



Universitat Autònoma de Barcelona
Facultat de Veterinària
Departament de Ciència Animal i dels Aliments

**DEL PASADO AL FUTURO DE LAS RAZAS BOVINAS DE
CARNE AUTÓCTONAS. ANÁLISIS GENEALÓGICO Y DE
MARCADORES SNP PARA LA IMPLEMENTACIÓN DE LA
SELECCIÓN GENÓMICA**

TESIS DOCTORAL

JHON JACOBO CAÑAS ÁLVAREZ

Director

Dr. Jesús Piedrafita Arilla

Programa de Doctorado en Producción Animal

Bellaterra, mayo de 2015



Universitat Autònoma de Barcelona
Departament de Ciència Animal i dels Aliments

El trabajo de investigación “**Del pasado al futuro de las razas bovinas de carne autóctonas. Análisis genealógico y de marcadores SNP para la implementación de la selección genómica**” ha sido realizado por Jhon Jacobo Cañas Álvarez y se presenta como requisito para optar por el grado de Doctor.

Bellaterra, quince de mayo de dos mil quince.

El director de Tesis,

El doctorando,

Jesús Piedrafita Arilla

Jhon Jacobo Cañas Álvarez

AGRADECIMIENTOS

Sin lugar a dudas son tantas las personas que durante todos estos años me han brindado su apoyo, que difícilmente lograré mencionarlos a todos. Intentaré resumir en mis agradecimientos a todos y cada uno de ellos.

Quisiera iniciar agradeciendo a la vida por la oportunidad que me otorgó de poder estar acá, por permitirme emprender una aventura en la cual no sabía que me esperaba pero en la que sí sabía qué fin buscaba. Ha sido una aventura de muchos años, de experiencias únicas, que me han llevado a conocer y disfrutar sitios inigualables, y de gente nueva que me ha enriquecido tanto cultural como personalmente.

Esta experiencia ha sido posible gracias al apoyo de Colciencias (Beca de formación doctoral “Francisco José de Caldas” 497/2009) y al gobierno de Colombia, quienes han visto en la educación la principal herramienta para lograr grandes cambios.

De la misma manera, esto no habría sido posible sin la ayuda de mi director, Jesús, quien me apoyo y creyó en mí. Gracias por la paciencia y el apoyo durante estos años de doctorado; sin lugar a dudas sus aportes me ayudaron a crecer académicamente.

A mi familia, por apoyarme siempre en la obtención de las metas que me he propuesto. Gracias a mi madre por ayudarme en todo momento, a crecer tanto espiritual como personalmente. Gracias por la educación que me has brindado durante todos estos años, por ser padre y madre al mismo tiempo, por ser esa “guerrera” en todo el sentido de la palabra. A mi abuelo, gracias por brindarme todo su apoyo y por sus sabios consejos. Abuela... en este difícil momento que estas pasando solo quiero decirte que esta tesis te la dedico complemente a ti. Gracias a todos por aguantar mi ausencia.

También quiero agradecer a mi compañera de aventuras, a mi gran amor. Gracias Liliana por compartir todas estas experiencias conmigo. Experiencias que nos han permitido crecer y aprender mucho uno del otro. Momentos que sin duda quiero seguir pasando contigo durante todos los años de mi vida.

Qué decir de ese gran amigo William, con quien inicié esta audaz aventura del doctorado, grandes momentos de diversión, estudio y anécdotas por contar ocurridas a lo largo de cuatro años. De esas personas especiales que conocí en Barcelona: Leidy, Fabio, Pedro, Genaro y Bruno, nunca me pude haber reído tanto y haber pasado tan buenos y grandiosos momentos a su lado. Espero nuestro pronto reencuentro, ya sea en Brasil, México, España o Colombia.

Igualmente no quiero dejar de mencionar a todos los compañeros de Veterinaria: Quim, Marta, Ceci, Anna, Sara, Jacopo, Lais, Samir. Grandes compañeros que me hicieron sentir parte de su grupo y me brindaron todo su conocimiento. A los amigos del Máster: Sergio, Sonia, Sandra, Adriana y Ester, con los cuales compartí grandes experiencias. A todos los compañeros de producción: Las Marías, Arturo, Montse, Román, Rosa, Sara, Sergio y Sondes me brindaron momentos de diversión increíbles. De la misma manera no puedo dejar de lado a las grandes personas que conocí en el CRAG: Anna Puig, gracias por esas charlas en el momento que más las necesité; Yulixaxis, gracias por tu sencillez y colaboración; Antonia, gracias por esa alegría que siempre traes contigo; a todos los demás Erika, Jordi, Rayner y Bete ¡qué gran grupo!

Finalmente quiero dar un agradecimiento muy especial a todo el grupo de Zaragoza: Luis, Juan, Aldemar, Sebastián, Elena y a todos los demás investigadores que hicieron parte de las publicaciones, sus aportes fueron muy valiosos.

Gracias también a las asociaciones de las distintas razas por las muestras y datos aportados, sin estos no hubiera sido posible lograr esta tesis.

RESUMEN

Durante muchas décadas las razas autóctonas han permitido dar cumplimiento a numerosas necesidades de las diferentes comunidades. La variación genética de estas poblaciones ha sido un aspecto fundamental en el mantenimiento de la biodiversidad. Esta variación permite a los individuos diferenciarse entre sí, y es gracias a esta diferenciación que es posible estimar la diversidad, la evolución y la divergencia entre poblaciones. En la actualidad, las técnicas de genotipado masivo de SNP han proporcionado una herramienta muy útil, tanto para determinar la diversidad así como para la mejora genética animal. La selección genómica se ha convertido en un instrumento muy valioso para detectar diferentes regiones del genoma relacionadas con variantes fenotípicas de interés. Sin embargo, en las razas españolas de vacuno de carne su aplicabilidad no ha sido tan evidente hasta ahora. Es por esta razón, que las preguntas más importantes que se plantearon en esta tesis doctoral fueron: 1) determinar la cantidad de variación genética que hay en las principales razas autóctonas de ganado de carne español; 2) estimar la estructuración de esa variación en las distintas poblaciones; 3) calcular las distancias genéticas y tener una visión global del grado de mixtura entre las distintas razas; y 4) explorar la historia y constitución genética de las razas a través del desequilibrio de ligamiento, la persistencia de fases y el tamaño efectivo ancestral con miras a una futura implementación de la selección genómica. Las razas objeto de estudio, el acrónimo y el número de tríos -padre, madre y descendiente- utilizados fueron los siguientes: Asturiana de los Valles (AV, 25), Avileña-Negra Ibérica (ANI, 24), Bruna dels Pirineus (BP, 25), Morucha (Mo, 25), Pirenaica (Pi, 24), Retinta (Re, 23) y Rubia Gallega (RG, 22).

Para dar respuesta a estas inquietudes se partió de la estimación de los parámetros demográficos y poblacionales evaluados mediante un análisis de pedigrís. Para este análisis se empleó la información registrada en los libros genealógicos desde su creación hasta el año 2009. Los resultados mostraron incrementos continuos de los censos poblacionales y una alta tasa de intercambio de machos reproductores entre rebaños en todas las razas evaluadas. La alta incidencia de la inseminación artificial en la raza RG incrementó el número promedio de progenie por cada toro a 299,5 descendientes; para las demás razas varió entre 35,9 y 89,8 descendientes. La compleción del pedigrí mostró

índices promedio del 92% una generación atrás y del 61% si se consideran seis generaciones atrás. Para los animales nacidos en 2009, los coeficientes de endogamia promedio variaron entre 0,6% (BP) y 7,2% (Re). Se observó un aumento general de la endogamia media en todas las razas. El tamaño efectivo de la población, basado en el incremento promedio de la tasa de endogamia individual para generaciones equivalentes, varió entre 19 (Mo) y 90 (AV) en el año 2009. El número efectivo de ancestros osciló entre 42 (RG) y 838 (BP) y fue menor de lo que era el número efectivo de fundadores en todas las razas, lo que sugiere la existencia de cuellos de botella en las poblaciones estudiadas. Estos hechos sugieren la necesidad de implementar una política para el manejo adecuado de los apareamientos que ayude a controlar la endogamia y no perder la variabilidad genética.

Posteriormente, y con el fin de complementar estos análisis, se evaluó la diversidad y divergencia genética por medio de marcadores moleculares de tipo SNP obtenidos a partir de un chip de Illumina de alta densidad (778 K). Los resultados mostraron una heterocigosis esperada de alrededor de 0,30. Por su parte, el análisis de varianza molecular (AMOVA) reveló que el 94,22% de la variación total se explica por las diferencias entre los individuos mientras que sólo el 4,46% fue el resultado de las diferencias entre las poblaciones. Estos resultados indicaron gran diversidad dentro de individuos y un bajo grado de divergencia entre las razas. Las distancias genéticas de Nei, estimadas entre cada par de poblaciones, oscilaron entre 0,009 (ANI-BP) y 0,016 (Pi-Re). El árbol filogenético de *neighbor joining* (N-J) y el análisis de componentes principales (PCA) mostraron dos grupos principales de razas: Pi y BP, por un lado, y ANI, Mo y Re, por el otro. Las razas AV y RG, por su parte, ocuparon una posición intermedia. El análisis de clúster asignó la mayoría de los individuos a grupos que coinciden con la población de origen, sin embargo también reveló cierto grado de mixtura entre las razas.

Finalmente, se estimó la magnitud del desequilibrio de ligamiento (LD) y la persistencia de las fases haplotípicas a partir de los marcadores moleculares, y se calculó el tiempo de divergencia, y el tamaño efectivo ancestral de la población. Las estimaciones promedio del LD para los marcadores adyacentes estuvieron en el entorno de 0,52 en las siete razas, y disminuyeron a medida que se incrementó la distancia entre los marcadores (0,07 y 0,05 para los marcadores separados 200 kb y 1000 kb, respectivamente). Estos resultados indican que con un panel de cómo mínimo 38.000 SNP sería suficiente para alcanzar un programa de selección genómica exitoso dentro de cada raza. De manera

similar, la persistencia de fases también disminuyó de un rango de 0,92-0,96, hasta 0,01-0,08 cuando la distancia entre marcadores aumentó de 5 a 1000 kb, respectivamente. Por lo tanto, y con el fin de tener una persistencia de fases mínima de 0,8 aplicable a la selección genómica, el tamaño del panel de SNP entre razas debe oscilar entre 50.000 y 83.000 SNP según la raza. Por otro lado, la divergencia entre razas se dio aproximadamente hace 129 (AV-Mo) y 206 (Pi-Re) generaciones. En general, los tamaños efectivos ancestrales hace 1500 generaciones oscilaron entre 1753 y 2893 para Pi y AV, respectivamente, mientras que hace una generación estos valores variaron entre 24 y 32 para RG y BP, respectivamente. Los altos valores de LD en las distancias cortas, así como la rápida disminución de LD, se reflejó en el bajo tamaño efectivo de la población de las poblaciones de ganado vacuno español. Los resultados de este estudio son relevantes para la futura implementación dentro y entre razas de programas de selección genómica en las poblaciones de ganado vacuno español.

SUMMARY

For many decades the autochthonous breeds allowed to fulfill many needs of different communities. Genetic variation of these populations has been a fundamental aspect in maintaining biodiversity. This variation allows individuals to differentiate themselves, and thanks to this distinction it is possible to estimate the diversity, evolution and divergence between populations. Currently, massive SNP genotyping techniques have provided a very useful tool, both to determine the diversity as well for breeding purposes. Genomic selection has become an invaluable tool to identify genomic regions associated with the phenotypic variation of traits of interest. However, the applicability in the Spanish beef cattle breeds has not been so evident so far. It is for this reason that the most important questions raised in this thesis were: 1) determining the genetic variation existing in the autochthonous Spanish beef cattle breeds; 2) to estimate the structure of this variation in the different populations; 3) to calculate the genetic distances and get an overview of the degree of admixture between breeds; and 4) explore the genetic history of the breeds through linkage disequilibrium, persistence of LD phase and the past effective population size to implement genomic selection in the future. The breeds, acronyms and number of trios –father, mother, offspring- used in this study were: Asturiana de los Valles (AV, 25), Avileña-Negra Ibérica (ANI, 24), Bruna dels Pirineus (BP, 25), Morucha (Mo, 25), Pirenaica (Pi, 24), Retinta (Re, 23) and Rubia Gallega (RG, 22).

To answer these concerns we started with the estimation of demographic and population parameters assessed by pedigree analysis. For this analysis the herdbook information from their foundation up to 2009 was used. The genetic structure estimated from pedigrees showed continued increases in population censuses and a high exchange rate of breeding males among herds in all breeds studied. The high incidence of artificial insemination in RG breed increased the average number of progeny per bull to 299.5; for the rest of breeds it ranged between 35.9 and 89.8. Pedigree showed average of completeness indexes of 92% one generation ago, and 61% six generations ago. For animals born in 2009, average inbreeding coefficients ranged from 0.6% (BP) to 7.2% (Re). There was a general increase in average inbreeding in all breeds. Effective population size based on averages of individual increase of inbreeding by equivalent generations

ranged from 19 (Mo) to 90 (AV). Effective number of ancestors ranged from 42 (RG) to 838 (BP), and was lower than was the effective number of founders in all breeds, suggesting the existence of bottlenecks in the populations studied. These facts suggest the necessity to implement a policy for proper management of matings to control inbreeding and the loss genetic variability.

Subsequently, in order to complement these analyses, the genetic diversity and the degree of divergence was evaluated using a high-density SNP chip obtained from Illumina (778 K). The expected heterozygosity was around 0.30. The analysis of molecular variance (AMOVA) 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. These results indicate great diversity within individuals and low degree of divergence among breeds. Nei's genetic distances, estimated between each pair of populations, ranged from 0.009 (ANI-BP) to 0.016 (Pi-Re). The neighbor joining (N-J) phylogenetic tree and the principal component analysis (PCA) showed two main groups of breeds: Pi and BP on the one hand, and ANI, Mo and Re, on the other. The AV and RG breeds occupied an intermediate position. A cluster analysis assigned all the individuals to the breed they belong to, but also revealed some degree of admixture among breeds.

Finally, the extent of linkage disequilibrium (LD) and the persistence of the haplotypic phases from molecular markers were estimated, as well as the time of divergence and the past effective population size. The average estimates of LD for adjacent markers was close to 0.52 in the seven breeds, and decreased with the increase of the distance between markers (0.07 and 0.05 for markers to 200 and 1000 kb apart, respectively). This decay suggests that with a panel having a low boundary of 38,000 SNPs would be enough to achieve a successful within breed genomic selection program. Furthermore, the persistence of phase decreased from the range of 0.92 to 0.96, to a range of 0.01 to 0.08 when the marker distance increased from 5 to 1000 kb, respectively. In order to have a minimum persistence of LD phase of 0.8 applicable to across breed genomic selection, the size of the SNP panels must be in the range of 50,000 to 83,000, depending upon the breed pairs. On the other hand, the divergence among breeds occurred about 129 (AV-Mo) and 206 (Re-Pi) generations ago. Estimates of past effective population sizes showed values ranging from 1753 to 2893 for 1500 generation ago in Pi and AV, respectively, and values ranging from 24 to 32 for RG and BP, respectively, one generation ago. The high values of LD at short distances and the rapid decline of LD was

reflected in the low past effective population size of the Spanish beef cattle populations. The results of this study are relevant for the future implementation of within breed and across breed genomic selection programs in the Spanish beef cattle populations.

TABLA DE CONTENIDO

LISTA DE TABLAS.....	xii
LISTA DE FIGURAS.....	xiv
LISTA DE ABREVIATURAS.....	xvi
CAPÍTULO 1. Introducción.....	3
1.1. Importancia de las razas bovinas autóctonas	3
1.2. Sistemas de producción del vacuno de carne en España	4
1.2.1. Sistema de dehesa.....	5
1.2.2. Sistema de montaña.....	5
1.2.3. Sistemas no montañosos húmedos	5
1.3. Programas de mejora genética de vacuno de carne en España.....	6
1.4. Nuevos paradigmas en el mejoramiento genético animal.....	7
1.5. Perspectivas para la aplicación de la selección genómica en razas bovinas de carne autóctonas españolas.....	10
1.5.1. Importancia de los análisis genealógicos	10
1.5.2. Utilización de herramientas moleculares.....	11
CAPÍTULO 2. Objetivos.....	19
CAPÍTULO 3. Revisión bibliográfica.....	23
3.1. Historia y origen del ganado vacuno español.....	23
3.2. Principales razas bovinas autóctonas españolas de carne	24
3.2.1. Asturiana de los Valles	25
3.2.2. Avileña-Negra Ibérica	26
3.2.3. Bruna dels Pirineus.....	28
3.2.4. Morucha.....	29
3.2.5. Pirenaica	30
3.2.6. Retinta.....	31
3.2.7. Rubia Gallega	32
3.3. Estructura genética poblacional.....	33
3.3.1. Estructura demográfica y genealógica de las poblaciones	33

3.3.1.1	Censo poblacional	34
3.3.1.2	Intervalos generacionales	34
3.3.1.3	Tamaños familiares.....	34
3.3.1.4	Estructura piramidal	35
3.3.1.5	Grado de compleción del pedigrí.....	35
3.3.1.6	Coefficiente de endogamia	36
3.3.1.7	Coefficiente de coascendencia.....	36
3.3.1.8	Tamaño efectivo de la población.....	37
3.3.1.9	Probabilidad de origen de los genes	39
3.3.2.	Estructura poblacional y diversidad genética	40
3.3.2.1.	Uso de los SNP en la caracterización de la diversidad genética	40
3.3.2.2.	Herramientas informáticas para el análisis de datos moleculares	42
3.3.2.3.	Medidas de estructura poblacional, diversidad genética y divergencia.....	45
3.3.2.4.	Herramientas multivariantes en el análisis de datos moleculares.....	51
3.3.3.	Algunas aplicaciones de los Análisis Poblacionales	52
3.3.3.1.	Desequilibrio de ligamiento	53
3.3.3.2.	Persistencia de fases	54
3.3.3.3.	Tamaños efectivos ancestrales de la población	55
CAPÍTULO 4. Monitoring changes in the demographic and genealogical structure of the main Spanish local beef breeds		67
CAPÍTULO 5. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high density SNP chip		93
CAPÍTULO 6. Linkage disequilibrium, persistence of phase and effective population size of seven Spanish beef cattle breeds		119
CAPÍTULO 7. Discusión General.....		155
Estructura demográfica de las poblaciones autóctonas. La importancia de los censos poblacionales		155
La endogamia y el tamaño efectivo. Parámetros claves en la variabilidad genética.....		156
Estructura y diversidad genética en los recursos genéticos locales.....		159
Diferenciación y divergencia entre poblaciones. ¿Qué tan cercanas son las razas autóctonas españolas?.....		162

El futuro de la implementación de la selección genómica. Desequilibrio de ligamiento, persistencia de fases y tamaño efectivo ancestral.....	164
CAPÍTULO 8. Conclusiones.....	175

LISTA DE TABLAS

CAPÍTULO 1. Introducción

Tabla 1.1 Objetivos y criterios de selección de las evaluaciones genéticas en las principales razas de ganado de carne autóctono	8
--	---

CAPÍTULO 2. Revisión bibliográfica

Tabla 3.1 Principales programas empleados en análisis poblacionales.....	43
Tabla 3.2 Estimación del AMOVA con tres niveles jerárquicos	48

CAPÍTULO 4. Demographic and genealogical structure in Spanish beef breeds

Table 1. Number of males and females giving birth in 1998, 2003, and 2009 and the corresponding effective sizes under random mating and unequal number of males and females.....	74
Table 2. Pyramid structure of the different breeds according to the use of own or purchased bulls. The periods analyzed went from the starting year of recording practices to the years 1998, 2003, and 2009	76
Table 3. Generation interval estimates for the 4 pathways corresponding to animals born in the period 2005 through 2009.....	77
Table 4. Average family sizes of breeding animals born in 1988, 1993, and 2000	78
Table 5. Average pedigree completeness index considering 1 to 6 generations deep for animals born in 2009 and average complete generation equivalent values for animals born up to 2009	79
Table 6. Average inbreeding coefficients of all animals and inbred animals born in 1998, 2003, and 2009	80
Table 7. Inbreeding and coancestry rates and effective population size by generation	82
Table 8. Averages of individual increase of inbreeding of animals born in 3 yr and realized effective population sizes computed from them.....	86
Table 9. Estimates of parameters of probability of gene origin	87

CAPÍTULO 5. Genetic diversity and divergence among Spanish beef breeds

Table 1. Genetic diversity within seven Spanish beef breeds	98
Table 2. Analysis of molecular variance in the seven populations	101
Table 3. Estimates of pairwise F_{ST} statistics and the Nei's D genetic distances among populations	102
Table S1. Specific F_{IS} indices in 7 Spanish beef cattle breeds	113

CAPÍTULO 6. Linkage disequilibrium, persistence of phase and past effective population size in Spanish beef breeds

Table 1. Average r^2 values and standard deviations across all autosomes at various marker distances in seven Spanish beef breeds	123
Table 2. Average r^2 between adjacent markers across all autosomes depending upon the proportion of SNPs retained in seven Spanish beef breeds	127
Additional file 3. Estimates of time since divergence between breeds from linear regression of the natural logarithm of the expected correlation of r.....	150
Additional file 4. Past effective population size at various generations ago in seven Spanish beef breeds.....	151
Additional file 5. Average of linkage disequilibrium between nonsyntenic SNPs, effective population sizes, and effective number of loci in seven Spanish beef breeds.....	152

LISTA DE FIGURAS

CAPÍTULO 3. Revisión bibliográfica

Figura 3.1 Troncos étnicos de los bovinos autóctonos españoles	24
Figura 3.2 Distribución geográfica de las principales razas de carne españolas	25
Figura 3.3 Raza Asturiana de los valles	26
Figura 3.4 Raza Avileña-Negra Ibérica	27
Figura 3.5 Raza Bruna dels Pirineus	28
Figura 3.6 Raza Morucha	29
Figura 3.7 Raza Pirenaica	30
Figura 3.8 Raza Retinta	31
Figura 3.9 Raza Rubia Gallega	33

CAPÍTULO 4. Demographic and genealogical structure in Spanish beef breeds

Figure 1. Inbreeding and coancestry coefficients evolution by year of birth in Rubia Gallega and Retinta breeds.	84
---	----

CAPÍTULO 5. Genetic diversity and divergence among Spanish beef breeds

Figure 1. Average proportions of SNPs within MAF ranges by population.	100
Figure 2. Neighbor joining representation of the pairwise Nei's D genetic distances among populations.	103
Figure 3. Different population groups defined by Principal component analysis.	105
Figure 4. Estimated membership coefficients for each individual for K = 4-7.	107
Figure S1. Different population groups as defined by a multidimensional scaling plot on the relationship matrix	114
Figure S2. Neighbor joining representation of the pairwise Nei's D genetic distances among populations in the whole dataset.	115
Figure S3. Different population groups defined with Principal component analysis in the whole dataset	116

CAPÍTULO 6. Linkage disequilibrium, persistence of phase and past effective population size in Spanish beef breeds

Figure 1. Evolution of average from markers 0 to 200 kb apart across all autosomes in seven Spanish beef breeds. 125

Figure 2. Evolution of pairwise persistence of phase across all autosomes in seven Spanish beef breeds..... 128

Figure 3. Persistence of phase matrices across all autosomes among seven Spanish beef breeds evaluated at 20 and 100 kb..... 130

Figure 4. Effective population size estimated using specific recombination rates per chromosome in the periods 0 to 1000 generations ago, and 0 to 10 generations ago. 133

Additional file 1. Figures A1-A7. Frequency distribution of linkage disequilibrium for adjacent markers in 7 Spanish beef breeds..... 143

Additional file 2. Figures B1-B7. Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in 7 Spanish beef breeds. 147

LISTA DE ABREVIATURAS

AFLP: Amplified fragment length polymorphism	H _o : Heterocigosis observada
AI: Artificial insemination	IBS: Identical by state
AMOVA: Analysis of molecular variance	LD: Linkage disequilibrium
AV: Raza Asturiana de los Valles	MAF: Minor allele frequency
ANI: Raza Avileña-Negra Ibérica	mtDNA: Polimorfismos de DNA mitocondrial
BP: Raza Bruna dels Pirineus	MDS: Multidimensional scaling
DIST: Matriz de distancias	MNA: Mean number of alleles
<i>F</i> : Coeficiente de endogamia	Mo: Raza Morucha
<i>f</i> : Coascendencia	N _e : Tamaño efectivo de la población
ΔF : Incremento de endogamia	N-J: Neighbor joining
Δf : Incremento de coascendencia	PCA: Principal components analysis
<i>f_e</i> : Número efectivo de fundadores	Pi: Raza Pirenaica
<i>f_a</i> : Número efectivo de antepasados	QTL: Quantitative trait locus
F _{IS} , F _{IT} , F _{ST} : Índices de fijación	Re: Raza Retinta
FREQ: Datos de frecuencias	RFLP: Restricted fragment length polymorphism
GEBV: Genomic estimated breeding value	RG: Raza Rubia Gallega
H-W: Hardy-Weinberg	SNP: Single nucleotide polymorphism
H _e : Heterocigosis esperada	STR: Short tandem repeat (microsatellites)

Capítulo 1

INTRODUCCIÓN

1. Introducción

1.1. Importancia de las razas bovinas autóctonas

El uso racional y sostenible de los recursos locales permite dar cumplimiento a diversas necesidades como son: la preservación de la fauna autóctona, la biodiversidad y el mantenimiento de los ecosistemas, además de permitir el desarrollo socioeconómico de una zona en particular (Gandini y Villa, 2003).

La biodiversidad es la característica básica más sobresaliente de las razas autóctonas. La importancia de esta condición radica principalmente en una mejor adaptación de las razas autóctonas a su hábitat y a los recursos existentes, los cuales transforman de forma eficiente en armonía con su medio natural (FAO, 2007). Los animales domésticos contribuyen también al mantenimiento de los ecosistemas en los que viven, proporcionando servicios tales como la dispersión de semillas, el ciclo de nutrientes o simplemente haciéndolos económicamente sustentables (Molina, 2010). De la misma manera, la alta variabilidad genética de las razas autóctonas les permite tener una alta capacidad de adaptarse a los cambios climáticos, y la evolución y selección han fomentado la rusticidad. Estas características les han concedido una alta resistencia a las enfermedades propias de cada zona y les ha mejorado la capacidad de aprovechamiento de recursos forrajeros de baja calidad (Hoffmann, 2010).

Desde el punto de vista socioeconómico, los animales domésticos también juegan un papel importante en la contribución del desarrollo y mantenimiento de la población rural. Éstos satisfacen más del 30% de las necesidades humanas en alimentación y agricultura, lo cual es fundamental en el mantenimiento de la seguridad alimentaria (Molina, 2010). Además de la relevancia socioeconómica, en la mayoría de las regiones del mundo se reconoce las funciones históricas y culturales del ganado autóctono. Las motivaciones culturales influyen en la utilización de los recursos zoogenéticos y con frecuencia, existen fuertes vínculos entre las comunidades y sus razas locales (FAO, 2007). Al igual que los restos arqueológicos, las razas autóctonas tienen el mismo derecho a ser preservadas, y más aun si consideramos que una raza es el fruto de más de 8.000 años de actuación del hombre sobre las poblaciones animales y que por lo tanto es un patrimonio vivo de la humanidad. Las razas son el espejo de las diferentes migraciones, invasiones y culturas de los distintos pueblos que han poblado un determinado territorio, ya que en los procesos

migratorios los seres humanos han ido acompañados de sus animales domésticos, los cuales al llegar a esos nuevos territorios se adaptan y/o se cruzan con los allí presentes, dando lugar a nuevas razas (Molina, 2010). Desde el punto de vista científico, las especies domesticas desempeñan un papel muy importante en los estudios fisiológicos o genéticos. La adaptación de los animales autóctonos a dietas raras, hábitats especiales o a parásitos específicos, puede constituir un material de inapreciable valor para el estudio de la naturaleza fundamental de esas adaptaciones (Sánchez et al., 1992).

La trascendencia económica de las razas bovinas autóctonas radica principalmente en los costes de producción y en la calidad. Los métodos de explotación en régimen extensivo han evolucionado en todo el mundo al igual que en España. En este sentido, el uso de las razas autóctonas trae consigo una disminución en los costes de explotación, ya que no requiere el mantenimiento de sistemas de explotación muy intensificados. Además, es importante mencionar las implicaciones sanitarias potenciales que tiene para el ser humano el consumo de productos ganaderos provenientes de sistemas intensificados y el deterioro ambiental derivado de la utilización de fertilizantes, y de residuos de plaguicidas y herbicidas (Ibáñez y Mas, 1997). Por otro lado, la demanda de productos ganaderos y los gustos de los consumidores cambian a lo largo del tiempo, de hecho en algunos países han surgido denominaciones específicas y marcas de calidad (Jimeno et al., 1996). De esta manera, actualmente se dibuja un escenario ideal para los modelos ganaderos que buscan que las granjas estén en conjunción con sistemas sostenibles y que no dependan de insumos externos (Molina, 2010). La sociedad exige cada vez más alimentos de calidad diferenciada, que llevan implícitos conceptos de respeto al medio ambiente que mitiguen el cambio climático. Así, la calidad se convierte en una de las principales características que diferencian a las razas autóctonas de las razas especializadas.

1.2. Sistemas de producción del vacuno de carne en España

Los diferentes sistemas productivos extensivos y semiextensivos en los que se desarrolla el ganado vacuno de carne de España se ajustan a tres modelos principalmente: sistema de dehesa, sistema de montaña y sistemas no montañosos húmedos.

1.2.1. Sistema de dehesa

En un sentido amplio se consideran dehesas a aquellas explotaciones de ganado en régimen extensivo de pastoreo. Es un sistema agro-silvo-pastoril que se encuentra sobre terrenos de baja aptitud agrícola, poco fértiles y con arbolado abierto. La ubicación geográfica de la dehesa es el oeste y suroeste de la Península Ibérica, sobre terrenos llanos u ondulados. En general, suele estar sobre suelos poco profundos, ácidos y con bajo contenido en materia orgánica (Daza, 1997). El máximo crecimiento de los pastizales se produce durante la primavera, mientras que durante el verano el crecimiento es muy bajo o nulo. El número medio de cabezas por explotación es de 125 y la superficie media que ocupa cada explotación es de 548 hectáreas (Milán et al., 2006). En la zona de dehesa las razas que se emplean son razas rústicas, adaptadas a las difíciles condiciones climáticas. Entre las razas autóctonas más usuales están: Avileña-Negra Ibérica, Morucha y Retinta, entre otras.

1.2.2. Sistema de montaña

Los sistemas montañosos como los Pirineos, Picos de Europa, Sistema Central, entre otros, han estado estrechamente relacionados con la producción de vacuno de carne. Son numerosas las razas que se explotan bajo este sistema, sin embargo las principales razas autóctonas son: Asturiana de la Montaña, Bruna dels Pirineus y Pirenaica, entre otras (Acero, 2009). Generalmente, las explotaciones que se encuentran en áreas de montaña, cuentan con tres tipos de producciones forrajeras bien diferenciadas (Revilla, 1997):

- **Praderas de siega.** Situadas normalmente en los fondos de valle, o en lugares cercanos a los pueblos y utilizadas, mediante siega y henificación, para obtener recursos forrajeros para la alimentación invernal.
- **Pastos de media montaña.** Se trata de superficies que se aprovechan mediante pastoreo directo durante la primavera y el otoño. En algunos casos son de propiedad privada y en otros son comunales.
- **Pastos de alta montaña (puertos).** Son superficies de aprovechamiento comunal en la mayoría de los casos, situadas en altitudes superiores a los 1.500 m.s.n.m. y utilizadas durante el verano.

