Published November 10, 2015

Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip[1]

J. J. Cañas-Álvarez,*[2] A. González-Rodríguez,†[3] S. Munilla,†‡ L. Varona,†§ C. Díaz,# J. A. Baro,‖ J. Altarriba,†§ A. Molina,¶ and J. Piedrafita*[4]

*Grup de Recerca en Remugants, Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193-Bellaterra, Barcelona, Spain; †Departamento de Anatomía, Embriología y Genética, Universidad de Zaragoza, 50013-Zaragoza, Spain; ‡Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, 1417, CABA, Argentina; §Instituto Agroalimentario de Aragón (IA2), 50013-Zaragoza, Spain; #Departamento de Mejora Genética Animal, INIA, 28040-Madrid, Spain; ‖Departamento de Ciencias Agroforestales, Universidad de Valladolid, 34004-Palencia, Spain; and ¶MERAGEM, Universidad de Córdoba, 14071-Córdoba, Spain

ABSTRACT: The availability of SNP chips for massive genotyping has proven to be useful to genetically characterize populations of domestic cattle and to assess their degree of divergence. In this study, the Illumina BovineHD BeadChip genotyping array was used to describe the genetic variability and divergence among 7 important autochthonous Spanish beef cattle breeds. The within-breed genetic diversity, measured as the marker expected heterozygosity, was around 0.30, similar to other European cattle breeds. The analysis of molecular variance revealed that 94.22% of the total variance was explained by differences within individuals whereas only 4.46% was the result of differences among populations. The degree of genetic differentiation was small to moderate as the pairwise fixation index of genetic differentiation among breeds (F_{ST}) estimates ranged from 0.026 to 0.068 and the Nei's D genetic distances ranged from 0.009 to 0.016. A neighbor joining (N-J) phylogenetic tree showed 2 main groups of breeds: Pirenaica, Bruna dels Pirineus, and Rubia Gallega on the one hand and Avileña-Negra Ibérica, Morucha, and Retinta on the other. In turn, Asturiana de los Valles occupied an independent and intermediate position. A principal component analysis (PCA) applied to a distance matrix based on marker identity by state, in which the first 2 axes explained up to 17.3% of the variance, showed a grouping of animals that was similar to the one observed in the N-J tree. Finally, a cluster analysis for ancestries allowed assigning all the individuals to the breed they belong to, although it revealed some degree of admixture among breeds. Our results indicate large within-breed diversity and a low degree of divergence among the autochthonous Spanish beef cattle breeds studied. Both N-J and PCA groupings fit quite well to the ancestral trunks from which the Spanish beef cattle breeds were supposed to derive.

Key words: admixture, beef breeds, genetic distances, heterozygosity, neighbor joining trees, single nucleotide polymorphism variability

J. Anim. Sci. 2015.93:5164–5174
doi:10.2527/jas2015-9271

[1]The research leading to these results has received funding of a Ministerio de Ciencia e Innovación AGL 2010-15903 grant from the Spanish government, and from the European Union's Seventh Framework Program for research, technological development and demonstration under grant agreement number 289592 – Gene2Farm. The collaboration of breed societies and Juan José Arranz in collecting samples and the support of FEAGAS is also acknowledged.

[2]J. J. Cañas-Álvarez acknowledges the financial support given by COLCIENCIAS through the Francisco José de Caldas fellowship 497/2009.

[3]A. González-Rodríguez acknowledges the financial support given by the Spanish government BES-2011-045434 fellowship.

[4]Corresponding author: jesus.piedrafita@uab.es

Received May 5, 2015.
Accepted September 9, 2015.

INTRODUCTION

European cattle breeds derive from the migration of cattle from the Near East. When this expansion reached the Iberian Peninsula, new crosses took place with animals from the African continent (Ammerman and Cavalli-Sforza, 1984; Decker et al., 2014). The formation of the autochthonous cattle breeds in Spain went through different stages. Initially, they were used as triple purpose (draft, milk, and beef) animals; then, depending on the region, their characteristics, and the geographical boundaries, they began to diverge to the present breeds. In a more recent era, the systematic application of modern breeding techniques led to the differentiation of the breeds and set each racial biotype (Sánchez-Belda, 1984) by isolation, drift, selection, and adaptation to their particular habitat.

Initially, the breeds were genetically characterized through immunogenetic markers and/or biochemical polymorphisms (Kidd et al., 1980; González et al., 1987; Blott et al., 1998). Later on, the genetic relationships among breeds have been studied through the use of microsatellite markers (MacHugh et al., 1997; Beja-Pereira et al., 2003; Martín-Burriel et al., 2011). However, microsatellite markers do not occur throughout the whole genome and, compared with SNP, have relatively high mutation rates per generation, what would misperceive the population history which would misinterpret the population history (Brumfield et al., 2003). On the other hand, SNP are biallelic, what facilitates high-throughput genotyping and minimizes recurrent substitutions at a single site because multiple mutations at a single site are unlikely. The availability of SNP has been particularly useful to genetically characterize populations of cattle and estimate phylogenetic relationships (Gautier et al., 2007; The Bovine HapMap Consortium, 2009; Decker et al., 2014). Therefore, the objective of this study was to reassess the genetic diversity, the degree of divergence, and the relationships of 7 autochthonous Spanish beef cattle breeds using a high-density SNP chip.

MATERIAL AND METHODS

Animals and Sampling

The breeds, acronyms, and sample sizes were as follows: Asturiana de los Valles (**AV**; $n = 50$), Avileña-Negra Ibérica (**ANI**; $n = 48$), Bruna dels Pirineus (**BP**; $n = 50$), Morucha (**Mo**; $n = 50$), Pirenaica (**Pi**; $n = 48$), Retinta (**Re**; $n = 46$), and Rubia Gallega (**RG**; $n = 44$). The animals studied, both males and females in equal numbers, were chosen from different and separated geographical areas, taking care to avoid known relationships. The animals were the parents of trios used in a study aimed at estimating linkage disequilibrium (**LD**) and effective populations sizes (Cañas-Álvarez, 2015). The blood samples were collected from the caudal vein of animals in tubes with EDTA as anticoagulant, following the recommendations of the Joint Working Group on Refinement (1993).

