Genomic diversity and population structure of Mexican and Spanish bovine Lidia breed

P. G. Eusebi, O. Cortés, S. Dunner and J. Cañón

Departamento de Producción Animal, Facultad de Veterinaria, Universidad Complutense de Madrid, Avenida Puerta de Hierro, s/n, 28040, Madrid, Spain.

Summary

The Lidia bovine breed is distinguished for its low genetic exchangeability given its selection on aggressive behavior, its management uniqueness and its subdivided structure. In this study, we present a comprehensive genome-wide analysis of genetic diversity, population structure and admixture of 468 animals from Mexican and Spanish Lidia breed populations and 64 samples belonging to 10 Spanish native and American-creole breeds using 37 148 single nucleotide polymorphisms. We found similar average inbreeding values in the Lidia breed, with different distributions within groups; variability of inbreeding values among Spanish lineages was significant and no differences were found among the Mexican sub-populations. Together, the high F_{IS} of the lineages and the behavior of the runs of homozygosity are consequences of the lineage's small effective population sizes, contributing to their inbreeding increase. Population admixture analysis discarded any influence on the genetic structure of the Lidia populations from the Spanish native and American-creole breeds. In addition, both Lidia populations depicted different genetic origins, with the exception of some Mexican individuals whose origins traced back to recent Spanish importations.

Keywords fighting bull, genetic variation, runs of heterozyosity

Retaining the aggressive features of its wild ancestors, the aurochs (*Bos taurus primigenius*), and distinguished by its extensive management, the Lidia bovine population could be considered a primitive breed which originated ~250 years ago in the Iberian Peninsula. In Mexico, most of the current Lidia population derive from a few Spanish animals imported in 1908 by two families of breeders (Llaguno and González) and a lesser proportion from more recent Spanish importations made during 1996 and 1997 (Niño de Rivera 2004).

Recently, the availability of SNP panels (Bovine HapMap Consortium 2009) has allowed a higher level of resolution when investigating livestock genetic diversity and structure. So, in this study we present for the first time a comprehensive genome-wide analysis of the genomic diversity and population structure of both Mexican and Spanish Lidia populations using the 50K Beadchip panel (Illumina Inc.). In addition, 10 local bovine breeds from Spain and North

America were included in the analysis to assess possible shared genetic origins with the Lidia breed.

A total of 468 DNA Lidia breed samples, 349 from 28 Spanish lineages classified according to Cañón *et al.* (2008) and 119 from the two Mexican lineages classified according to Eusebi *et al.* (2016), were analyzed (Table 1). Samples were genotyped for 54 609 SNPs using the Bovine 50K SNP BeadChip following standard protocols (http://www.illumina.com; data are available via the Figshare repository, https://doi.org/10.6084/m9.figshare.5394895.v3). We also included existing genotypic data kindly provided by Decker *et al.* (2014) of 64 animals from 10 breeds from the Iberian Peninsula and North America that may have possible shared genetic origins with the Lidia breed (Table S1). The animals were classified into four groups: two from the Lidia breed populations (Mexican and Spanish) divided into their corresponding lineages and two from the non-Lidia breeds (Spanish native and American-creole).

After standardizing our Lidia genotypic data with the data provided by Decker *et al.* (2014) into the UMD 3.1 assembly, we used the PLINK V 1.07 software (Purcell *et al.* 2007) to exclude individuals with more than 20% missing genotypes, SNPs located on sex chromosomes, those with a minimum allele frequency less than 0.01 and markers that did not

Table 1 Description of the groups analyzed by their origin, lineages of the Lidia groups, number of breeders (NB), Number of samples (NS) and individual inbreeding values averaged (F) computed from PLINK as $1 - Ho/He$.

