and the evolutionary history of the populations (Hayes et al., 2003), and identifying regions influenced by natural selection. The extent of LD was first evaluated for each adjacent syntenic SNP pair. The average distance between adjacent SNP pairs for the entire autosomal genome was about 61 and 63 kb for the CIN and MOD breeds, respectively (Table 2). Average r^2 between adjacent SNP was 0.162 ± 0.222 for CIN and 0.200 ± 0.247 for MOD, and some variations in the LD values within chromosomes were observed in both populations. The value of r^2 ranged from 0.138 ± 0.201 for BTA23 to 0.184 ± 0.240 for BTA7 in CIN and from 0.168 ± 0.212 for BTA23 to 0.234 ± 0.274 for BTA4 in MOD (Table 2). Differences in LD among chromosomes have already been reported in Holstein cattle, and these can be attributed to the recombination rate, heterozygosity, genetic drift, and the effect of selection (Qanbari et al., 2010). Other authors reported similar average LD

Table 3. Mean linkage disequilibrium (r^2) among syntenic SNP over different map distances in Cinisara (CIN) and Modicana (MOD) breeds

Distance range (kb)	r^2	
	CIN	MOD
<50	0.192	0.234
50–100	0.113	0.154
100–200	0.071	0.111
200–500	0.046	0.084
500–1,000	0.038	0.071
1,000–2,000	0.034	0.062
2,000–5,000	0.030	0.050
>5,000	0.021	0.027

values between adjacent SNP pairs among the 29 different autosomes (Bohmanova et al., 2010; Flury et al., 2010; Beghain et al., 2013). The comparison of LD levels in different studies is not straightforward because of differences in several factors, such as sample size, type of LD measure (D' or r^2), marker types (microsatellites or SNP), marker density and distribution, and population demography (Qanbari et al., 2010). So far, studies of the extent of LD have been reported mostly for dairy and beef cattle breeds under selection, but there is little information about the degree of genome-wide LD in local cattle breeds and populations. Considering the LD levels between adjacent markers, the average r^2 in the MOD breed was comparable to those reported for] indigenous Swiss Eringer ($r^2 = 0.24$; Flury et al., 2010),
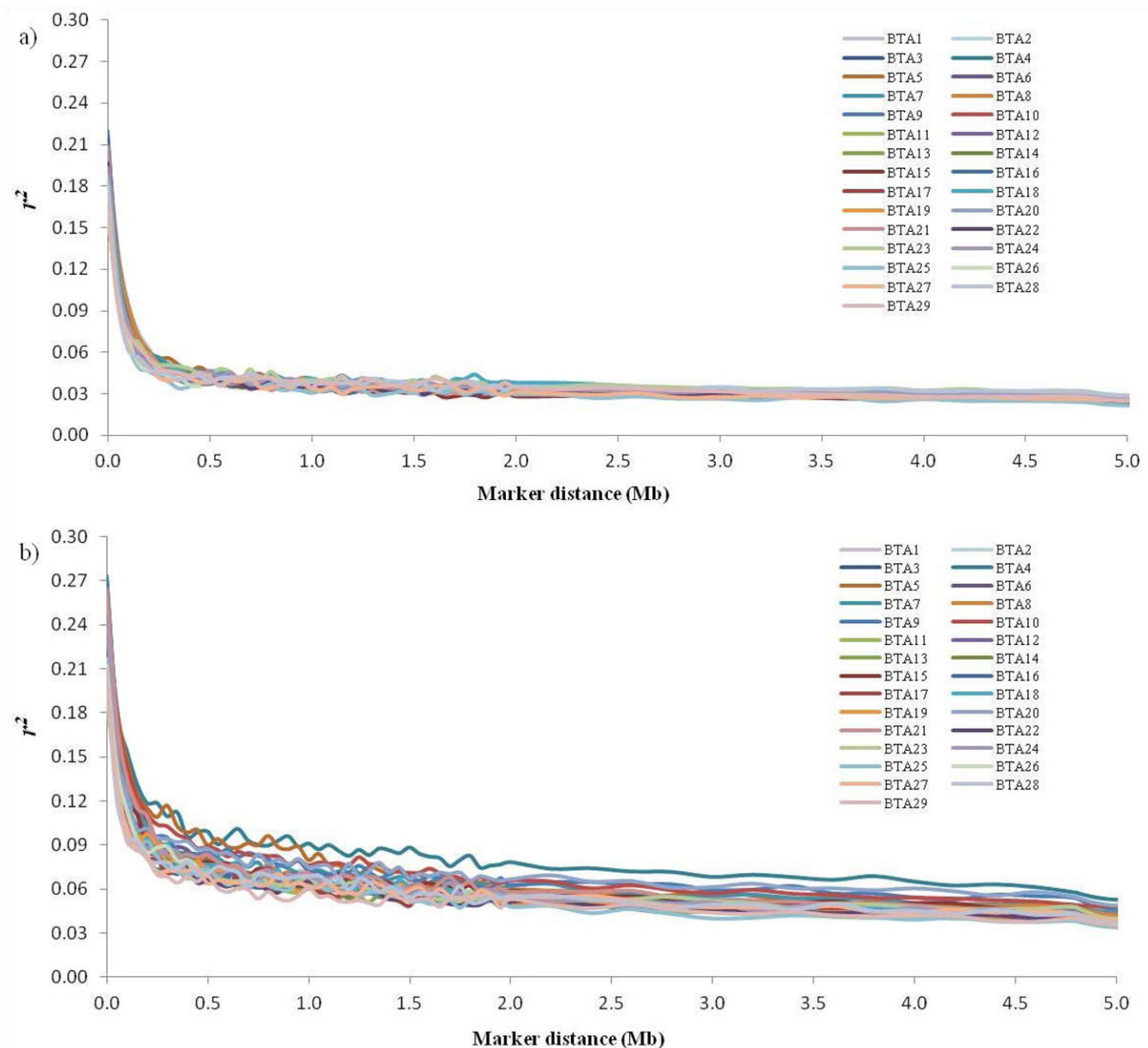
Figure 7. The r^2 distribution between syntenic SNP pairs on each autosome as a function of physical distance in (a) Cinisara and (b) Modicana cattle breeds. BTA = number of SNP per chromosome. See online version for figure in color.