Single Nucleotide Polymorphism Genotyping, Data Pruning, and Genetic Diversity Assessment

The samples were processed following the protocol described in the PrepFiler Forensic DNA Extraction Kit of Applied Biosystems (Foster City, CA), using MagMAX Express-96 Magnetic Particle Processor automated equipment of Applied Biosystems (Foster City, CA). High-density SNP genotyping was performed according to the protocol of the manufacturer by using the BovineHD BeadChip (Illumina Inc., 2012) designed to genotype 777,962 SNP, at a commercial laboratory (Xenética Fontao, Lugo, Spain). The SNP mapped to the UMD3.1 assembly (Zimin et al., 2009). Only SNP mapped to autosomal chromosomes were used in this study. The SNP that have the same genomic location (3,014 SNP, 0.387% of the total, the same number for all breeds) and those with Mendelian error rates greater than 5% were removed. These initial quality controls retained 735,239 SNP for each breed. To assess the genetic diversity, we computed the following statistics: 1) percentage of markers genotyped in more than 95% of the samples, 2) percentage of markers with a minor allele frequency (**MAF**) > 0.05, 3) expected heterozygosity, 4) mean number of alleles, and 5) percentage of loci in Hardy–Weinberg equilibrium ($P > 0.01$) using an exact Hardy–Weinberg equilibrium test (Wigginton et al., 2005).

Before analyzing the divergence among breeds, we made an additional pruning of the marker data set according to the following criteria. First, an individual call rate ≥ 0.95 and a SNP call rate ≥ 0.95 were required. Next, considering that the SNP in strong LD can affect both principal component analysis (**PCA**) and clustering analysis, we thinned the marker set by excluding SNP in strong LD (pairwise genotypic correlation $r^2 > 0.1$) in a window of 50 SNP, sliding the window by 5 SNP at a time as in Moorjani et al. (2013). As a result of this data pruning, the same 57,674 SNP were left for the analysis of divergence among populations. After pruning, all individuals were kept in the study. The average distance between markers was 43.5 kb and the average r^2 was 0.025, ranging from 0.023 (AV and BP) to 0.03 (RG). All quality controls and data pruning were performed using PLINK software version 1.07 (Purcell et al., 2007). Nevertheless, in addition to the analyses based in pruned data, we perform a sensitivity analysis with the complete data set to check for potential differences.

Analysis of Molecular Variance

Levels of genetic variation within and among populations were estimated using an analysis of molecular variance (**AMOVA**; Excoffier et al., 1992). The analysis was performed with Arlequin software version 3.5 (Excoffier et al., 2005), where significance levels for variance components were tested using nonparametric permutation procedures (1,000 permutations). To convert files between PLINK and Arlequin formats, PGDSpider 2.0.5.1 software (Lischer and Excoffier, 2012) was used.

Distance Measures

To assess the divergence among breeds, we computed 2 measures of distance using breed allele frequencies. First, genetic differentiation among breeds ($\mathbf{F_{ST}}$) fixation indices were calculated through the Arlequin software version 3.5 (Excoffier et al., 2005), by using 20,000 permutations and with a significance level of 0.05. Next, Nei's D genetic distances (Nei, 1972) between all pairs of breeds were estimated by means of the PHYLIP software version 3.69 (Felsenstein, 1989). A neighbor joining (**N-J**) tree (Saitou and Nei, 1987) was built from Nei's D genetic distances, using the packages *seqboot*, *gendist*, *neighbor*, and *consense* from the PHYLIP software (Felsenstein, 1989). To evaluate the robustness of the phylogenetic tree, a bootstrap from 1,000 replicates was performed. This bootstrapping approach implied generating a tree for each replicate and then computing the percentage of trees repeated for each node. The tree was plotted in an R environment (R Core Team, 2014) with the APE R package (Paradis et al., 2004).

Principal Components and Multidimensional Scaling Analysis

To achieve a different approach to characterize divergence, a PCA was applied to the relationship matrix, built up from the pairwise identity by state (**IBS**) between all individuals using PLINK (Purcell et al., 2007). Each entry of this relationship matrix relates any 2 individuals genotyped and is computed as 1 minus the pairwise IBS, averaged across markers. In addition, a multidimensional scaling (**MDS**) plot on the same relationship matrix was also performed to complement the PCA. All estimates and plots were performed using self-written code developed under an R environment (R Core Team, 2014).

Genetic Structure and Levels of Admixture

Finally, a cluster analysis using the ADMIXTURE software version 1.23 (Alexander et al., 2009) was performed to characterize genetic structure across all breeds. The program implements a maximum likelihood method to infer the genetic ancestry of each individual from a mixture of K predefined ancestral groups. The number of clusters (K) tested ranged from 2 to 7. A preferable value of K will exhibit a low cross-validation error compared with other K values.

RESULTS AND DISCUSSION

Genetic Diversity Analysis

Genetic diversity across breeds was assessed through several statistics computed from 735,239 SNP in autosomal chromosomes after eliminating the SNP having Mendelian error rates greater than 5%. The proportion of markers genotyped on 95% of the samples was around 97 to 98% in all breeds, which suggests the suitability of the chip to genotype the breeds studied (Table 1). The proportion of SNP with a MAF > 0.05 ranged between 86 (Pi) and 89% (AV), indicating that most SNP are segregating in all breeds. In fact, the mean number of alleles was similar among breeds, ranging from 1.87 (Pi and Re) to 1.89 (AV). To have further insight on the degree of polymorphism observed across breeds, we represented the average proportion of SNP for different ranges of MAF (Fig. 1). The spectrum of MAF was very similar among breeds. The percentage of monomorphic loci varied between 11 (RG) and 15% (Re). In general, these results suggest similar within-breed variability, an observation also reinforced by the similar expected heterozygosity across breeds (Table 1).

The degree of polymorphism observed in this sample was greater than that obtained from a sample of several European (73–83%) and African (47–71%) cattle breeds using 696 SNP (Gautier et al., 2007). It was also greater than the average 83% reported by Salomon-Torres et al. (2014) for bovine breeds using the same BovineHD chip with MAF > 0.01. In addition, Edea et al. (2013) found a similar degree of polymorphism in Ethiopian and Hanwoo cattle populations, ranging from 81.63 to 95.21%, using the Illumina's Bovine 8K SNP chip (MAF > 0.05; Boichard et al., 2012). The differences between estimates could be due in part to ascertainment bias, that is, the low amount of polymorphic markers found in these breeds due to the relationship between the breed assayed and the ones used in the design of the chip (Matukumalli et al., 2009). Porto-Neto et al. (2014) studied different breeds in Australia and compared them with those involved in the bovine HapMap (Gibbs et al., 2009). They found that similarly to what happens in the bovine HapMap populations, Indicine breeds as well as composite breeds showed a greater level of polymorphism than taurine breeds. The level of polymorphisms

Table 1. Genetic diversity within 7 Spanish beef breeds

Breed[1]	Markers genotyped on >95% of the samples, %	Markers with MAF[2] > 0.05, %	Mean number of alleles	Markers in HWE[3] ($P > 0.01$), %	Expected heterozygosity (SD)
AV	97.74	89.28	1.89	99.27	0.319 (0.172)
ANI	97.94	87.17	1.88	99.05	0.306 (0.177)
BP	97.05	88.17	1.88	99.20	0.309 (0.175)
Mo	97.60	87.72	1.88	98.70	0.310 (0.176)
Pi	98.03	86.65	1.87	99.32	0.299 (0.180)
Re	97.74	87.18	1.87	98.77	0.304 (0.178)
RG	97.97	87.79	1.88	99.28	0.308 (0.176)

[1]AV = Asturiana de los Valles; ANI = Avileña-Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

[2]MAF = minor allele frequency.