Group and lineage	Acronym	NB	NS	F
Lidia breed				
Mexico, $F_{IS} = 0.068$, $F_{ST} = 0.125$				
González	GON	3	16	0.29
Llaguno	LLA	17	101	0.30
Spain, $F_{IS} = 0.079$, $F_{ST} = 0.205$				
Albaserrada	ALB	3	14	0.41
Anastasio Martín	ANA	1	6	0.19
Antonio Pérez	ANT	1	9	0.32
Araúz de Robles	ARA	1	10	0.34
Atanasio Fernández	ATA	3	14	0.36
Baltasar Iban	BAL	2	12	0.28
Carlos Núñez	CAR	4	9	0.25
Santa Coloma	COL	8	36	0.35
Contreras	CON	3	10	0.32
Conde de la Corte	COR	1	10	0.39
José Marzal	CRM	1	9	0.19
Cuadri	CUA	1	7	0.44
Domecq	DOM	5	29	0.36
Félix Gómez	FEL	1	9	0.29
Gamero Cívico	GAM	3	16	0.43
Hidalgo Barquero	HID	3	15	0.31
Manuel Arranz	MAN	1	9	0.30
Conde de la Maza	MAZ	1	3	0.24
Miura	MIU	1	9	0.26
Murube	MUR	4	16	0.34
Pablo Romero	PAB	1	9	0.32
Pedrajas	PED	2	10	0.39
Saltillo	SAL	3	15	0.33
Concha y Sierra	SIE	1	10	0.22
Urcola	URC	1	7	0.26
Veragua	VER	2	16	0.19
Vega Villar	VEG	4	17	0.38
Marqués de Villamarta	VIL	2	13	0.26
Non-Lidia breeds				
Spanish native, $F_{IS} = 0.065$, $F_{ST} = 0.093$	SPA	8[1]	25	0.15
American-creole, $F_{IS} = 0.010$, $F_{ST} = 0.024$	AME	2[1]	39	0.07

[1]Non-Lidia breed groups NB correspond to the number of breeds, as defined in Table S1.

match Hardy-Weinberg equilibrium (0.001) to finally obtain 37 148 SNPs with which to perform the analyses.

Using PLINK v1.07, we calculated individual inbreeding values (F) and then analyzed their variability across and within groups. We also performed two analyses of molecular variance using ARLEQUIN V 3.5 software (Excoffier *et al.* 2005), adjusting a hierarchical model at three levels (groups, lineages and individuals) to assess for the different sources of genetic variation. We also computed runs of homozygosity (ROH), as described by Purfield *et al.* (2012), with a sliding window of 30 SNPs, with less than 100 kb between two consecutive homozygous SNPs, more than two missing genotypes, one possible heterozygous SNP and a minimum length of 500 kbp. ROH were classified into five length categories.

The subdivision level of the Lidia breed in terms of F_{IS} (0.076) was similar to the values found in both the Spanish (0.079) and the Mexican group (0.068) (Table 1). However, the individual inbreeding values within each group showed different distribution patterns. The country of origin of the Lidia group explained 34% of the variability of the individual inbreeding values; but, although 42% of the variability was explained by the lineages within the Spanish Lidia, differences within the Mexican group were not significant.

Previous studies associated high average number and length of ROH to practices of mating related animals (Upadhyay *et al.* 2017); this is consistent with our results, which evidenced higher number and size of ROH segments in the Spanish and Mexican Lidia than in the non-Lidia groups (Spanish native and American-creole) (Table 2, Figs S1 & S2). Both the genomic ROH achieved and high F_{IS} values in the Lidia breed are reflections of high sub-division in the lineages and its main consequences, reduced effective population sizes and high inbreeding levels (Cortés *et al.* 2014).

The genetic variability explained by the Lidia breed lineages (Mexican and Spanish) was 19% (Table S2). In the Mexican Lidia population, the lower genetic distances among breeders are the consequence of a relatively frequent exchange of sires, a common practice in Mexico but less usual in Spain. Thus, the F_{ST} value within the Spanish Lidia group was significantly higher (0.21) (Table 1), and these distances were more than twice the average values of the Spanish native (0.09) and American-creole groups (0.02).

Non-significant correlation between diversity in the origin of the Spanish and Mexican Lidia lineages and their contemporary expected diversities was found (Table S3).