Blonde d'Aquitaine ($r^2 = 0.20$; Beghain et al., 2013), and Chinese and Nordic Holsteins ($r^2 = 0.20$ and 0.21, respectively; Zhou et al., 2013), whereas the CIN breed had lower values. The results may be explained by considering the influence of selection on LD; in fact, the CIN breed is not subject to breeding programs, whereas the MOD breed is characterized by low selection pressure. Other factors, such as MAF and sample size, may be affecting the extent of LD within the genome. In both breeds, more than 66% of SNP had MAF greater than 0.2, suggesting that the effect of low MAF on the overall LD estimate should be small. Moreover, Khatkar et al. (2008) pointed out that a sample size near 70 individuals was sufficient for unbiased and accurate LD estimates (r^2) across marker intervals spanning <1 kb to >50 Mb.

Supplemental Data Table 3 shows the average values of r^2 estimated for all pairwise SNP combinations on each chromosome per breed. The mean values of r^2 pooled over all the autosomes were 0.038 ± 0.024 for CIN and 0.065 ± 0.032 for MOD. The level of pairwise LD decreased with physical distance between SNP (Fig. 7), with average r^2 ranging from 0.192 (CIN) and 0.234 (MOD) for SNP located up to 50 kb apart to 0.021 (CIN) and 0.027 (MOD) for SNP separated by more than 5,000 kb (Table 3). The LD decay in a genome determines the power of QTL detection in association mapping studies and helps us define the number of markers that are required for successful association mapping and genomic selection. Meuwissen et al. (2001), in a simulation to predict genomic breeding values from dense markers across the whole genome with accuracies up to 0.85, found a required r^2 level of 0.2. Qanbari et al. (2010) considered a threshold of 0.25 as a useful LD value for association studies. Therefore, these results support the need for more dense SNP arrays for a high-power association mapping and genomic selection efficiency, particularly for CIN cattle breeds. Moreover, the lowest value of LD detected in the CIN cattle breed could be explained by an effect of the sampling method, given that individuals from different farms were collected, or by ascertainment bias, given that the CIN and MOD cattle breeds were not included in the design of the BeadChip array. In addition, the possibility of crossbreeding with other breeds in the past could have broken the patterns of LD.

Conclusion

This study has reported, for the first time, population structure estimates, levels of inbreeding and coancestry, and LD from a genome-wide perspective in the CIN and MOD cattle breeds. Our results indicate that animals from the 2 breeds formed 2 different clusters with some degree of gene exchange between them. The high levels of inbreeding and coancestry as well as the low N_e obtained in this study could compromise the vi-

ability of these populations and indicate the necessity of implementing conservation programs to preserve these breeds. Avoiding mating among relatives (i.e., minimizing coancestry) is a strategy to control the increase of inbreeding, and it is the responsibility of all breeders to participate in pedigree recording to perform the appropriate mating. On the basis of the LD analyses, these results support the need for more dense SNP panels for a high-power association mapping and efficient genomic selection, particularly for CIN cattle breeds. In fact, the level of polymorphisms gives insight into how the Illumina Bovine BeadChip performed for the Sicilian breeds, and this insight has implications for its application for Genome Wide Association Study and genomic selection in these breeds.