1.2.3. Sistemas no montañosos húmedos

Se trata de áreas de España que no tienen grandes accidentes montañosos. Son zonas muy benignas para la producción de pastos y, por lo tanto, para la cría de vacas. El máximo

exponente de este tipo de explotación es la comunidad autónoma de Galicia, aunque también está presente en otras comunidades de la cornisa cantábrica, como Asturias y Cantabria. En estas zonas hay excelentes prados naturales (que en ocasiones se mejoran mediante fertilización), aunque también se siembran prados artificiales compuestos de varias especies, normalmente la siembra combina 4 ó 5 especies que aseguran una persistencia importante. Es frecuente que se realice pastoreo rotacional, para lograr unos mejores resultados en la producción de forraje (Acero, 2009). Las razas autóctonas que generalmente se explotan bajo este sistema son: Asturiana de los Valles y Rubia Gallega.

1.3. Programas de mejora genética de vacuno de carne en España

En España, las diferentes asociaciones de ganaderos han tenido como principal actividad la gestión de los libros genealógicos. No fue hasta la década de los 90 que muchas de ellas toman conciencia de la necesidad de establecer controles de rendimientos para la recogida sistemática de la información productiva, y de la utilidad de dicha información a la hora de predecir el valor genético de los mismos como herramienta para seleccionar los reproductores (Díaz y Quintanilla, 2002).

Pese a que cada raza posee su propio esquema de selección según las pautas marcadas por sus respectivas asociaciones, la totalidad de las razas sometidas a selección siguen programas comunes que pueden resumirse de la siguiente manera (Sánchez-Belda, 2002):

- 1) Definición de un objetivo general, que dependerá de lo que busque cada asociación de criadores en cada raza en particular.
- 2) Obtención de información individualizada de datos productivos con la correspondiente genealogía.
- 3) Predicción de valor genético individual de cada carácter medido en cada animal. Generalmente se aplica el método BLUP del modelo animal.
- 4) Elaboración de índices que ponderen los valores genéticos obtenidos.
- 5) Realización y publicación de catálogos de los valores genéticos y/o de los índices obtenidos anteriormente.
- 6) Alternativamente algunas razas realizan pruebas de desempeño e inseminación artificial en planteles o centros especializados.

Los objetivos y criterios de selección de las principales razas autóctonas españolas se resumen en la tabla 1.1. En la actualidad, en la gran mayoría de los programas de vacuno autóctono de carne, la evaluación genética se basa en los caracteres de crecimiento, y más concretamente al peso ajustado a edades tipo y en ocasiones la morfología. En algunos casos, las características maternas se declaran como criterio de elección de sementales y como criterios de selección de las madres en casi todos los casos. En cuanto al registro de los datos, en todas las razas se registran pesos pre y post destete y características morfológicas, en varias de ellas se registran características de la canal, y solo en algunas de ellas se registran medidas de conformación, estado de engrasamiento y pH 24 horas después del sacrificio. A excepción de la facilidad de parto, en pocos casos se registran datos relativos a la reproducción de las hembras con fines de selección (precocidad, intervalo entre partos) (Serradilla, 2008).

En la actualidad, el uso de la información molecular no se hace parte, de una manera formal, de los programas de selección de vacuno de carne autóctono en España. Sin embargo, esta información en ocasiones es utilizada de modo individual por determinados productores en apoyo a sus decisiones a la hora de seleccionar los animales. Por ejemplo, la información sobre el genotipo culón es utilizada en la raza Asturiana de Valles y en la mayoría de las razas se hacen controles de paternidad a los sementales con pruebas de ADN.

1.4. Nuevos paradigmas en el mejoramiento genético animal

El principal objetivo que persigue la mejora genética de las poblaciones es obtener avances en características económicamente importantes. Esto trae consigo el aumento de la frecuencia de genes favorables para la manifestación de una característica en particular (Falconer y Mackay, 1996). La mayor ganancia genética de las especies domésticas se debe, sin lugar a dudas, al mejoramiento genético tradicional basado en el uso del fenotipo registrado de cada animal junto con el conocimiento de su pedigrí para estimar su valor genético. El método estadístico que ha sido utilizado con mayor frecuencia para estas valoraciones ha sido el BLUP (*best linear unbiased prediction*) (Henderson, 1984).

Los microsatélites fueron los primeros tipos de marcadores genéticos que cubrían el genoma, y por lo tanto, tenían la posibilidad de detectar QTL sin importar dónde estuvieran localizados. Georges et al. (1995) utilizaron entre 100 y 200 microsatélites para

Tabla 1.1 Objetivos y criterios de selección de las evaluaciones genéticas en las principales razas de ganado de carne autóctono

	Asturiana de los Valles	Avileña-Negra Ibérica	Bruna dels Pirineus	Morucha	Pirenaica	Retinta	Rubia Gallega	
Objetivos selección	Crecimiento y desarrollo muscular hasta 14 meses	Crecimiento, conformación y reproducción	Crecimiento y reproducción	Desarrollo muscular y esquelético	Crecimiento	Crecimiento, conformación y reproducción	Crecimiento, conformación y reproducción	
Caracteres registrados	Fertilidad	No	Sí	No	No	No	Sí	No
	Pesos predestete	Medidas corporales al destete	Destete	Nacimiento y destete	Nacimiento y destete	Nacimiento y destete	Pesadas periódicas	Nacimiento y destete
	Pesos postdestete	Crecimiento postdestete	En cebo	No	No	No	Pesadas periódicas	No
	Canal Calificación	Sí	Sí	No	No	Sí	Sí	No
	Calificación morfológica	Lineal	Sí	Sí	Sí	Sí	Sí	Sí
Criterios selección	Sementales	Crecimientos pre y postdestete, facilidad al parto	Múltiples criterios	Peso al nacimiento y al destete, facilidad al parto	Peso al nacimiento y al destete	Índice compuesto nacimiento, aptitud cárnica	Peso a los 4 y 6 meses	Índice compuesto del peso al nacimiento y al destete
	Madres de sementales	Crecimiento predestete (hijos)	Múltiples criterios	Aptitud maternal	Índice aptitud maternal	Índice compuesto materno	Índice Global (maternos, reproductivos, morfología)	Caracteres maternos

Fuente: Serradilla, 2008

cubrir el genoma, y detectaron QTL mediante ligamiento dentro de familias de medios hermanos o hermanos completos. Sin embargo, las limitaciones de estos estudios fueron la asignación imprecisa de los QTL y la variación entre familias de la fase de ligamiento. En consecuencia, la fase de ligamiento tuvo que ser determinada dentro de cada familia antes de que el marcador pudiera ser utilizado para la selección.

Los estudios de mapeo de QTL demostraron que muchos QTL afectan a un rasgo cuantitativo típico (Hayes y Goddard, 2001). Por lo tanto, un nuevo tipo de marcadores que detectara todos los QTL era necesario para la selección asistida, además de que no requiriera estimar una fase de ligamiento para cada familia. En la actualidad, la disponibilidad de los marcadores densos de SNP ha abierto nuevas posibilidades en las evaluaciones genéticas de individuos. La precisiones teóricas alcanzadas por Meuwissen et al. (2001) de 0,85 en la predicción del valor genético total, son cercanas a la exactitudes que se obtienen en una prueba de progenie. Esto significa que los programas de mejora de ganado, basados en la selección genómica, podría duplicar el progreso genético generalmente alcanzado por las pruebas de progenie convencionales con un 92% menos de coste, esto específicamente en ganado lechero (Schaeffer, 2006).

La potencial aplicación de la selección genómica ha sido investigada en numerosos estudios. En aves por ejemplo, González-Recio et al. (2009) analizaron la exactitud de las predicciones genómicas para la tasa de conversión de alimento en pollos de engorde. En acuicultura, Sonesson y Meuwissen (2009) investigaron las ganancias genéticas, la endogamia y la exactitud de la selección genómica. En vacuno lechero, Luan et al. (2009) estimaron la exactitud y el posible sesgo de la predicción de los valores genómicos en toros Noruegos. Dodds et al. (2014) evaluaron la aplicación de la selección genómica en ovejas doble propósito de Nueva Zelanda. Por su parte, Lillehammer et al. (2013) compararon diferentes implementaciones de la selección genómica en un esquema de cría de cerdos. Los diferentes resultados de estos estudios sugieren un claro potencial en la aplicación de la selección genómica en diferentes especies dedicadas a la producción animal.

El vacuno lechero es la especie que en la actualidad tiene los mayores avances en la implementación de la selección genómica en los tradicionales programas de evaluación genética. El beneficio que la selección genómica ha traído para esta especie radica principalmente en la reducción de los tamaños generacionales. Los toros y vaquillas pueden seleccionarse a muy temprana edad lo que conduce a duplicar la ganancia genética por año (Schaeffer, 2006; König et al, 2009). Esto cambia radicalmente el diseño de los

programas de mejoramiento lechero, los cuales se han basado en las pruebas de progenie. Mediante el uso de marcadores genéticos se puede seleccionar los mejores toros cuando ellos tienen un año de edad, en lugar de esperar hasta que hayan completado la prueba de progenie a los 5 años de edad. Las poblaciones de toros lecheros son las que más rápidamente se genotipan en todo el mundo. Los genotipos de toros Holstein con información de sus hijas ya están disponibles para más de 50.000 toros. En 2012, Interbull recibió 57.902 toros Holstein jóvenes con estimaciones de valores de cría genómico (Sullivan y Jacobsen, 2012). De hecho, en la mejora del ganado Frisón, en muchos países los toros que entran en inseminación artificial han sido seleccionados genómicamente y existe la tendencia a genotipar cada vez más vacas (Mantysaari, 2014).

Sin embargo, la implementación de la selección genómica en vacuno de carne es aún materia de discusión. Pese a que algunos trabajos ya han comenzado a evaluar diferentes metodologías para su futura aplicación (Carvalho, 2014; Gunia et al., 2014), existen algunas razones que aún los limitan. Entre los principales motivos están: el limitado número de individuos con valores genéticos precisos, las grandes diferencias entre los sistemas de manejo y el uso limitado de la inseminación artificial, y el limitado tamaño de las poblaciones de raza pura (Garrick, 2009).

1.5. Perspectivas para la aplicación de la selección genómica en razas bovinas de carne autóctonas españolas

En la convocatoria de 2010, un grupo integrado por distintas universidades y centros de investigación, liderado por el Dr. Luis Varona, presentó una propuesta titulada “Exploring the use of genomic selection tools to improve autochthonous local beef cattle breeds in Spain” que fue financiada por el Ministerio de Ciencia e Innovación, con la referencia AGL 2010-15903. En dicha propuesta se contemplaban una serie de estudios genealógicos y a partir de datos moleculares, algunos de los cuales se incluyen en esta tesis.

1.5.1. Importancia de los análisis genealógicos

Gutiérrez et al. (2003) estudiaron la estructura genética de algunas de las razas incluidas en esta tesis doctoral, con datos del pedigrí desde su fundación hasta el año 1995. Más recientemente, en la raza Retinta, Avilés et al. (2006) estimaron el número efectivo de ancestros y el número efectivo de rebaños fundadores. No obstante, en las razas locales, el

tamaño de la población y los métodos de selección cambian, incluso en cortos periodos de tiempo (Boichard et al., 1997), lo que hace conveniente analizar la estructura poblacional de forma periódica con el fin de evaluar la dinamica y la posible pérdida de diversidad de las razas. En nuestro caso el momento era particularmente interesante pues, por una parte, podíamos analizar los efectos de la implementación de distintos programas de mejora basados en la aplicación de la metodología BLUP y, por otra parte, el conocimiento de la situación actual de las razas sería un elemento importante para decidir la implementación de la selección genómica y tener una referencia para evaluar a medio plazo sus consecuencias en la dinámica poblacional.

Los análisis genealógicos son una importante herramienta que permite describir la variabilidad genética a través de las generaciones. La tasa de endogamia y el tamaño efectivo de la población son los parámetros más comúnmente utilizados que permiten cuantificar la pérdida de variabilidad y la tasa de deriva genética (Wright, 1922, 1931). Por su parte, el análisis de la probabilidad de origen de los genes, es una aproximación complementaria que permite determinar la contribución de los fundadores y ancestros en las poblaciones. Estos pueden ser combinados con el número efectivo de fundadores (Lacy, 1989) y el número efectivo de ancestros (Boichard et al., 1997), siendo este último el que permite determinar cuellos de botella en las poblaciones analizadas.

Adicionalmente, los análisis genealógicos permiten estimar censos poblacionales, intervalos generacionales y la estructura piramidal de los hatos en cada una de las razas. Los intervalos generacionales condicionan el progreso genético, mientras que la estructura piramidal explica el flujo de reproductores entre los diferentes rebaños que hacen parte de las razas, además de detectar aquellos rebaños donde se concentra la selección de los animales.

1.5.2. Utilización de herramientas moleculares

La diversidad genética ha sido definida como la variedad de alelos y genotipos presentes en una población y esto se refleja en diferencias morfológicas, fisiológicas y de comportamiento entre los individuos y poblaciones (Frankham et al., 2002). En esencia, las mutaciones del ADN, recombinaciones y el azar, así como los cambios dirigidos en las frecuencias alélicas son analizadas para reconstruir los procesos evolutivos, ecológicos, históricos o demográficos (divergencia, expansiones, cuellos de botella, aislamiento genético, introgresión, adaptación y selección) (Hey y Machado 2003). Desde principios

de la década de 1990 hasta ahora, los marcadores microsatélites y los SNP han liderado la caracterización de la diversidad genética de las poblaciones, logrando el desarrollo de numerosas herramientas moleculares que han permitido complementar los análisis genealógicos descritos anteriormente (Conrad y Hurler, 2007). En particular, para este trabajo se ha contado con una poderosa herramienta, novedosa en su aplicación al estudio de las razas españolas, como es *el panel o chip* de 778 000 SNP (Illumina BovineHD Beadchip genotyping chip; Illumina Inc., 2012).

Generalmente, la diversidad genética se mide por las frecuencias de genotipos y alelos, la proporción de loci polimórficos, la heterocigosis observada y esperada o la diversidad alélica. Igualmente, los marcadores moleculares permiten estimar numerosos índices que ayudan a determinar niveles de estructuración poblacional y diversidad genética, entre los cuales se destacan los índices de fijación (Wright, 1951, 1965) y el análisis de varianza molecular (Excoffier et al., 1992). Por su parte, las medidas moleculares empleadas en la diferenciación de las poblaciones se basan generalmente en las distancias genéticas de las frecuencias de alelos entre poblaciones (Laval et al., 2002; Nei, 1987). Por su parte, las representaciones multivariantes de las relaciones genéticas entre las poblaciones y los análisis de clúster son herramientas adicionales que permiten tener otra perspectiva de los niveles de agrupación entre razas. En poblaciones bovinas españolas, se han llevado a cabo análisis de este tipo utilizando marcadores de tipo microsatélite (Beja-Pereira et al., 2003; Martín-Burriel et al., 2011). El análisis de un panel de marcadores SNP de alta densidad supone un salto cuantitativo enorme en cuanto al número de marcadores analizados, que puede tener consecuencias en lo que respecta a la evaluación de la diversidad y de las relaciones genéticas entre las razas.

El panel utilizado es igualmente de interés para el análisis del desequilibrio de ligamiento (LD, en siglas inglesas). Se espera que con respecto a lo estimado con microsatélites y paneles de menor densidad, el panel de alta densidad permita el análisis preciso del LD en distancias cortas. Esta información, junto con la estimación de la persistencia de fases entre razas sirve de guía para predecir la densidad de marcadores requeridos en la implementación de la selección genómica tanto dentro de razas como entre razas (de Roos et al. 2008), así como para obtener una aproximación del tiempo de divergencia de las razas y de sus tamaños efectivos (Sved, 1971).

Bibliografía

- Acero, P. (2009). *Planificación y manejo de la explotación de Vacuno*. Valladolid: Consejería de Agricultura y Ganadería, Junta de Castilla y León.
- Avilés, C., Álvarez, F., Azor, P.J., Fernández, I., Pérez, J.A., Rodero, A. y Molina, A. (2006). Estructura del sistema de cría y variabilidad genética del núcleo de control cárnico del esquema de selección de la raza Retinta. *Feagas*, 30, 64–72.
- Beja-Pereira, A., Alexandrino, P., Bessa, I., Carretero, Y., Dunner, S., Ferrand, N., Jordana, J., Laloe, D., Moazami-Goudarzi, K., Sanchez, A. y Cañón, J. (2003). Genetic characterization of southwestern European bovine breeds: a historical and biogeographical reassessment with a set of 16 microsatellites. *Journal of Heredity*, 94, 243–250.
- Boichard, D., Maignel, L. y Verrier, É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution*, 29, 5–23.
- Carvalho, R. (2014). Genomic selection in Nelore cattle in Brazil. En *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver*. Recuperado de https://asas.org/docs/default-source/wcgalp-proceedings-oral/258_paper_10329_manuscript_1314_0.pdf?sfvrsn=2
- Conrad, D.F. y Hurler, M.E. (2007). The population genetics of structural variation. *Nature Genetics*, 39, S30–S36.
- Daza, A. (1997). Sistemas de producción de vacuno de carne en zonas de dehesa. En C. Buxadé (Ed.), *Vacuno de carne: aspectos clave* (pp. 191–211). Madrid: Mundi-Prensa.
- de Roos, A.P.W., Hayes, B.J., Spelman, R.J. y Goddard, M.E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, 179, 1503–1512.
- Díaz, C. y Quintanilla, R. (2002). Estado y nuevas demandas de los programas de mejora de vacuno de carne. En *XI Reunión nacional de mejora genética animal, Pamplona*. Recuperado de http://acteon.webs.upv.es/CONGRESOS/XI%20Reunion%20MG%20PAMPLONA%202002/Docs%20XI/CDiaz_RQuintanilla.pdf
- Dodds, K.G., Auvra, B., Lee, M., Newman, S.A.N. y McEwan, J.C. (2014). Genomic selection in New Zealand dual purpose sheep. En *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver*. Recuperado de https://asas.org/docs/default-source/wcgalp-proceedings-oral/333_paper_10352_manuscript_1331_0.pdf?sfvrsn=2

- Excoffier, L., Smouse, P.E. y Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.
- Falconer, D.S. y Mackay, T.F.C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow: Addison Wesley Longman Limited.
- FAO. (2007). *The state of the World's animal genetic resources for food and agriculture*. En B. Rischkowsky y D. Pilling (Eds.). Rome: Food and Agriculture Organization of the United Nations.
- Frankham, R., Ballou, J.D. y Briscoe, D.A. (2002). *Introduction to conservation genetics*. Cambridge: University Press.
- Gandini, G.C. y Villa, E. (2003). Analysis of the cultural value of local livestock breeds: a methodology. *Journal of Animal Breeding and Genetics*, 120, 1–11.
- Garrick, D. (2009). The nature and scope of some whole genome analyses in US beef cattle. En Proceedings of the Beef Improvement Federation 41st Annual Research Symposium, Sacramento. (pp. 92–102) Recuperado de <http://beefimprovement.org/content/uploads/2013/07/BIF-Proceedings3.pdf>
- Georges, M., Nielsen, D., Mackinnon, M., Mishra, A., Okimoto, R., Pasquino, A.T., Sargeant, L.S., Sorensen, A., Steele, M.R., Zhao, X., Womack, J.E. y Hoeschele, I. (1995). Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics*, 139, 907–920.
- González-Recio, O., Gianola, D., Rosa, G.J.M., Weigel, K.A. y Kranis, A. (2009). Genome-assisted prediction of a quantitative trait measured in parents and progeny: application to food conversion rate in chickens. *Genetics Selection Evolution*, 41, 3.
- Gunia, M., Saintilan, R., Venot, E., Hozé, C., Fouilloux, M.N. y Phocas, F. (2014). Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers. *Journal of Animal Science*, 92, 3258–3269.
- Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J. y Piedrafita, J. (2003). Pedigree analysis of eight Spanish beef cattle breeds. *Genetics Selection Evolution*, 35, 43–63.
- Hayes, B.J. y Goddard, M.E. (2001). The distribution of the effects of genes affecting quantitative traits in livestock. *Genetics Selection Evolution*, 33, 209–229.
- Hayes, B.J., Visscher, P.M., McPartlan, H.C. y Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13, 635–643.
- Henderson, C.R. (1984). *Applications of linear models in animal breeding*. Guelph: University of Guelph Press.

- Hey, J. y Machado, C.A. (2003) The study of structured populations – new hope for a difficult and divided science. *Nature Reviews Genetics*, 4, 535–543.
- Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, 41(Suppl. 1), 32–46.
- Ibáñez, M. y Mas, B. (1997). Razas bovinas autóctonas de interés. En C. Buxadé (Ed.), *Vacuno de carne: aspectos clave* (pp. 115–133). Madrid: Mundi-Prensa.
- Illumina Inc. (2012). BovineHD Genotyping BeadChip Datasheet. Recuperado de http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf
- Jimeno, V., Aguado, J.A. y Buxadé, C. (1996). Carne de vacuno de calidad. *Bovis*, 71, 49–63.
- König, S., Simianer, H. y Willam, A. (2009). Economic evaluation of genomic breeding programs. *Journal of Dairy Science*, 92, 382–391.
- Lacy, R.C. (1989). Analysis of founder representation in pedigrees. Founder equivalents and founder genome equivalents. *Zoo Biology*, 8, 111–123.
- Laval, G., SanCristobal, M. y Chevalet, C. (2002). Measuring genetic distances between breeds: use of some distances in various short term evolution models. *Genetics Selection Evolution*, 34, 481–507.
- Lillehammer, M., Meuwissen T.H.E. y Sonesson, A.K. (2013). Genomic selection for two traits in a maternal pig breeding scheme. *Journal of Animal Science*, 91, 3079–3087.
- Luan, T., Woolliams, J.A., Lien, S., Kent, M., Svendsen, M. y Meuwissen, T.H.E. The Accuracy of Genomic Selection in Norwegian Red Cattle Assessed by Cross-Validation. *Genetics*, 183, 1119–1126.
- Mantysaari, E.A. (2014). Challenges in industry application of genomic prediction experiences from dairy cattle. En *Proceedings of the 10th World Congress of Genetics Applied to Livestock Production, Vancouver*. Recuperado de https://asas.org/docs/default-source/wcgalp-proceedings-oral/079_paper_10272_manuscript_1276_0.pdf?sfvrsn=2
- Martín-Burriel, I., Rodellar, C., Cañón, J., Cortés, O., Dunner, S., Landi, V., Martínez-Martínez, A., Gama, L.T., Ginja, C., Penedo, M.C.T., Sanz, A., Zaragoza, P. y Delgado, J.V. (2011). Genetic diversity, structure, and breed relationships in Iberian cattle. *Journal of Animal Science*, 89, 893–906.
- Meuwissen, T.H.E., Hayes, B.J. y Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819–1829.
- Milán, M.J., Bartolomé, J., Quintanilla, R., García-Cachán, M.D., Espejo, M., Herráiz, P.L., Sánchez-Recio, J.M. y Piedrafita, J. (2006). Structural characterisation and typology of beef cattle farms of Spanish wooded rangelands (dehesas). *Livestock Science*, 99, 197–209.

- Molina, A. (2010). Biodiversidad y conservación de razas autóctonas de animales domésticos. *Ambienta*, 91, 109–125.
- Nei, M. (1987). *Molecular evolutionary genetics*. New York: Columbia University Press.
- Revilla, R. (1997). Sistemas de explotación del ganado reproductor en zonas de montaña. En C. Buxadé (Ed.), *Vacuno de carne: aspectos clave* (pp. 229–249). Madrid: Mundi-Prensa.
- Sánchez, L., Vallejo, M., Iglesias, A., Álvarez, F., Fernández, M. y Salgado, J.M. (1992). *Razas bovinas autóctonas de Galicia*. Pontevedra: Dirección Xeral da Producción Agropecuaria e Industrias Agroalimentarias, Consellería de Agricultura, Gandería e Montes, Xunta de Galicia.
- Sánchez-Belda, A. (2002). *Razas ganaderas españolas. I. Bovinas*. Madrid: FEAGAS y MAPA.
- Schaeffer, L.R. (2006). Strategy for applying genome-wide selection in dairy cattle. *Journal of Animal Breeding and Genetics*, 123, 218–223.
- Serradilla, J.M. (2008). Objetivos, organización y factores limitantes de los programas de selección de las razas autóctonas en España. En *XIV Reunión nacional de mejora genética animal, Sevilla*. Recuperado de http://acteon.webs.upv.es/CONGRESOS/XIV%20Reunion%20MG%20SEVILLA%202008/Docs%20XIV/Serradilla_XIV.doc
- Sonesson, A.K. y Meuwissen, T.H.E. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genetics Selection Evolution*, 41, 37.
- Sullivan, P.G. y Jakobsen, J.H. (2012). Robust GMACE for young bulls – methodology. *Interbull Bulletin*, 45, 3–7.
- Sved, J.A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, 2, 125–141.
- Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*, 56, 330–338.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159.
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, 15, 323–354.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19, 395–420.

Capítulo 2

OBJETIVOS

2. Objetivos

El objetivo general de esta tesis es conocer la estructura genética actual de las principales razas de vacuno de carne autóctonas españolas, así como explorar la historia genética y constitución de las mismas mediante el uso de marcadores SNP de cara a la implementación de la selección genómica dentro y entre poblaciones.

Para lograr este objetivo general, se proponen tres estudios, que a su vez constituyen los objetivos específicos de la tesis:

1. Evaluar la estructura genética de siete razas de ganado vacuno de carne autóctono español mediante la estimación de parámetros demográficos y poblacionales por medio del análisis de las genealogías registradas.
2. Evaluar la diversidad y divergencia genética de siete poblaciones autóctonas de ganado vacuno de carne español usando un panel de alta densidad (778.000 marcadores SNP).
3. Determinar las relaciones entre loci y entre poblaciones a partir del cálculo del desequilibrio de ligamiento y la persistencia de fases, y asimismo el tamaño efectivo ancestral, como herramientas básicas en la implementación de la selección genómica, en las principales razas de bovinas de carne españolas empleando un panel denso de SNP.

Como se ha señalado, esta tesis doctoral hace parte del proyecto titulado “Exploring the use of genomic selection tools to improve autochthonous local beef cattle breeds in Spain”, donde el objetivo principal del proyecto busca evaluar la factibilidad del uso de la información genómica para objetivos de selección en razas autóctonas españolas de ganado de carne. Específicamente con esta tesis se buscó dar solución a los objetivos específicos siguientes:

- Actualizar los parámetros demográficos y poblacionales estimados en las poblaciones de ganado de carne español.
- Explorar la historia genética de las poblaciones a través del análisis de las frecuencias génicas y el desequilibrio de ligamiento de estas poblaciones a partir de datos de SNP.

Capítulo 3

REVISIÓN BIBLIOGRÁFICA

3. Revisión bibliográfica

3.1. Historia y origen del ganado vacuno español

Según estudios paleontológicos relacionados con el origen de los bóvidos, el ganado vacuno se encuentra clasificado dentro del género *Bos*, que comprende bovinos de gran tamaño con cuernos en ambos sexos. Se divide en dos especies: los taurinos (*Bos taurus taurus*) que tuvieron su origen en Europa, y los cebúes (*Bos taurus indicus*) que provienen de la India y se caracterizan por una joroba en la cruz (Loftus et al., 1994). Con tecnología más reciente, la división en estas dos grandes especies es explicada por una profunda bifurcación encontrada en el DNA mitocondrial bovino, indicativo de una divergencia que ocurrió hace más de 100.000 años (antes de la domesticación) (Bradley et al., 1996; Loftus et al., 1994).

El Uro o Auroch, el cual en la actualidad se encuentra extinto, es conocido como la forma ancestral de todas las especies bovinas actuales. Este ha sido clasificado como *Bos primigenius primigenius* y está más estrechamente relacionado con los *B. taurus* modernos que con los *B. indicus* (Troy et al., 2001). Los primeros signos de domesticación del Uro se dieron en el valle medio del río Éufrates en la edad neolítica y en el valle alto del río Tigris hace alrededor de 10.300 - 10.800 años (Helmer et al., 2005; Hongo et al., 2009). Posteriormente, el ganado doméstico comenzó a aparecer en el occidente de Anatolia y el sudeste de Europa (hace 8.800 años), sur de Italia (hace 8.500 años) y Europa central (hace 8.000 años), debido principalmente a la creciente importancia económica del ganado vacuno para la producción de carne y leche (Vigne y Helmer, 2007; Vigne, 2008).

Las grandes corrientes migratorias hacia Europa dieron lugar a dos itinerarios distintos. El primero de ellos recorrió Europa central hasta llegar a la península Ibérica, y el segundo llegó desde el continente africano vía Egipto, y posteriormente la costa mediterránea (Ammerman y Cavalli-Sforza, 1984). Ambos movimientos, sumado a las influencias del medio, dieron origen a los tres troncos étnicos que en la actualidad clasifican a los bovinos autóctonos españoles: *B. taurus ibéricus*, *B. taurus cantábricus* y *B. taurus turdetanus*, que en el ámbito ganadero se definen como: Tronco Cóncavo castaño, Tronco Recto negro y Tronco Convexo rojo (Figura 3.1) (Sánchez-Belda, 1984).

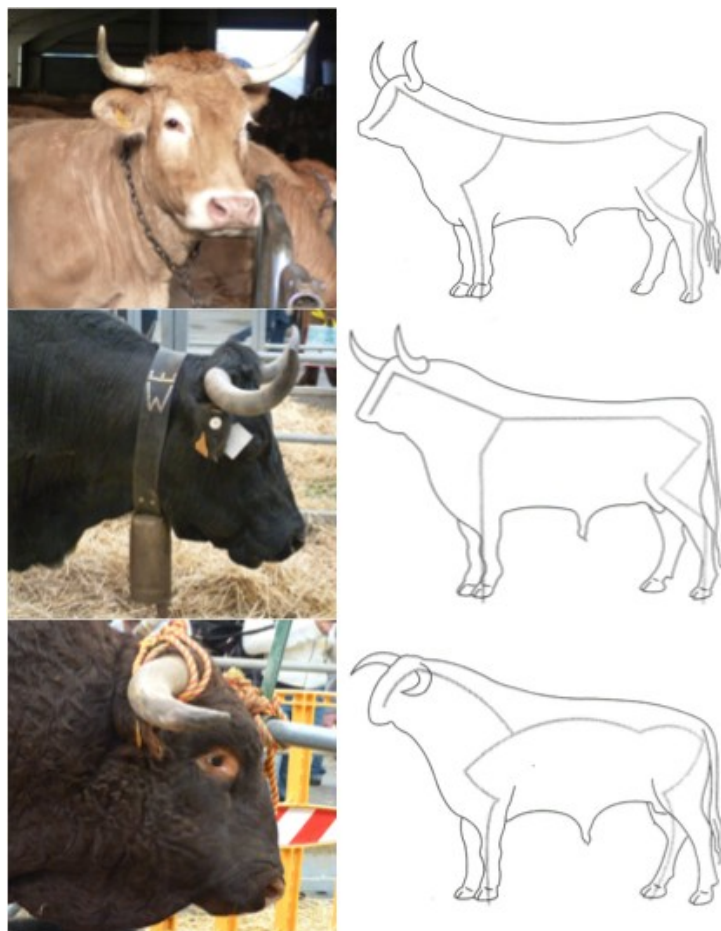


Figura 3.1 Troncos étnicos de los bovinos autóctonos españoles

De arriba a abajo, Cóncavo castaño, Recto negro y Convexo rojo. Tomado y modificado de Sánchez-Belda (2002).