We used ADMIXTURE v1.23 software (Alexander *et al.* 2009) to analyze the genetic structure and PLINK v1.07 to perform a multi-dimensional scaling analysis. We did not detect shared genomic origins between the Spanish native and American-creole groups and the Lidia breed (Fig. 1). In addition, the genetic origins of the Mexican Lidia lineages rarely coincided with that of the Spanish lineages. The spatial separation on the multi-dimensional scaling analysis of the Lidia groups (Fig. 2) confirms this differentiation.

There are arguments that explain the clear genetic differentiation between Lidia groups. In the early 20th century, a few Spanish Lidia animals were imported by both the Gonzalez and Llaguno families of breeders; each family followed different breeding strategies, mating the recent imports with the extant aggressive selected bovines or among them respectively (Niño de Rivera 2004). Additionally, we hypothesize the probability of an admixture with local cimarron genes. Cimarrons are 'run-away' individuals that escaped from their original environment and returned to its wild state; in this context, the *cimarronage* is considered a typical phenomenon of the livestock colonization in the New World (Maudet 2010).

Group	Number of ROH				Total length of genome >1 Mb			
	Mean	SD	Min.	Max.	Mean	SD	Min.	Max.
Spanish Lidia	104	24.4	9	151	736.2	240.3	25.1	1630
Mexican Lidia	116	27.4	46	162	683.9	221	182.6	1264
Spanish native	43	31.1	5	125	332.2	295.9	8.3	1183.9
American-creole	22	14	6	60	160.7	109.5	14.3	438.6

Table 2 Descriptive statistics for the number of runs of heterozyosity (ROH) and total length (Mb) of genome in ROH for the four cattle groups.
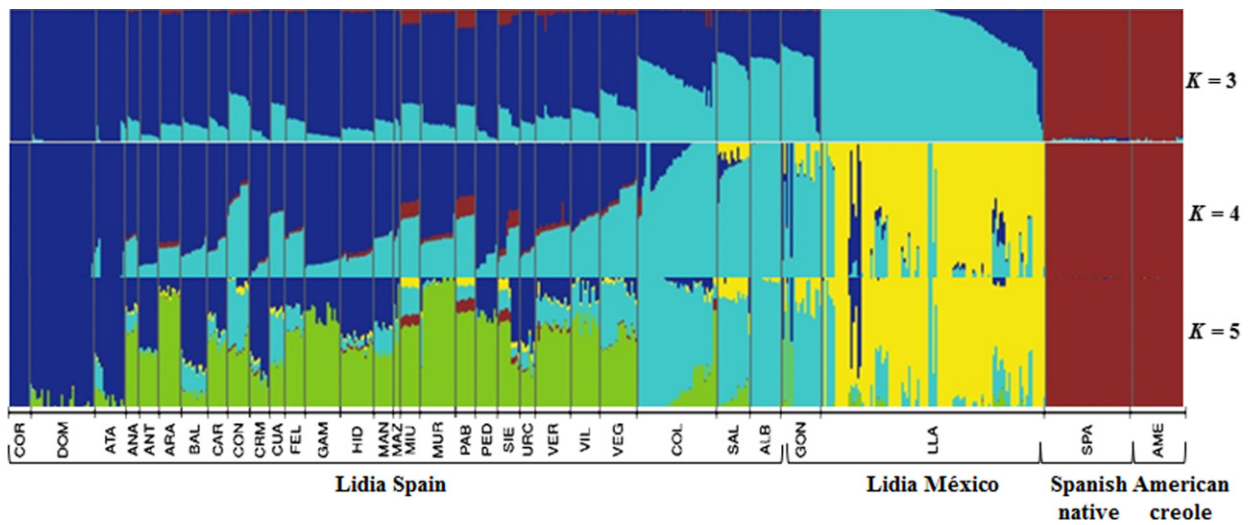