[3]HWE = Hardy–Weinberg equilibrium.

found in the Spanish breeds was similar to Australian taurine breeds (0.85–0.90).

In general, no significant departures from the Hardy–Weinberg equilibrium were observed among the polymorphic markers in any breed, and the percentage of markers in Hardy–Weinberg equilibrium ($P > 0.01$) was always over 98%. Therefore, the expected and observed heterozygosities were similar, what may suggest a lack of stratification within breeds (Fernández et al., 2008). However, the analysis performed in these breeds based on the pedigree books showed a different picture (Cañas-Álvarez et al., 2014). The rate of coancestry by year of birth among individuals in the populations, except in one, was smaller than the rate of inbreeding, pointing to hidden structures in them (Cañas-Álvarez et al., 2014). Depth of pedigrees trace back a limited number of generations (between 3 and 6); therefore, an excess of IBS that could not be identified as identical by descent has been detected because the SNP chip mostly recalls ancestral events (Lachance and Tishkoff, 2013). Expected heterozygosity across breeds averaged 0.307, with estimates ranging from 0.299 in Pi to 0.319 in AV, indicating that high within-breed diversity exists. Our results were fairly close to those observed in European breeds by Gautier et al. (2007), who found average percentage of markers in Hardy–Weinberg equilibrium, expected heterozygosity, and mean number of alleles of 82.89%, 0.30, and 1.85, respectively. In turn, expected heterozygosity values around 0.35 in *Bos indicus* and 0.4 in *Bos taurus* were reported in indigenous cattle from Ethiopia and Korea (Edea et al., 2013) and about 0.4 in 2 Bangladeshi zebu cattle populations with a 80K SNP chip (GeneSeek, Lincoln, Nebraska). All these estimates (including our own) are potentially biased upward due to the ascertainment of SNP markers that tend to be at intermediate allele frequencies (Lachance and Tishkoff, 2013).

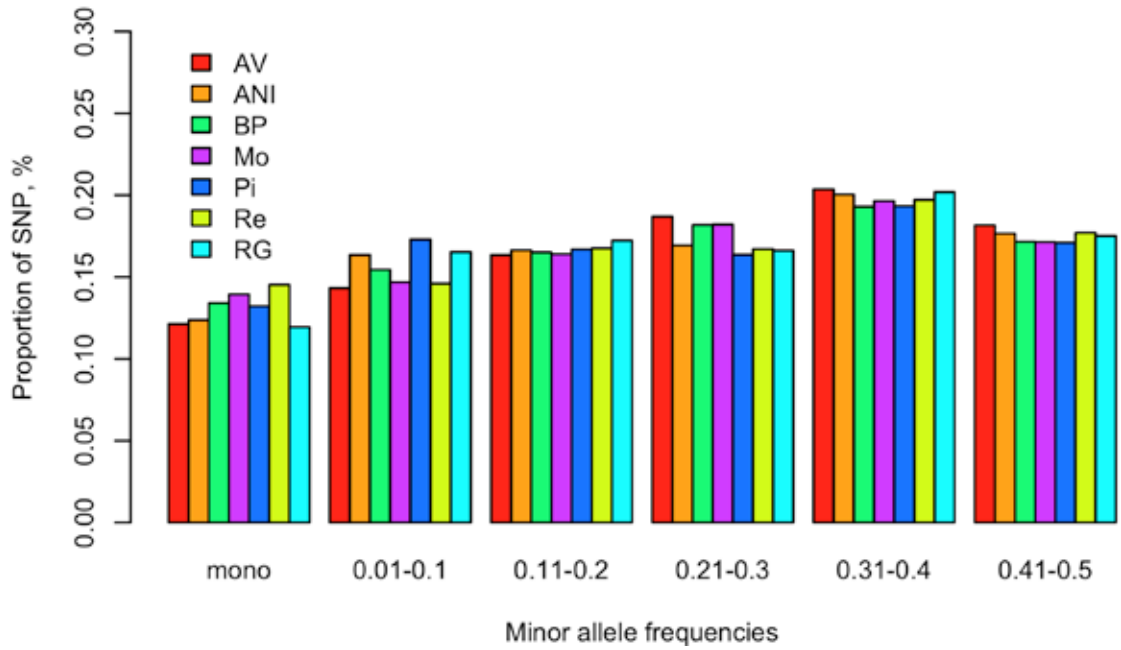

Figure 1. Average proportions of SNP within minor allele frequency ranges by population. mono = monomorphic; AV = Asturiana de los Valles; ANI = Avileña-Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

Table 2. Analysis of molecular variance in the 7 populations

Source of variation	df	Variance components	Percentage of variation	Fixation indices[1]
Among populations	6	258.165	4.46	F_{ST} = 0.044**
Among individuals within populations	329	76.543	1.32	F_{IS} = 0.014 NS[2]
Within individuals	336	5,456.510	94.22	F_{IT} = 0.058**
Total	671	5,791.218		

**$P < 0.001$.

[1]F_{ST} = genetic differentiation among breeds; F_{IS} = within population inbreeding; F_{IT} = total inbreeding.

[2]NS = not significant.

Analysis of Divergence among Breeds

Before analyzing the divergence among breeds, we thinned the marker set by excluding SNP in strong LD (pairwise genotypic correlation $r^2 > 0.1$). As a result of this data pruning, the same 57,674 SNP were left for the analysis of divergence among populations, and the average r^2 was 0.025, ranging from 0.023 (AV and BP) to 0.03 (RG). This average value is close to the background LD level estimated from nonsyntenic markers, which was around 0.01 across breeds (Cañas-Álvarez, 2015).

The pattern of divergence among the 7 Spanish beef cattle breeds was studied following 4 different approaches: 1) AMOVA and fixation indices, 2) distance measures and phylogenetic trees, 3) PCA on molecular relationship matrices, and 4) genetic structure and levels of admixture. All the measures of divergence were based on a pruned subset of 57,674 autosomal SNP, with a LD lower than 0.1 among any of them.