3.2. Principales razas bovinas autóctonas españolas de carne

España cuenta con una larga tradición en la cría y explotación de ganado vacuno particularmente adaptado a su climatología. En la actualidad, el ganado vacuno autóctono de carne está distribuido por las diferentes regiones del país, alternando con otros sistemas de producción lechero (Sánchez-Belda, 2002).

Dentro de las razas autóctonas de ganado de carne españolas más relevantes están: Asturiana de los Valles (AV), Avileña-Negra Ibérica (ANI), Bruna dels Pirineus (BP), Morucha (Mo), Pirenaica (Pi), Retinta (Re) y Rubia Gallega (RG), todas ellas clasificadas actualmente dentro de la categoría de “Razas de Fomento”, con la excepción de la Bruna dels Pirineus que pertenece a la categoría “Raza Autóctona en Peligro de Extinción”. En la

Figura 3.2 se ilustra la distribución actual de las principales razas vacunas autóctonas de carne en España.

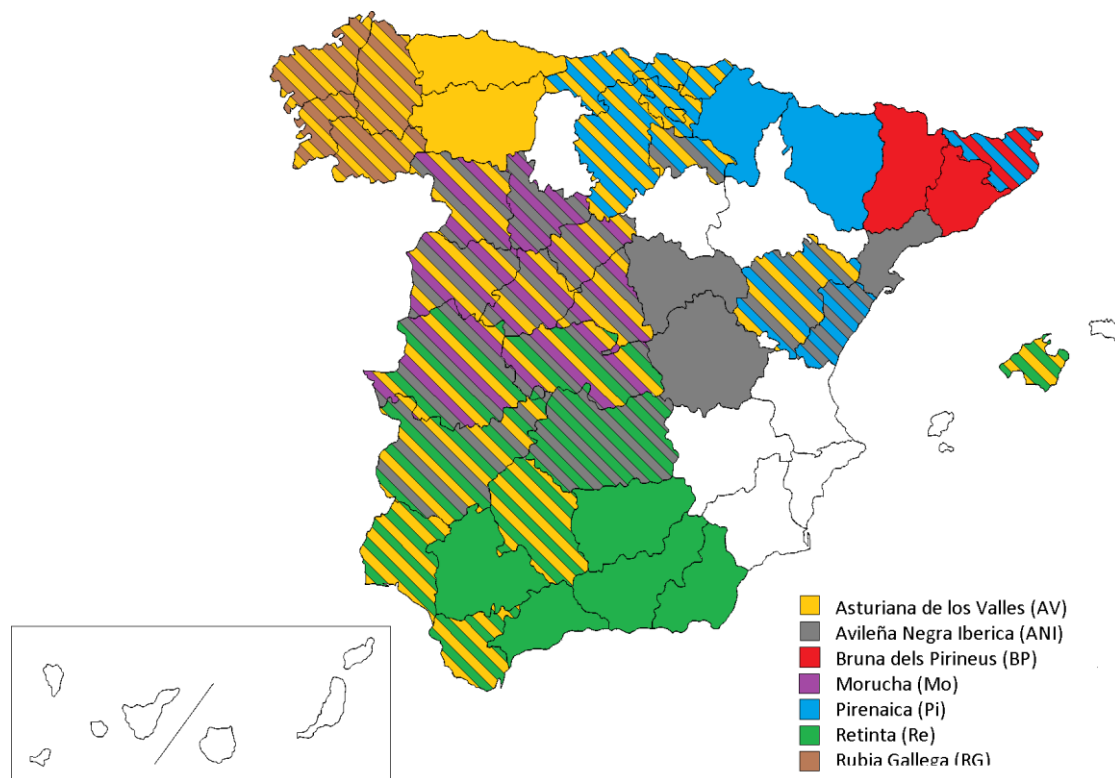


Figura 3.2 Distribución geográfica de las principales razas de carne españolas

3.2.1. Asturiana de los Valles

La raza Asturiana de los Valles se encuentra ubicada en las zonas bajas del Principado de Asturias, en las estribaciones de la cordillera Cantábrica. En la actualidad, los rebaños de esta raza se han extendido por todas las comunidades autónomas de la Cornisa Cantábrica, además en Castilla-León, Aragón, Madrid, Castilla-La Mancha y Extremadura. En su explotación predomina el sistema extensivo con pastoreo en épocas de primavera, verano y otoño, o bien en forma de heno y ensilado en los meses de invierno (sistema valle/puerto). En las zonas bajas próximas a la costa, se da un sistema semintensivo, donde se combina la estabulación con pastoreo (Sánchez-Belda, 2002).

La filogenia de la raza Asturiana de los Valles tiene directa dependencia del Tronco Cantábrico. Posee un perfil cóncavo, es hipermétrica y longimorfa, de mucha masa, buen hueso y de capa castaña (Cima-García, 1986) (Figura 3.3). Inicialmente fue explotada por su triple aptitud, leche, carne y trabajo, y aunque con ciertos antiguos intentos mejorantes de la primera, hoy en día está especializada en la producción de carne. Se diversifica en

dos tipos, normal y culón, siendo este último portador de hipertrofia muscular hereditaria que afecta al 50% de las vacas y al 15% de los sementales (Cima-García, 1996).



Figura 3.3 Raza Asturiana de los valles

Foto cedida por Ignacio. Licencia: Creative Commons

En la actualidad, esta raza representa el mayor censo de todas las razas autóctonas con cerca de 44.000 reproductoras inscritas en el libro genealógico. El proceso de selección está basado en el libro genealógico y los controles relativos a la inscripción y permanencia en el mismo. Igualmente es practicado el testaje de machos jóvenes para elección de los futuros sementales destinados a los centros de inseminación artificial. Éstos disponen de catálogos de sementales con información sobre morfología, valoración genética y grado de cularidad (Sánchez-Belda, 2002).

3.2.2. Avileña-Negra Ibérica

La Avileña-Negra Ibérica es una raza bovina autóctona ubicada en las zonas de montaña del centro de la península Ibérica. Procede del Tronco Negro Ibérico (*Bos taurus ibericus*). Posee gran capacidad para aprovechar medios muy difíciles con buenos rendimientos productivos, elevada fertilidad y cualidades maternas muy acusadas. La raza se explota de manera exclusivamente bajo un sistema extensivo, permaneciendo los animales todo el año en distintas praderas. En algunas zonas de montaña se sigue manteniendo un sistema de producción vinculado a la trashumancia (Sánchez-Belda, 2002).



Figura 3.4 Raza Avileña-Negra Ibérica

Inicialmente, el bovino negro castellano se desarrolló bajo condiciones favorables de cría y sin influencia e infusión de razas foráneas en un proceso endogámico mantenido desde remotas épocas hasta nuestros días. A partir de la segunda mitad del siglo XIX se redujo el área geográfica del gran bovino negro ibérico, recluyéndolo a terrenos montañosos de medio difícil, acosado por la expansión de otro tipo de bovinos más productivos hasta el punto de desaparecer en muchos de sus reductos y, en los que permanecía, manifestaba claros signos de degeneración. Con el tiempo la Avileña se impuso como ganado mejorado, contribuyendo a sostenimiento de la Negra Ibérica. Posteriormente y dada la identidad de rasgos, sistemas de explotación y coincidencia de objetivos de selección hacia la explotación de carne, se optó por agruparlas bajo una misma denominación (Sánchez-Belda, 1983).

La raza Avileña-Negra Ibérica posee una capa uniforme y es de tamaño medio a grande (Figura 3.4). Su censo es de alrededor de 36.000 reproductoras en el libro genealógico. En la actualidad se lleva a cabo un control de rendimientos, que reúne datos sobre pesos al nacimiento, destete y en cebaderos, incluyendo fertilidad, longevidad y facilidad de parto. Adicionalmente cuenta con centro de selección y orientación donde el principal objetivo de la selección se dirige a favorecer el mantenimiento de las cualidades

maternas, mejorando las características productivas y de calidad de carne (Sánchez-Belda, 2002).

3.2.3. Bruna dels Pirineus

La vaca Bruna dels Pirineus es una raza de aptitud cárnica extendida por diversas comarcas del Pirineo y el pre-pirineo catalán. Se trata de una población que tiene como principales características la rusticidad, la longevidad y la buena aptitud materna, siendo explotada habitualmente mediante un sistema extensivo, con trashumancia o simple “subida a puertos” en época favorable. La explotación es de carácter intensivo en los terneros para su terminación hacia el mercado (Sánchez-Belda, 2002).



Figura 3.5 Raza Bruna dels Pirineus

Foto cedida por Marta Fina.

En el plano general, la definición etnológica de la raza Bruna dels Pirineus corresponde al tipo ortoide, eumétrica, mesomorfa, de buena masa, hueso fuerte y de capa parda (Sánchez-Belda, 2002) (Figura 3.5). La Bruna dels Pirineus proviene de una antigua fusión entre un ecotipo local de la raza Pirenaica y bovinos de raza Brown Swiss procedentes de Suiza. En sus orígenes la vaca Bruna tenía aptitud triple (carne, leche y trabajo), pero más adelante se ha ido seleccionando y mejorando para la producción cárnica (Mujal, 1998).

Su censo aproximado es de 30.000 vacas de más de dos años, de las cuales cerca de 12.000 están inscritas en el libro genealógico y cerca de 5.500 hacen parte del control de rendimientos. Existen además, un gran número de animales cruzados, con un porcentaje elevado de sangre de esta raza (Fina, 2013). Desde el año 1990 hasta la actualidad se está llevando a cabo un programa de mejora genética donde los objetivos principales son, para las madres, mejorar la fertilidad, obtener buenas cualidades maternas y buenas características de rusticidad. Para los terneros se pretende elevar los pesos al destete mejorando así los índices de crecimiento. Además, se procura obtener un peso al nacimiento medio que no afecte la dificultad al parto (Jordana y Piedrafita, 1993).

3.2.4. Morucha

La Morucha es una antigua raza rústica, hoy mejorada para la producción de carne. Su nombre deriva de la coloración degradada de la pigmentación negra. La Morucha posee un perfil recto, proporciones y tamaño medio y discreta proporción de hueso (Figura 3.6). El área geográfica donde habita la raza Morucha es la zona occidental del país, en las provincias de Cáceres, Ávila, Toledo, Valladolid, entre otras. Su forma de cría es extensiva y de permanencia a campo abierto la mayor parte del tiempo y con muy limitados desplazamientos para aprovechar recursos temporales (Sánchez-Belda, 2002).



Figura 3.6 Raza Morucha

Foto cedida por la Asociación de Criadores de Raza Morucha. Licencia: Creative Commons

La raza Morucha procede del tronco Ibérico (*Bos taurus ibericus*). Originalmente se utilizó para el trabajo, aunque también fue empleada como animales de doble propósito carne-lidia. Hoy en día su explotación principal es la producción de carne.

Para el año 2000 la raza contaba con alrededor de 25.000 reproductoras inscritas en el libro genealógico. El esquema de mejora genética se concentra en tres aspectos: el primero es el control de rendimientos cárnicos con pesajes en los rebaños y valoraciones del desarrollo muscular; el segundo implica pruebas de valoración individual de machos en estación de testaje; y el tercero es el establecimiento de toros de referencia en un centro técnico (Sánchez-Belda, 2002).

3.2.5. Pirenaica

La raza Pirenaica tiene proporciones alargadas, mucha masa y predominante aptitud cárnica. Posee una capa de color rubio uniforme (Figura 3.7). Está ubicada principalmente en los valles navarros, registrando en las últimas décadas un proceso expansivo que la ha llevado hasta zonas de Aragón, Guipúzcoa, Vizcaya y Cataluña. Esta raza está adaptada al régimen mixto de pastoreo-estabulación, pero de intenso componente forrajero, alternando herbajes de alta montaña en verano, graderíos de ladera en otoño y primavera, y suministro de henos, ensilado y paja en invierno (Sánchez-Belda, 2002).



Figura 3.7 Raza Pirenaica

La raza Pirenaica se encuentra adscrita al tronco bovino rubio (*Bos taurus turdetanus*). En su origen era una raza de triple aptitud, animal de trabajo, productora de carne y leche. La mecanización rural y la introducción de razas lecheras en su medio tuvo consecuencias negativas y la puso en peligro de desaparecer. Sin embargo, las potencialidades de esta raza en el pastoreo extensivo, entre otros tantos factores, jugaron a su favor como la mejor opción (Mendizábal-Aizpuru et al., 2005).

En la actualidad, el número de vacas pirenaicas inscritas en el libro genealógico supera las 23.000 cabezas (Aranguren, 2002). Por su parte, el servicio de control de rendimientos cuenta con alrededor de 15.000 ejemplares. La selección recae preferiblemente sobre cualidades maternas como fertilidad, facilidad al parto y capacidad lechera, igualmente se selecciona por aptitudes cárnicas (Aranguren, 2002; Sánchez-Belda, 2002).

3.2.6. Retinta

La raza Retinta agrupa animales macizos y fuertes, de perfil corporal subconvexo con proporciones medias, buena masa y equilibrado hueso. Su capa es roja oscura (Figura 3.8). Destaca por su resistencia al calor que permite rendimientos cárnicos destacados. El origen de la raza se enmarca dentro del territorio de las Comunidades de Extremadura y Andalucía (Aljama-Gutiérrez, 1982).



Figura 3.8 Raza Retinta

Según Sánchez-Belda (1984) la raza Retinta se deriva del tronco bovino Rojo Convexo (*Bos taurus turdetanus*). Sin embargo, Jordana et al. (1991) sitúan a la raza Retinta en una posición intermedia entre los grupos Turdetano e Ibérico.

Durante siglos esta raza se mantuvo como bovino de doble aptitud carne-trabajo. Posteriormente, las influencias ambientales y su poco empleo en la agricultura por tratarse de áreas con poca vocación para los cultivos, confluyó en su encuadre como animales cárnicos, generando a finales del siglo XIX una entidad racial única. Su alta adaptación a los terrenos que ocupa y su continua superación, la han llevado a un próspero estado actual (Aljama-Gutiérrez, 1982).

Los objetivos del esquema de selección se orientan a mejorar las cualidades de producción de carne, la conformación de los animales y las características reproductivas, manteniendo la rusticidad y adaptación. El esquema se basa en tres aspectos: valoración de los terneros al destete en los rebaños, valoraciones individuales en centros de testaje y valoración por descendencia. El número de vacas adscritas al libro genealógico es de aproximadamente 17.000 (Sánchez-Belda, 2002).

3.2.7. Rubia Gallega

La raza Rubia Gallega es una raza de aptitud cárnica de buenas cualidades maternas, ubicada y explotada mayoritariamente en Galicia. Su color es fundamentalmente rubio, trigüeño o canela, su conformación es generalmente larga y profunda (Figura 3.9). El sistema de explotación está basado en el pastoreo, dada su facilidad de manejo y rusticidad (Sánchez-Belda, 2002).

La historia de la raza Rubia Gallega comienza cuando una rama humana celta procedente de Francia ocupó Galicia con su ganado e influyó sobre los bovinos autóctonos existentes en la zona. Posteriormente, a comienzos del siglo XX, se inició una etapa ininterrumpida de actividad en materia de selección que ha llevado a esta raza a su estado actual. En la actualidad, es definida como una raza de producción de carne de calidad con buenas cualidades maternas (Sánchez-Belda, 2002).

El censo de la raza Rubia Gallega supera las 57.000 cabezas, de las cuales el 55% corresponde a reproductores. El esquema de selección combina la valoración de las vacas, la evaluación de sementales jóvenes por los rendimientos individuales en la estación de testaje y la evaluación de la descendencia de toros por aptitud cárnica (Sánchez-Belda, 2002). Dentro de las características específicas a evaluar se encuentran: facilidad de parto,

ganancia media diaria durante el periodo de predestete y la conformación de los animales en vivo (ACRUGA, 2011).



Figura 3.9 Raza Rubia Gallega

3.3. Estructura genética poblacional

La estructura de las poblaciones puede ser estudiada partiendo de dos puntos de vista interrelacionados entre sí: el análisis demográfico y el análisis molecular. El análisis demográfico se realiza mediante censos, observaciones y según las características del sistema de apareamiento (análisis del pedigrí). Por su parte, el análisis molecular comprende el patrón de variación genético, lo que se traduce en evaluar los genotipos de los diferentes individuos (Slatkin, 1994).

3.3.1. Estructura demográfica y genealógica de las poblaciones

Una de las herramientas más importantes de que se dispone para el manejo genético de poblaciones es el análisis del pedigrí. Con pedigríes detallados es posible obtener información genética, tales como el nivel de endogamia y la distribución de genes fundadores dentro de los miembros actuales de la población (Hedrick y Miller, 1992).

El estudio de la estructura genética de una población mediante análisis genealógico ha sido realizado en la mayoría de las especies, como el bovino de carne y leche, ovinos y caballos, entre otras (Boichard et al., 1996; Cañas-Álvarez et al., 2014; Goyache et al., 2003; Gutiérrez et al., 2003; Pinheiro et al., 2013). A continuación se describen los

principales parámetros empleados para caracterizar la estructura demográfica de una población.

3.3.1.1 Censo poblacional

El número de reproductores machos y hembras por año, en un momento dado, determina la estructura genética de la población en las generaciones posteriores (Groeneveld et al., 2009). El censo de reproductores es uno de los parámetros clave para evaluar el riesgo que pone en peligro a una población local. Bajo condiciones naturales la proporción de machos y hembras es de 1:1. Sin embargo, esta condición no se cumple en poblaciones de granja bajo selección genética, y más aún con implementación de inseminación artificial.

3.3.1.2 Intervalos generacionales

El concepto de intervalo generacional es definido como el tiempo promedio entre el nacimiento de los padres y el nacimiento de sus descendientes destinados a constituir la próxima generación parenteral. La transferencia de genes de los padres a las crías se puede obtener para las cuatro vías posibles: padre-hijo, padre-hija, madre-hijo y madre-hija. Así, el intervalo generacional se puede estimar por cuatro vías y es expresado en años (James, 1977).

El intervalo generacional es uno de los factores claves que afectan la tasa de progreso genético, ya que las pérdidas o ganancias de variabilidad genética y los avances en selección se producen por generación (Groeneveld et al., 2009). De este modo, si el intervalo generacional es corto, las pérdidas de variabilidad genética se producen a mayor velocidad por unidad de tiempo.

3.3.1.3 Tamaños familiares

El tamaño de la familia se refiere al número de descendientes de un individuo que se convierten en individuos reproductores en la próxima generación (Falconer y Mackay, 1996). Bajo condiciones ideales, los padres tienen la misma oportunidad de contribuir a la descendencia en la siguiente generación. En la práctica, particularmente en producción animal, la contribución genética de los padres no es la misma. Esta contribución desigual conduce a las diferencias o variaciones en los tamaños familiares, lo que a su vez puede tener consecuencias en el incremento de la tasa de endogamia y en la reducción del tamaño efectivo de la población (Groeneveld et al., 2009).

En los tamaños familiares el número de descendientes puede ser categorizado principalmente en dos grupos: el tamaño de familia representado en todos los descendientes machos y hembras nacidos en la población, y el tamaño de familia representado en los descendientes machos y hembras que han sido seleccionados y que tienen al menos un registro de servicio (Groeneveld et al., 2009).

3.3.1.4 Estructura piramidal

La estructura piramidal es empleada para la clasificación de los rebaños según el porcentaje de uso de machos reproductores propios. La estructura piramidal se clasifica en: rebaños núcleo si nunca utilizan machos reproductores nacidos en otros rebaños; rebaños multiplicadores si compran y venden los machos reproductores (Multiplicador 1 o Multiplicador 2 si utilizan o no sus propios machos reproductores, respectivamente); y los rebaños comerciales si compran pero no venden machos reproductores (Comercial 1 o Comercial 2 si utilizan o no sus propios machos reproductores, respectivamente) (Vassallo et al., 1986).

3.3.1.5 Grado de completión del pedigrí

La información sobre la calidad de pedigrí, es decir, la descripción de la integridad de cada uno de los antepasados en el árbol genealógico se describe mediante el índice de integridad (MacCluer et al., 1983). Este índice calcula la proporción de padres, abuelos, bisabuelos y otros antepasados por medio de una representación gráfica del porcentaje de antepasados conocidos, distinguiendo las vías paterna y materna. Se estima mediante las siguientes fórmulas:

$$I_d = \frac{4I_{d_{pat}}I_{d_{mat}}}{I_{d_{pat}} + I_{d_{mat}}} \quad [1]$$

$$I_{d_k} = \frac{1}{d} \sum_{i=1}^d a_i \quad [2]$$

donde k representa la línea paterna (pat) o materna (mat) de un individuo, a_i es la proporción de antepasados conocidos en la generación i y d es el número de generaciones consideradas en el cálculo de la completión del pedigrí.

Por otra parte, el número equivalente de generaciones completas (Boichard et al., 1997) se calcula como la suma de $(1/2)^n$, donde n es el número de generaciones que separan el individuo a cada antepasado conocido.

La finalidad de conocer el grado de profundidad del pedigrí es identificar el nivel de integridad en el que se encuentra el pedigrí de la población bajo estudio a lo largo de las diferentes generaciones, a efectos de comparar con otras poblaciones. Además, muchos otros índices como la endogamia, el coeficiente de parentesco y el número efectivo de fundadores y ancestros dependen de que tan completo se encuentre el pedigrí (Cassell et al., 2003).

3.3.1.6 Coeficiente de endogamia

La endogamia es el resultado del apareamiento entre individuos que están emparentados entre sí por ascendencia. La consecuencia esencial de que dos individuos tengan un antepasado común es que ambos pueden llevar copias de uno de los genes presentes en el antepasado y, si se aparean, estas copias pueden pasar a sus hijos (Falconer y Mackay, 1996). Así, el coeficiente de endogamia de un individuo (F) es la probabilidad de que dos alelos que porta dicho individuo en cualquier locus sean idénticos por descendencia (Wright, 1922).

El coeficiente de endogamia es esencialmente una comparación entre la población en cuestión y una población base especificada. Por lo tanto, solo tiene sentido si se especifica algún tiempo en el pasado, antes del cual los antepasados no sean investigados. Este punto es la población base y por definición los individuos que la formen tienen un coeficiente de endogamia igual a cero. El coeficiente de endogamia de una generación subsecuente expresa la cantidad del proceso disperso que ha tenido lugar desde la población base (Falconer y Mackay, 1996).

3.3.1.7 Coeficiente de coascendencia

El coeficiente de coascendencia (f) entre dos individuos cualesquiera (también llamado coeficiente de parentesco) es idéntico al coeficiente de endogamia de sus hijos si dichos individuos se aparearan. De esta forma, el coeficiente de coascendencia de dos individuos es la probabilidad de que dos gametos tomados al azar, uno de cada individuo, lleven alelos que sean idénticos por descendencia (Malécot, 1948).

La coascendencia promedio de todos los posibles pares de individuos en la población es el coeficiente de endogamia potencial que podría ser obtenido si se hicieran todos los apareamientos al azar de todos los miembros de la población.

3.3.1.8 *Tamaño efectivo de la población*

Sewall Wright mencionó por primera vez el término tamaño efectivo poblacional (N_e) en su trabajo sobre deriva génica (Wright, 1931). El N_e tiene gran importancia debido a la relación inversa con los incrementos de endogamia y con las pérdidas de variabilidad genética debidas a deriva genética (Falconer y Mackay, 1996). Este es considerado como el factor determinante de la tasa de decaimiento de la heterocigosis, y fue descrito como el número de individuos que daría lugar a la varianza del muestreo o a la tasa de endogamia apropiadas a las condiciones bajo consideración, si dichos individuos se reprodujeran en la forma como se haría en la población ideal (Falconer y Mackay, 1996).

Sin embargo, es necesario precisar primeramente, aquello que es llamado como población ideal. Esta es definida como aquella población que inicialmente es grande en la que el apareamiento es aleatorio, y esta población se subdivide en un gran número de subpoblaciones (Fisher, 1930; Wright, 1931). Cuando se habla de la población ideal el tamaño efectivo es igual al tamaño censal. Debido a que esto es difícil de encontrar y que las condiciones especificadas para la población ideal no se cumplen en las poblaciones reales, se habla entonces del tamaño efectivo de la población (N_e) (Caballero, 1994). El N_e está asociado al tamaño poblacional que es relevante en términos evolutivos, esto es, el número de individuos reproductivos, ya que son éstos los que contribuyen a la generación siguiente en términos demográficos y sobre todo genéticos (Hedrick, 2010).

Una de las condiciones de la población ideal es que el número de individuos reproductores es constante en los dos sexos. Sin embargo, en animales domésticos, los sexos están representados desigualmente entre los individuos reproductores. Por lo tanto, la varianza de muestreo atribuible a los dos sexos debe calcularse separadamente. Así, el censo poblacional es $N = N_m + N_f$, donde N_m y N_f son los números de padres y madres, respectivamente. La media armónica es $1/\left[\frac{1}{2}\left(1/N_m + 1/N_f\right)\right]$, de manera que el tamaño efectivo puede ser estimado por:

$$\frac{1}{N_e} = \frac{1}{4N_m} + \frac{1}{4N_f} \quad [3]$$

$$N_e = \frac{4N_m N_f}{N_m + N_f} \quad [4]$$

El N_e también se puede estimar a partir del coeficiente de endogamia. De esta forma, el coeficiente de endogamia se compone de dos partes: la primera denominada “incremento”, $1/2N$, donde N es el tamaño poblacional, que se atribuye a la consanguinidad nueva; y la segunda denominada “remanente”, que procede de la consanguinidad previa y que depende del coeficiente de consanguinidad de la generación anterior (Falconer y Mackay, 1996; Wright, 1931). Así, el incremento de la endogamia se puede representar por la siguiente fórmula:

$$\Delta F = \frac{1}{2N} \quad [5]$$

En este caso el incremento de la endogamia (ΔF) se relaciona con el censo poblacional N . Igualmente, el censo efectivo N_e también se puede relacionar de la misma manera como:

$$N_e = \frac{1}{2\Delta F} \quad [6]$$

donde el incremento de la endogamia por generación ΔF es definido por la siguiente fórmula:

$$\Delta F = \frac{F_n - F_{n-1}}{1 - F_{n-1}} \quad [7]$$

donde F_n es la endogamia en una generación n y F_{n-1} es la endogamia en la generación anterior.

Bajo apareamiento aleatorio, la tasa de endogamia (ΔF) es igual a la tasa de coascendencia (Δf). Así, el tamaño efectivo N_e puede ser obtenido por medio de la fórmula [6], o por medio de:

$$N_e = \frac{1}{2\Delta f} \quad [8]$$

Por lo tanto, la discrepancia entre los dos tamaños efectivos indicaría una desviación de un sistema de apareamiento al azar.

Otra forma alternativa de estimar el N_e es calculando el aumento de la endogamia para cada individuo (ΔF_i), que es igual a $1 - (1 - F_i)^{1/t}$, en el cual F_i es el coeficiente de endogamia (González-Recio et al., 2007) y t puede ser, para cada individuo, el número equivalente de generaciones completas, el número máximo de generaciones conocidas o el número de generaciones completas (Gutiérrez et al., 2008). Promediando estos ΔF_i , se

puede estimar ΔF , como $\overline{\Delta F}$ en la población de referencia, y luego el N_e se calcula como $\overline{N_e} = 1/2 \overline{\Delta F}$ (Cervantes et al., 2008).

3.3.1.9 Probabilidad de origen de los genes

La probabilidad de origen de los genes hace referencia a cómo la contribución desequilibrada de genes de ciertos reproductores puede reducir la reserva genética de la población. Puede ser estimada mediante los siguientes parámetros.

Número efectivo de fundadores

Un fundador es definido como un antepasado con padres desconocidos. El número efectivo de fundadores (f_e), es definido como el número de fundadores que, contribuyendo en igual medida, producirían la diversidad genética existente en una población (Lacy, 1989). Este es estimado como:

$$f_e = 1 / \sum_{k=1}^f q_k^2 \quad [9]$$

donde cada fundador k se caracteriza por su contribución q_k a la población, es decir, la probabilidad de que un gen muestreado aleatoriamente en esta población se origina a partir del fundador k .

Cuando cada fundador tiene la misma contribución ($1/f$), el número efectivo de fundadores es igual al número real de fundadores. El número efectivo de fundadores es habitualmente menor que el número real de fundadores (Boichard et al., 1997).

Número efectivo de antepasados

Una limitación importante del f_e es que ignora los posibles cuellos de botella en el pedigrí producido por el uso desequilibrado de individuos reproductores. El número efectivo de antepasados (f_a) es el mínimo número de antepasados (fundadores o no) necesarios para explicar la diversidad genética completa de una población. Este se estima como:

$$f_a = 1 / \sum_{j=1}^a q_j^2 \quad [10]$$

donde q_j es la contribución marginal del antepasado j , es decir, la contribución genética hecha por un antepasado que no se explica por otros antepasados elegidos previamente.

Por medio de este parámetro se logra obviar gran parte de la sobreestimación de animales fundadores que se puede obtener con el número efectivo de fundadores. Esta

sobrestimación es particularmente fuerte en programas de selección muy intensivos, donde el germoplasma es limitado a un número muy bajo de animales reproductores, por ejemplo en la inseminación artificial (Boichard et al., 1997).

Número efectivo de rebaños fundadores

El número efectivo de rebaños hace referencia a los rebaños que producen padres, abuelos, bisabuelos y tatarabuelos (Robertson, 1953). Este parámetro se calcula como la inversa de la probabilidad de que dos animales tomados al azar en la población procedan de animales del mismo rebaño y resulta informativo sobre la concentración del origen de los animales.

3.3.2. Estructura poblacional y diversidad genética

En la incursión a la estructura genética de las poblaciones que se ha efectuado hasta ahora se han considerado los procesos relacionados con la estructura demográfica y genealógica. En este apartado se describirán algunos procesos que resumen la diversidad genética de las poblaciones basados en las diferencias de las secuencias de DNA.

La caracterización fenotípica proporciona una información inicial de las variantes funcionales de los genes que porta un individuo o población. Sin embargo, en ausencia de datos fenotípicos, o para complementar los existentes, la diversidad genética puede ser estudiada mediante el análisis de polimorfismos de marcadores genéticos. Estos han sido útiles para investigar el origen y la domesticación de especies y sus migraciones posteriores, así como para dar información sobre relaciones evolutivas (Stoneking y Krause, 2011).

La variación genética identificable dentro de una población puede ser estudiada empleando técnicas de marcadores genéticos, siendo los polimorfismos del DNA los de mayor elección. Los marcadores más utilizados en estudios de diversidad genética son: amplificación de fragmentos polimórficos (AFLP), polimorfismo de longitud de fragmentos de restricción (RFLP), polimorfismos de DNA mitocondrial (mtDNA), microsatélites o repeticiones simples en tándem (STR) y los polimorfismos de un solo nucleótido (SNP) (Awise, 2004).