Figure 1 Graphic representation of the proportions of the sampled individual genomes belonging to 3, 4 and 5 (*K*) inferred clusters.
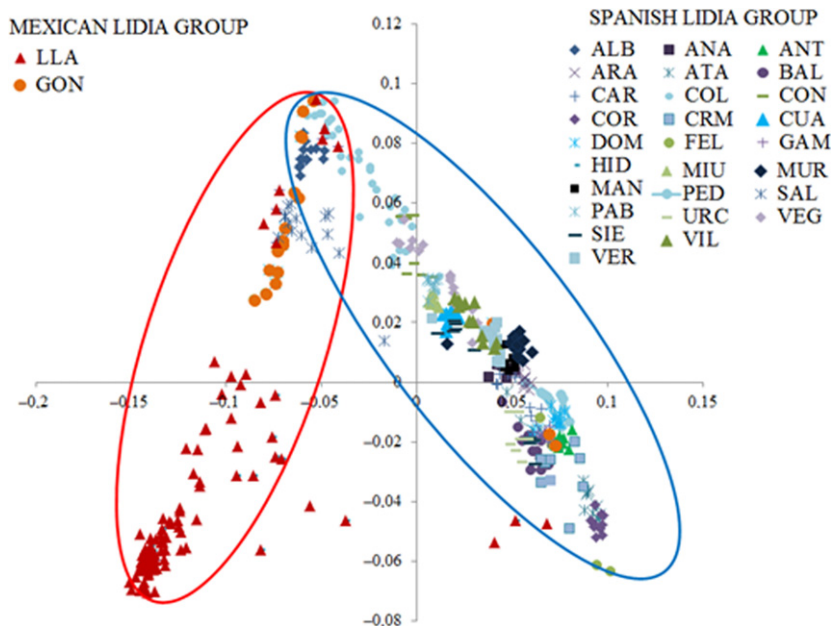

Figure 2 Multidimensional scaling plot based on the matrix of genome-wide pairwise identity-by-state distances inferred with PLINK. The graphic shows the genetic relationships between the Lidia lineages from Mexico (inside the red circle) and Spain (inside the blue circle). See Table 1 for acronym definitions.

Acknowledgements

References

Alexander D.H., Novembre J. & Lange K. (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* **19**, 1655–64.

The Bovine HapMap Consortium, Gibbs, R.A., Taylor, J.F. *et al.* (2009) Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science* **324**, 528–532.

Cañón J., Tupac-Yupanqui I., Garcia-Atance M.A., Cortés O., Garcia D., Fernandez J. & Dunner S. (2008) Genetic variation within the Lidia bovine breed. *Animal Genetics* **39**, 439–45.

Cortés O., Sevane N., Baro J.A. & Cañón J. (2014) Pedigree analysis of a highly fragmented population, the Lidia cattle breed. *Livestock Science* **167**, 1–8.

Decker J.E., McKay S.D., Rolf M.M. *et al.* (2014) Worldwide Patterns of Ancestry, Divergence, and Admixture in Domesticated Cattle. *PLoS Genetics* **10**: e1004254.

Eusebi P.G., Cortés O., Dunner S. & Cañón J. (2016) Genetic diversity of the Mexican Lidia bovine breed and its divergence from the Spanish population. *Journal of Animal Breeding and Genetics* **134**, 287–350.

Excoffier L., Laval G. & Schneider S. (2005) ARLEQUIN (version 3.0): an integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**, 47.

Maudet J.B. (2010) *Terres de Taureaux: les Jeux Taurins de L'Europe à L'Amerique*. Casa de Velázquez, Madrid, Spain.

Niño de Rivera L. (2004). *Sangre de Llaguno, la Razón de Ser del Toro Bravo Mexicano*. Punto de Lectura, México City, México.

Purcell S.M., Neale B., Todd-Brown K. *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics* **81**, 559–75.

Purfield D.C., Berry D.P., McParland S. & Bradley D.G. (2012) Runs of homozygosity and population history in cattle. *BMC Genetics* **13**, 70.

Upadhyay M.R., Chen W., Lenstra J.A. *et al.* (2017) Genetic origin, admixture and population history of aurochs (*Bos primigenius*) and primitive European cattle. *Heredity* **118**, 169–76.