1. ***Analysis of Molecular Variance and Fixation Indices.*** The genetic variation was partitioned by means of an AMOVA. The most important part of the variation (94.22%; $P < 0.001$) was attributable to the variation within individuals (Table 2). The variation between populations accounted for just 4.46% ($P < 0.001$), and the variation among individuals within populations was even lower (1.32%; $P = 0.078$). Our results indicate that the Spanish breeds have low levels of between-population genetic variation, pointing toward a higher common genetic background and probably some admixture between populations.

   The fixation indices F_{IT}, F_{IS}, and F_{ST} across all loci, estimated from the AMOVA, were also examined (Table 2). We found a total inbreeding (F_{IT} value) of 0.058, within-population inbreeding (F_{IS} value) of 0.014 (nonsignificant), and a genetic differentiation among breeds (F_{ST} value) of 0.044. The low and nonsignificant F_{IS} value confirms that the allele frequencies of the populations are in the expected Hardy–Weinberg proportions. The population-specific F_{IS} indices in each breed also showed values close to 0, ranging from –0.012 to 0.052 (Supplementary Table S1; see the online version of the article at http://journalofanimalscience.org). The F_{ST} value of 0.044 indicated the existence of limited population structuring among the studied breeds, according to the rating of Wright (1965). This low differentiation among breeds could be attributed to a lack of selection pressure or the existence of a moderate gene flow (migration) among these populations. Gautier et al. (2007), studying a limited panel of SNP, observed average values of F_{ST} of 0.099 in European breeds, including both beef and dairy breeds. Meanwhile, Edea et al. (2013) observed a low level of differentiation between the Ethiopian cattle populations (F_{ST} = 0.01). In the context of Iberian breeds, we found a lower value than the one reported by Martín-Burriel et al. (2011) using a panel of 30 microsatellites, who found an F_{ST} value of 0.086. However, Martín-Collado et al. (2013) obtained the same magnitude of F_{ST} estimates between clusters of individuals in the ANI populations. Estimates obtained with microsatellites could not be strictly compared with ours because the range of F_{ST} estimates depends on the frequency of the most frequent allele, and this is always higher for biallelic SNP than for multiallelic microsatellites. Therefore, the use of SNP should provide higher F_{ST} estimates (Jakobsson et al., 2013). Given that our estimate is clearly lower than the one obtained in Martín-Burriel et al. (2011), it is suggested that the whole set of Iberian breeds are much more differentiated, probably due to genetic drift because small effective sizes, than the subset of breeds we are studying in this work. It should be noted that, in this study, the populations were chosen as the most representative and with a higher census among the authochtonous Spanish beef cattle populations. On the other hand, the study of Martín-Burriel et al. (2011) uses up to 27 Spanish and 13 Portuguese populations, some of them with a very limited census. Nevertheless, the differences also can be attributed to a scale effect due to the density and the way the sampling was done in both studies.

Table 3. Estimates of the pairwise genetic differentiation statistic (F_{ST} statistics; below the diagonal) and the Nei's D genetic distance (above the diagonal) among populations

Breed[1]	AV	ANI	BP	Mo	Pi	Re	RG
AV		0.0105	0.0086	0.0092	0.0120	0.0109	0.0095
ANI	0.0345		0.0140	0.0092	0.0159	0.0119	0.0132
BP	0.0257	0.0529		0.0131	0.0123	0.0145	0.0116
Mo	0.0284	0.0299	0.0481		0.0150	0.0110	0.0121
Pi	0.0427	0.0644	0.0452	0.0588		0.0165	0.0144
Re	0.0375	0.0436	0.0563	0.0391	0.0680		0.0137
RG	0.0290	0.0481	0.0401	0.0419	0.0552	0.0509	

[1]AV = Asturiana de los Valles; ANI = Avileña-Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

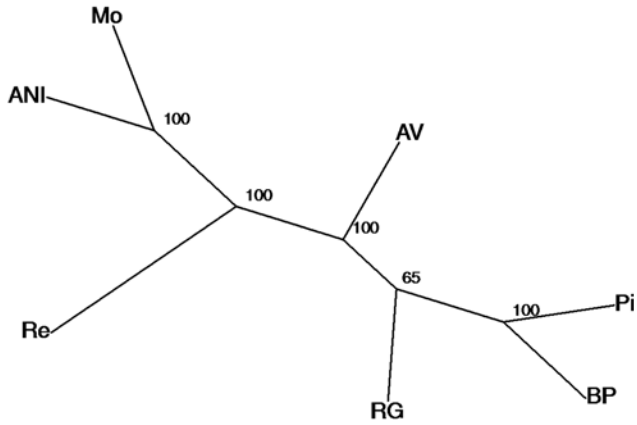

Figure 2. Neighbor joining representation of the pairwise Nei's D genetic distances among populations. The numbers at the nodes are the percentage of bootstrap replicates that resulted in the depicted topology. AV = Asturiana de los Valles; ANI = Avileña-Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

2. Distance Measures and Phylogenetic Trees. The pairwise F_{ST} statistic among populations is also a measure of the genetic distance among subpopulations (Excoffier et al., 2005). The pairwise F_{ST} estimates between breeds showed values that ranged from 0.026 to 0.068 (Table 3). The lowest pairwise F_{ST} estimates were observed between AV and the rest of the breeds, whereas the highest were between Pi and both ANI and Re (>0.06). In addition to pairwise F_{ST}, we estimated the Nei's D genetic distance (Nei, 1972) among all populations. The pairwise Nei's D genetic distance showed a pattern similar to the one obtained with the F_{ST} statistics, with values ranging from 0.009 to 0.016. Analyzing European breeds with SNP, Gautier et al. (2007) found pairwise F_{ST} values ranging from 0.035 (French breeds Salers and Aubrac) to 0.132 (Normande and Holstein). Our higher F_{ST} distances were closer to the lower bound found by these authors.

A phylogenetic tree was constructed from Nei's D genetic distances (Nei, 1972), using the N-J method (Saitou and Nei, 1987; Fig. 2). It is worth mentioning that bootstrap percentages computed to assess the robustness of the phylogenetic tree, depicted in the internal nodes of Fig. 2, showed values of 100% except for the classification of RG, which showed a bootstrap value of 65%. These values clearly support the classification represented in the tree, compared with the low to moderate bootstrap percentages reported in previous studies (Martín-Burriel et al., 1999; Beja-Pereira et al., 2003; Martín-Burriel et al., 2011). In those studies, a limited set of markers were analyzed, and this might have afflicted the percentages obtained (Soltis and Soltis, 2003). The tree shows 2 main groups of closely related breeds. The first of these groups includes ANI and Mo breeds, sharing the same node, with the Re breed located very close to them. These breeds are raised in Mediterranean forest characterized by oak trees ("dehesas") of central and southwestern Spain (Milán et al., 2006). Breeds living in Pyrenean mountain areas, BP and Pi, and a breed living in the northwestern Spain, RG, formed the second group in the opposite side of the tree. In turn, AV occupied an intermediate position. The Mantel test (Mantel, 1967) did not revealed significant correlation between genetic and geographical distances (not shown in tables).