3.3.2.1. Uso de los SNP en la caracterización de la diversidad genética

Cada marcador tiene diferentes propiedades y ventajas para la evaluación de las variaciones en el genoma de acuerdo a los objetivos de la investigación (Anne, 2006;

Sunnucks, 2000). En los años 90, los marcadores tipo microsatélite fueron ampliamente utilizados en la genética por su alto polimorfismo, reproducibilidad y automatización para la detección. Sin embargo, desventajas como el tiempo, costo elevado y dificultades en la detección precisa de la longitud de repetición limitaron su utilidad. En la actualidad, la disponibilidad de nuevas tecnologías de detección de polimorfismos a lo largo de todo el genoma son mucho más populares (Liu et al., 2005; Xing et al., 2005).

Los SNP son variaciones en nucleótidos únicos en el genoma que muestra una mutación entre las secuencias derivadas de una población (Vignal et al., 2002). Existen SNP en todo el genoma. Son muy abundantes en el genoma humano, a razón de un SNP por cada 1.000 pares de bases (The International SNP Map Working Group, 2001). Al ser marcadores bialélicos, los SNP poseen un polimorfismo limitado en comparación con los microsatélites (Helyar et al., 2011) y por tal razón hay que usar muchos para obtener un buen nivel de información. Por ejemplo, en estimaciones de distancias genéticas o en verificación de parentescos se sugiere que alrededor de 100 SNP son tan informativos como 20 marcadores microsatélites (Baruch y Weller, 2008; Kalinowski, 2002).

En la actualidad, los SNP son marcadores atractivos para aplicarlos en estudios de diversidad genética debido a que son abundantes, estables y el proceso de obtención es altamente automatizado. Sin embargo, la fase preliminar del descubrimiento de los SNP a partir de bases de datos es crítica, ya que pueden sufrir un sesgo causado por el proceso para estimar y seleccionar los SNP. Este hecho se da cuando, en la construcción del panel, los SNP se estiman en una muestra pequeña de individuos (Clark et al., 2005). Por lo tanto, el efecto del sesgo tenderá a ignorar SNP con baja frecuencia en poblaciones no relacionadas con las incluidas en el panel de descubrimiento (Albrechtsen et al., 2010). No obstante, las nuevas tecnologías moleculares permiten obtener genotipos de alrededor de 778.000 SNP en bovinos, diseñados a partir de una amplia gama de razas incluyendo *Bos taurus taurus*, *Bos taurus indicus* y razas híbridas (Illumina Inc., 2012).

Numerosos estudios que han implementado marcadores SNP han proporcionado información valiosa en la caracterización de la diversidad genética en distintas razas de ganado vacuno (Gautier et al., 2007; Salomon-Torres et al., 2014; The Bovine HapMap Consortium, 2009; Villa-Angulo et al., 2009).

3.3.2.2. Herramientas informáticas para el análisis de datos moleculares

En la secuenciación del genoma humano (International Human Genome Sequencing Consortium, 2001; Venter et al., 2001), y particularmente en nuestro caso, del ganado bovino (The Bovine Genome Sequencing and Analysis Consortium, 2009; The Bovine HapMap Consortium, 2009) se ha generado una enorme cantidad de información que es necesario procesar, almacenar y distribuir de forma adecuada para poderse analizar convenientemente. Generalmente esta información es voluminosa, compleja y se encuentra dispersa en diferentes formatos y fuentes (Excoffier y Heckel, 2006). La mayoría de las veces se requiere el uso de varios programas o una combinación de varios enfoques que combine programas particulares de genética de poblaciones y líneas de programación (*scripts*) bajo diferentes plataformas, especialmente cuando se trata de grandes cantidades de datos.

En muchas ocasiones los procesos de depuración y análisis se realizan bajo la plataforma UNIX (Linux-Macintosh) empleando el intérprete de comandos Shell, siendo este último el que le permite comunicarse al usuario con el sistema. La ventaja de este tipo de plataformas radica en la multifuncionalidad que poseen muchos de los comandos que emplea, como es el caso particular del lenguaje de programación AWK, el cual permite procesar, modificar y reorganizar datos o ficheros por medio de líneas de comandos resumidos en scripts. Adicionalmente, en un mismo script, se puede interactuar con muchos programas de genética de poblaciones, inclusive con el programa estadístico R (R Core Team, 2014).

En genética de poblaciones existen numerosos programas que permiten calcular los estadísticos básicos que describen la diversidad genética dentro y entre poblaciones, así como otros análisis más elaborados o especializados. En la Tabla 3.1 se resumen los principales programas empleados en los capítulos 5 y 6 de esta tesis. A continuación se realiza una descripción de estos programas con algunas de sus principales propiedades.

ADMIXTURE. Fast ancestry estimation

Es un software que permite la estimación por máxima verosimilitud de las ascendencias individuales a partir de un conjunto de datos de SNP. Utiliza el mismo modelo estadístico de STRUCTURE pero las estimaciones se calculan utilizando un algoritmo de optimización numérica mucho más rápido. Los archivos de entrada son generalmente bajo el formato de PLINK (.bed, .ped) y se requiere además, una idea de K, es decir el número

de poblaciones ancestrales posibles. Para más información, ver <https://www.genetics.ucla.edu/software/admixture/>.

Tabla 3.1 Principales programas empleados en análisis poblacionales

Nombre	Versión	Interfaz gráfica	Plataforma	Tipo de datos aceptados	Referencia
ADMIXTURE	1.23	No	Linux y MacOS	SNP	(Alexander et al., 2009)
ARLEQUIN	3.5.1.3	Si	Win, versión ARLECORE para Linux y MacOS	STR, SNP, RFLP, AFLP, FREQ	(Excoffier y Lischer, 2010)
BEAGLE	4.0	No	Win, Linux y MacOS (bajo interprete Java)	SNP	(Browning y Browning, 2007)
PGDSpider	2.0.7.4	Si	Win, Linux y MacOS (bajo interprete Java)	RNA, STR, SNP, RFLP, AFLP	(Lischer y Excoffier, 2012)
PHYLIP	3.696	No	Win, Linux y MacOS	RNA, STR, SNP, FREQ, DIST	(Felsenstein, 1989)
PLINK	1.07	No	MS-DOS, Linux y MacOS	SNP, STR	(Purcell et al., 2007)

AFLP, amplificación de fragmentos polimórficos; DIST, matriz de distancias; FREQ, datos de frecuencias; RFLP, polimorfismo de longitud de fragmentos de restricción; SNP, polimorfismos de un solo nucleótido; STR, microsatélites o repeticiones simples en tándem; Win, Windows.

Tabla modificada según Excoffier y Heckel (2006).

ARLEQUIN. An integrated software for population genetics data analysis

Arlequin permite estimar un numeroso conjunto de métodos básicos y pruebas estadísticas en genética de poblaciones, entre las cuales se destacan: los índices de diversidad genética, los estadísticos F, las distancias genéticas entre poblaciones, la prueba exacta del equilibrio de Hardy-Weinberg, el desequilibrio de ligamiento, el análisis de la estructura genética (AMOVA), entre otros. Arlequin posee un formato específico con extensión (.arp). Más información puede encontrarse en <http://cmpg.unibe.ch/software/arlequin35/>.

Beagle

El software Beagle es utilizado para deducir las fases haplotípicas de individuos no relacionados, parejas de padre e hijos y tríos de padre-madre-hijo. Además, permite inferir el genotipo de datos faltantes esporádicos e imputar marcadores no genotipados obtenidos

en un panel de referencia. Beagle tiene un formato propio, donde las filas son variables y las columnas son individuos. Para más información ver <http://faculty.washington.edu/browning/beagle/beagle.html>.

PGDSpider

PGDSpider es una herramienta de conversión automática de datos de gran alcance para programas de genética y genómica poblacional que permite conectar diferentes programas. Facilita las posibilidades de intercambio de datos entre numerosos programas en una amplia gama de tipos de datos. PGDSpider se basa en un formato XML (lenguaje utilizado para almacenar datos de forma legible) y, por tanto, es independiente de cualquier sistema operativo en particular. Información complementaria puede hallarse en <http://www.cmpg.unibe.ch/software/PGDSpider/>

PHYLIP. The PHYLogeny Inference Package

PHYLIP es un paquete de programas que permite inferir filogenias (árboles evolutivos). Los métodos que están disponibles en el paquete incluyen matriz de distancias genéticas y métodos de similitud, incluyendo el *bootstrapping* y los árboles de consenso. Los programas se controlan a través de un menú que permite modificar las diferentes opciones según el análisis que se requieran. Los datos pueden ser ingresados al programa a partir de un archivo de texto, elaborado a partir de cualquier procesador o editor de texto. Más información puede hallarse en <http://evolution.genetics.washington.edu/phylip.html>.

PLINK. Whole genome association analysis toolset

PLINK es un conjunto de herramientas, de código abierto, que permiten realizar análisis de del genoma completo. Está diseñado para llevar a cabo una serie de análisis a gran escala de una manera eficiente computacionalmente. PLINK ofrece un sin número de funciones, entre las cuales destacan:

- Recodificar, combinar y reordenar datos o archivos,
- Extraer un subconjunto de datos (SNP o individuos),
- Determinar frecuencias de los genotipos,
- Estimar el desequilibrio de ligamiento,
- Análisis de escalamiento multidimensional para visualizar estructura,
- Pruebas de asociación, entre muchos otros más.

Al igual que la mayoría de los programas, PLINK también maneja un formato propio con extensión (.ped y .map). Más información en el manual del programa, <http://pngu.mgh.harvard.edu/~purcell/plink/>.

3.3.2.3. *Medidas de estructura poblacional, diversidad genética y divergencia*

La variación encontrada en las secuencias de DNA puede ser medida y comparada a través de las frecuencias alélicas y genotípicas de los individuos que componen las poblaciones (Hedrick, 2010).

Existen numerosos métodos para estimar la diversidad y estructura genética en una población. Generalmente, las diferentes medidas de estructura genética están relacionadas entre sí, y simplemente se basan en analizar las diferencias en las frecuencias alélicas. A continuación se describen algunas herramientas utilizadas para caracterizar la diversidad genética de una población.

Equilibrio de Hardy–Weinberg

El equilibrio de Hardy-Weinberg explica que en una población panmíctica (es decir, donde los individuos se aparean al azar), de gran tamaño y donde todos los individuos son igualmente viables, las frecuencias génicas y genotípicas se mantienen constantes, es decir en equilibrio. El equilibrio de Hardy-Weinberg significa que la reproducción sexual no causa una reducción constante de la variación genética en cada generación; por el contrario, la cantidad de variación permanece constante generación tras generación, en ausencia de otras fuerzas evolutivas (Griffiths et al., 2007).

La relación en general entre frecuencias alélicas y genotípicas puede describirse en términos algebraicos. Si p es la frecuencia del alelo A , q es la frecuencia del alelo a , se cumple que $p + q = 1$ si no existe más que esos dos alelos. Las frecuencias genotípicas vienen dadas por:

$$\begin{array}{ccc} A/A & A/a & a/a \\ p^2 & 2pq & q^2 \end{array} \quad [11]$$

La comprobación del equilibrio de Hardy-Weinberg se realiza por medio de la comparación de los datos observados y esperados mediante un test estadístico de χ^2 (Hartl y Clark, 2007), basado en el valor:

$$\chi^2 = \sum \frac{(obs - esp)^2}{esp} \quad [12]$$

donde *obs* y *esp* indican el número de individuos observados y esperados de cada clase genotípica.

Heterocigosis

Una medida de variación genética es la cantidad de heterocigosis para un gen en la población, que viene dada por la frecuencia de individuos heterocigotos para ese gen. Normalmente, la variación en los genes es necesaria para permitir a los organismos adaptarse a las cambiantes condiciones ambientales. Sin embargo, es la variación alélica la que es crítica. La frecuencia de la aparición de un alelo cambia regularmente como resultado de mutaciones, deriva genética y selección, mencionados anteriormente.

Esta heterocigosis puede ser directamente observada contando los individuos heterocigotos en la población (heterocigosis observada, H_o), o se puede calcular a partir de las frecuencias alélicas usando las proporciones del equilibrio de Hardy-Weinberg (heterocigosis esperada, H_e). Las desviaciones de H_o a partir de la H_e es un indicador de la dinámica poblacional. Si un alelo se encuentra a una frecuencia muy elevada, y todos los demás tienen frecuencias próximas a cero, habrá muy poca heterocigosis, porque muchos individuos serán homocigotos para el alelo más común. Se espera que la heterocigosis sea máxima cuando un gen tiene múltiples alelos y todos ellos tengan la misma frecuencia (Griffiths et al., 2007).

Índices de fijación

Los índices de fijación, o también conocidos como estadísticos F (Wright, 1951, 1965), consisten en una aproximación de tres diferentes coeficientes utilizados para asignar la variabilidad genética a nivel de la población total (T), las subpoblaciones (S), y los individuos (I). Estos tres coeficientes, F_{ST} , F_{IT} , y F_{IS} , se interrelacionan de manera que:

$$1 - F_{IT} = (1 - F_{ST})(1 - F_{IS}) \quad [13]$$

$$F_{ST} = \frac{F_{IT} - F_{IS}}{1 - F_{IS}} \quad [14]$$

El F_{ST} es una medida de la diferenciación genética entre las diferentes subpoblaciones, su valor varía entre 0 y 1. Se interpreta como la reducción de la heterocigosis de las subpoblaciones respecto a la población total. El F_{IS} y el F_{IT} son medidas de desviación de las proporciones de Hardy-Weinberg dentro de las subpoblaciones y en la población total, respectivamente, donde los valores positivos

indican una deficiencia de heterocigotos y los valores negativos indican un exceso de heterocigotos. El F_{IS} se interpreta como la reducción de la heterocigosis de un individuo con respecto a su subpoblación y el F_{IT} como la reducción de la heterosis de un individuo con respecto a la población total.

Análisis de varianza molecular (AMOVA)

La estructura genética de la población puede ser investigada por medio de un análisis de varianza, tal como se define inicialmente por Cockerham (1973), y se extendió en otros estudios (Long, 1986; Weir y Cockerham, 1984). El Análisis de Varianza Molecular (AMOVA) (Excoffier et al., 1992) es esencialmente similar a otros enfoques basados en los análisis de varianza de las frecuencias génicas, pero toma en cuenta el número de mutaciones entre los haplotipos moleculares. El AMOVA emplea métodos de permutación que elimina la suposición de normalidad que es convencional para el análisis de la varianza pero no apropiado para los datos moleculares.

El AMOVA es un análisis jerárquico que subdivide la varianza total en componentes de covarianza debido a las diferencias dentro de individuos, entre individuos, y/o diferencias entre la población. Los componentes de covarianza pueden ser utilizados para calcular índices de fijación, definidos originalmente por Wright (1951, 1965) en términos de coeficientes de consanguinidad, y que fueron explicados en el apartado anterior.

Formalmente, suponemos que la frecuencia del vector del i -ésimo alelo del j -ésimo individuo en la k -ésima población es una ecuación lineal de la forma:

$$x_{kj(i)} = x + a_k + b_{k(j)} + c_{kj(i)} \quad [15]$$

donde $x_{kj(i)}$ es un valor entre 0 y 1 para el i -ésimo alelo del j -ésimo individuo en la k -ésima población, el vector x es el promedio. Los efectos son a para la población, b para el individuo dentro de la población, y c para los alelos dentro de los individuos de la población. La varianza molecular total (σ^2) es la suma de los componentes de covarianza debido a las diferencias entre los alelos dentro de un individuo (σ_c^2), el componente de covarianza debido a las diferencias entre los individuos dentro de una población (σ_b^2), y el componente de covarianza debido a las diferencias entre las poblaciones (σ_a^2).

La significación de los índices de fijación se contrasta utilizando un enfoque de permutación no paramétrico (Excoffier et al., 1992), que consiste en permutar alelos, individuos o poblaciones, entre los individuos o las poblaciones. Después de cada ronda de permutación, se recalculan todos los estadísticos para obtener su distribución nula.

A continuación, en la Tabla 3.2 se detalla un tipo de AMOVA que consta de tres niveles jerárquicos, donde se describe la forma en que se divide la suma de cuadrados total, como se obtienen los componentes de covarianza y los índices de fijación.

Tabla 3.2 Estimación del AMOVA con tres niveles jerárquicos

Fuente de variación	Grados de libertad	Suma de cuadrados (SSD)	Cuadrados medios esperados
Entre poblaciones	$P - 1$	SSD (AP)	$n\sigma_a^2 + 2\sigma_b^2 + \sigma_c^2$
Entre individuos dentro de poblaciones	$N - P$	SSD (AI/WP)	$2\sigma_b^2 + \sigma_c^2$
Dentro de individuos	N	SSD (WI)	σ_c^2
Total	$2N - 1$	SSD (T)	σ_T^2

A partir de estos valores, n y los índices de fijación están definidos por:

$$n = \frac{2N - \sum_{p \in P} \frac{2N_p^2}{N}}{P - 1} \quad [16]$$

$$F_{ST} = \frac{\sigma_a^2}{\sigma_T^2} \quad [17]$$

$$F_{IT} = \frac{\sigma_a^2 + \sigma_b^2}{\sigma_T^2} \quad [18]$$

$$F_{IS} = \frac{\sigma_b^2}{\sigma_b^2 + \sigma_c^2} \quad [19]$$

donde,

SSD (T): Suma total de las desviaciones al cuadrado.

SSD (AP): Suma de las desviaciones al cuadrado entre poblaciones.

SSD (WI): Suma de las desviaciones al cuadrado dentro de individuos.

SSD (AI/WP): Suma de las desviaciones al cuadrado entre individuos, dentro de poblaciones.

P : Número total de poblaciones.

N : Número total de individuos.

N_p : Número de individuos en cada población p .

Distancias genéticas

Las medidas de distancias genéticas permiten consolidar una gran cantidad de datos obtenidos a través de muchos loci en proporciones manejables, permitiendo la visualización de las relaciones generales entre las poblaciones. La medida de distancia genética es generalmente análoga a la distancia geométrica, es decir, una distancia genética de cero es equivalente a ninguna diferencia entre los grupos, y el aumento de la distancia genética se traduce en una mayor diferencia entre los grupos (Hedrick, 2010).

Hay diferentes medidas de distancia genética entre pares de poblaciones. Así por ejemplo, se puede calcular directamente la F_{ST} entre dos poblaciones y usarla como distancia genética. Sin embargo, la más comúnmente usada es la distancia genética estándar de Nei, la cual pretende estimar el número de mutaciones que a nivel nucleotídico se han acumulado en las secuencias de dos linajes a partir de tiempo que ha transcurrido desde su divergencia (Nei, 1972). El promedio en el número neto de sustituciones nucleotídicas, D , estaría dado por:

$$D = -\ln(I) \quad [20]$$

donde I es la identidad genética, que es igual a:

$$I = \frac{J_{xy}}{(J_x J_y)^{1/2}} \quad [21]$$

donde

$$J_{xy} = \sum_{i=1}^n p_{i \cdot x} p_{i \cdot y} \quad J_x = \sum_{i=1}^n p_{i \cdot x}^2 \quad J_y = \sum_{i=1}^n p_{i \cdot y}^2 \quad [22]$$

y donde $p_{i \cdot x}$ y $p_{i \cdot y}$ son las frecuencias del i -ésimo alelo en las poblaciones x e y .

I tiene un rango que va desde cero, cuando no hay alelos compartidos entre las dos poblaciones, hasta uno, cuando las dos poblaciones tienen frecuencias alélicas idénticas. Y D tiene un rango desde cero para poblaciones idénticas, hasta infinito cuando las dos poblaciones no comparten ningún alelo.

Aislamiento por distancia

Para determinar si una especie sigue un modelo de aislamiento por distancia se puede determinar si existe una correlación entre las distancias genéticas y geográficas (Wright, 1943). Esta comparación se puede realizar por medio de una prueba de Mantel, la cual consiste en un análisis de correlación entre dos matrices del mismo rango (Mantel, 1967),

en este caso empleada para comparar una matriz de distancias genéticas y una matriz de distancias geográficas entre los mismos individuos o poblaciones.

Métodos de agrupamiento en árboles filogenéticos

Los árboles filogenéticos son representaciones gráficas que consisten en nodos (unidades taxonómicas) y ramas (vías que conectan los nodos) que resumen las relaciones evolutivas entre los organismos. Los árboles filogenéticos pueden ser graficados de múltiples maneras. Un árbol es a escala cuando las longitudes de las ramas se dibujan de manera proporcional al número de cambios genéticos; de lo contrario, es sin escala. Un árbol es con raíz cuando un nodo interno es especificado como aquel que representa el ancestro común de todos, de lo contrario, es un árbol sin raíz y se observa como una red (Avisé, 2004).

Las distancias genéticas entre pares de poblaciones pueden ser usadas para reconstruir las genealogías de poblaciones mediante métodos de agrupamiento como el “*Unweighted Pair Group Method with Arithmetic Averages*” (UPGMA) (Sneath y Sokal, 1973), o el método de “*Neighbor-joining*” (Saitou y Nei, 1987). Ambos métodos consisten en unir inicialmente los dos “vecinos” que tengan la menor distancia genética. Luego, este par inicial se considera como una sola entidad, y se busca el siguiente “vecino” que tenga la menor distancia genética con éste. El procedimiento continúa hasta unir todos los terminales al árbol filogenético. El método de *Neighbor-joining* se diferencia del UPGMA en que utiliza una matriz de distancias corregida, donde la idea es cambiar la noción de “vecino”. Vecinos serán aquellos que están próximos pero también alejados de los demás (Hartl y Clark, 2007).

En los árboles filogenéticos se recomienda incluir niveles de confianza acerca de las clasificaciones reveladas en una reconstrucción filogenética. Una aproximación común es la técnica del *bootstrapping* (Felsenstein, 1985). Se trata de la creación de un nuevo conjunto de datos mediante el muestreo de N caracteres al azar con reemplazo de datos, de modo que el conjunto de datos resultante tiene el mismo tamaño que el original, pero algunos caracteres han quedado fuera y otros están duplicados. La variación aleatoria de los resultados de los análisis de estos conjuntos de datos re-muestreados puede demostrar estadísticamente ser típico de la variación que se podrían obtener de la recolección de nuevos conjuntos de datos. El resultado final se muestra gráficamente como un número

junto a cada nodo que indica el porcentaje de veces que ese clúster está presente entre los árboles remuestreados (Hartl y Clark, 2007).

3.3.2.4. Herramientas multivariantes en el análisis de datos moleculares

La gran cantidad de datos obtenidos a partir de marcadores moleculares tipo SNP requiere la implementación de análisis multivariados para la interpretación de los datos de una manera eficiente. La principal ventaja de estas metodologías es la reducción de la dimensionalidad de los datos en nuevas variables sintéticas (Jombart et al., 2009). Existen múltiples metodologías multivariantes aplicadas a la información genética y a continuación se explican algunas de ellas.

Análisis de componentes principales y escalamiento multidimensional

El análisis de componentes principales (PCA usando el acrónimo inglés) se ha convertido en una herramienta estándar para explorar y visualizar la estructura genética de las poblaciones. Cavalli-Sforza y sus colaboradores fueron los pioneros en el uso del análisis de componentes principales (PCA) en la genética de poblaciones, empleando mapas que resumen la variación genética humana en todas las regiones continentales (Menozzi et al., 1978). El método de PCA por lo general se aplica en los datos a nivel de población, sin embargo también puede ser aplicado a los individuos. Este método no intenta clasificar a todos los individuos en poblaciones discretas o combinaciones lineales de las poblaciones. En lugar de ello, el PCA emite coordenadas de cada individuo a lo largo de ejes de variación (Patterson et al., 2006).

Por su parte, el método de escalamiento multidimensional (MDS en siglas inglesas) es una técnica de reducción de dimensión con la que se busca representar todos los individuos en unas pocas dimensiones. El objetivo de escalamiento multidimensional es un gráfico de dispersión en el que cada punto es un individuo donde se muestra información acerca de cómo los elementos se relacionan entre sí, y que puede ser útil para identificar cualquier agrupación o si proporciona alguna otra interpretación válida de los datos (Rencher, 2002).

Ambos métodos pueden emplear matrices de distancias genéticas observadas entre individuos. Estas distancias se pueden generar a partir de:

$$1 - IBS \quad [23]$$

donde *IBS* representa la identidad por estado (*Identity by State*, IBS).

La matriz de *IBS* es un procedimiento en el cual se examinan pares de SNP entre dos individuos y se determina la proporción media de alelos compartidos. Los valores de esta matriz oscilan, en teoría, entre 0 y 1. En la práctica, no se esperaría observar valores cercanos 0, incluso entre individuos no emparentados, de los cuales se esperaría que compartan una gran proporción del genoma. Un valor de 1 indicaría un par de gemelos, o una duplicación de la muestra (Purcell et al., 2007).

Métodos de agrupamiento y niveles de mixtura

El análisis de clústeres se ha consolidado como una de las técnicas más utilizadas para evaluar la estructura de la población y, junto con el PCA y el MDS, presentan un resumen de las características de la población a partir de los datos genéticos. Estos métodos permiten llevar a cabo una reconstrucción de acontecimientos históricos, por medio de la estimación de ancestrías y niveles de mixtura. En algunos casos, la demarcación y estructuración de las poblaciones es fácil de deducir a partir de la ubicación geográfica de los individuos, en otros casos, la subdivisión de las poblaciones en subpoblaciones reproductoras locales menos definidas, puede llegar a ser problemática (Manel et al., 2005).

En estos casos, diferentes métodos de agrupamiento se han implementado (Alexander et al., 2009; Corander et al., 2003; Falush et al., 2003; Pritchard et al., 2000). El método de Pritchard y sus colaboradores (STRUCTURE, Pritchard et al., 2000) utiliza un marco bayesiano para inferir el número de grupos (poblaciones) mediante la comparación de la probabilidad posterior para diferentes números de poblaciones especificadas por el usuario. STRUCTURE proporciona un medio para identificar el "mejor" valor de K (número de poblaciones ancestrales), basado en computar la evidencia del modelo para cada valor posible K a través de probabilidades posteriores vía MCMC (*Markov Chain Monte Carlo*). Por su parte, ADMIXTURE (Alexander et al., 2009), no intenta estimar la evidencia del modelo. Más bien, incluye un procedimiento de validación cruzada que permite al usuario identificar el valor de K para el que el modelo tiene mejor exactitud de predicción.

3.3.3. Algunas aplicaciones de los Análisis Poblacionales

La selección genómica es una nueva tecnología que está revolucionando la cría de ganado. Esta se propuso como estrategia para predecir los valores genéticos a partir de un conjunto

masivo de marcadores SNP en todo el genoma (Meuwissen et al., 2001). Aunque la idea de utilizar marcadores de DNA para mejorar la tasa de ganancia genética en el ganado lechero ha existido por décadas (Rothschild y Soller, 1997; Soller y Beckmann, 1983), la adopción de la selección asistida por marcadores se ha limitado a la última década.

Para que la selección genómica funcione, los SNP deben estar en desequilibrio de ligamiento (LD) con un QTL, de tal manera que los marcadores predigan los efectos de los QTL a través de la población (de Roos et al., 2008). Adicionalmente, el LD entre marcador y el QTL, y el alcance del LD, deben ser consistentes entre la población utilizada para la estimación (población de referencia) y la población en la que la selección genómica será practicada. Esta relación entre poblaciones se puede obtener por medio de la estimación de la persistencia de fases del LD entre poblaciones (Dekkers y Hospital, 2002). A continuación se detallan estas dos metodologías.

3.3.3.1. Desequilibrio de ligamiento

El desequilibrio de ligamiento (LD, *linkage disequilibrium*) es la propiedad que tienen algunos genes de no segregar de forma independiente, es decir, la asociación no aleatoria de alelos de diferentes loci. Puede definirse también como la diferencia entre la frecuencia observada de un gameto y la que debería tener si estuviera en equilibrio (Griffiths et al., 2007).

El LD se puede ver afectado por numerosas causas. La selección, por ejemplo, puede generar desequilibrio de ligamiento debido a que altera las frecuencias de los gametos. Inclusive, si la selección es lo suficientemente fuerte y reciente, señales de selección pueden ser detectadas por medio del LD (Hartl y Clark, 2007). Igualmente, el cruce de dos líneas o razas con diferentes frecuencias alélicas, pese a que éstas se encuentren en equilibrio gamético, generarán un mayor LD en la población resultante (Dekkers y Hospital, 2002). Otras causas asociadas al incremento del LD son la mutación y la migración reciente.

Cuando dos marcadores moleculares se encuentran a poca distancia en el mismo cromosoma se dice que están ligados. Por esta razón, se espera que el LD sea alto a poca distancia genética y que decrezca a medida que esta distancia se incrementa, debido principalmente al efecto de la recombinación (Hartl y Clark, 2007). El descenso del LD asociado a la distancia entre marcadores ha sido evaluado en numerosas razas de ganado de carne y leche, inclusive en razas compuestas (de Roos et al., 2008; Espigolan et al.,

2013; Gautier et al., 2007; Salomon-Torres et al., 2014; Villa-Angulo et al., 2009). Estos trabajos han indicado que los marcadores moleculares son capaces de detectar niveles de LD entre SNP cercanos y hasta varias decenas de centimorgans.

El nivel de LD entre marcadores se puede cuantificar por medio del parámetro r^2 (Hill y Robertson, 1968), representado como:

$$r^2 = \frac{(p_{11} - p_1 q_1)^2}{(p_1 q_1 p_2 q_2)} \quad [24]$$

donde p_1 y p_2 son las frecuencias de los alelos menor y mayor en el SNP1 respectivamente, q_1 y q_2 son las frecuencias de los alelos menor y mayor en el SNP2 respectivamente, y p_{11} corresponde a la frecuencia observada entre los dos alelos menores en el mismo individuo a lo largo de toda la población.

3.3.3.2. Persistencia de fases

Como se mencionó anteriormente, la persistencia de fases tiene implicaciones importantes en la implementación de la selección genómica, ya que busca la similitud en los patrones de LD a través de diferentes poblaciones. A través de la persistencia de fases se puede determinar un panel de SNP adecuado para aumentar la precisión de la selección genómica. Si la persistencia de fases entre las dos poblaciones de interés es baja, entonces la combinación de éstas poblaciones para formar un panel de SNP tendría un impacto negativo sobre la precisión resultante de la selección genómica (de Roos et al., 2009).

La persistencia de fases entre dos pares de SNP en dos poblaciones puede ser estimada como un coeficiente de correlación entre alelos, es decir, la raíz cuadrada de r^2 ($\sqrt{r^2}$), o partiendo de la fórmula [24] como: $r = p_{11} - p_1 q_1 / \sqrt{p_1 q_1 p_2 q_2}$. Esta última aproximación permite utilizar el signo del numerador para calcular los coeficientes de correlación entre los marcadores de dos poblaciones.

Diferentes estudios muestran que la persistencia de fases entre razas es alta en distancias cortas, pero disminuye rápidamente a medida que la distancia aumenta (de Roos et al., 2008; Gautier et al., 2007). El nivel de decaimiento de esta persistencia, o bien llamada pendiente, tiene una fuerte implicación en la determinación del tiempo en el que posiblemente las razas divergieron. Esta divergencia puede ser estimada mediante la siguiente aproximación: $r = e^{-2cT}$, donde r es la persistencia de fases, c es la distancia de recombinación entre los marcadores (en Morgans) y T es el tiempo en que las razas divergieron medido en generaciones (de Roos et al., 2008).