Supporting information

Additional supporting information may be found online in the supporting information tab for this article:

Figure S1 Relationship between the number of ROH >1 Mb and the total length (Mb) of the genome in those ROH for individuals from each group.

Figure S2 Differences between groups of the total length of genome in ROH divided into different length categories and ROH length.

Table S1 Data of the Spanish native and American-creole breeds included in the analyses according Decker *et al.* (2014).

Table S2 Analysis of molecular variance results: (a) between Mexican and Spanish Lidia groups and (b) each group analyzed separately.

Table S3 Diversity in origin (*Div*) (calculated as $1 - \Sigma(qk)2$, where *qk* is an average fraction of the breed's genetic ancestry from the *K* separate genetic clusters at the optimal *K*, identified in the ADMIXTURE analysis) and expected heterozygosity (*He*) values estimated per lineage of the Lidia groups and for the Spanish native and American-creole groups.

Supporting information

Table S1 Data of the Spanish native and American-creole breeds included in the analyses according Decker et al. (2013).

Group	Breed	Number of Samples	Continent	Geographic Origin
American creole	Corriente	5	America	Sonora, Mexico
	Texas Longhorn	20	America	Texas, United States
Spanish native	Berrenda en Negro	5	Europe	Ciudad Real, Jaen, Cordoba, Sevilla, and Huelva, Spain
	Berrenda en Colorado	5	Europe	Cordoba, Sevilla, Huelva, and Cadiz, Spain
	Cardena Andaluza	5	Europe	Sierra Morena, Spain
	Mostrenca	5	Europe	National Park of Donana, southwestern Spain
	Morucha	5	Europe	Salamanca
	Negra Andaluza	5	Europe	Sierra Morena Mountains, Cordoba, and Sevilla Spain
	Retinta	4	Europe	Southwest of Spain and bordering Portugal
	Terreña	5	Europe	Vasconcades mountainous region of Alava, Spain

Table S2 AMOVA analyses results: (A) between Mexican and Spanish Lidia groups and (B) each group analyzed separately.

(A)

Source of variation	Variance components	% of variation
Among groups	137.2	2.1
Among lineages	1246.3	19.4
Among individuals	385.6.	6
Residual	4650.3	72.4

(B)

Source of variation	Variance components		% of variation	
	Spain	Mexico	Spain	Mexico
Among lineages	1295.6	631	20.6	12.5
Among individuals	396.5	201.4	6.3	6
Residual	4599.7	4795	73.1	81.5

Table S3 Diversity in origin (*Div*) (calculated as $1-\Sigma(qk)^2$ where qk is an average fraction of the breed´s genetic ancestry from the *K* separate genetic clusters at the optimal *K*, identified in the ADMIXTURE analysis) and expected heterozygosity (He) values estimated per lineage of the Lidia groups and for the Spanish native and American-creole groups.