Our results were not entirely consistent with the classification of Iberian cattle in 3 different morphological trunks proposed by Sánchez-Belda (1984) but fitted quite well with the observations of Jordana et al. (1991), who analyzed 29 morphological traits. According to the last authors, BP, Pi, and RG breeds are classified within the Red Convex (Turdetanus) trunk, Mo and ANI breeds are placed in the Black Orthoid (Iberian) trunk, and the AV breed belongs to the Brown Concave (Blond-brown Cantabrian) trunk. The Re breed, having a red coat, occupied an intermediate position between the Iberian and the Turdetanus trunks. The relationship of the Re breed to the ANI and Mo breeds found in this study was consistent with that observed by Martín-Burriel et al. (2011). These 3 breeds are subjected to fully extensive production systems and have similar breeding objectives. Incidentally, these last 3 breeds showed the highest oxidative activity and average intramuscular fat contents of the Spanish breeds (Gil et al., 2001), which might suggest some proximity in physiological and genetic backgrounds.

3. Principal Component Analysis of Populations. In a different approach to characterize divergence, a

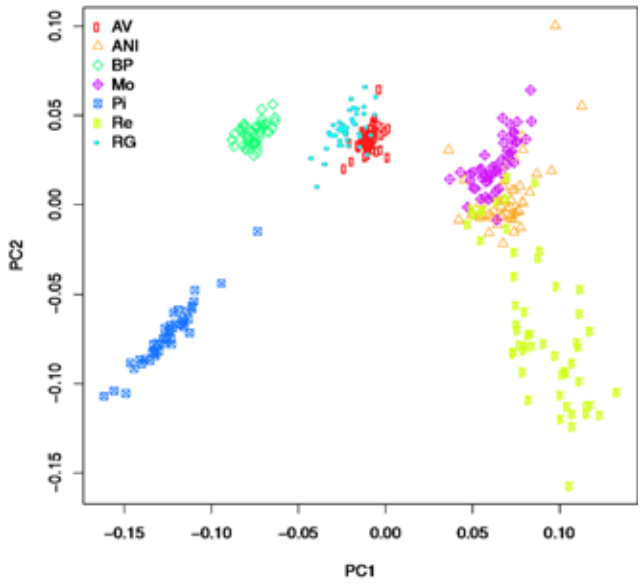

Figure 3. Population groups defined by principal component analysis. PC1 and PC2 correspond to principal components 1 and 2, respectively. AV = Asturiana de los Valles; ANI = Avileña-Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

PCA was applied to a distance matrix built up from the marker IBS relationship matrix (Fig. 3). This analysis allows us to represent each particular animal on the basis of the PCA coordinates. The first and second PCA axis accounted for 11.9 and 5.4% of the variance, respectively. In general, the PCA groupings suggested a classification pattern similar to that observed in the N-J tree. Several features of the PCA results, however, must be highlighted. First, the animals from Pi are distanced from other populations, confirming the high F_{ST} and Nei's distances values. Second, the plot shows a central placement of AV and RG with some degree of mixture of individuals between these populations, indicating its genetic closeness to the other breeds. Third, there is a clear dispersion of the relationship values of the Re breed. Finally, the mixing of some animals between ANI, Mo, and Re breeds suggests a certain gene flow among these breeds that have had a geographic proximity during their history and share a similar production system.

For large-scale SNP data, the PCA and MDS plots have been widely used to summarize the structure of genetic variation (Wang et al., 2012). For this reason, in addition to the PCA analysis, a MDS was also performed to determinate some differences between both methods. In general, the groups of populations showed the same distribution in both methods (Fig. 3; Supplementary Fig. S1; see the online version of the article at http://journalofanimalscience.org) as each of them represent a sampled individual as a point in a Euclidean vector space in such a manner that the placement of points carries information about the similarity of the genotypes in the underlying individuals or populations (Wang et al., 2010).

4. ***Genetic Structure and Levels of Admixture.*** As an ultimate approach to characterize the divergence among the Spanish cattle breeds, we performed a cluster analysis by means of a maximum likelihood method that infers the genetic ancestry of every individual from a mixture of predefined ancestral groups. Clustering by ADMIXTURE relies on a likelihood function that assumes absence of LD and nonrelated individuals. The average LD in the pruned data set was low (0.025) and close to the background LD. In turn, the within-breed average genomic relationships among individuals were, in general, also low across the breeds studied. Supplementary Table S2 (see the online version of the article at http://journalofanimalscience.org) shows a number of statistics on the genomic relationship among individuals of the same breed by using VanRaden's methodology (VanRaden, 2007). The averages ranged from 0.0129 (AV) to 0.1388 (Pi). These statistics probably reflect the limited effective population sizes of the breeds (Cañas-Álvarez et al., 2014). The genetic relationship can add a covariate term to the likelihood function, but we suspect that the low average relationship among individuals and the low LD among markers can only cause a nonrelevant distortion in the clustering process.

The ancestral groups tested ranged from K = 2 to K = 7. The lowest cross validation errors obtained were 0.39894 for K = 6 and 0.39907 for K = 7, indicating that these were the most parsimonious numbers of clusters. For both K = 6 and K = 7, the maximum likelihood estimation of ancestries assigned all individuals to clusters that coincide with the population of origin, although some admixture among populations was also revealed (Fig. 4). The most striking difference between both clusterings was related to the AV breed. For K = 6, AV appears as a mixture of other breeds, mainly of RG, Mo, and BP and of ANI and Re to a lower extent. This confirms the PCA results, where the AV breed was placed in a central location, closer to the other breeds (Fig. 3). For K = 7, AV presents its own identity with signals of admixture with other breeds. For the rest of the breeds, the results are similar for both K = 6 and K = 7. The contribution of ANI to Mo and Re is important, possibly related to transhumance (seasonal migrations) of the ANI breed. Almost 20% of ANI breeders have adopted transhumance