3.3.3.3. *Tamaños efectivos ancestrales de la población*

Anteriormente se había hecho mención del cálculo de los tamaños efectivos (N_e) a partir del incremento de la consanguinidad (ΔF) obtenida a partir del análisis genealógico. Sin embargo, esta estimación tiene la desventaja de que no es posible estimar muchas generaciones atrás debido a las limitaciones del pedigrí. En la actualidad, gracias a la alta densidad de marcadores genéticos tipo SNP, se puede conocer la historia de las poblaciones e inferir el tamaño de la población muchas generaciones atrás. De esta manera, el N_e ancestral proporciona información útil sobre la evolución de una población, permite mejorar los modelos de mapeo de una enfermedad y se toma como referencia para el seguimiento de la pérdida de la variabilidad genética (Hayes et al., 2003; Tenesa et al., 2007).

El LD se puede emplear para estimar el N_e ancestral si se conoce la tasa de recombinación por medio de la siguiente aproximación (Sved, 1971):

$$r^2 = 1/(1 + 4N_e c) \approx N_e = (1/4c) * (1/r^2 - 1) \quad [25]$$

donde N_e es el tamaño efectivo de la población, r^2 es el desequilibrio de ligamiento, y c es la tasa de recombinación que se sustituye por la distancia de mapa en Morgans.

Para estimar N_e ancestral en diferentes generaciones del pasado se utiliza la siguiente aproximación (Hayes et al., 2003):

$$T = 1/2c \quad [26]$$

donde T representa las diferentes generaciones en el pasado y c es la tasa de recombinación mencionada anteriormente.

Pese a que generalmente se tiene a aplicar la generalización 1 Mb ~ 1 cM para calcular N_e ancestral, se pueden estar cometiendo errores de estimación, debido a que la tasa de recombinación en bovinos varía considerablemente entre cromosomas (Arias et al., 2009). Por esta razón se ha sugerido emplear tasas de recombinación específicas para cada cromosoma (Corbin et al., 2010; Qanbari et al., 2010; Tenesa et al., 2007; Villa-Angulo et al., 2009).

Adicionalmente, el N_e ancestral se puede estimar por medio de un modelo no lineal de mínimos cuadrados aplicando la fórmula [25], detallado de la siguiente manera:

$$r^2 = 1/(a + 4bd) + e \approx b = (1/4d) * (1/r^2 - a) \quad [27]$$

donde r^2 es el desequilibrio de ligamiento, d es la distancia en Morgans, y los parámetros a estimar de forma iterativa por mínimos cuadrados son a y b , que corresponden a la tasa de mutación y el tamaño efectivo de la población, respectivamente (Corbin et al., 2010; Tenesa et al., 2007).

Bibliografía

- ACRUGA. (2011). Programa de mejora de la raza bovina Rubia Gallega. Departamento de Anatomía, Embriología y Genética. Facultad de Veterinaria. Universidad de Zaragoza.
- Albrechtsen, A., Nielsen, F.C. y Nielsen, R. (2010). Ascertainment biases in SNP chips affect measures of population divergence. *Molecular Biology and Evolution*, 27, 2534–2547.
- Alexander, D.H., Novembre, J. y Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19, 1655–1664.
- Aljama-Gutiérrez, P. (1982). *La Raza Retinta*. Córdoba: Monte de Piedad y Caja de Ahorros de Córdoba.
- Anne, C. (2006). Choosing the right molecular genetic markers for studying biodiversity: From molecular evolution to practical aspects. *Genetica*, 127, 101–120.
- Aranguren, F.J. (2002). *La raza Pirenaica*. Pamplona: ASPINA-ITG Ganadero.
- Arias, J.A., Keehan, M., Fisher, P., Coppieters, W. y Spelman, R. (2009). A high density linkage map of the bovine genome. *BMC Genetics*, 10, 18.
- Avise, J.C. (2004). *Molecular markers, natural history and evolution* (2nd ed.). Sunderland: Sinauer Associates.
- Baruch, E. y Weller, J.I. (2008). Estimation of the number of SNP genetic markers required for parentage verification. *Animal Genetics*, 39, 474–479.
- Boichard, D., Maignel, L. y Verrier, É. (1996). Analyse généalogique des races bovines laitières françaises. *INRA Prod. Anim*, 9, 323–335.
- Boichard, D., Maignel, L., y Verrier, É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution*, 29, 5–23.
- Bradley, D.G., MacHugh, D.E., Cunningham, P. y Loftus, R.T. (1996). Mitochondrial diversity and the origins of African and European cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 93, 5131–5135.

- Browning, S.R. y Browning, B.L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, 81, 1084–1097.
- Caballero, A. (1994). Developments in the prediction of effective population size. *Heredity*, 73, 657–679.
- Cañas-Álvarez, J.J., González-Rodríguez, A., Martín-Collado, D., Avilés, C., Altarriba, J., Baro, J.A., De la Fuente, L.F., Díaz, C., Molina, A., Varona, L. y Piedrafita, J. (2014). Monitoring changes in the demographic and genealogical structure of the main Spanish local beef breeds. *Journal of Animal Science*, 92, 4364–4374.
- Cassell, B.G., Adamec, V. y Pearson, R.E. (2003). Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. *Journal of Dairy Science*, 86, 2967–2976.
- Cervantes, I., Goyache, F., Molina, A., Valera, M. y Gutiérrez, J.P. (2008). Application of individual increase in inbreeding to estimate realized effective sizes from real pedigrees. *Journal of Animal Breeding and Genetics*, 125, 301–310.
- Cima-García, M. (1986). *Biotipología de las razas bovinas Asturiana de los Valles y Asturiana de la Montaña: definición y calificación racial de los animales por el tipo*. Oviedo: Principado de Asturias. Consejería de Agricultura y Pesca.
- Cima-García, M. (1996). *El ganado vacuno de la raza Asturiana de los Valles: pasado, presente y futuro*. Oviedo: ASEAVA.
- Clark, A.G., Hubisz, M.J., Bustamante, C.D., Williamson, S.H. y Nielsen, R. (2005). Ascertainment bias in studies of human genome-wide polymorphism. *Genome Research*, 15, 1496–1502.
- Cockerham, C.C. (1973). Analysis of gene frequencies. *Genetics*, 74, 679–700.
- Corander, J., Waldmann, P. y Sillanpaa, M.J. (2003). Bayesian analysis of genetic differentiation between populations. *Genetics*, 163, 367–374.
- Corbin, L.J., Blott, S.C., Swinburne, J.E., Vaudin, M., Bishop, S.C. y Woolliams, J.A. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics*, 41(Suppl. 2), 8–15.
- de Roos, A.P.W., Hayes, B.J., Spelman, R.J. y Goddard, M.E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, 179, 1503–1512.
- de Roos, A.P.W., Hayes, B.J. y Goddard, M.E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, 183, 1545–1553.
- Dekkers, J.C.M. y Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews. Genetics*, 3, 22–32.

- Espigolan, R., Baldi, F., Boligon, A.A., Souza, F.R., Gordo, D.G., Tonussi, R.L., Cardoso, D.F., Oliveira, H.N., Tonhati, H., Sargolzaei, M., Schenkel, F.S., Carneiro, R., Ferro, J.A. y Albuquerque, L.G. (2013). Study of whole genome linkage disequilibrium in Nelore cattle. *BMC Genomics*, 14, 305.
- Excoffier, L., Smouse, P.E. y Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.
- Excoffier, L. y Heckel, G. (2006). Computer programs for population genetics data analysis: a survival guide. *Nature Reviews. Genetics*, 7, 745–758.
- Excoffier, L. y Lischer, H.E.L. (2010). Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources*, 10, 564–567.
- Falconer, D.S. y Mackay, T.F.C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow: Addison Wesley Longman Limited.
- Falush, D., Stephens, M. y Pritchard, J.K. (2003). Inference of population structure using multilocus genotype data: Linked loci and correlated allele frequencies. *Genetics*, 164, 1567–1587.
- Felsenstein, J. (1985). Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*, 39, 783–791.
- Felsenstein, J. (1989). PHYLIP - Phylogeny inference package (Version 3.2). *Cladistics*, 5, 164–166.
- Fina, M. (2013). *Optimització de l'avaluació genètica de la raça bovina Bruna dels Pirineus*. (Tesis Doctoral). Universitat Autònoma de Barcelona, Departament de Ciència Animal i dels Aliments. Barcelona, España.
- Fisher, R.A. (1930). *The genetical theory of natural selection*. Oxford: Clarendon Press.
- Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J.G., Boichard, D., Lathrop, G.M., Gut, I.G. y Eggen, A. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*, 177, 1059–1070.
- González-Recio, O., López de Maturana, E. y Gutiérrez, J.P. (2007). Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *Journal of Dairy Science*, 90, 5744–5752.
- Goyache, F., Gutiérrez, J.P., Fernández, I., Gomez, E., Alvarez, I., Díez, J. y Royo, L.J. (2003). Using pedigree information to monitor genetic variability of endangered populations: The Xalda sheep breed of Asturias as an example. *Journal of Animal Breeding and Genetics*, 120, 95–105.
- Griffiths, A.J.F., Wessler, S.R., Lewontin, R.C. y Carroll, S.B. (2007). *Introduction to genetic analysis* (9th ed.). Freeman, W. H. y Company.

- Groeneveld, E., Westhuizen, B.D., Maiwashe, A., Voordewind, F. y Ferraz, J.B.S. (2009). POPREP: a generic report for population management. *Genetics and Molecular Research*, 8, 1158–1178.
- Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J. y Piedrafita, J. (2003). Pedigree analysis of eight Spanish beef cattle breeds. *Genetics Selection Evolution*, 35, 43–63.
- Gutiérrez, J. P., Cervantes, I., Molina, A., Valera, M. y Goyache, F. (2008). Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genetics Selection Evolution*, 40, 359–378.
- Hartl, D.L. y Clark, A.G. (2007). *Principles of population genetics* (4th ed.). Sunderland: Sinauer Associates.
- Hayes, B.J., Visscher, P.M., McPartlan, H.C. y Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13, 635–643.
- Hedrick, P.W. (2010). *Genetics of populations* (4th ed.). Tempe, Arizona: Jones and Bartlett Publishers.
- Hedrick, P.W. y Miller, P.S. (1992). Conservation genetics: techniques and fundamentals. *Ecological Applications*, 2, 30–46.
- Helmer, D., Gourichon, L., Monchot, H., Peters, J. y Saña Seguí, M. (2005). Identifying early domestic cattle from Pre-Pottery Neolithic sites on the Middle Euphrates using sexual dimorphism. En J.D. Vigne, D. Helmer, y J. Peters (Eds.), *The First Steps of Animal Domestication: New Archaeozoological Approaches* (pp. 86–95). London: Oxbow Books.
- Helyar, S.J., Hemmer-Hansen, J., Bekkevold, D., Taylor, M.I., Ogden, R., Limborg, M.T., Cariani, A., Maes, G.E. Diopere, E., Carvalho, G.R. y Nielsen, E.E. (2011). Application of SNPs for population genetics of nonmodel organisms: New opportunities and challenges. *Molecular Ecology Resources*, 11, 123–136.
- Hill, W.G. y Robertson, A. (1968). Linkage disequilibrium in finite populations. *Theoretical and Applied Genetics*, 38, 226–231.
- Illumina Inc. (2012). BovineHD Genotyping BeadChip Datasheet. Recuperado de http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf
- Hongo, H., Pearson, J., Öksüz, B. y Ilgezdi, G. (2009). The process of ungulate domestication at Cayönü, Southeastern Turkey: a multidisciplinary approach focusing on *Bos sp.* and *Cervus elaphus*. *Anthropozoologica*, 44, 63–78.
- International Human Genome Sequencing Consortium. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860–921.
- James, J.W. (1977). A note on selection differentials and generation length when generations overlap. *Animal Production*, 24, 109–112.

- Jombart, T., Pontier, D. y Dufour, A.B. (2009). Genetic markers in the playground of multivariate analysis. *Heredity*, 102, 330–341.
- Jordana, J., Pelegrín, M. y Piedrafita, J. (1991). Relaciones genéticas en bovinos españoles obtenidas a partir del estudio de caracteres morfológicos. *Información Técnica Económica Agraria*, 87A, 50–64.
- Jordana J. y Piedrafita J. (1993). Programa de mejora genética de la agrupación racial bovina “Bruna dels Pirineus”. En *V Reunión Nacional de Mejora Genética Animal*. Córdoba: Consejería de Agricultura y Pesca. Junta de Andalucía.
- Kalinowski, S.T. (2002). How many alleles per locus should be used to estimate genetic distances? *Heredity*, 88, 62–65.
- Lacy, R.C. (1989). Analysis of founder representation in pedigrees. Founder equivalents and founder genome equivalents. *Zoo Biology*, 8, 111–123.
- Lischer, H.E.L. y Excoffier, L. (2012). PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. *Bioinformatics*, 28, 298–299.
- Liu, N., Chen, L., Wang, S., Oh, C. y Zhao, H. (2005). Comparison of single-nucleotide polymorphisms and microsatellites in inference of population structure. *BMC Genetics*, 6(Suppl 1), S26.
- Loftus, R.T., MacHugh, D.E., Bradley, D.G., Sharp, P.M. y Cunningham, P. (1994). Evidence for two independent domestications of cattle. *Proceedings of the National Academy of Sciences of the United States of America*, 91, 2757–2761.
- Long, J.C. (1986). The allelic correlation structure of Gainj- and Kalam-speaking people. I. The estimation and interpretation of Wright’s F-statistics. *Genetics*, 112, 629–647.
- Malécot, G. (1948). *Les mathématiques de l'hérédité*. Paris: Masson & Cie.
- MacCluer, J.W., Boyce, A.J., Dyke, B., Weitkamp, L.R., Pfennig, D.W. y Parsons, C.J. (1983). Inbreeding and pedigree structure in Standardbred horses. *Journal of Heredity*, 74, 394–399.
- Manel, S., Gaggiotti, O.E. y Waples, R.S. (2005). Assignment methods: Matching biological questions with appropriate techniques. *Trends in Ecology and Evolution*, 20, 136–142.
- Mantel, N. (1967). The detection of disease clustering and a generalized regression approach. *Cancer Research*, 27, 209–220.
- Mendizábal-Aizpuru, J.A., Ibarbia-Barreras, J.R. y Etxaniz-Makazaga, J.M. (2005). Aportaciones a la historia de la raza vacuna pirenaica. Paradigma de la zootecnia española. *Archivos de zootecnia*, 54, 39–50.
- Menozi, P., Piazza, A. y Cavalli-Sforza, L. (1978). Synthetic maps of human gene frequencies in Europeans. *Science*, 201, 786–792.

- Meuwissen, T.H.E., Hayes, B.J. y Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819–1829.
- Mujal, M.M. (1998). *Anàlisi demogràfica i genètica de la vaca bruna dels Pirineus: amb una breu introducció històrica*. (Tesis Doctoral). Universitat Autònoma de Barcelona, Departament de Ciència Animal i dels Aliments. Barcelona, España.
- Nei, M. (1972). Genetic distance between populations. *American Naturalist*, 106, 283–292.
- Patterson, N., Price, A.L. y Reich, D. (2006). Population structure and eigenanalysis. *PLoS Genetics*, 2, 2074–2093.
- Pinheiro, M., Kjöllnerström, H.J. y Oom, M.M. (2013). Genetic diversity and demographic structure of the endangered Sorraia horse breed assessed through pedigree analysis. *Livestock Science*, 152, 1–10.
- Pritchard, J.K., Stephens, M. y Donnelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, 155, 945–959.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J. y Sham, P.C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, 81, 559–575.
- Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. y Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics*, 41, 346–356.
- Rencher, A.C. (2002). *Methods of multivariate analysis* (2nd ed.). Hoboken: Wiley-Interscience.
- Robertson, A. (1953). A numerical description of breed structure. *Journal of Agricultural Science*, 43, 334–336.
- Rothschild, M.F. y Soller, M. (1997). Candidate gene analysis to detect genes controlling traits of economic importance in domestic livestock. *Probe*, 8, 13–20.
- Saitou, N. y Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, 4, 406–425.
- Salomon-Torres, R., Matukumalli, L.K., Van Tassell, C.P., Villa-Angulo, C., Gonzalez-Vizcarra, V.M. y Villa-Angulo, R. (2014). High density LD-based structural variations analysis in cattle genome. *PLoS ONE*, 9(7), e103046.
- Sánchez-Belda, A. (1983). *La raza bovina Avileña-Negra Ibérica*. Madrid: Ministerio de Agricultura, Pesca y Alimentación, Dirección General de la Producción Agraria.
- Sánchez-Belda, A. (1984). *Razas bovinas españolas*. Madrid: Publicaciones de Extensión Agraria.

- Sánchez-Belda, A. (2002). *Razas ganaderas españolas. I. Bovinas*. Madrid: FEAGAS y MAPA.
- Slatkin, M. (1994). Gene flow and population structure. En L.A. Real (Ed.), *Ecological Genetics* (pp. 3–17). Princeton: Princeton University Press.
- Sneath, P.H.A. y Sokal, R.R. (1973). *Numerical taxonomy: the principles and practice of numerical classification*. San Francisco: W. H. Freeman and Company.
- Soller, M., y Beckmann, J.S. (1983). Genetic polymorphism in varietal identification and genetic improvement. *Theoretical and Applied Genetics*, 67, 25–33.
- Stoneking, M. y Krause, J. (2011). Learning about human population history from ancient and modern genomes. *Nature Reviews. Genetics*, 12, 603–614.
- Sunnucks, P. (2000). Efficient genetic markers for population biology. *Trends in Ecology and Evolution*, 15, 199–203.
- Sved, J. A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, 2, 125–141.
- Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E. y Visscher, P.M. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Research*, 17, 520–526.
- The Bovine Genome Sequencing and Analysis Consortium. (2009). The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science*, 324, 522–528.
- The Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324, 528–532.
- The International SNP Map Working Group. (2001). A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409, 928–933.
- Troy, C.S., MacHugh, D.E., Bailey, J.F., Magee, D.A., Loftus, R.T., Cunningham, P., Chamberlain, A.T., Sykes, B.C. y Bradley, D.G. (2001). Genetic evidence for Near-Eastern origins of European cattle. *Nature*, 410, 1088–1091.
- Vassallo, J.M., Díaz, C. y García-Medina, J.R. (1986). A note on the population structure of the Avileña breed of cattle in Spain. *Livestock Production Science*, 15, 285–288.
- Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., et al. (2001). The sequence of the human genome. *Science*, 291, 1304–1351.
- Vignal, A., Milan, D., SanCristobal, M. y Eggen, A. (2002). A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, 34, 275–305.
- Vigne, J.D. (2008). Zooarchaeological aspects of the Neolithic diet transition in the Near East and Europe, and their putative relationships with the Neolithic demographic

- transition. En J.P. Bocquet-Appel y O. Bar-Yosef (Eds.), *The Neolithic demographic transition and its consequences* (pp. 179–205). New York: Springer.
- Vigne, J.D. y Helmer, D. (2007). Was milk a “secondary product” in the Old World Neolithisation process? Its role in the domestication of cattle, sheep and goats. *Anthropozoologica*, 42, 9–40.
- Villa-Angulo, R., Matukumalli, L.K., Gill, C.A., Choi, J., Van Tassell, C.P. y Grefenstette, J.J. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics*, 10, 19.
- Weir, B.S. y Cockerham, C.C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, 38, 1358–1370.
- Wright, S. (1922). Coefficients of Inbreeding and Relationship. *The American Naturalist*, 56, 330–338.
- Wright, S. (1931). Evolution in Mendelian populations. *Genetics*, 16, 97–159.
- Wright, S. (1943). Isolation by Distance. *Genetics*, 28, 114–138.
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, 15, 323–354.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19, 395–420.
- Xing, C., Schumacher, F.R., Xing, G., Lu, Q., Wang, T. y Elston, R.C. (2005). Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis. *BMC Genetics*, 6, S29.

Capítulo 4

DEMOGRAPHIC AND GENEALOGICAL STRUCTURE IN SPANISH BEEF BREEDS

Monitoring changes in the demographic and genealogical structure of the main Spanish local beef breeds¹

J. J. Cañas-Álvarez,^{*2} A. González-Rodríguez,^{†3} D. Martín-Collado,[‡] C. Avilés,[§] || J. Altarriba,[†] J. A. Baro,[#] L. F. De la Fuente,[¶] C. Díaz,[‡] A. Molina,^{||} L. Varona,[†] 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 Mejora Genética Animal, INIA, 28040-Madrid, Spain; §Departamento de Producción Animal, Universidad de Córdoba, 14071-Córdoba, Spain; ||MERAGEM, Universidad de Córdoba, 14071 Córdoba, Spain; #Departamento de Ciencias Agroforestales, Universidad de Valladolid, 34004 Palencia, Spain; and ¶Departamento de Producción Animal, Universidad de León, 24071-León, Spain

¹We are thankful for the financial support of Spanish AGL 2010-15903 grant and the breeders' societies for providing the herdbook data.

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

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

⁴Corresponding author: jesus.piedrafita@uab.es

J ANIM SCI 2014, 92:4364-4374

doi: 10.2527/jas.2013-7420

ABSTRACT: Demographic and pedigree analyses describe the structure and dynamics of livestock populations. We studied information recorded in the herdbooks of Asturiana de los Valles (AV; N = 458,806), Avileña-Negra Ibérica (ANI; N = 204,623), Bruna dels Pirineus (BP; N = 62,138), Morucha (Mo; N = 65,350), Pirenaica (Pi; N = 217,428), Retinta (Re; N = 135,300), and Rubia Gallega (RG; N = 235,511) beef breeds from their creation until 2009. All breeds have increased in the number of registered cows in recent years. In all breeds, herds do not behave as isolated entities and a high rate of exchange of breeding males between herds exists. A percentage of herds (12–52%) make some type of selection and sell bulls to other herds. There were large differences in average number of progeny per bull, ranging from 15.6 (AV) to 373.7 animals (RG, with a high incidence of AI). Generation interval estimates ranged from 4.7 (AV) to 7.6 (RG) yr in the sire pathway and from 5.95 (AV) to 7.8 (Mo) yr in the dam pathway. Density of pedigrees varied among breeds, with Pi, ANI, and Re having the more dense pedigrees, with average completeness indexes of more than 96% in the first generation and 80% when 6 generations were considered. A general increase in average inbreeding was observed in all breeds in the years analyzed. For animals born in 2009, average inbreeding coefficients ranged from 0.6 (BP) to 7.2% (Re) when all animals were considered and from 3.6 (Pi) to 17.6% (BP) when only inbred animals were considered. Due to the lack of completeness of pedigrees in most populations, inbreeding coefficients may be considered as a lower bound of the true parameters. The proportion of inbred animals tended to increase in the periods analyzed in all breeds. Differences between inbreeding and coancestry rates (except in RG) suggest the presence of population structure. Effective population size (N_e) based on the inbreeding rate estimated by regression ranged from 43 to 378 for Re and BP, whereas N_e estimates based on coancestry were greater, with a range of 100 for RG to 9,985 for BP. These facts suggest that an adequate mating policy can help to monitor inbreeding so as not to lose genetic variability. Effective number of ancestors in 2009 for 6 of the breeds ranged from 42 (RG) to 220 (AV), with BP having much a greater value, and was lower than was the effective number of founders in all breeds, suggesting the existence of bottlenecks.

Key words: beef cattle, inbreeding, local breeds, pedigree

INTRODUCTION

Breeds are dynamic resources, changing with each generation through interaction with the physical environment and according to the selection criteria of their keepers (FAO, 1998). Well-designed breeding programs can exploit the existing genetic variation of beef cattle breeds to use local available resources, produce under harsh conditions, support specific labeling of high-added-value that prevents health scandals (Tolosana et al., 2005), and contribute to maintaining rural culture diversity (Gandini and Villa, 2003). Demographic and pedigree analysis are important tools that describe the structure and dynamics of livestock populations (Vu Tien Khang, 1983). Population structure determines the development of inbreeding, whose trend is the tool most frequently used to quantify the rate of genetic drift (Wright, 1922, 1931). Pedigree analysis has been implemented for most species, such as beef and dairy cattle and sheep, among others (Boichard et al., 1997; Goyache et al., 2003; Gutiérrez et al., 2003). However, in most local breeds, the population size and selection methods are frequently subject to changes, even in short periods of time (Boichard et al., 1997). Therefore, the demographic and pedigree analysis of populations should be done periodically to assess the effect of the breeding program in the inbreeding and genetic diversity of the breed. The genetic structure of some of the Spanish breeds was studied by Gutiérrez et al. (2003), including pedigree data until the year 1995. Since then, most of the breeds have intensified performance recording and developed more or less intensive genetic selection programs based on BLUP genetic evaluation, whose impact in the breed's genetic structure is unknown. In this context, the main objective of this study is to assess the current genetic structure of 7 Spanish local beef cattle breeds by analyzing their pedigrees. This can illustrate the general behavior of local beef populations undergoing a selection process.

MATERIALS AND METHODS

Herdbook information of 7 Spanish beef breeds was analyzed from their foundation to the last year with complete records (2009). Specifically, we analyzed the following breeds and animals in the pedigree within breed: Asturiana de los Valles (AV; N = 458,806), Avileña-Negra Ibérica (ANI; N = 204,623), Bruna dels Pirineus (BP; N = 62,138), Morucha (Mo; N = 65,350), Pirenaica (Pi; N = 217,428), Retinta (Re; N = 135,300), and Rubia Gallega (RG; N = 235,511). These breeds represent 72% of the total

census of local beef breeds in Spain. All breeds started BLUP based selection programs in the last 20 yr. The breeds differ not only in production and carcass traits (Piedrafita et al., 2003) but also in meat quality traits (Gil et al., 2001). They are reared in harsh environments, including mountains to rangelands, and the degree of crossbreeding with specialized beef meat breeds is variable as well as is the extent of AI. Further details of the breeds can be found at <http://dad.fao.org>. We analyzed several genetic parameters regarding demographic structure and genetic status of the populations based on pedigree information, which are briefly described in the following sections.

Demographic Structure

Breed Censuses. Breed censuses are assessed as the number of reproductive males and females giving progeny in 3 specific years: 1998 (a year close to the previous study), 2003 (the year in which common rules for direct support schemes for farmers under the Common Agricultural Policy of the European Union were established – Council Regulation EC number 1782/2003), and 2009.

Effective Population Size. Under random mating with unequal numbers of males and females, effective population size (N_e) is calculated as $N_e = 4N_mN_f/(N_m + N_f)$, in which N_m and N_f are the number of males and females, respectively (Falconer and Mackay, 1996).

Pyramid Structure. The pyramid structure is defined by the percentage of use of own reproductive males as the criterion for classifying herds, designated as nucleus herds if they never use reproductive males born in other herds, as multipliers if they buy and sell reproductive males (Multiplier 1 or Multiplier 2 if they use or not their own bulls, respectively), and as commercial herds if they buy but do not sell bulls (Commercial 1 or Commercial 2 if they use or not their own bulls, respectively) (Vassallo et al., 1986).

Generation Intervals. Generation intervals are defined as the average age of parents at the birth of their progeny kept for reproduction (James, 1977). This parameter was computed for the 4 classic pathways (sire–son, sire–daughter, dam–son, and dam–daughter) using the animals born between 2003 and 2009.

Family Size. Family size refers to the number of offspring of an individual that become breeding individuals in the next generation (Falconer and Mackay, 1996).

Pedigree Information Quality, Inbreeding Coefficient, and Effective Population Size

Pedigree Completeness Level. The pedigree completeness level is described by computing the average of the MacCluer et al. (1983) index considering 1 to 6 generations deep. The pedigree completeness index of an individual is $I_d = 4I_{d_{pat}}I_{d_{mat}}/(I_{d_{pat}} + I_{d_{mat}})$ and $I_{d_k} = (1/d)\sum_{i=1}^d a_i$, $k = pat, mat$, in which k represents the paternal (*pat*) or maternal (*mat*) line, a_i is the proportion of known ancestors in generation i , and d is the number of generations considered in the calculation of the completeness index.

In addition, the complete generation equivalent value (Boichard et al., 1997) was estimated by averaging over the sum of $(1/2)^n$, in which n is the number of generations separating the individual from each known ancestor.

Inbreeding Coefficient. The inbreeding coefficient (F) is the probability of having 2 genes that are identical by descent (Wright, 1922).

Coancestry. The coancestry (f) is defined as the probability that a progeny of 2 parents carries 2 alleles identical by descent. Under random mating, the rate of inbreeding (ΔF) equals the rate of coancestry (Δf).

Effective Population Size. The N_e is a measure of genetic diversity within a population (Wright, 1931). The N_e was calculated using $N_e = 1/(2\Delta F)$. The rate of inbreeding per generation (ΔF) was calculated using $\Delta F = (F_n - F_{n-1})/(1 - F_{n-1})$, in which F_n and F_{n-1} are the average inbreeding of offspring and their parents, respectively (Falconer and Mackay, 1996). Animals born a number of years equal to the generation interval were considered as parents (the previous generation) of a particular cohort (animals born in the same year).

An alternative way to estimate N_e is by computing the increase of inbreeding for each individual (ΔF_i) that equals $1 - (1 - F_i)^{1/t}$, in which F_i is the inbreeding coefficient (González-Recio et al., 2007) and t can be, for each individual, the equivalent complete generations, the maximum number of generations known, or the complete generations (Gutiérrez et al., 2008). By averaging these ΔF_i , we obtain an estimate of ΔF , $\overline{\Delta F}$, in the reference population, and then a realized N_e is computed as $\overline{N_e} = 1/2 \overline{\Delta F}$ (Cervantes et al., 2008).

Genetic Representation and Probabilities of Gene Origin

Effective Number of Founders. The effective number of founders (f_e) is the number of equally contributing founders that would be expected to produce the same genetic diversity as in the population under study (Lacy, 1989). It is calculated as $f_e = 1/\sum_{k=1}^f q_k^2$, in which q_k is the probability of gene origin of the k ancestor, that is, the proportional contribution of founder k . When founders contribute unequally, the effective number of founders is smaller than is the actual number.

Effective Number of Ancestors. The previous method does not account for bottlenecks in the population, which is a major cause of gene loss in cattle populations. Effective number of ancestors (f_a) overcomes, in part, this shortcoming. It is the minimum number of ancestors, founders or not, necessary to explain the complete genetic diversity of the population under study (Boichard et al., 1997) and is computed as $f_a = 1/\sum_{j=1}^a q_j^2$.

In this case, q_j is the marginal contribution of the j ancestor, that is, the genetic contribution made by an ancestor that is not explained by other ancestors chosen previously.

Effective Number of Founder Herds. The initial contribution of founders can be added into each herd founder contribution, and the inverse of their added squared value gives an effective number of founder herds (Robertson, 1953).

Software

POPREP software (Groeneveld et al., 2009) was used to analyze the census trend, family sizes, average pedigree completeness traced back until the sixth generation, inbreeding, coancestry, and N_e calculated from the last 2 parameters. The software ENDOG, version 4.6 (Gutiérrez and Goyache, 2005), was used to calculate the pyramid structure, generation interval, complete generation equivalent value, individual increase of inbreeding and the corresponding realized N_e , and probabilities of gene origin: effective number of founders, effective number of ancestors, and effective number of founder herds.

RESULTS AND DISCUSSION***Censuses***

The number of breeding females is one of the key parameters to assess the risk of endangerment of a local bovine population. Table 1 shows that all breeds but the Mo breed

increased their censuses over time. In fact, all of the breeds have experienced an increase in the number of registered cows in relation to the period studied by Gutiérrez et al. (2003), and the support schemes for farmers established in 2003 had a variable impact on the evolution of the censuses. The group of ANI, Mo, Pi, and Re had moderate increases, ranging from 16 to 41% in the female censuses, whereas AV, BP, and RG have much larger increases (158 to 381%), probably due to a strong support of the regional governments for the selection schemes.