The information generated in this study has important implications from economic and scientific perspectives and highlights the necessity to implement a genomic-driven conservation program for these Sicilian autochthonous cattle breeds.

LITERATURE CITED

AIA. 2014. Associazione Italiana Allevatori. www.aia.it/aia-website/it/attivita/area-tecnica/ufficio-sviluppo/razze-bovine-autoctone.

Alexander, D. H., and K. Lange. 2011. Enhancements to the ADMIXTURE algorithm for individual ancestry estimation. BMC Bioinform. 12:246.

Allendorf, F. W., P. A. Hohenlohe, and G. Luikart. 2010. Genomics and the future of conservation genetics. Nat. Rev. Genet. 11:697–709.

Barrett, J. C., B. Fry, J. Maller, and M. J. Daly. 2005. Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265.

Beghain, J., S. Boitard, B. Weiss, M. Boussaha, I. Gut, and D. Rocha. 2013. Genome wide linkage disequilibrium in the Blonde d'Aquitaine cattle breed. J. Anim. Breed. Genet. 130:294–302.

Bohmanova, J., M. Sargolzaei, and F. S. Schenkel. 2010. Characteristics of linkage disequilibrium in North American Holsteins. BMC Genomics 11:421.

Bovine HapMap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532.

Bozzi, R., I. Alvarez, A. Crovetti, I. Fernandez, D. De Petris, and F. Goyache. 2012. Assessing priorities for conservation in Tuscan cattle breeds using microsatellites. Animal 6:203–211.

Caballero, A., and M. A. Toro. 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet. 3:289–299.

Decker, J. E., S. D. McKay, M. M. Rolf, J. Kim, A. M. Alcalá, T. S. Sonstegard, O. Hanotte, A. Götherström, C. M. Seabury, L. Praharani, M. E. Babar, L. C. d. A. Regitano, M. A. Yildiz, M. P. Heaton, W. Liu, C. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel, and J. F. Taylor. (2013) Data from: Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. Dryad Digital Repository. doi:10.5061/dryad.th092.

Decker, J. E., S. D. McKay, M. M. Rolf, J. Kim, A. M. Alcalá, T. S. Sonstegard, O. Hanotte, A. Götherström, C. M. Seabury, L. Praharani, M. E. Babar, L. C. d. A. Regitano, M. A. Yildiz, M. P. Heaton, W. Liu, C. Lei, J. M. Reecy, M. Saif-Ur-Rehman, R. D. Schnabel, and J. F. Taylor. 2014. Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. PLoS Genet. 10(3):e1004254.

Flury, C., M. Tapio, C. Sonstegard, C. Drogemuller, T. Leeb, H. Simianer, O. Hanotte, and S. Rieder. 2010. Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium. J. Anim. Breed. Genet. 127:339–347.

Frankham, R. 1995. Conservation genetics. Annu. Rev. Genet. 29:305–327.

Frankham, R., J. D. Ballou, and D. A. Briscoe. 2002. Introduction to conservation genetics. Cambridge Univ. Press, Cambridge.

Gautier, M., D. Laloë, and K. Moazami-Goudarzi. 2010. Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds. PLoS ONE 5:e13038.

Ginja, C., M. C. T. Penedo, M. F. Sobral, J. Matos, C. Borges, D. Neves, T. Rangel-Figueiredo, and A. Cravador. 2010. Molecular genetic analysis of a cattle population to reconstitute the extinct Algarvia breed. Genet. Sel. Evol. 42:18.

Guastella, A. M., S. Sorbolini, A. Zuccaro, E. Pintus, S. Bordonaro, D. Marletta, and N. P. P. Macciotta. 2011. Melanocortin 1 receptor (MC1R) gene polymorphisms in three Italian cattle breeds. Anim. Prod. Sci. 51:1039–1043.

Hayes, B. J., P. M. Visscher, H. C. McPartlan, and M. E. Goddard. 2003. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. Genome Res. 13:635–643.

Khatkar, M. S., F. W. Nicholas, A. R. Collins, K. R. Zenger, J. A. Cavanagh, W. Berris, R. D. Schnabel, J. F. Taylor, and H. W. Raadsma. 2008. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high-density SNP panel. BMC Genomics 9:187.