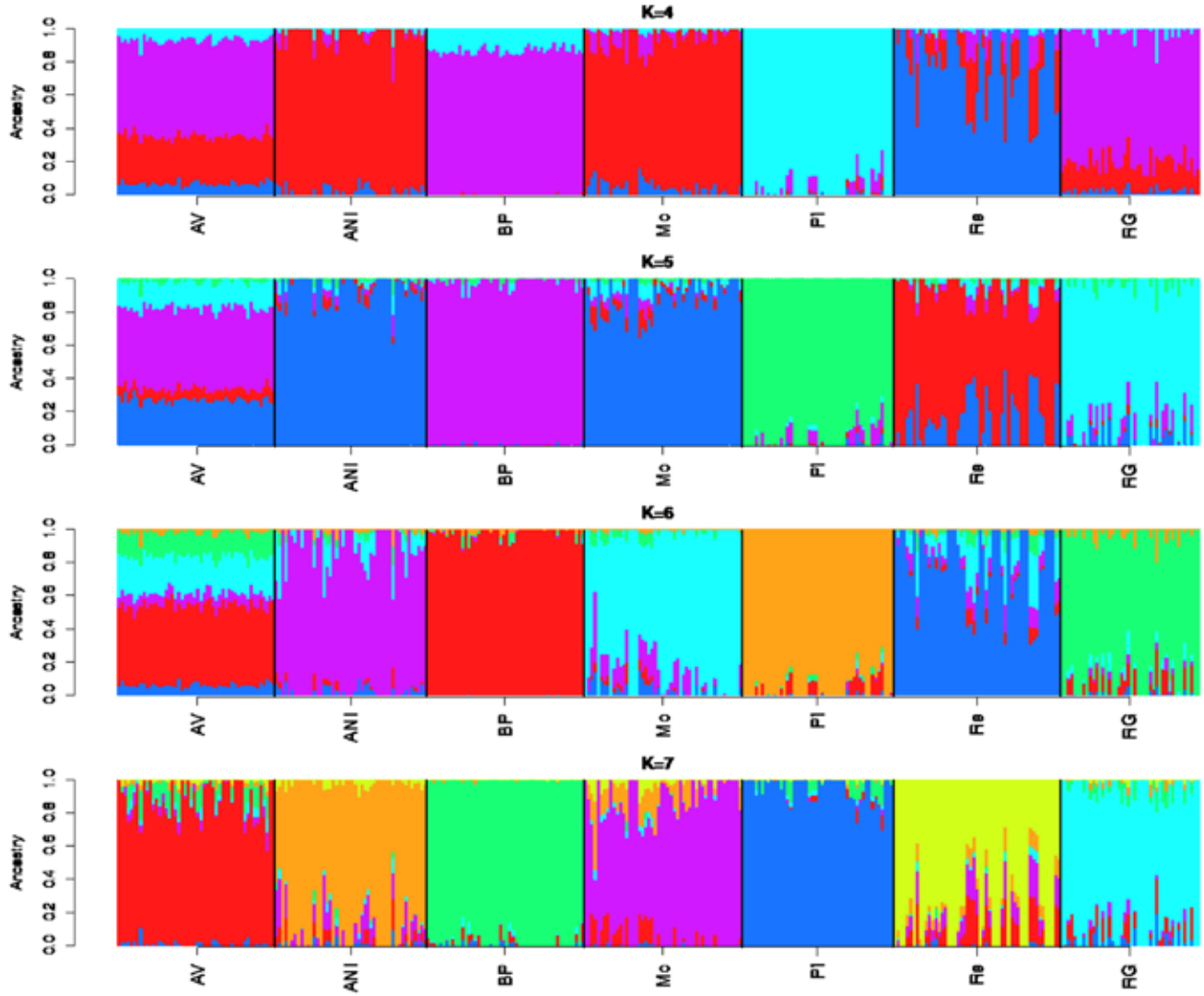

Figure 4. Estimated membership coefficients for each individual for K = 4 through 7. AV = Asturiana de los Valles; ANI = Avileña-Negra Ibérica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega.

as a system of production to take advantage of the seasonal lag between regions. In this way, ANI breeders who do not own land or breeders who own land in different geographic locations (Martín-Collado et al., 2014) diminish the cost of feeding, maximizing the pasture availability. With regard to the Pi breed, most individuals were unequivocally assigned to 1 cluster, with some exchange of genes with AV and BP breeds, the 2 breeds more closely related to the Pi breed. The contributions between BP and Pi breeds could be explained by interchanges due to their geographical proximity in the Pyrenean Mountains. In turn, the admixture with AV confirms a previous result for Spanish beef cattle breeds (Martín-Burriel et al., 2011) that detected the contribution of Brown Swiss cattle, from which the BP breed originated, to the AV breed. In comparison with this latter study (Martín-Burriel et al., 2011), based on a microsatellite analysis, our clusters appear more clearly defined, suggesting that high-density genotyping is a more powerful tool to unravel the relationship among breeds.

Sensitivity Analysis

Given that many population-level statistics assume independence of loci, the divergence analyses were performed using a pruned data set (57,674 SNP), to eliminate the bias in the test statistics that may result from substantial breed-specific differences in LD (Petersen et al., 2013). However, we repeated the analyses of divergence characterization with the whole data set after pruning for an individual call rate ≥ 0.95 and a SNP call rate ≥ 0.95 (717,172 SNP), to check for the potential differences. The results for the N-J tree and the PCA are included in Supplementary Fig. S2 and S3 (see the online version of the article at http://journalofanimalscience.org),

respectively. We found slight increases in the values of Nei's D genetic distances among breeds, with a higher range from 0.013 between AV and BP to 0.026 between Pi and Re. The N-J phylogenetic tree was rotated without suffering changes in the distribution of the groups previously shown in the data pruned for LD. Bootstrap reached levels of 100% for all nodes (Supplementary Fig. S2; see the online version of the article at http://journalofanimalscience.org). In the PCA, the variance explained by the first and second axis increased in a relatively small proportion from 11.9 to 12% and from 5.4 to 7.3%, respectively, due to covariances among SNP. The Pi breed remained differentiated from other populations and some separation of individuals between populations of AV and RG occurred as well as among ANI, Mo, and Re populations (Supplementary Fig. S3; see the online version of the article at http://journalofanimalscience.org). For another analysis of divergence, higher variations were not observed. In general, the results did not suffer major changes and the classification of the populations did not differ between pruned and whole databases.

Conclusions

High-density SNP markers were used to describe the genetic variability and divergence among autochthonous Spanish beef cattle breeds. Our results indicate a large degree of diversity among individuals within populations, as assessed by the BovineHD BeadChip. In turn, the global F_{ST} value and the low genetic distances observed reveal the existence of limited population structuring. Signals of admixture among all breeds were also detected. Both N-J trees and PCA analysis show defined clusters representing the ancestral trunks from which the breeds are supposed to derive: the Turdetanus trunk, including Pirenaica, Bruna dels Pirineus, and Rubia Gallega breeds; Asturiana de los Valles as representative of the Cantabrian trunk, occupying an intermediate position; and the Iberian trunk, which includes Avileña-Negra Ibérica and Morucha, with Retinta breed in a different but close node.