The number of breeding females per bull was below 10 in AV, above 30 in RG, and between 14 and 22 in the rest of the populations. This confirms the extended use of natural service bulls, with AI being fairly residual, with the exception of RG. In AV, the low number of cows per bull can be due to the small size and geographical isolation of herds.

The effective number, N_e , computed by assuming a random mating population with unequal number of males and females, increased in all breeds, except Mo in 2009, in parallel to the census (Table 1). In 2009, N_e ranged from 470 in the Mo breed to 8,752 in AV breed. The differences in N_e between AV and RG (8,752 and 1,668, respectively), which had a similar number of females in 2009, are due to the extensive use of AI in the last breed. These values will be contrasted below with other estimates of the N_e that allow for more relaxed conditions.

Pyramid Structure

It is well established that genetic improvement efforts are usually made in a fraction of the population and from these animals is later on disseminated to the rest of the population. Our results show that there are not isolated herds (Table 2), and none of the herds of any breeds used only their own males for breeding, which means that no herd was classified as nucleus. In consequence, all herds were assigned either to the multiplier (1 and 2) or commercial (1 and 2) categories, depending on the breed. A herd is defined as multiplier when it sells bulls to other herds, although it can use its own or purchased breeding males for mounting. In fact, in herds catalogued as Multiplier 1, the proportion of purchased bulls can be very high (close to 80% in AV and RG).

The proportion of multiplier (1 and 2) and Commercial 1 herds has increased markedly in BP at the expense of the Commercial 2 herds in the period studied, with this trend also being observed in AV, ANI, and Mo, although to a lesser extent. This evolution implies a greater involvement in the selection process of more breeders and could be

related to the increment of herd sizes and the impact of genetic evaluations, which facilitate both the use of highly ranked own bulls and also the exchange of bulls with other herds. The herd structure of Pi, Re, and RG has remained more stable in the period 1998 to 2009. In the last year analyzed, the proportion of multiplier herds oscillates between 12 and 24% for ANI, Re, and RG and 43 to 52% for BP, Mo, and Pi, with an intermediate value (35%) for AV.

Table 1. Number of males and females giving birth in 1998, 2003, and 2009 and the corresponding effective sizes (N_e) under random mating and unequal number of males and females

Breed ¹	Year	Males	Females	N_e
AV	1998	1,152	8,064	4,032
	2003	1,792	15,501	6,425
	2009	2,444	20,884	8,752
ANI	1998	232	5,318	889
	2003	310	6,856	1,186
	2009	389	7,504	1,479
BP	1998	77	1,369	292
	2003	154	2,643	582
	2009	335	6,593	1,275
Mo	1998	159	1,588	578
	2003	167	3,355	636
	2009	125	1,947	470
Pi	1998	551	7,372	2,051
	2003	677	9,612	2,530
	2009	810	10,531	3,009
Re	1998	239	3,933	901
	2003	290	4,246	1,086
	2009	345	4,573	1,283
RG	1998	168	5,154	651
	2003	207	9,190	810
	2009	433	16,512	1,688

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

This indicates that in the first 3 breeds, the selection efforts are concentrated in a small number of herds. In the other breeds, the proportion of herds that made some type of

selection and sold bulls to other herds is greater (40–50%). A lower number of multipliers can facilitate the implementation and development of a selection program and, in the future, the genomic selection, although at the risk of increasing inbreeding if not properly monitored. What is interesting to note is that in Vassallo et al. (1986), some herds were classified as nucleus herds, differing from the current results. This can be due to the fact that the selection program applied has changed the dynamic of the population or that the sample used for the previous study did not represent the information registered in the herdbook of the breed.

Generation Intervals and Family Size

Average generation intervals for the 4 pathways associated with the animals born in the period 2005 through 2009 are shown in Table 3. The estimates ranged from 4.7 to 7.6 yr in the sire pathway and from 5.95 to 7.8 yr in the dam pathway. In general, the average values were shorter in the sire pathway, with ANI, AV, and Mo having a more rapid replacement of sires, whereas RG sires were used almost 2 additional years on average. This fact can be due to the greater practice of AI in RG and the longer use of proven sire's semen. The average generation interval in the dam pathway was around 7 yr, due partially to a later start (30–36 mo) and longer reproductive life of the females. In addition, the greater reliability of the genetic evaluations of bulls achieved at an early age facilitates earlier selection decisions in young bulls than in cows.

The average estimates were homogeneous across breeds, with the exception again of AV, which showed a shorter interval suggesting a relative acceleration of the selection process in this breed, and the high sire–daughter interval in RG, related to the use of semen of older proven bulls as fathers of replacement cows. These estimates represent a considerable increase with respect to the generation intervals estimated by Gutiérrez et al. (2003) in AV, ANI, BP, Mo, and Pi beef breeds from the parents of the 1995 calf crop. This increase could be related to the expansion of herd censuses associated with a longer use of the cows. The current estimates, however, are in the order of the results found in other beef breeds such as Irish Charolais, Limousin, Hereford, Angus, and Simmental cattle (Mc Parland et al., 2007). Furthermore, a longer generation interval in females than in males has previously been reported in other breeds, such as Australian Shorthorn (Herron and Pattie, 1977) or British Hereford (Özkütük and Bichard, 1977), whereas in

Irish Hereford only slight differences were found and in Limousin the sire–offspring pathway was 2 yr greater than the dam–offspring pathway (Mc Parland et al., 2007).

Table 2. Pyramid structure of the different breeds according to the use of own or purchased bulls. The numbers correspond to the percentage of herds belonging to the category, with the percentage of purchased bulls included between parentheses. The periods analyzed went from the starting year of recording practices to the years 1998, 2003, and 2009

Breed ¹	Year	Type of herd ²				
		Nucleus	Multiplier 1	Multiplier 2	Commercial 1	Commercial 2
AV	1998	0	18.2 (72.02)	5.68	7.62 (81.59)	68.67
	2003	0	21.05 (75.56)	8.42	6.57 (84.84)	63.78
	2009	0	23.05 (77.70)	10.58	6.77 (84.19)	59.60
ANI	1998	0	10.14 (51.88)	3.32	7.34 (81.72)	79.20
	2003	0	10.36 (51.51)	4.33	9.12 (77.04)	76.20
	2009	0	13.64 (53.13)	5.34	9.42 (78.18)	71.59
BP	1998	0	7.89 (65.87)	0.88	4.97 (77.32)	86.26
	2003	0	11.70 (66.46)	5.34	12.48 (72.03)	70.48
	2009	0	39.85 (56.15)	3.42	20.54 (58.11)	36.19
Mo	1998	0	38.21 (45.90)	3.65	19.93 (46.99)	38.21
	2003	0	44.98 (38.21)	2.13	20.06 (42.04)	32.83
	2009	0	48.53 (35.64)	3.53	21.47 (39.2)	26.47
Pi	1998	0	21.41 (75.18)	19.19	5.66 (75.66)	53.74
	2003	0	22.62 (73.91)	18.59	5.98 (80.94)	52.81
	2009	0	23.49 (73.36)	18.68	6.19 (81.8)	51.64
Re	1998	0	18.22 (48.12)	3.78	18.44 (63.49)	59.56
	2003	0	17.82 (47.93)	3.94	17.07 (64.06)	61.16
	2009	0	18.77 (46.67)	4.61	16.72 (68.76)	59.90
RG	1998	0	5.1 (85.72)	7.22	1.19 (86.16)	86.49
	2003	0	4.66 (84.39)	5.97	1.49 (82.46)	87.88
	2009	0	5.33 (81.32)	6.42	1.98 (81.53)	86.27

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

²Nucleus: herds using their own bulls and selling bulls for breeding; Multiplier 1: herds using purchased bulls and their own bulls, and also selling bulls for breeding; Multiplier 2: herds using purchased bulls and selling bulls for breeding; Commercial 1: herds using purchased bulls and their own bulls; Commercial 2: herds using only purchased bulls.

Family size estimates from breeding animals born in 1988, 1993, and 2000 are presented in Table 4. The year 2000 was chosen back as the first one that presented estimates similar to previous years and corresponds to a reference year where the breeding

animals have been able to complete their reproductive career. There were large differences in the average progeny per bull, ranging from 15.6 (AV) to 373.7 animals (RG). As before, this large progeny in RG is related to the use of AI. Most of the breeds increased the average number of progeny per sire. The average number of selected offspring per sire was in the range between 8.3 (AV) and 44 (RG), which together accounted for a percentage ranging from 15 (Pi and RG) to 26% (AV) of the total of all offspring born, this last being consistent with the lower generation interval of this breed. In addition to this between-breed variability in family size, there exists a within breed variability (not shown in tables) that can increase the rate of inbreeding and reduce N_e .

Table 3. Generation interval estimates (with SE in parentheses) for the 4 pathways corresponding to animals born in the period 2005 through 2009

Breed ¹	Sire-son	Sire-daughter	Dam-son	Dam-daughter
AV	4.70 (0.01)	4.77 (0.01)	6.24 (0.01)	5.95 (0.01)
ANI	4.79 (0.10)	5.04 (0.09)	6.87 (0.16)	6.80 (0.17)
BP	5.86 (0.16)	5.66 (0.15)	7.66 (0.24)	7.05 (0.25)
Mo	4.88 (0.21)	5.49 (0.25)	7.83 (0.43)	7.39 (0.41)
Pi	5.85 (0.15)	5.86 (0.14)	7.21 (0.13)	7.32 (0.14)
Re	5.41 (0.05)	5.27 (0.01)	7.13 (0.07)	6.86 (0.02)
RG	6.56 (0.20)	7.60 (0.22)	6.81 (0.19)	7.28 (0.20)

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

Pedigree Completeness Level

The estimates of inbreeding are dependent on pedigree completeness (Lutaaya et al., 1999; Cassell et al., 2003), and a large proportion of missing parents in the pedigree leads to an important underestimation of the inbreeding level. In this study, pedigree knowledge was assessed through the completeness index (MacCluer et al., 1983). This index is an indicator of the proportion of ancestors known in a predefined number of generations. In this study, we analyzed the results when 1, 2, 3, 4, 5, or 6 generations back were considered. The results varied among breeds. Pirenaica, ANI, and Re breeds have deeper pedigrees, with average completeness indexes of more than 96% in the first generation. Pedigree completeness decreased as more generations were considered, albeit even when 6 generations were considered, those 3 breeds had a pedigree completeness of around 80% (Table 5). In general, there has been an increase in the pedigree records across breeds from

1995 (Gutiérrez et al., 2003). This pedigree records increase might be due to the need of reliable genealogical information to implement selection programs based on BLUP predictions of breeding values. Nevertheless, the genealogy of the BP breed underwent a severe contraction in the late 1990s, from which it is slowly recovering, which led to the low proportion of known ancestors beyond grandparents.

Table 4. Average family sizes of breeding animals born in 1988, 1993, and 2000

Breed ¹	Year	All offspring ²		Selected offspring ³	
		Sires	Dams	Sires	Dams
AV	1988	15.6	3.1	8.3	1.9
	1993	24.3	3.4	11.9	1.8
	2000	35.9	3.6	9.3	1.5
ANI	1988	92.9	4.3	29.6	2.1
	1993	83.1	4.5	24.9	1.9
	2000	89.8	3.2	18.9	1.5
BP	1988	36.5	3.1	11.0	1.6
	1993	40.5	3.0	10.5	1.5
	2000	52.0	2.8	9.3	1.2
Mo	1988	25.6	2.6	19.4	1.6
	1993	70.5	3.3	32.4	1.7
	2000	55.1	2.1	13.8	1.3
Pi	1988	40.6	5.2	15.0	2.2
	1993	43.7	5.4	13.7	2.0
	2000	65.1	4.2	10.0	1.3
Re	1988	63.0	3.5	21.3	1.9
	1993	68.9	3.9	22.9	1.9
	2000	67.6	3.3	13.8	1.3
RG	1988	79.1	3.2	39.8	1.8
	1993	373.7	3.6	87.7	1.7
	2000	299.5	3.7	44.0	1.4

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

²All offspring born in the population.

³Offspring that have at least a service record.

In correspondence with the average completeness indexes estimated, 3 breeds have the highest complete generation equivalent values (3.77, 3.99, and 4.62 for Re, ANI, and Pi). To compute these values, all animals born up to 2009 were used. These figures are in the order of those published by Boichard et al. (1997) in Abondance, Normande, and Limousine French breeds but lower than the estimates found in Danish dairy cattle (Sørensen et al., 2005).

Evolution of Inbreeding and Effective Population Sizes

A general increase in average inbreeding over time was observed in all breeds through the years analyzed. The average inbreeding coefficients of animals born in 2009 ranged from 0.6% in BP to 7.2% in Re, when all animals were considered (Table 6). In all breeds, the SE of the estimates were large. It should be highlighted that the inbreeding coefficients may be probably underestimated due to the lack of completeness of pedigrees in most populations. This can be particularly important in the breeds with the lowest quality of pedigree information.

Table 5. Average pedigree completeness index (%) considering 1 to 6 generations deep for animals born in 2009 and average complete generation equivalent values for animals born up to 2009

Breed ¹	Number of animals	Number of generations known						Average complete generation equivalents
		1	2	3	4	5	6	
AV	21,418	91.7	83.4	75.8	67.6	58.6	50.2	2.18
ANI	8,052	96.7	91.9	88.5	85.8	82.9	79.3	3.99
BP	6,902	83.2	59.1	42.1	32.0	25.7	21.5	1.01
Mo	2,223	86.8	77.9	70.8	64.8	59.4	53.6	2.17
Pi	10,845	99.0	97.9	96.8	95.2	92.0	86.7	4.62
Re	4,928	96.3	91.7	88.6	86.5	84.6	80.9	3.77
RG	17,124	94.4	83.2	75.1	69.0	63.5	57.9	3.08

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

The proportion of inbred animals tended to increase in all breeds. In the year 2009 it had a range from 38 (AV) to 95% (Pi), with the exception of BP (3.5%). The greater percentages belong to the breeds with better pedigree knowledge (Pi, ANI, and Re).

The average inbreeding coefficient of inbred animals was below 11%, with the exception of BP and Mo. The SE of the estimates were also large. In all breeds, individuals exceeding an inbreeding coefficient of 40% could be found. In general, no correlation existed between the percentage of inbred animals and the average inbreeding of inbred animals. Whereas Pi is the breed with the most inbred animals, the *F* average stayed at around 3.5%, suggesting that breeders avoid matings between close relatives. Another cause of this low *F* average can be a larger number of Multiplier 1 herds than in ANI and

Re, populations that have similar pedigree completeness. On the contrary, the particularly high average inbreeding of inbred animals born in 2009 in BP (17.6%) suggests that a relevant number of matings between close relatives occurs, which should be subjected to careful monitoring to avoid the loss of genetic variability in this population.

Table 6. Average inbreeding coefficients of all animals and inbred animals born in 1998, 2003, and 2009

Breed ¹	Year	All animals			Inbred animals	
		Number	Maximum <i>F</i>	Average <i>F</i> (SD)	Number (%)	Average <i>F</i> (SD)
AV	1998	11,454	0.4248	0.0102 (0.0428)	1,270 (11.1)	0.0921 (0.0947)
	2003	18,361	0.4219	0.0112 (0.0429)	3,586 (19.5)	0.0573 (0.0823)
	2009	21,418	0.5361	0.0142 (0.0474)	8,175 (38.2)	0.0371 (0.0710)
ANI	1998	6,145	0.4629	0.0472 (0.0746)	3,980 (64.8)	0.0728 (0.0820)
	2003	7,924	0.4681	0.0518 (0.0751)	5,910 (74.6)	0.0711 (0.0794)
	2009	8,052	0.4911	0.0611 (0.0805)	6,522 (81.0)	0.0754 (0.0832)
BP	1998	3,142	0.2500	0.0030 (0.0224)	80 (2.5)	0.1162 (0.0811)
	2003	3,592	0.3125	0.0035 (0.0228)	132 (3.7)	0.0951 (0.0740)
	2009	6,902	0.4375	0.0062 (0.0372)	243 (3.5)	0.1758 (0.0972)
Mo	1998	2,170	0.5098	0.0485 (0.0883)	799 (36.8)	0.1316 (0.1013)
	2003	3,811	0.4805	0.0645 (0.0909)	2,015 (53.0)	0.1221 (0.0927)
	2009	2,223	0.4893	0.0593 (0.0919)	1,233 (55.5)	0.1116 (0.1009)
Pi	1998	7,825	0.3821	0.0261 (0.0532)	6,508 (83.2)	0.0313 (0.0509)
	2003	10,174	0.5000	0.0330 (0.0601)	9,112 (89.6)	0.0368 (0.0643)
	2009	10,845	0.4975	0.0345 (0.0581)	10,303 (95.0)	0.0363 (0.0590)
Re	1998	4,942	0.5434	0.0628 (0.0891)	2,781 (55.7)	0.1127 (0.0930)
	2003	5,278	0.5449	0.0620 (0.0874)	3,161 (59.8)	0.1036 (0.0920)
	2009	4,928	0.4591	0.0722 (0.0870)	3,599 (73.0)	0.0988 (0.0879)
RG	1998	7,303	0.3984	0.0140 (0.0427)	2,612 (35.8)	0.0390 (0.0641)
	2003	11,166	0.4688	0.0211 (0.0502)	5,952 (53.3)	0.0396 (0.0632)
	2009	17,124	0.5039	0.0269 (0.0529)	10,130 (59.2)	0.0455 (0.0623)

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

In general, all breeds have experienced an increase of inbreeding in relation to the 1995 estimates (Gutiérrez et al., 2003), both for the whole populations and for inbred animals. This is probably a consequence of the increase of depth in pedigree information.

Our estimates are much greater than are the ones found in Irish cattle, where the estimates of mean inbreeding of animals born in 2004 were in the range of 0.54 to 2.19% for Charolais and Hereford breeds, respectively, with complete generation equivalents around 5 (Mc Parland et al., 2007).

Other interesting parameters to monitor the evolution of populations are the rate of inbreeding and the potential inbreeding that could be obtained if all matings were made at random from all members of the population. This last parameter is the average coancestry of all possible pairs of individuals in the population, and its evolution is the rate of coancestry. Estimates of both rates are presented in Table 7. The estimates of the rate of inbreeding and the N_e from animals born in the years 1998, 2003, and 2009 were highly heterogenous, and no clear trend can be inferred (Table 7A). The estimates of the rate of coancestry and the corresponding effective numbers were also diverse, but with the exception of RG and some particular years in different breeds, the increase of coancestry was much lower than was that of inbreeding. This fact was confirmed when a longer period was analyzed, and both rates were estimated as the regression on the year of birth (Table 7B). The N_e based on the increment of inbreeding ranged from 43 to 378 for Re and BP, whereas N_e based on coancestry were greater, with a range of 100 for RG to 9,985 for BP. Note that the coancestry rate of Pi was even negative and the corresponding effective size could not be calculated and also that the very high value of the N_e of BP may be overestimated due to the low completeness of its pedigree.

The estimates for N_e in Table 7B are much lower than are those presented in Table 1. This indicates that the breeds studied are far from the conditions of an idealized population where bulls and cows can mate at random. In fact, the differences between inbreeding and coancestry rates suggest the presence of population structure (Fernández et al., 2008), either by geographical reasons or selection policies within farms or groups of farms. These differences indicate, however, the potential feedback for the implementation of a mating system to control possible losses of variability in the populations studied. Two examples illustrating what has been just discussed are presented in Figure 1. In the RG breed, the evolution of inbreeding and coancestry was very close, probably due to the greater deployment of AI that exists in this breed, whereas Re, similarly to the other breeds studied, presents a steeper increase of the average inbreeding coefficient, suggesting the existence of some kind of mating structure.

Table 7. Inbreeding (ΔF) and coancestry (Δf) rates and effective population size (N_e) by generation

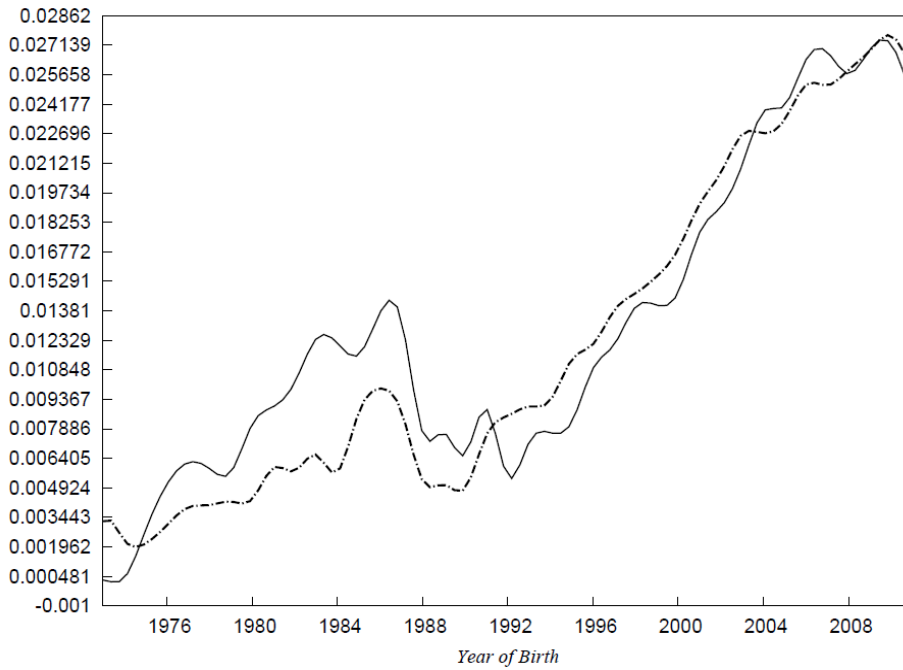
A. Estimated in 3 different periods (1998, 2003, and 2009)					
Breed ¹	Year	ΔF	$N_e (\Delta F)$	Δf	$N_e (\Delta f)$
AV	1998	0.00682	73	0.00148	338
	2003	0.00242	207	0.00046	1,087
	2009	0.00253	198	0.00025	2,000
ANI	1998	0.00199	251	0.00050	1,000
	2003	0.00095	526	0.00088	568
	2009	0.00981	51	0.00427	117
BP	1998	0.00120	417	0.00022	2,273
	2003	0.00080	625	-0.00084	–
	2009	0.00291	172	0.00008	6,250
Mo	1998	0.00823	61	0.00141	355
	2003	0.02359	21	0.00090	556
	2009	-0.00309	–	0.00097	515
Pi	1998	0.00815	61	0.00269	186
	2003	0.01105	45	0.00265	189
	2009	0.00155	323	0.00272	184
Re	1998	0.01295	39	0.00233	215
	2003	0.00372	134	0.00379	132
	2009	0.01087	46	0.00550	91
RG	1998	0.00855	58	0.00620	81
	2003	0.01031	48	0.01067	47
	2009	0.00937	53	0.00786	64
B. Estimated by regression of the average inbreeding and coancestry coefficients on year of birth					
Breed	Period	ΔF	$N_e (\Delta F)$	Δf	$N_e (\Delta f)$
AV	1975–2011	0.00266	188	0.00045	1,123
ANI	1970–2011	0.01085	46	0.00220	228
BP	1979–2011	0.00132	378	0.00005	9,985
Mo	1976–2011	0.01141	44	0.00201	249
Pi	1934–2010	0.00227	221	-0.00180	–
Re	1960–2011	0.01157	43	0.00161	310
RG	1973–2011	0.00437	114	0.00499	100

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

González-Recio et al. (2007) and Gutiérrez et al. (2008) presented an alternative method for the estimation of realized effective size from pedigrees based on the definition of the individual increase in inbreeding coefficient (ΔF_i). Although still dependent on the completeness of the pedigree, the realized N_e appears more stable than does that estimated based on regressions and is less affected by changes in the mating policy or the accumulation of pedigree. Depending on the value of t used for computing ΔF_i , maximum number of generations known, complete generations, or complete equivalent generations, we can obtain an upper estimate, a lower estimate, or the actual estimate of the realized N_e . Table 8 presents the estimates in the 7 breeds in the 3 reference years. In general, the changes of estimates from different years were small, although 3 breeds showed an increase of realized N_e and 2 showed a decrease. For the year 2009, the estimates ranged from 19 for Mo breed to 90 for AV breed. The values estimated by computing F_i using complete generations were lower in general than were those computed by regression. It is difficult to assess what estimate better reflects the true evolution of inbreeding according to the particular characteristics of pedigree regarding the degree of completeness, mating strategies, and presence or not of bottlenecks. Cervantes et al. (2008) have tested the properties of the method and they conclude that the procedure provides stable estimates when considering pedigrees with 4 or more equivalent generations known, which would make the method desirable only for ANI, Pi, and Mo. Estimates based on linkage disequilibrium of SNP markers might contribute new insights to the estimation of N_e , as in de Roos et al. (2008).

For genetic improvement, an inbreeding rate of 1% per generation ($N_e = 50$) is accepted because such slow inbreeding allows selection to remove deleterious alleles without endangering the population (Franklin, 1980). Disregarding the method of estimation, 3 of the breeds (ANI, Mo, and Re) presented an inbreeding rate exceeding 1%, which led to N_e lower than 50. Fortunately, the N_e estimated from the increment of coancestry (Δf) were bigger, which suggest that the populations may well maintain genetic variability if appropriate mating policies were implemented.

a)



b)

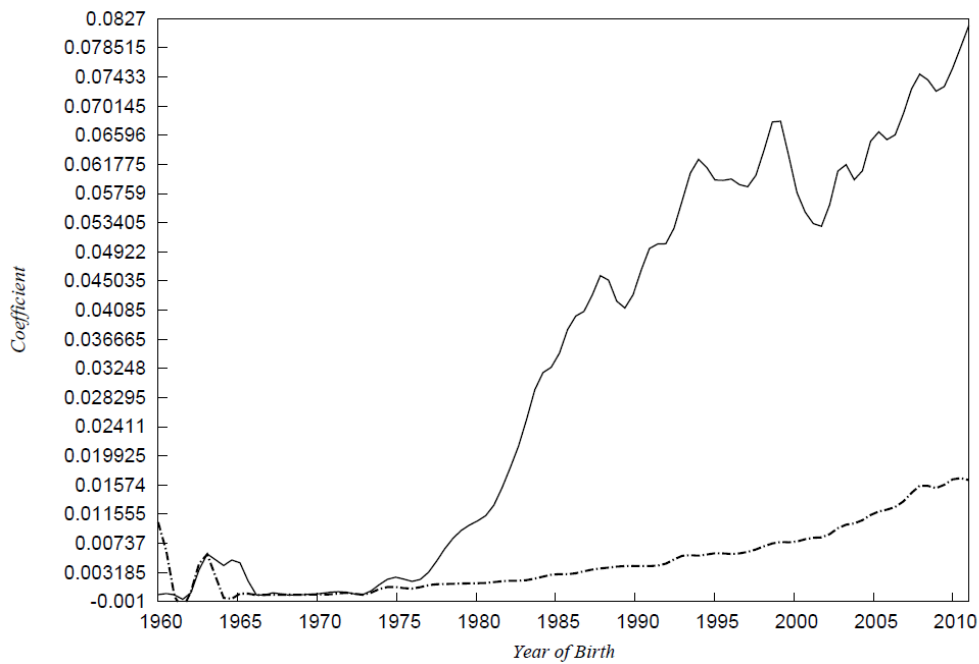


Figure 1. Inbreeding (–) and coancestry (---) coefficients evolution by year of birth in Rubia Gallega (a) and Retinta (b) breeds.

Low effective numbers are common in livestock. In beef cattle, Mc Parland et al. (2007) have found effective sizes of 64 and 127 in Irish Hereford and Simmental, respectively, whereas Cleveland et al. (2005) found that the effective size of the U.S. Hereford was 85 and Boichard et al. (1997) estimated the effective sizes of Abondance and

Normande breeds at 106 and 47, respectively. In 3 Danish dairy cattle breeds, Sørensen et al. (2005) estimated much lower effective sizes ranging from 47 to 53, whereas in U.S. Holsteins the effective size was 39 (Caraviello, 2004). There is no agreement about the effective size of the populations to avoid genetic diversity loss. Meuwissen and Woolliams (1994) reported critical N_e for maintaining fitness ranging from 25 to 255, Meuwissen (1999) suggested that a N_e below 100 animals implies a reduction of population fitness, and the Food and Agriculture Organization of the United Nations (1998) and FAO (2007) recommended that an effective size of at least 50 animals be maintained. As was said before, some of the studied breeds have effective numbers lower than these limits, which would suggest that these populations would require the implementation of strategies to control the rate of inbreeding.

Probability of Gene Origin

The parameters of probability of gene origin are less sensitive to the pedigree completeness level than are the parameters related to inbreeding (Boichard et al., 1997). These parameters indicate how the overrepresentation of certain males (unbalanced contribution of genes) can reduce the gene pool of the population, leading to effective number of herds, ancestors, and founders lower than the actual values.

The unbalanced contribution of genes from herds, founders, and ancestors is common among the different breeds studied, as the effective numbers of these parameters are lower than the observed values (not shown in tables). The effective number of founder herds has diminished in 4 of the breeds in the period studied, probably related to the selection practiced and the extinction by the concentration of some herds. In the other 3 breeds, however, there has been an increase in the effective number of founder herds, probably due to the fact that they maintained the herdbooks open. In 2009, the number of founder herds ranged from 6 to 196 for ANI and AV but, in general, a limited number of herds, below 56, have provided the gene pool of the different breeds (Table 9). The good reputation of some breeders considered as leaders could have been responsible for this fact.

The effective number of founders presents an irregular evolution across years, ranging from 76 (ANI) to 228 (Re) in 2009, with the exception of BP (838) and AV (694), both breeds having the lowest complete generation equivalent values.

Table 8. Averages of individual increase of inbreeding of animals born in 3 yr (1998, 2003, and 2009) and realized effective population sizes (N_e) computed from them

Breed ¹	Year	By maximum generations		By complete generations		By equivalent generations	
		ΔF	N_e	ΔF	N_e	ΔF	N_e
AV	1998	0.0026	194	0.0106	47	0.0068	73
	2003	0.0020	254	0.0097	52	0.0059	85
	2009	0.0018	281	0.0093	54	0.0056	90
ANI	1998	0.0080	63	0.0209	24	0.0141	35
	2003	0.0071	70	0.0200	25	0.0131	39
	2009	0.0067	74	0.0199	25	0.0126	40
BP	1998	0.0025	199	0.0086	58	0.0062	80
	2003	0.0026	189	0.0102	49	0.0074	68
	2009	0.0028	181	0.0099	50	0.0078	64
Mo	1998	0.0126	40	0.0379	13	0.0265	19
	2003	0.0115	43	0.0407	12	0.0278	18
	2009	0.0101	49	0.0398	13	0.0263	19
Pi	1998	0.0012	408	0.0107	47	0.0071	71
	2003	0.0014	348	0.0106	47	0.0073	69
	2009	0.0017	300	0.0103	48	0.0073	68
Re	1998	0.0074	67	0.0258	19	0.0191	26
	2003	0.0069	72	0.0244	20	0.0175	29
	2009	0.0065	77	0.0229	22	0.0162	31
RG	1998	0.0026	192	0.0100	50	0.0059	85
	2003	0.0028	177	0.0116	43	0.0066	76
	2009	0.0030	167	0.0143	35	0.0080	62

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

The effective number of ancestors is a measure of genetic diversity that accounts for recent bottlenecks and, as a consequence, accounts, in part, for the loss of allelic diversity from the foundation population. In the years analyzed, the effective number of ancestors generally has been slightly reduced, with the exception of Re, where the reduction was more pronounced, whereas BP experienced a large increase due to having maintained the herdbook open. The effective number of ancestors in 2009 for 6 of the breeds ranged from 42 (RG) to 220 (AV), with BP having much greater value. In general, the effective number

of ancestors was lower than was the effective number of founders in all breeds, suggesting the existence of more or less strong bottlenecks. In particular, Pi, RG, and AV could have undergone the stronger bottlenecks, whereas the ANI breed presents no evidence of this phenomenon. The number of ancestors explaining 50% of the variability ranged from 17 (RG) to 89 (Mo), again excluding AV and BP, which showed much greater values. In all breeds, both the effective number of founders and the effective number of ancestors increased in comparison to the 1995 study (Gutiérrez et al., 2003). This can be due to the increase of herds involved in the herdbook.