Kijas, J. W., J. A. Lenstra, B. Hayes, S. Boitard, L. R. Porto Neto, M. San Cristobal, B. Servin, R. McCulloch, V. Whan, K. Gietzen, S. Paiva, W. Barendse, E. Ciani, H. Raadsma, J. McEwan, B. Dalrymple, other members of the International Sheep Genomics Consortium 2012. Genome-Wide Analysis of the world's sheep breeds reveals high levels of historic mixture and strong recent selection. PLoS Biol. 10(2):e1001258.

Kijas, J. W., D. Townley, B. P. Dalrymple, M. P. Heaton, J. F. Maddox, A. McGrath, P. Wilson, R. G. Ingersoll, R. McCulloch, S. McWilliam, D. Tang, J. McEwan, N. Cockett, V. Hutton Oddy, F. W. Nicholas, H. Raadsma for the International Sheep Genomics Consortium. 2009. A genome wide survey of SNP variation reveals the genetic structure of sheep breeds. PLos ONE 4(3):e4668.

Li, M. H., I. Strandén, T. Tiirikka, M. L. Sevón-Aimonen, and J. Kantanen. 2011. A comparison of approaches to estimate the inbreeding coefficient and pairwise relatedness using genomic and pedigree data in a sheep population. PLoS ONE 6:e26256.

Maretto, F., J. Ramljak, F. Sbarra, M. Penasa, R. Mantovani, A. Ivankovic, and G. Bittante. 2012. Genetic relationships among Italian and Croatian Podolian cattle breeds assessed by microsatellite markers. Livest. Sci. 150:256–264.

Mastrangelo, S., M. T. Sardina, V. Riggio, and B. Portolano. 2012. Study of polymorphisms in the promoter region of ovine β-lactoglobulin gene and phylogenetic analysis among the Valle del Belice breed and other sheep breeds considered as ancestors. Mol. Biol. Rep. 39:745–751.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4:e5350.

Meuwissen, T. H. E., and M. E. Goddard. 2000. Fine mapping of quantitative trait loci using linkage disequilibrium with closely linked marker loci. Genetics 155:421–430.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome wide dense marker maps. Genetics 155:945–959.

Miller, S. A., D. D. Dykes, and H. F. Polesky. 1988. A simple salting out procedure for extracting DNA from human nucleated cells. Nucleic Acids Res. 16:1215.

Pritchard, J. K., M. Stephens, and P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly, and P. C. Sham. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81:559–575.

Qanbari, S., E. C. Pimentel, J. Tetens, G. Thaller, P. Lichtner, A. R. Sharifi, and H. Simianer. 2010. The pattern of linkage disequilibrium in German Holstein cattle. Anim. Genet. 41:346–356.

R Development Core Team. 2011. R: A language and environment for statistical computing. R Found. Stat. Comput., Vienna.

Saura, M., A. Fernández, M. C. Rodríguez, M. A. Toro, C. Barragán, A. I. Fernández, and B. Villanueva. 2013. Genome-wide estimates of coancestry and inbreeding depression in an closed herd of ancient Iberian pigs. PLoS ONE 8(10):e78314.

Sechi, T., M. G. Usai, S. Miari, L. Mura, S. Casu, and A. Carta. 2007. Identifying native animals in crossbred populations: The case of the Sardinian goat population. Anim. Genet. 38:614–620.

Tishkoff, S. A., F. A. Reed, F. R. Friedlaender, C. Ehret, A. Ranciaro, A. Froment, J. B. Hirbo, A. A. Awomoyi, J.-M. Bodo, O. Doumbo, M. Ibrahim, A. T. Juma, M. J. Kotze, G. Lema, J. H. Moore, H. Mortensen, T. B. Nyambo, S. A. Omar, K. Powell, G. S. Pretorius, M. W. Smith, M. A. Thera, C. Wambebe, J. L. Weber, and S. M. Williams. 2009. The genetic structure and history of Africans and African Americans. Science 324:1035–1044.

Villanueva, B., R. Pong-Wong, J. A. Woolliams, and S. Avendaño. 2004. Managing genetic resources in selected and conserved populations. In: G. Simm, B. Villanueva, K. D. Sinclair, and S. Townsend, editors, Farm animal genetic resources. Nottingham Univ. Press, Nottingham, UK. p. 113–132.

Wright, S. 1922. Coefficients of inbreeding and relationship. Am. Nat. 56:330–338.

Zhou, L., X. Ding, Q. Zhang, Y. Wang, M. S. Lund, and G. Su. 2013. Consistency of linkage disequilibrium between Chinese and Nordic Holsteins and genomic prediction for Chinese Holsteins using a joint reference population. Genet. Sel. Evol. 45:7.