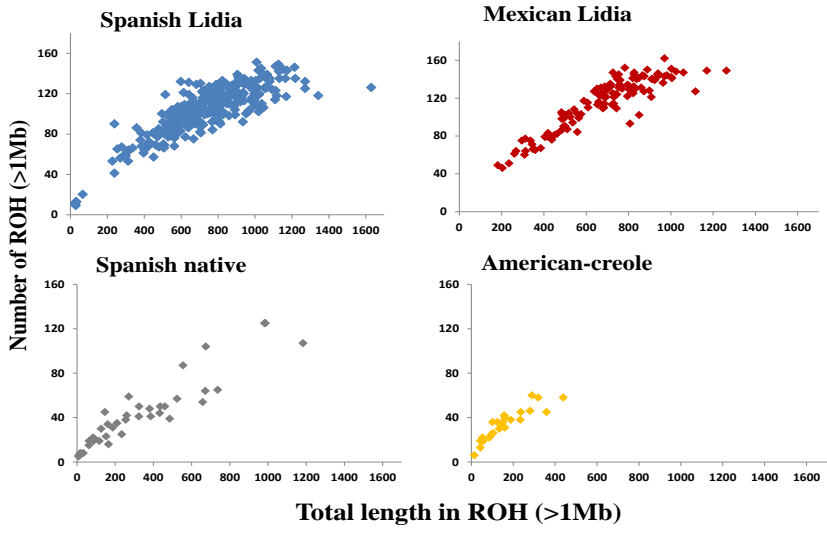| Group	Acronym	*Div*	*He*
Spanish Lidia	ALB	0.03	0.364
	ANA	0.83	0.364
	ANT	0.70	0.364
	ARA	0.00	0.364
	ATA	0.25	0.364
	BAL	0.52	0.364
	CAR	0.82	0.364
	COL	0.62	0.364
	CON	0.25	0.364
	COR	0.01	0.364
	CRM	0.78	0.364
	CUA	0.00	0.364
	DOM	0.42	0.364
	FEL	0.18	0.364
	GAM	0.14	0.363
	HID	0.01	0.364
	MAN	0.01	0.364
	MAZ	0.73	0.364
	MIU	0.00	0.364
	MUR	0.15	0.364
	PAB	0.00	0.364
	PED	0.33	0.364
	SAL	0.09	0.364
	SIE	0.35	0.364
	URC	0.85	0.364
	VEG	0.34	0.364
	VER	0.56	0.364
	VIL	0.47	0.364
Mexican Lidia	GON	0.59	0.364
	LLA	0.37	0.364
Spanish Lidia	SPA	0.49	0.364
American-creole	AME	0.10	0.364

Figure S1 Relationship between the number of ROH>1 Mb and the total length (Mb) of the genome in those ROH for individuals from each group.
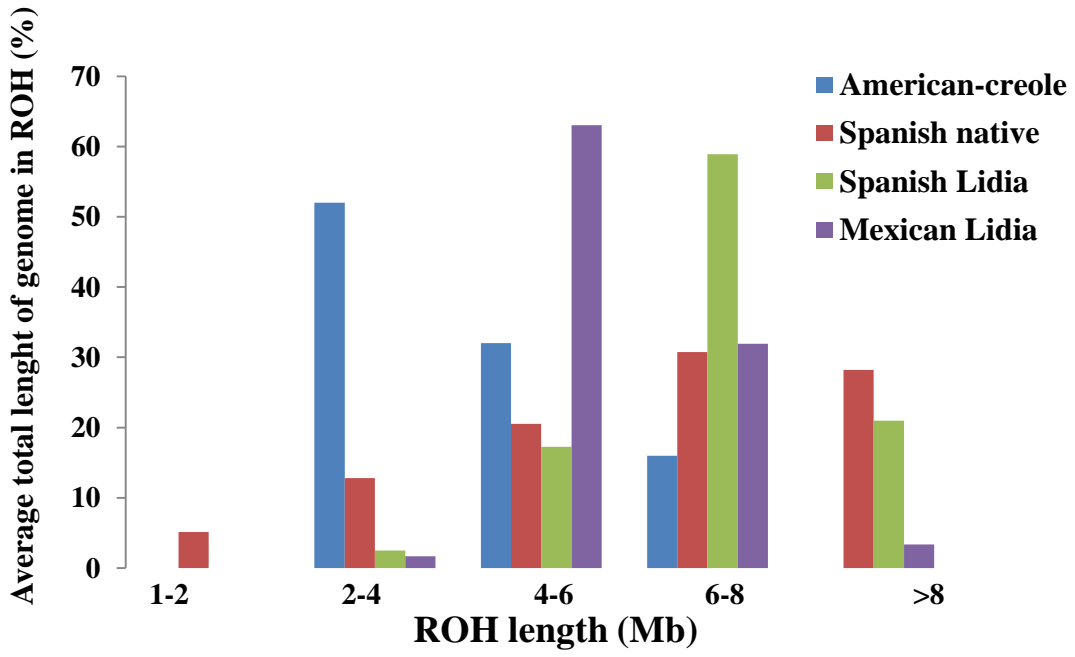

Figure S2 Differences between groups of the total length of genome in ROH divided into different length categories and ROH length.