Table 9. Estimates of parameters of probability of gene origin

Breed ¹	Year	Reference population size	Effective number of founder herds	Effective number of founders	Effective number of ancestors	Ancestors explaining 50%
AV	1998	48,188	64	1,301	211	422
	2003	119,015	119	871	215	382
	2009	237,320	196	694	220	304
ANI	1998	76,879	7.3	88	84	52
	2003	111,597	6.8	81	77	46
	2009	151,380	6.3	76	72	40
BP	1998	9,467	4.3	185	170	111
	2003	16,206	5.6	273	250	175
	2009	46,580	17.1	838	739	417
Mo	1998	11,185	36.0	177	166	85
	2003	26,883	39.7	185	158	87
	2009	39,287	43.6	199	159	89
Pi	1998	84,711	66.6	182	87	60
	2003	128,946	60.2	176	83	49
	2009	197,027	56.2	162	74	40
Re	1998	59,346	22.0	334	196	126
	2003	81,092	20.3	280	154	102
	2009	108,637	18.2	228	117	81
RG	1998	49,832	52.9	102	62	30
	2003	87,127	44.8	102	50	20
	2009	166,255	33.0	104	42	17

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

Reduced values for the number of ancestors similar to the ones found in this study have been frequently found in beef and dairy cattle populations. Values of 25, 40, and 360 were estimated for Abondance, Normande, and Limousine (Boichard et al., 1997); a range between 35 and 82 were estimated for Irish Charolais, Limousine, Simmental, Hereford, and Aberdeen Angus (Mc Parland et al., 2007); and even lower values were reported in Danish dairy cattle (20.6, 23.8, and 34.6 for Holstein, Jersey, and Danish Red, respectively).

Conclusions

We have studied the population dynamics of 7 Spanish beef breeds under BLUP-based selection programs. Notwithstanding the differences in herd sizes, rearing environments, and degree of crossbreeding with other breeds, the pedigree analysis of the breeds shows that 1) data recording is in expansion in most of the breeds; 2) family sizes suggest a low use of AI, except in the case of the RG breed; 3) there is an extensive exchange of males among farms of the same breed and none of the farms behaves as though belonging to a closed selection nucleus; 4) the pedigree completeness level varies among breeds but shows that in all breeds, the records are very complete in recent generations; 5) the evolution of inbreeding has led to effective population size estimates, in some breeds, that are lower than are the minimum recommended for the maintenance of genetic variability; 6) there is evidence of population structure that restricts matings within herds or groups of herds, as suggested by the fact that the increase of inbreeding is greater than is that of coancestry; and 7) most of the breeds have experienced some bottlenecks as the effective number of founders exceeds the number of ancestors. All of these findings strongly suggest the necessity of monitoring inbreeding in the populations studied to maintain future improvements by genetic selection.

LITERATURE CITED

- Boichard, D., L. Maignel, and E. Verrier. 1997. The value of using probabilities of gene origin to measure genetic variability in a population. *Genet. Sel. Evol.* 29:5–23.
- Caraviello, D. 2004. Inbreeding in dairy cattle. Dairy update, number 615. The Babcock Institute, University of Wisconsin, Madison, WI.
- Cassell, B. G., V. Adamec, and R. E. Pearson. 2003. Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. *J. Dairy Sci.* 86:2967–2976.

- Cervantes, I., F. Goyache, A. Molina, M. Valera, and J. P. Gutiérrez. 2008. Application of individual increase in inbreeding to estimate effective sizes from real pedigrees. *J. Anim. Breed. Genet.* 125:301–310.
- Cleveland, M. A., H. D. Blackburn, R. M. Enns, and D. J. Garrick. 2005. Changes in inbreeding of U.S. Herefords during the twentieth century. *J. Anim. Sci.* 83:992–1001.
- de Roos, A. P. W., B. J. Hayes, R. J. Spelman, and M. E. Goddard. 2008. Linkage disequilibrium and persistence of phase in Holstein- Friesian, Jersey and Angus cattle. *Genetics* 179:1503–1512.
- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to quantitative genetics*. Longman Group, Harlow, UK.
- 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.
- Food and Agriculture Organization of the United Nations (FAO). 1998. *Secondary guidelines for the development of national farm animal genetic resources management plans: Management of small populations at risk*. FAO, Rome, Italy.
- Food and Agriculture Organization of the United Nations (FAO). 2007. *The State of the World's Animal Genetic Resources for Food and Agriculture*, B. Rischkowsky and D. Pilling, editors. FAO, Rome, Italy.
- Franklin, I. R. 1980. Evolutionary changes in small populations. In: M. E. Soule and B. A. Wilcox, editors, *Conservation biology – An evolutionary ecological perspective*. Sinauer, Sunderland, MA. p. 135–149.
- Gandini, G., and E. Villa. 2003. Analysis of the cultural value of local livestock breeds: A methodology. *J. Anim. Breed. Genet.* 120:1–11.
- Gil, M., X. Serra, M. Gispert, M. A. 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.
- González Recio, O., E. López de Maturana, and J. P. Gutiérrez. 2007. Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *J. Dairy Sci.* 90:5744–5752.
- Goyache, F., J. P. Gutiérrez, I. Fernández, J. Gómez, I. Álvarez, J. Díez, and L. J. Royo. 2003. Using pedigree information to monitor genetic variability of an endangered population: The Xalda sheep breed of Asturias as an example. *J. Anim. Breed. Genet.* 120:95–103.
- Groeneveld, E., B.D. Westhuizen, A. Maiwashe, F. Voordewind, and J. B. S. Ferraz. 2009. POPREP: a generic report for population management. *Genet. Mol. Res.* 8:1158–1178.
- Gutiérrez, J. P., J. Altarriba, C. Díaz, R. Quintanilla, J. Cañón, and J. Piedrafita. 2003. Genetic analysis of eight Spanish beef cattle breeds. *Genet. Sel. Evol.* 35:43–64.

- Gutiérrez, J. P., I. Cervantes, A. Molina, M. Valera, and F. Goyache. 2008. Individual increase in inbreeding allows estimating effective population sizes from pedigrees. *Genet. Sel. Evol.* 40:359–378.
- Gutiérrez, J. P., and F. Goyache. 2005. A note on ENDOG: A computer program for analysing pedigree information. *J. Anim. Breed. Genet.* 122:172–176.
- Herron, N. D., and W. A. Pattie. 1977. Studies of the Australian Illawarra Shorthorn breed of dairy cattle. II. *Aust. J. Agric. Res.* 28:1119–1132.
- James, J. W. 1977. A note on selection differentials and generation length when generations overlap. *Anim. Prod.* 24:109–112.
- Lacy, R. C. 1989. Analysis of founder representation in pedigrees: Founder equivalent and founder genome equivalents. *Zoo Biol.* 8:111–123.
- Lutaaya, E., I. Misztal, J. K. Bertrand, and J. W. Mabry. 1999. Inbreeding in populations with incomplete pedigrees. *J. Anim. Breed. Genet.* 116:475–480.
- MacCluer, J., B. Boyce, L. Dyke, D. Weitzkamp, A. Pfenning, and C. Parsons. 1983. Inbreeding and pedigree structure in Standardbred horses. *J. Hered.* 74:394–399.
- Mc Parland, S., J. F. Kearney, M. Rath, and D. P. Berry. 2007. Inbreeding trends and pedigree analysis of Irish dairy and beef cattle populations. *J. Anim. Sci.* 85:322–331.
- Meuwissen, T. H. E. 1999. Operation of conservation schemes. In: J. K. Oldenbroek, editor, *Genebanks and the conservation of farm animal genetic resources*. Dienst Landbouwkundig Onderzoek, Lelystad, The Netherlands. p. 91–112.
- Meuwissen, T. H. E., and J. A. Woolliams. 1994. Effective sizes of livestock populations to prevent a decline in fitness. *Theor. Appl. Genet.* 89:1019–1026.
- Özkütük, K., and M. Bichard. 1977. Studies of pedigree Hereford cattle breeding. 1. Herdbook analyses. *Anim. Prod.* 24:1–13.
- Piedrafita, J., R. Quintanilla, C. Sañudo, J. L. Olleta, M. M. Campo, B. Panea, G. Renand, F. Turin, S. Jabet, K. Osoro, M. C. Oliván, G. Noval, P. García, M. D. García-Cachán, M. A. Oliver, M. Gispert, X. Serra, M. Espejo, S. García, M. López, and M. Izquierdo. 2003. Carcass quality of 10 beef cattle breeds of the Southwest of Europe in their typical production systems. *Livest. Prod. Sci.* 82:1–13.
- Robertson, A. 1953. A numerical description of breed structure. *J. Agric. Sci.* 43:334–336.
- Sørensen, A. C., M. K. Sørensen, and P. Berg. 2005. Inbreeding in Danish dairy cattle breeds. *J. Dairy Sci.* 88:1865–1872.
- Tolosana, A. M. O., Z. Whebi, and E. Manrique. 2005. Quality perception and consumer attitudes to specific quality beef in Aragon, Spain. *Span. J. Agric. Res.* 3:418–428.
- Vassallo, J. M., C. Díaz, and J. R. García-Medina. 1986. A note on the population structure of the Avileña breed of cattle in Spain. *Livest. Prod. Sci.* 15:285–288.
- Vu Tien Khang, J. 1983. Méthodes d'analyse des données démographiques et généalogiques dans les populations d'animaux domestiques. (In French.) *Genet. Sel. Evol.* 15:263–298.
- Wright, S. 1922. Coefficient of inbreeding and relationship. *Am. Nat.* 56:330–338.
- Wright, S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.

Capítulo 5

GENETIC DIVERSITY AND DIVERGENCE AMONG SPANISH BEEF BREEDS

Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high density SNP chip¹

J. J. Cañas-Álvarez,^{*2} A. González-Rodríguez,^{†3} S. Munilla,^{†‡} L. Varona,^{† §} C. Díaz,^{||} J. A. Baro,[#] J. Altarriba,^{† §} A. Molina,[¶] 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 de Biocomputación y Física de los Sistemas Complejos (BIFI). 50018. 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

¹The research leading to these results has received funding of AGL 2010-15903 grant from the Spanish government, and from the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement n° 289592 - Gene2Farm. The collaboration of Breed societies in collecting samples and the support of FEAGAS is also acknowledged.

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

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

⁴Corresponding author: jesus.piedrafita@uab.es

JOURNAL OF ANIMAL SCIENCE, submitted

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 seven 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 (AMOVA) 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 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 two 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 two 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 to assign 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, SNP variability.

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 (draught, milk and beef) animals; then, depending on the region, their characteristics and 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 to SNPs, have relatively high mutation rates per generation, what would confound the population history (Brumfield et al., 2003). On the other hand, SNPs are bi-allelic, 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 SNPs 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 relationships of seven autochthonous Spanish beef cattle breeds, using a high-density SNP chip.

MATERIAL AND METHODS

Animals and sampling

The animals studied, both males and females in equal numbers, were chosen to capture the existing variability in seven Spanish beef cattle populations. 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 blood samples were collected from the caudal vein of animals in tubes with EDTA as anticoagulant, following the recommendations of the Joint Working Group of Refinement (1993).

SNP genotyping, data pruning and genetic diversity assessment

The samples were processed following the protocol described in Prefiler™ Forensic DNA Kit of Applied Biosystems, using Mag-Max™ Express 96-Magnetic Particle Processor automated equipment. High-density SNP genotyping was performed according to the protocol of the manufacturer by using the BovineHD Beadchip (Illumina Inc, USA, 2012) designed to genotype 777,962 SNPs, at a commercial laboratory (Xenética Fontao,

Lugo, Spain). The SNPs mapped to the UMD3.1 assembly (Zimin et al., 2009). Only SNPs mapped to autosomal chromosomes were used in this study. The SNPs that were in repeated positions and those with Mendelian error rates greater than 5% were removed. These initial quality controls retained 735,239 SNPs for each breed. In order 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 (MNA); 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 dataset 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 SNPs in strong linkage disequilibrium (LD) can affect both PCA and clustering analysis (Moorjani et al., 2013), we thinned the marker set by excluding SNPs in strong LD (pairwise genotypic correlation $r^2 > 0.1$) in a window of 50 SNPs, sliding the window by 5 SNPs at a time. As a result of this data pruning, 57,674 SNPs were left for the analysis of divergence among populations. In addition to the analysis based in pruned data, we perform a sensitivity analysis with the whole dataset in order to check for potential differences. All quality controls and data pruning were performed using PLINK software v1.07 (Purcell et al., 2007).

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 carried out with ARLEQUIN software v3.5 (Excoffier et al., 2005), where significance levels for variance components were tested using non-parametric 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 two measures of distance using breed allele frequencies. First, pairwise F_{ST} statistics were calculated through the ARLEQUIN software v3.5 (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 v3.69 (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

In order to achieve a different approach to characterize divergence, a principal component analysis (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 two individuals genotyped, and is computed as 1 minus the pairwise IBS, averaged across markers. In addition, a multidimensional scaling plot (MDS) on the same relationship matrix was also carried out 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 v 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 pre-defined 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 to other K values.

RESULTS AND DISCUSSION

Genetic diversity analysis

Genetic diversity across breeds was assessed through several statistics computed from 735,239 SNPs in autosomal chromosomes after eliminating the SNPs having Mendelian error rates greater than 5%. The proportion of markers genotyped on 95% of the samples was around 97%-98% in all breeds, which suggests the suitability of the chip to genotype the breeds studied (Table 1). The proportion of SNPs with a MAF > 0.05 ranged between 86% (Pi) and 89% (AV), indicating that most SNPs 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 a further insight on the degree of polymorphism observed

across breeds, we represented the average proportion of SNPs for different ranges of minor allele frequency (Figure 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).

Table 1. Genetic diversity within seven Spanish beef breeds

Breed ¹	Markers genotyped on > 95% of the samples, %	Markers with MAF > 0.05, %	Mean number of alleles (MNA)	Markers in HWE ($P > 0.01$), %	Expected heterozygosity with (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)

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

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 SNPs (Gautier et al., 2007). It was also greater than the average 83% reported by Salomon-Torres et al. (2014) for bovine breeds employing 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). The differences between estimates could be due in part to ascertainment bias, i.e., 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. 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 found in the Spanish breeds was similar to Australian taurine

breeds (0.85-0.90) what may help to conclude that if ascertainment bias exist that would affect all breeds in a similar way and the bias could be more related to the design of the SNP chip than to the breeds participating in the design.

In general, no significant departures from the H-W equilibrium were observed among the polymorphic markers in any breed, and the percentage of markers in H-W equilibrium ($P > 0.01$) was always over 98%. Thus, 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 out 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 IBD has been detected because the SNP chip mostly recall 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 H-W equilibrium, expected heterozygosity and MNA of 82.89%, 0.30 and 1.85, respectively. In turn, expected heterozygosity values around 0.35 in *B. indicus* and 0.4 in *B. taurus* were reported in indigenous cattle from Ethiopia and Korea (Edea et al., 2013), and about 0.4 in two Bangladeshi zebu cattle populations with a 80K SNP chip (Uzzaman et al, 2014). All these estimates (including our own) are potentially biased upwards due to the ascertainment of SNP markers that tend to be at intermediate allele frequencies (Lachance and Tishkoff, 2013).

Analysis of divergence among breeds

The pattern of divergence among the seven Spanish beef cattle breeds was studied following four different approaches: 1) analysis of molecular variance (AMOVA) and fixation indices; 2) distance measures and phylogenetic trees; 3) principal components analysis (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 SNPs, 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 analysis of molecular variance (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 compared to other breeds, pointing towards a higher common genetic background and probably some admixture between populations.

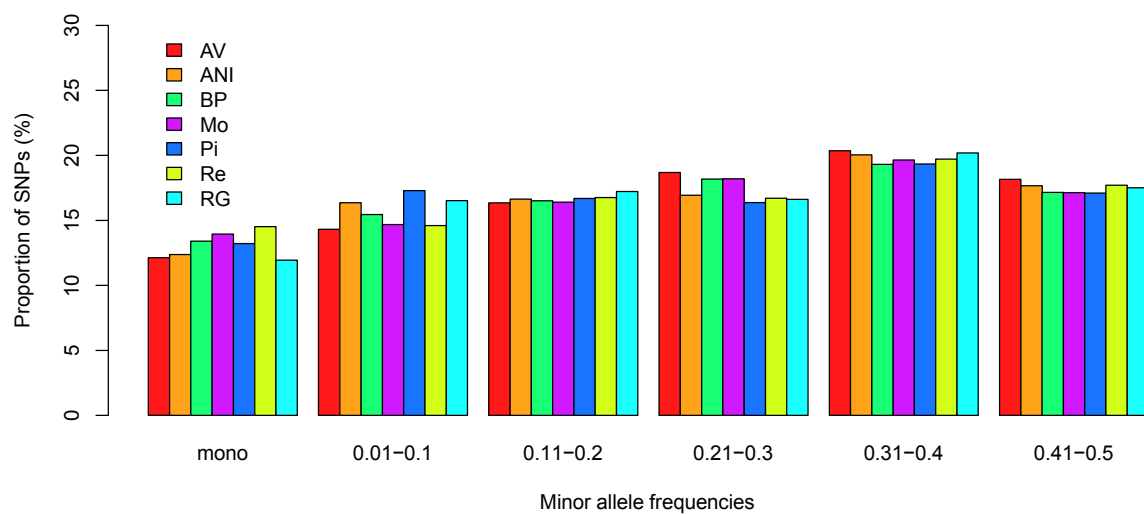


Figure 1. Average proportions of SNPs within MAF ranges by population.

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

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 (non significant), and a genetic differentiation among breeds (F_{ST} value) of 0.044. The low and non-significant 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 zero, ranging from -0.012 to 0.052 (Supplementary Table S1). 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 SNPs, 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. (2014) 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 to ours because the range of F_{ST} estimates depends upon the frequency of the most frequent allele, and this is always higher for biallelic SNPs than for multiallelic microsatellites. Thus, the use of SNPs 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 higher census among the autochthonous Spanish beef cattle populations. On the contrary, 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 can be also attributed to a scale effect due to the density and the way the sampling was done in both studies.

Table 2. Analysis of molecular variance (AMOVA) in the seven populations

Source of variation	d.f.	Variance components	Percentage of variation	Fixation indices
Among populations	6	258.165	4.46	$F_{ST} = 0.044^{**}$
Among individuals within populations	329	76.543	1.32	$F_{IS} = 0.014^{NS}$
Within individuals	336	5456.510	94.22	$F_{IT} = 0.058^{**}$
Total	671	5791.218		

NS, not significant; ** $P < 0.001$

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 similar pattern to the one obtained with the F_{ST} statistics, with values ranging from 0.009 to 0.016. Analyzing European breeds with SNPs, 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.

Table 3. Estimates of pairwise F_{ST} statistics (below the diagonal) and the Nei's D genetic distances (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	

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

A phylogenetic tree was constructed from Nei's D genetic distances (Nei, 1972), using the neighbor joining (N-J) method (Saitou and Nei, 1987) (Figure 2). It is worth to mention that bootstrap percentages computed to assess the robustness of the phylogenetic tree, depicted in the internal nodes of Figure 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, as compared to 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 two main groups of closely related breeds. The first of these groups includes ANI and Mo breeds, sharing the same node, with 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).

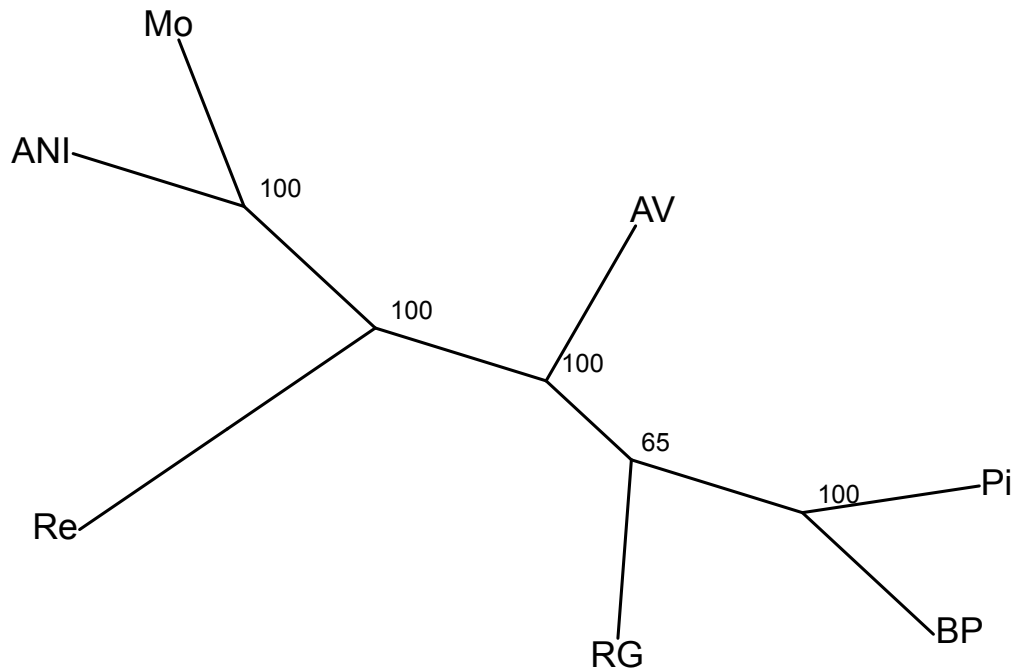


Figure 2. Neighbor joining representation of the pairwise Nei's D genetic distances among populations.

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.

Our results were not entirely consistent with the classification of Iberian cattle in three 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 three breeds are subjected to fully extensive production systems and have similar breeding objectives. Incidentally, these last three 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 components analysis of populations

In a different approach to characterize divergence, a principal component analysis (PCA) was applied to a distance matrix built up from the marker identity-by-state (IBS) relationship matrix (Figure 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 similar classification pattern 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 between these breeds that have had a geographic proximity during their history and share a similar production system. Furthermore, when the PCA analysis was performed on the raw genotypes, the results were similar (results not shown).

For large-scale SNP data, the PCA and multidimensional scaling (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 carried out to determinate some differences between both methods. In general, the groups of populations showed the same distribution in both methods (Figure 3 and Supplementary Figure S1) as each of them represent 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).

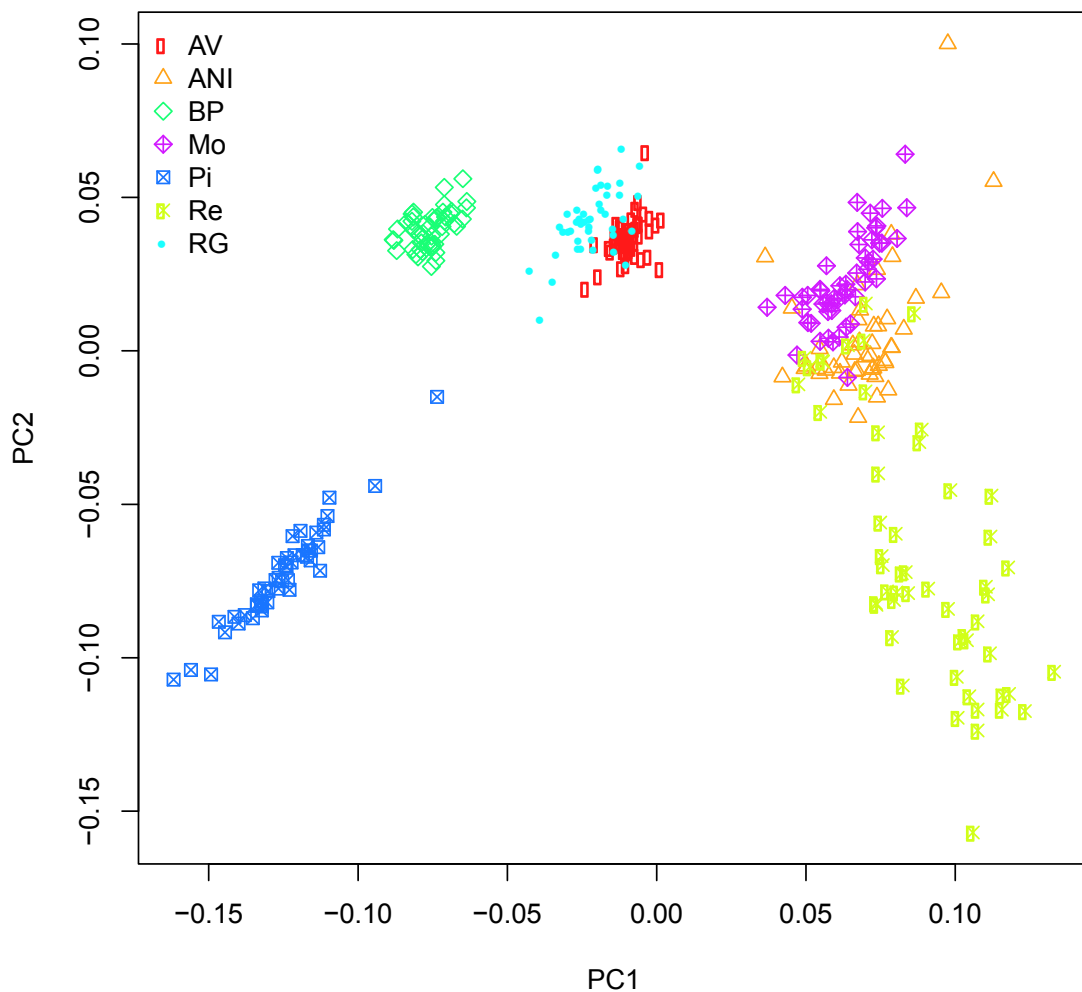


Figure 3. Different population groups defined by Principal component analysis.

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

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 pre-defined ancestral groups. The ancestral groups tested ranged from $K = 2$ to $K = 7$. The lowest cross validation errors (CVE) obtained were 0.39894 for $K = 6$ and 0.39907 for $K = 7$, indicating that these were the most parsimonious number of clusters. In particular, when $K = 7$, 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 (Figure 4). For example, AV contributed to all breeds, but mainly to Re, Mo, and RG. This confirms the PCA results, where the AV breed was placed in a central

location, closer to the other breeds (Figure 3). The contribution of ANI to Mo and Re is also important, possibly related to transhumance (seasonal migrations) of the ANI breed. Almost 20% of ANI breeders have adopted transhumance 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 one cluster, with some exchange of genes with AV and BP breeds, the two breeds more closely related to it. 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 where the BP breed was originated, to the AV breed. In comparison to 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 SNPs), 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 dataset after pruning for an individual call rate ≥ 0.95 and a SNP call rate ≥ 0.95 (717,172 SNPs), in order to check for the potential differences. The results for the N-J tree and the PCA were included in Supplementary Figures S2 and S3, 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 pruned data for LD. Bootstrap reached levels of 100% for all nodes (Supplementary Figure S2). In the principal component analysis, 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 SNPs. 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 Figure S3). For other analysis of divergence, higher variations were not observed. In general, the analysis did not suffer major changes and the classification of the populations did not differ between the two types of databases.

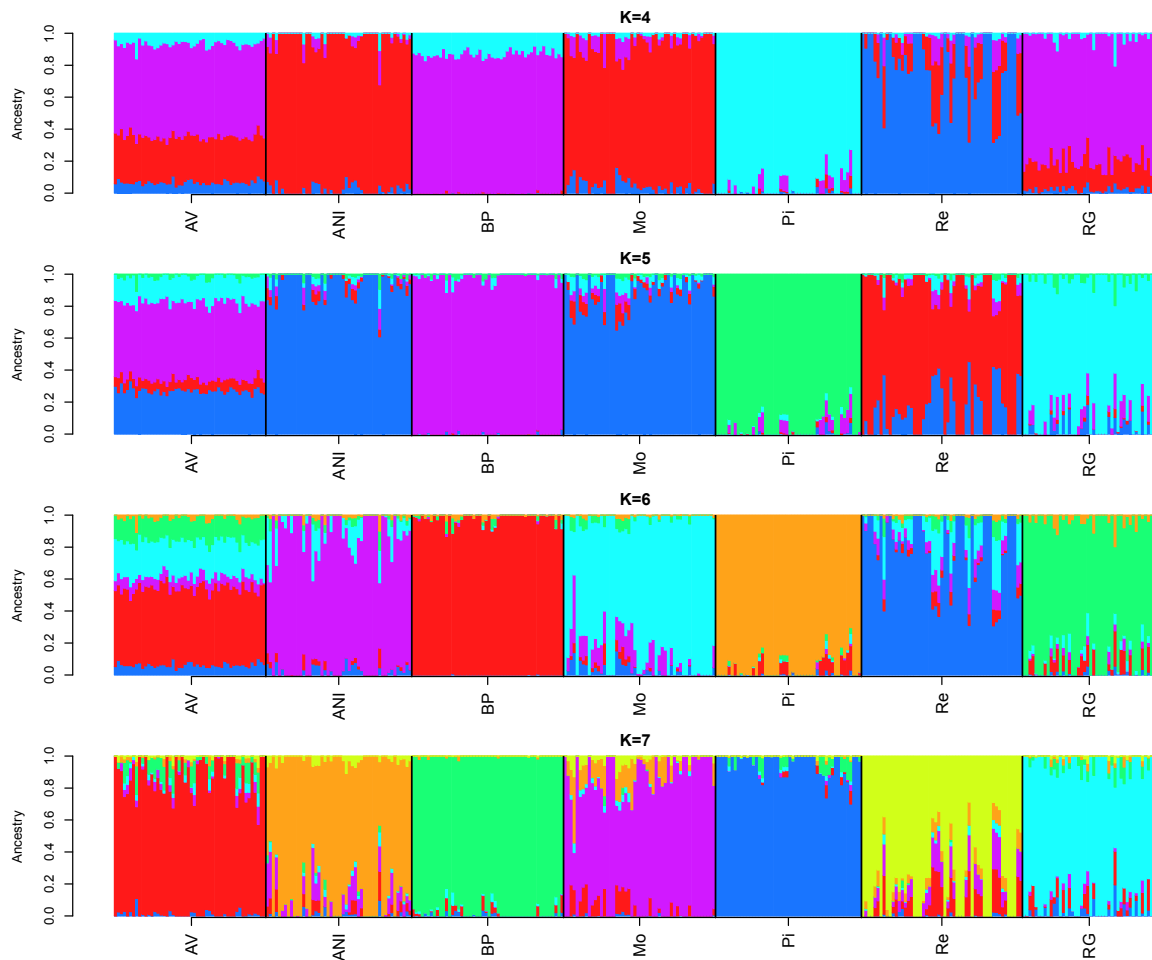


Figure 4. Estimated membership coefficients for each individual for $K = 4-7$.