LITERATURE CITED

Alexander, D. H., J. Novembre, and K. Lange. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome Res. 19:1655–1664. doi:10.1101/gr.094052.109.

Ammerman, A. J., and L. L. Cavalli-Sforza. 1984. The Neolithic transition and the genetics of populations in Europe. Princeton Univ. Press, Princeton, NJ.

Beja-Pereira, A., P. Alexandrino, I. Bessa, Y. Carretero, S. Dunner, N. Ferrand, J. Jordana, D. Laloe, K. Moazami-Goudarzi, A. Sanchez, and J. Cañón. 2003. Genetic characterization of southwestern European bovine breeds: A historical and biogeographical reassessment with a set of 16 microsatellites. J. Hered. 94:243–250. doi:10.1093/jhered/esg055.

Blott, S. C., J. L. Williams, and C. S. Haley. 1998. Genetic relationships among European cattle breeds. Anim. Genet. 29:273–282. doi:10.1046/j.1365-2052.1998.00327.x.

Boichard, D., H. Chung, R. Dassonneville, X. David, A. Eggen, S. Fritz, K. J. Gietzen, B. J. Hayes , C. T. Lawley, T. S. Sonstegard, C. P. Van Tassell, P. M. VanRaden, K. A. Viaud-Martinez, G. R. Wiggans, for the Bovine LD Consortium. 2012. Design of a bovine low-density SNP array optimized for imputation. PLoS ONE 7:e34130. doi: 10.1371/journal.pone.0034130.

Brumfield, R. T., P. Beerli, D. A. Nickerson, and S. V. Edwards. 2003. The utility of single nucleotide polymorphisms in inferences of population history. Trends Ecol. Evol. 18:249–256. doi:10.1016/S0169-5347(03)00018-1.

Cañas-Álvarez, J. J. 2015. Del pasado al futuro de las razas bovinas de carne autóctonas. Análisis genealógico y de marcadores SNP para la implementación de la selección genómica. (In Spanish.) PhD Diss., Universitat Autònoma de Barcelona, Bellaterra, Spain.

Cañas-Álvarez, J. J., A. González-Rodríguez, D. Martín-Collado, C. Avilés, J. Altarriba, J. A. Baro, L. F. De la Fuente, C. Díaz, A. Molina, L. Varona, and J. Piedrafita. 2014. Monitoring changes in the demographic and genealogical structure of the main Spanish local beef breeds. J. Anim. Sci. 92:4364–4374. doi:10.2527/jas.2013-7420.

Decker, J. E., S. McKay, M. M. Rolf, J. Kim, A. Molina-Alcala, T. S. Sonstegard, O. Hanotte, A. Götherström, C. M. Seabury, L. Praharani, M. E. Babar, L. C. A. Regitano, M. A. Yildiz, M. P. Heaton, W. Lui, 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:e1004254. doi:10.1371/journal.pgen.1004254.

Edea, Z., H. Dadi, S.-W. Kim, T. Dessie, T. Lee, H. Kim, J. J. Kim, and K. S. Kim. 2013. Genetic diversity, population structure and relationships in indigenous cattle populations of Ethiopia and Korean Hanwoo breeds using SNP markers. Front. Genet. 4:35. doi:10.3389/fgene.2013.00035.

Excoffier, L., G. Laval, and S. Schneider. 2005. Arlequin (version 3.0): An integrated software package for population genetics data analysis. Evol. Bioinform. Online 1:47–50.

Excoffier, L., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: Application to human mitochondrial DNA restriction data. Genetics. 131:479–491.

Felsenstein, J. 1989. PHYLIP – Phylogeny inference package (version 3.2). Cladistics. 5:164–166.

Fernández, J., M. A. Toro, and A. Caballero. 2008. Management of subdivided populations in conservation programs: Development of a novel dynamic system. Genetics. 179:683–692. doi: 10.1534/genetics.107.083816.

Gautier, M., T. Faraut, K. Moazami-Goudarzi, V. Navratil, M. Foglio, C. Grohs, A. Boland, J.-G. Garnier, D. Boichard, G. M. Lathrop, I. G. Gut, and A. Eggen. 2007. Genetic and haplotypic structure in 14 European and African cattle breeds. Genetics 177:1059–1070. doi:10.1534/genetics.107.075804.

Gibbs, R. A., J. F. Taylor, C. P. Van Tassell, and The Bovine HapMap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science. 324:528-532. doi:10.1126/science.1167936.

Gil, M., X. Serra, M. Gispert, M. Àngels Oliver, C. Sañudo, B. Panea, J. L. Olleta, M. Campo, M. Oliván, K. Osoro, M. D. García-Cachán, R. Cruz-Sagredo, M. Izquierdo, M. Espejo, M. Martín, and J. Piedrafita. 2001. The effect of breed-production systems on the myosin heavy chain 1, the biochemical characteristics and the colour variables of longissimus thoracis from seven Spanish beef cattle breeds. Meat Sci. 58:181–188. doi:10.1016/S0309-1740(00)00150-9.

González, P., M. J. Tuñon, and M. Vallejo. 1987. Genetic relationships between seven Spanish native breeds of cattle. Anim. Genet. 18:249–256. doi:10.1111/j.1365-2052.1987.tb00765.x.

Illumina Inc. 2012. BovineHD Genotyping BeadChip. http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf. (Accessed 15 May 2014.)

Jakobsson, M., M. D. Edge, and N. A. Rosenberg. 2013. The relationship between F_{ST} and the frequency of the most frequent allele. Genetics. 193:515–528. doi:10.1534/genetics.112.144758.

Joint Working Group on Refinement. 1993. Removal of blood from laboratory mammals and birds. First report of the BVA/FRAME/RSPCA/UFAW joint working group on refinement. Lab. Anim. 27:1–22.

Jordana, J., M. Pelegrín, and J. Piedrafita. 1991. Relaciones genéticas en bovinos españoles obtenidas a partir del estudio de caracteres morfológicos. (In Spanish.) ITEA 87A:50–64.

Kidd, K. K., W. H. Stone, C. Crimella, C. Carenzi, M. Casati, and G. Rognoni. 1980. Immunogenetic and population genetic analyses of Iberian cattle. Anim. Blood Groups Biochem. Genet. 11:21–38. doi:10.1111/j.1365-2052.1980.tb01489.x.

Lachance, J., and S. Tishkoff. 2013. SNP ascertainment bias in population genetic analyses: Why it is important, and how to correct it. BioEssays 35(9):780–786. doi:10.1002/bies.201300014.

Lischer, H. E. L., and L. Excoffier. 2012. PGDSpider: An automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28:298–299. doi:10.1093/bioinformatics/btr642.

MacHugh, D. E., M. D. Shriver, R. T. Loftus, P. Cunningham, and D. G. Bradley. 1997. Microsatellite DNA variation and the evolution, domestication and phylogeography of taurine and zebu cattle (*Bos taurus* and *Bos indicus*). Genetics 146:1071–1086.

Mantel, N. 1967. The detection of disease clustering and a generalized regression approach. Cancer Res. 27(2):209–220.

Martín-Burriel, I., E. García-Muro, and P. Zaragoza. 1999. Genetic diversity analysis of six Spanish native cattle breeds using microsatellites. Anim. Genet. 30:177–182. doi:10.1046/j.1365-2052.1999.00437.x.

Martín-Burriel, I., C. Rodellar, J. Cañón, O. Cortés, S. Dunner, V. Landi, A. Martínez-Martínez, L. T. Gama, C. Ginja, M. C. T. Penedo, A. Sanz, P. Zaragoza, and J. V. Delgado. 2011. Genetic diversity, structure, and breed relationships in Iberian cattle. J. Anim. Sci. 89:893–906. doi:10.2527/jas.2010-3338.

Martín-Collado, D., K. J. Abraham, S. T. Rodríguez-Ramilo, M. A. Toro, M. J. Carabaño, and C. Díaz. 2013. Using two different approaches to infer the genetic structure of populations with complex relationships: The case of the Avileña-Negra Iberica. In: 20th Conf. New Zealand Assoc. Adv. Anim. Breed. Genet., Napier, New Zealand. p. 134–137.

Martín-Collado, D., K. Soini, A. Määki-Tanila, M. A. Toro, and C. Díaz. 2014. Defining farmer typology to analyze the current state and development prospects of livestock breeds: The Avileña-Negra Ibeìrica beef cattle breed as a case study. Livest. Sci. 169:137–145. doi:10.1016/j.livsci.2014.09.003.

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. doi:10.1371/journal.pone.0005350.

Milán, M. J., J. Bartolomé, R. Quintanilla, M. D. García-Cachán, M. Espejo, P. L. Herráiz, J. M. Sánchez-Recio, and J. Piedrafita. 2006. Structural characterisation and typology of beef cattle farms of Spanish wooded rangelands (dehesas). Livest. Sci. 99:197–209. doi:10.1016/j.livprodsci.2005.06.012.

Moorjani, P., N. Patterson, P.-R. Loh, M. Lipson, P. Kisfali, B. I. Melegh, M. Bonin, Ľ. Kádaši, O. Rieß, B. Berger, D. Reich, and B. Melegh. 2013. Reconstructing Roma history from genome-wide data. PLoS ONE 8:e58633. doi:10.1371/journal.pone.0058633.

Nei, M. 1972. Genetic distance between populations. Am. Nat. 106:283–292. doi:10.1086/282771.

Paradis, E., J. Claude, and K. Strimmer. 2004. APE: Analyses of phylogenetics and evolution in R language. Bioinformatics 20:289–290. doi:10.1093/bioinformatics/btg412.

Petersen, J. L., J. R. Mickelson, E. G. Cothran, L. S. Andersson, J. Axelsson, E. Bailey, D. Bannasch, M. M. Binns, A. S. Borges, P. Brama, A. da Câmara Machado, O. Distl, M. Felicetti, L. Fox-Clipsham, K. T. Graves, G. Guérin, B. Haase, T. Hasegawa, K. Hemmann, E. W. Hill, T. Leeb, G. Lindgren, H. Lohi, M. S. Lopes, B. A. McGivney, S. Mikko, N. Orr, M. C. T. Penedo, R. J. Piercy, M. Raekallio, S. Rieder, K. H. Røed, M. Silvestrelli, J. Swinburne, T. Tozaki, M. Vaudin, C. M. Wade, and M. E. McCue. 2013. Genetic diversity in the modern horse illustrated from genome-wide SNP data. PLoS ONE 8:e54997. doi:10.1371/journal.pone.0054997.

Porto-Neto, L. R., J. W. Kijas, and A. Reverter. 2014. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. Genet. Sel. Evol. 46:22. doi:10.1186/1297-9686-46-22.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. R. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. W. 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. doi:10.1086/519795.

R Core Team. 2014. R: A language and environment for statistical computing. R foundation for Statistical Computing, Vienna, Austria. URL: http://www.R-project.org/.

Saitou, N., and M. Nei. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4:406–425.

Salomon-Torres, R., L. K. Matukumalli, C. P. Van Tassell, C. Villa-Angulo, V. M. Gonzalez-Vizcarra, and R. Villa-Angulo. 2014. High density LD-based structural variations analysis in cattle genome. PLoS ONE 9:e103046. doi:10.1371/journal.pone.0103046.

Sánchez-Belda, A. 1984. Razas bovinas españolas. (In Spanish.) Publicaciones de Extensión Agraria, Madrid, Spain.

Soltis, P. S., and D. E. Soltis. 2003. Applying the bootstrap in phylogeny reconstruction. Stat. Sci. 18:256–267. doi:10.1214/ss/1063994980.

The Bovine HapMap Consortium. 2009. Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. Science 324:528–532. doi:10.1126/science.1167936.

VanRaden, P. M. 2007. Genomic measures of relationship and inbreeding. Interbull Bull. 37:33–36.

Wang, C., Z. A. Szpiech, J. H. Degnan, M. Jakobsson, T. J. Pemberton, J. A. Hardy, A. B. Singleton, and N. A. Rosenberg. 2010. Comparing spatial maps of human population-genetic variation using procrustes analysis. Stat. Appl. Genet. Mol. Biol. 9:13. doi:10.2202/1544-6115.1493.

Wang, C., S. Zöllner, and N. A. Rosenberg. 2012. A quantitative comparison of the similarity between genes and geography in worldwide human populations. PLoS Genet. 8:e1002886. doi:10.1371/journal.pgen.1002886.

Wigginton, J. E., D. J. Cutler, and G. R. Abecasis. 2005. A note on exact tests of Hardy-Weinberg equilibrium. Am. J. Hum. Genet. 76:887–893. doi:10.1086/429864.

Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19:395–420. doi:10.2307/2406450.

Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, R. Hanrahan, G. Pertea, C. P. Van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. Genome Biol. 10:R42. doi:10.1186/gb-2009-10-4-r42.