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

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 within individuals, 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 University Press, Princeton.
- 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.
- 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
- 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 Genetics* 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. doi:10.1111/j.1755-0998.2010.02847.x
- 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.
- 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
- 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.
- 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. *ITEA- Informacion Tecnica Economica Agraria* 87A:50–64.
- Kidd, K. K., W. H. Stone, C. Crimella, C. Carezni, M. Casati, and G. Rognoni. 1980. Immunogenetic and population genetic analyses of Iberian cattle. *Anim. Blood Groups Biochem. Genet.* 11:21–38.
- 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. Soini, A. Mäki-Tanila, M. A. Toro, C. Díaz. 2014. Defining farmer typology to analyze the current state and development prospects of livestock breeds: The Avileña-Negra Ibérica beef cattle breed as a case study. *Livest. Sci.* 169:137–145. doi:10.1016/j.livsci.2014.09.003
- Martin-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. 20th Conference of the New Zealand Association for the Advancement of Animal Breeding and Genetics. Napier, New Zealand.
- 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. Kiszfali, B. I. Melegh, M. Bonin, L. 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.
- 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
- 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. Publicaciones de Extensión Agraria, Madrid.
- 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
- Uzzaman, R., Z. Edea, S. A. Bhuiyan, and J. Walker. 2014. Genome-wide single nucleotide polymorphism analyses reveal genetic diversity and structure of wild and domestic cattle in Bangladesh. *Asian-Australasian J. Anim. Sci.* 27:1381–1386. doi:10.5713/ajas.2014.14160
- 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:http://dx.doi.org/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.

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-r4.

SUPPLEMENTARY TABLES**Table S1. Specific F_{IS} indices in 7 Spanish beef cattle breeds**

Breed ¹	F_{IS}
AV	-0.00469
ANI	0.03218
BP	-0.00433
Mo	0.05187
Pi	-0.00856
Re	0.03811
RG	-0.01183

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

SUPPLEMENTARY FIGURES

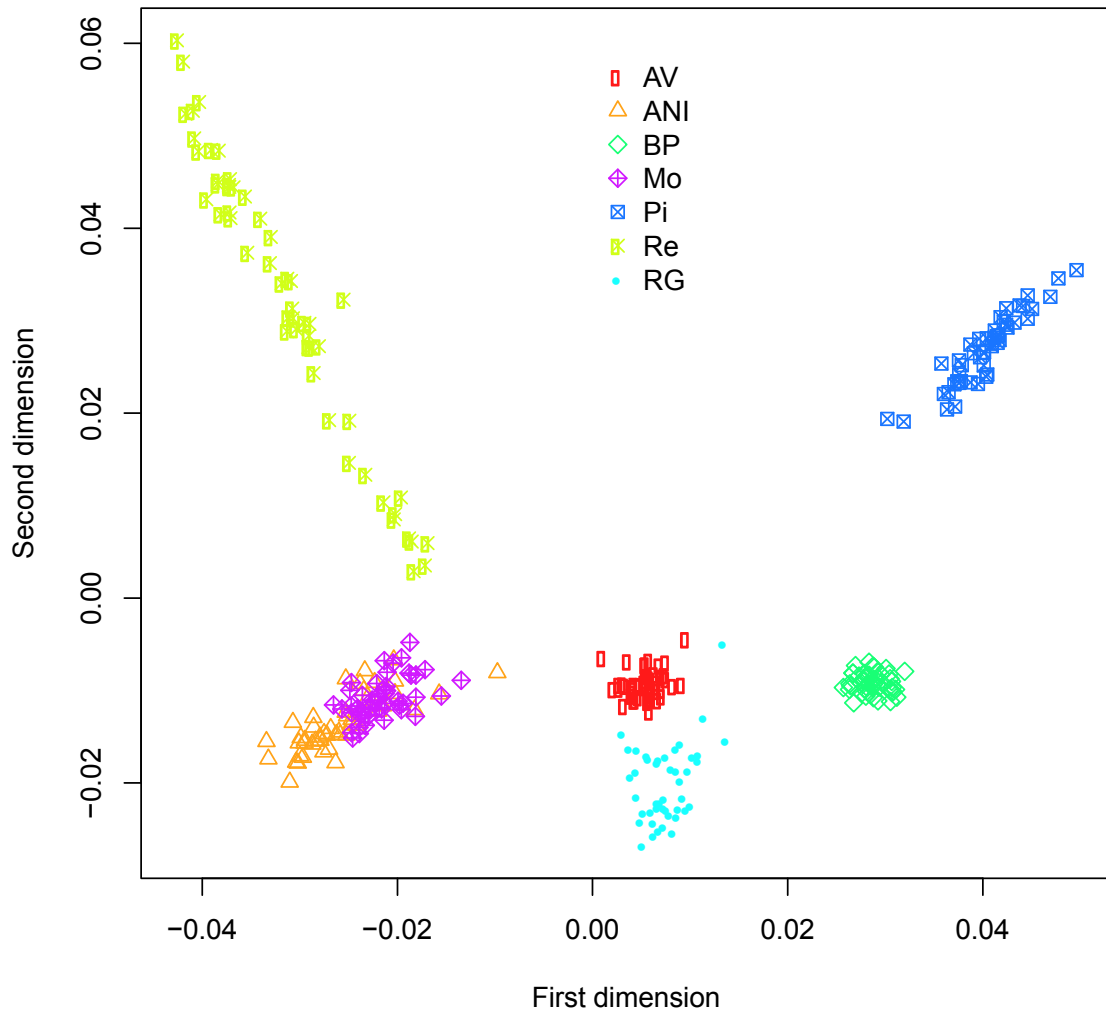


Figure S1. Different population groups as defined by a multidimensional scaling plot (MDS) on the relationship matrix

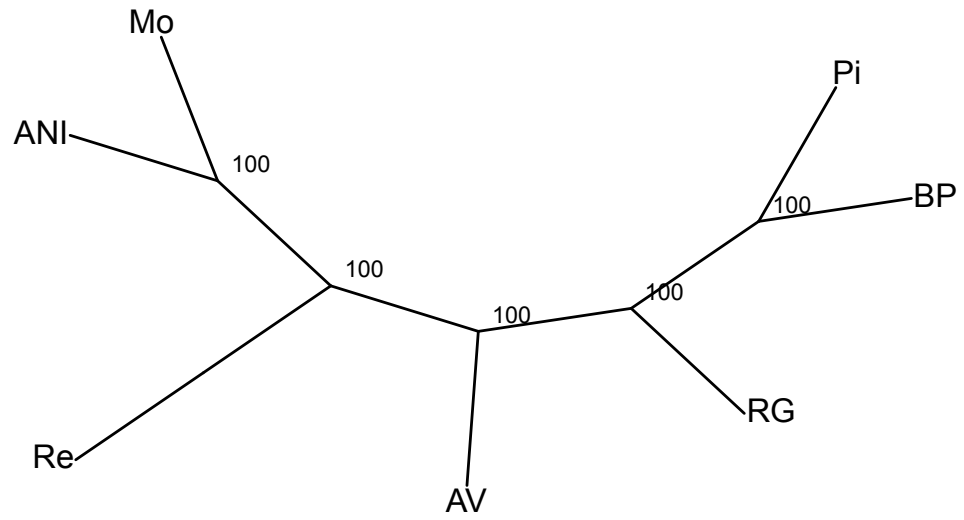


Figure S2. Neighbor joining representation of the pairwise Nei's D genetic distances among populations in the whole dataset

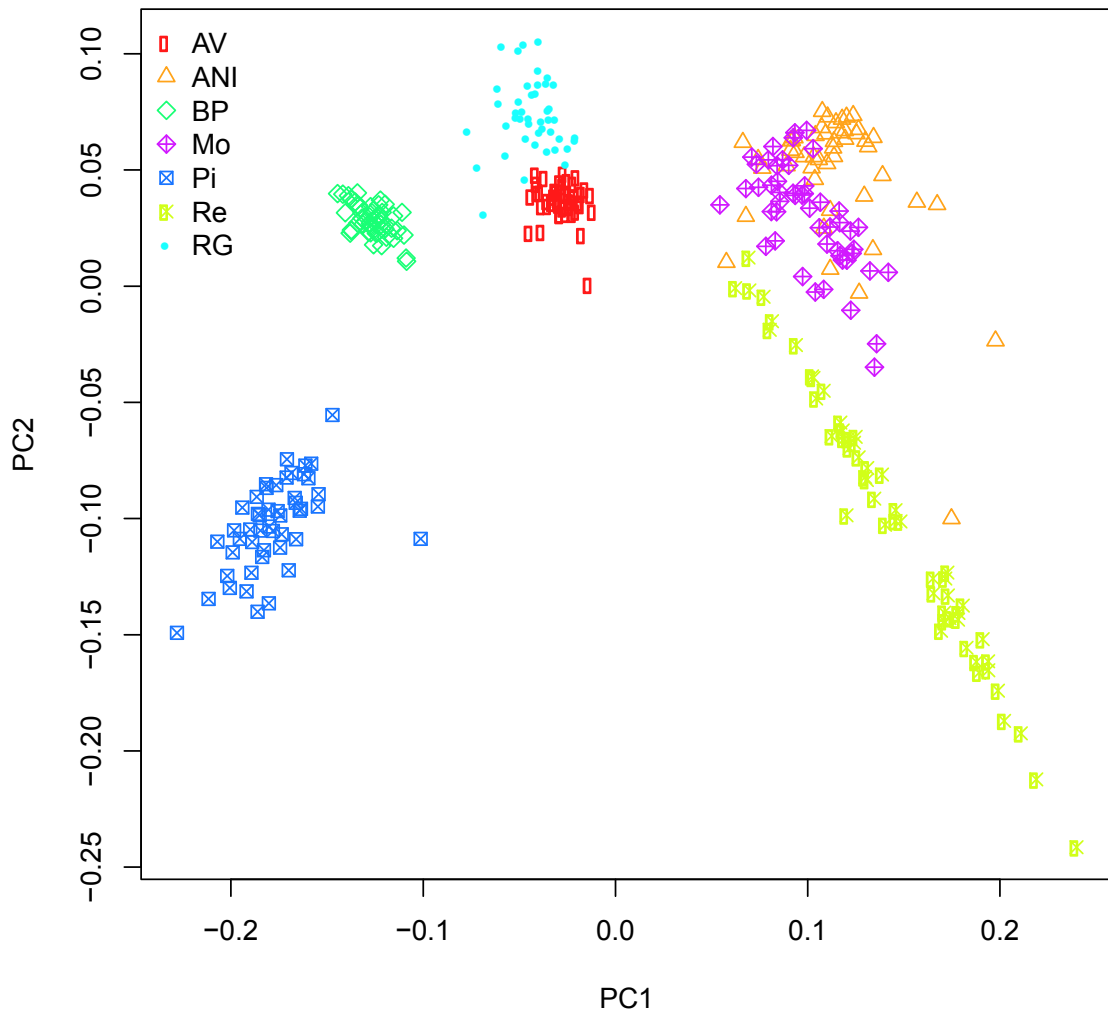


Figure S3. Different population groups defined with Principal component analysis in the whole dataset

Capítulo 6

LINKAGE DISEQUILIBRIUM, PERSISTENCE OF PHASE AND PAST EFFECTIVE POPULATION SIZE IN SPANISH BEEF BREEDS

Linkage disequilibrium, persistence of phase and effective population size of seven Spanish beef cattle breeds

Jhon Jacobo Cañas-Álvarez¹, Elena Flavia Mouresan², Luis Varona^{2,3}, Clara Díaz⁴, Antonio Molina⁵, Jesús Ángel Baro⁶, Juan Altarriba^{2,3}, María Jesús Carabaño⁴, Joaquim Casellas¹ and Jesús Piedrafita^{1*}

* Correspondence: jesus.piedrafita@uab.cat

¹Departament de Ciència Animal i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain

²Departamento de Anatomía, Embriología y Genética, Universidad de Zaragoza, 50013-Zaragoza, Spain

³Instituto de Biocomputación y Física de los Sistemas Complejos (BIFI). 50018. Zaragoza, Spain

⁴Departamento de Mejora Genética Animal, INIA, 28040-Madrid, Spain

⁵MERAGEM, Universidad de Córdoba, 14071 Córdoba, Spain

⁶Departamento de Ciencias Agroforestales, Universidad de Valladolid, 34004-Palencia, Spain

BMC Genomics, in preparation

Abstract

Background: Linkage disequilibrium (LD) and persistence of phase are fundamental tools for exploring the genetic basis of economically important traits in cattle, including the identification of QTLs for genomic selection and the estimation of effective population size (N_e) to determine the size of the training populations. In this study we have used the Illumina BovineHD (778 K) chip in 168 trios of seven Spanish beef cattle breeds to obtain an overview of the magnitude of LD and the persistence of LD phase through the physical distance between markers. Also, we have estimated the time of divergence based on persistence of LD phase and calculate past N_e from LD estimates using different alternatives to define the recombination rate.

Results: Estimates of average r^2 (as a measure of LD) for adjacent markers were close to 0.52 in the seven breeds, and decreased with the distance between markers, although in long distances still some LD remained (0.07 and 0.05 for markers 200 kb and 1000 kb apart, respectively). A panel with a minimum boundary of 38,000 SNPs would be necessary to reach a successful within breed genomic selection program. Persistence of phase, measured as the pairwise correlations between estimates of r in two breeds at short distances (10 kb) were in the range 0.89 to 0.94, and decreased from a range 0.33 to 0.52, to a range of 0.01 to 0.08 when marker distance increased from 200 to 1000 kb, respectively. The magnitude of the persistence of phase between the Spanish beef breeds was similar to those found in dairy breeds. For across breeds genomic selection, the size of the SNP panels must be in the range of 50,000 to 83,000 SNPs. Estimates of past effective population sizes showed values ranging from 26 to 31 for a generation ago in all breeds and models evaluated. Breeds diverged between 129 and 207 generations ago.

Conclusions: Linkage disequilibrium and persistence of phase estimates tended to decrease with increasing genetic distance in all studied populations. The high values of LD at short distances, as well the rapid decline of LD, was reflected in the low past effective population size of the Spanish beef cattle populations. The results of this study are relevant for the future implementation of within breed and across breed genomic selection programs in the Spanish beef cattle populations. They show that a reduced size of the SNP panel would be enough to achieve an adequate precision of the genomic predictions.

Keywords: Local breeds, Beef, High-density SNP chip, Linkage Disequilibrium, Persistence of phase, Effective population size, Time since divergence.

Background

Genomic selection was proposed as a strategy to predict breeding values from a massive set of markers across the genome (Meuwissen et al., 2001). It has been shown that the performance of genomic selection depends upon the existence of enough linkage disequilibrium (LD) between markers and QTLs such that the marker allele-QTL phase persists across generations (de Roos et al., 2008). In this context it is also important to determine if the extent of the LD is consistent between the reference population used for estimation, and the population in which selection is to be practiced (Dekkers and Hospital, 2002; Goddard et al., 2006). In addition, the persistence of LD phase across physical distance between markers for two populations can be used to predict the required marker density for a genome-wide association study or genomic selection (Andreescu et al., 2007; de Roos et al., 2008). SNPs have been used as markers and SNP chips of different densities have been developed in all domestic species, particularly in cattle (Matukumalli et al., 2009). As a previous step towards the implementation of genomic selection programs, the extent and persistence of LD has been assessed through SNP chips in several dairy and beef cattle, zebu and African cattle and also in composite breeds, (Gautier et al., 2007; de Roos et al., 2008; Villa-Angulo et al., 2009). However only a few studies have documented the behaviour of LD using the expanded set of 778 000 SNPs available on the BovineHD platform (Illumina Inc, San Diego) (Espigolan et al., 2013; Salomon-Torres et al., 2014; Porto-Neto et al., 2014). One of the significant advances of this denser chip is that it allows for an accurate estimation of LD over short physical distances as it contains many more marker pairs separated by 10 kb or less.

Past effective population size (N_e) provides useful information on the evolution of the population. It is taken as a reference to track the loss of genetic variability (Hayes et al., 2003) and is useful for calculating the size of the training population in genomic selection schemes (Goddard, 2009). Usually, N_e has been estimated from the analysis of pedigrees registered in herdbooks. Those estimates have a disadvantage because it is not possible to estimate N_e many generations ago due to the limitations of the database. However, we can infer N_e depending on the strength of LD at different genetic distances between loci (Sved, 1971). A number of studies have been conducted to evaluate the current and historical N_e of a variety of breeds of cattle (de Roos et al., 2008; Qanbari et al., 2010) from SNP data, but up to now, the estimates of N_e in populations of native

Spanish beef cattle are limited to those calculated from pedigree data (Gutiérrez et al., 2003; Cañas-Álvarez et al., 2014).

The main objective of this research is to assess the magnitude of LD and the persistence of phase among seven Spanish autochthonous beef breeds, and to estimate past effective population size from LD, measured from data of a High-density (778 K) chip. This can provide information from the behavior of these parameters in populations of beef cattle of limited censuses and under a non-intensive directional selection. From this overview we also attempt to arrive to recommendations on the SNP panel size needed for both within and across-breed genomic selection.

Results and discussion

Linkage disequilibrium

Linkage disequilibrium maps are fundamental tools for exploring the genetic basis of economically important traits in livestock species (Andreescu et al., 2007; McKay et al., 2007). To estimate LD, we used the parental haplotypes of the autosomes of 168 sire/dam/offspring trios of seven Spanish autochthonous beef breeds. After quality control, we had 629,869 markers per breed distributed across the 29 autosomal chromosomes (80.96% of the initial information). Those covered a total of 2,508,474.8 kb with one marker each 3.98 kb on average. LD was estimated as the pairwise r^2 statistic among all SNPs on the same chromosome located at distances between 0 and 2 Mb. Average r^2 at a variety of distances was computed for each breed by stacking all SNPs in classes of 5 kb across all autosomes. The number of haplotype pairs considered to estimate average r^2 ranged from ~ 920,000 for markers 5 kb apart, to ~ 760,000 for markers 2 Mb apart.

Average r^2 tended to decrease with increasing marker distance in all studied populations, the most rapid decline being seen over the first 100 kb. On average, LD decreased from 0.58, up to 0.32, 0.18, 0.11, 0.07 and 0.045 for 5 kb, 20 kb, 50 kb, 100 kb, 200 kb and 1Mb of marker distance, respectively (Table 1). From 1Mb to 2Mb of marker distance, the estimates almost did not change (not shown). The standard deviations of LD estimates were similar among breeds, ranging around 0.39 at distances of 5 kb, and around 0.05 at distances of 2 Mb (Table 1). The standard deviations diminished in parallel to the LD estimates, although they were greater in comparison to the LD estimates at longer distances. The larger standard deviations at shorter distances can be related to the shape of the distribution of r^2 values as a function of distance (Additional file 1). For adjacent

markers, 27-32% of complete LD ($r^2 > 0.95$) estimates were found across breeds, followed by markers close to complete equilibrium ($r^2 < 0.05$), with percentages in the range 16-17%. However, the other estimates were almost evenly distributed between these extreme values. As the distance between markers increases, the proportion of pairwise markers in complete LD decreases and beyond 100 kb no markers in complete LD are found.

Table 1 Average r^2 values and standard deviations across all autosomes at various marker distances in seven Spanish beef breeds

Breed¹	5 kb²	20 kb	50 kb	100 kb	200 kb	1 Mb	2 Mb
	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)	(sd)
AV	0.571 (0.388)	0.309 (0.328)	0.165 (0.230)	0.092 (0.152)	0.052 (0.088)	0.031 (0.044)	0.029 (0.040)
ANI	0.582 (0.395)	0.328 (0.341)	0.188 (0.250)	0.116 (0.174)	0.076 (0.115)	0.047 (0.067)	0.042 (0.058)
BP	0.579 (0.392)	0.320 (0.336)	0.179 (0.242)	0.107 (0.167)	0.066 (0.105)	0.038 (0.054)	0.033 (0.046)
Mo	0.577 (0.395)	0.319 (0.337)	0.180 (0.243)	0.108 (0.167)	0.067 (0.105)	0.041 (0.058)	0.037 (0.052)
Pi	0.593 (0.396)	0.340 (0.348)	0.201 (0.260)	0.129 (0.187)	0.088 (0.129)	0.058 (0.080)	0.051 (0.071)
Re	0.583 (0.396)	0.328 (0.342)	0.190 (0.251)	0.120 (0.177)	0.079 (0.120)	0.053 (0.075)	0.047 (0.066)
RG	0.581 (0.394)	0.325 (0.338)	0.185 (0.245)	0.115 (0.172)	0.075 (0.114)	0.053 (0.073)	0.049 (0.067)

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

²The averages correspond to intervals including a range of 5 kb below the specified number.

Some authors have investigated the potential effect of MAF on LD estimates (Lu et al., 2010; Espigolan et al., 2013). In fact, it has been shown that SNPs of divergent MAFs on average have different LD properties (Pritchard and Przeworski, 2001). Thus, to determine how estimates of LD could be affected by MAF in these populations a more restrictive MAF (<0.15) was imposed for the quality control. As expected, larger LD estimates were obtained (around 15% to 20% of increase for distances of 5 to 100 kb apart and less for longer distances, results not shown in tables), because the differences among MAF of pairs of loci are lower. A similar pattern was observed by Espigolan et al. (2013)

in Nellore cattle where they found an increase in LD estimates depending upon the MAF criterion. Furthermore, our results confirm that the dependence of r^2 on distance is stronger than its dependence on the difference in frequencies (results not shown), as found by Qanbari et al. (2010).

There were differences in average r^2 at a particular marker distance depending upon chromosomes. This is shown in Additional file 2, where average r^2 in all breeds for each chromosome at different marker distances are represented as circles of different colours. This variability among chromosomes leads to some changes in the ranking of LD among breeds when only the markers at a particular chromosome are considered. Such differences at chromosome level have been observed in different species like cattle (Lu et al., 2010) and chicken (Andreescu et al., 2007), and to some extent could be attributed to changes in density of markers among different chromosomes (Andreescu et al., 2007). However, differences among chromosomes also seemed to depend on breeds. If this is the case, it seems that causes of LD acted differently at specific genomic regions generating genetic distances at singular locations among these populations (González-Rodríguez et al., 2014). A wide variation in autosomal recombination rates can lead to a marked diversity in the pattern of LD in different genomic regions (Arias et al., 2009) and chromosomes, but the differences can be also due to heterozygosity levels, genetic drift or selection (Qanbari et al., 2010). The differences found suggest additional studies to identify particular regions that may exhibit different degrees of LD and to evaluate its potential use for detection of QTLs and signatures of selection (Perez O'Brien et al., 2014) and their consequences for the implementation of a genomic selection program.

Although LD in principle seems fairly simple in concept, in a more general context, the actual LD estimates and their relationship to the physical distance between loci are further influenced by a series of interrelated factors such as demographic history and genetic drift, admixture or migrations, population subdivision, selection (hitchhiking effect, epistatic selection), variable mutation rates and gene conversion (Ardlie et al., 2002). The differences in LD estimates across Spanish beef breeds were very small at short distances, however they had a mild increase with distances greater than 50 kb although the rank of the average r^2 was kept the same across distance with Pi as the breed with the greatest LD and AV with the lowest (Table 1, Figure 1). This suggests that the demographic and evolutionary history of the breeds has been rather similar, which can be

related to the results of a parallel study of variability where AV and Pi had the highest and lowest values of expected heterozygosity, respectively (Cañas-Álvarez et al., 2015).

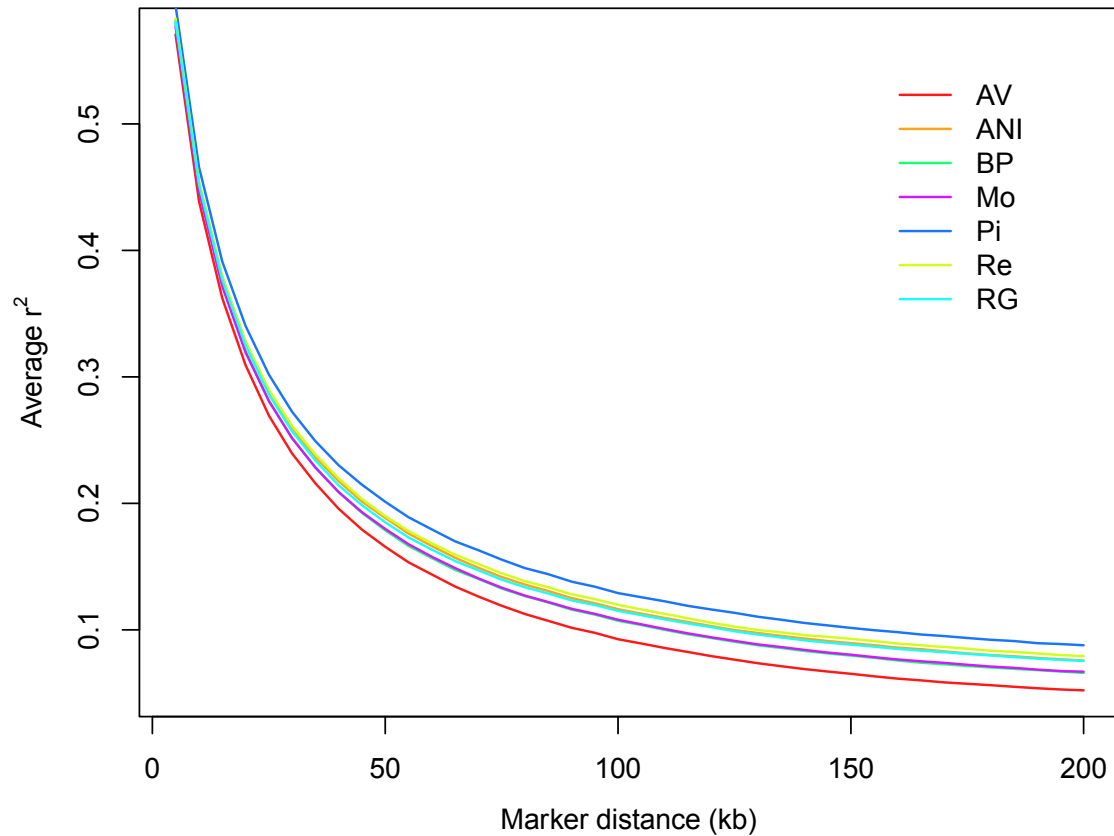


Figure 1 Evolution of average r^2 from markers 0 to 200 kb apart across all autosomes in seven Spanish beef breeds.

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

Our results are in the range of those obtained in Australian taurine cattle (Porto-Neto et al., 2014) and other cattle populations. A study of 19 breeds including European and British dairy and beef cattle, African and Indicine cattle, and composite breeds showed LD values ranging from 0.37 to 0.65 at 5 kb of distance between markers and from 0.08 to 0.25 for markers 10 kb apart (Salomon-Torres et al., 2014). In Nellore, the LD estimates were lower than ours at short intervals (5 kb) and similar at a distance between markers 100 kb apart (Espigolan et al., 2013). In both studies, they used a 778 Illumina chip. With a smaller SNP chip (50 K), Qanbari et al. (2010) found in German Holstein-Friesian cattle a lower LD than ours at short distances, but greater at distances of 75-120 kb. The causes of the differences can be due to some ascertainment bias, mainly in the Nellore study, but

these differences also probably reflect different evolutionary histories and selection practices. On the other hand, the extent of LD in our study was close to the one estimated in humans (Tenesa et al., 2007), but lower than in horses (Corbin et al., 2010) and pigs (Badke et al., 2012) at distances until 100 kb, 500 kb, and 1000 kb, respectively. The comparison of LD levels between studies however, should be taken with caution, as LD depends on sample sizes, marker densities and recent and historical population demographics (Pritchard and Przeworski, 2001).

The decay of LD in a genome determines the power of QTL detection in genome wide association studies and indicates the required marker density. It was shown that in indirect association studies, the sample size must be increased by roughly $1/r^2$ when compared with the sample size for detecting the causal mutation directly (Kruglyak, 1999; Pritchard and Przeworski, 2001). In the context of genomic selection, the accuracy depends both on the amount of LD between QTL and markers and the number of records available to estimate marker effects (Toosi et al., 2010). We computed LD between adjacent markers depending upon the proportion of the number of SNPs retained in the analysis (Table 2). The average r^2 across all chromosomes was 0.52 for adjacent markers in the complete dataset, with an inter-marker distance of 4 kb, and decreased to 0.30, 0.22 and 0.10 when 10%, 5% and 1% of the markers, randomly chosen, were included in the analysis. As expected, great differences were not observed among the seven breeds evaluated.

An accuracy of predicted breeding values from dense markers up to 85% can be obtained with an average r^2 between adjacent markers of 0.2 (Meuwissen et al., 2001). In our breeds, a similar average r^2 between adjacent markers would be obtained by using only 5% of the markers of the HD panel (38,000 markers), which corresponds to an average genomic distance of 80 kb (Table 2). However, given that markers with $r^2 = 1$ will be excluded in genomic selection, and because of the high variability of r^2 at small distances (Corbin et al., 2010), this is probably an underestimation of the actual number of SNPs needed. Furthermore, this number can vary according to the relationship between training and evaluation animals (Meuwissen, 2009).

Persistence of LD phase

For genomic selection, the value of markers in a population other than the reference population will depend on the persistence of LD phase between the reference population and the second population (Dekkers and Hospital, 2002). The persistence of phase between

two populations for a given marker distance has been calculated as the correlation of the r estimates, i.e., the signed squared roots of r^2 . The correlations of r between populations are a result of their genetic relationship and history.

Table 2 Average r^2 (average distance in kb) between adjacent markers across all autosomes depending upon the proportion of SNPs retained in seven Spanish beef breeds

Breed ¹	100%	10%	5%	1%
AV	0.509	0.284	0.208	0.087
	(4.03)	(40.20)	(80.19)	(401.63)
ANI	0.521	0.302	0.228	0.110
	(4.09)	(40.85)	(81.34)	(407.54)
BP	0.518	0.295	0.218	0.098
	(4.08)	(40.65)	(81.07)	(405.90)
Mo	0.516	0.295	0.218	0.099
	(4.09)	(40.75)	(81.27)	(405.30)
Pi	0.533	0.314	0.239	0.120
	(4.13)	(41.26)	(82.20)	(411.16)
Re	0.523	0.304	0.229	0.113
	(4.11)	(40.97)	(81.57)	(406.95)
RG	0.520	0.299	0.226	0.110
	(4.08)	(40.74)	(81.02)	(406.34)

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

We computed the correlation of r for each marker pair across all chromosomes among all pairwise populations over a series of genomic distance intervals using bins of 10 kb from 0 to 100 kb of inter-marker distance, and of 100 kb from 100 kb to 1000 kb. The observed correlation of r between two populations decreased with increasing marker distance, with a similar behavior to LD (Figure 2). At 10 kb apart the persistence was high and positive, ranging from 0.89 to 0.94, which means that at short distances there is an excess of similar haplotypes in the same phases among all populations. With increasing distance the persistence diminished to the ranges 0.70-0.82, 0.49-0.66 and 0.34-0.52 for markers 50 kb, 100 kb and 200 kb apart, respectively. It is worth to mention also that the estimates of persistence of phase are averages over a certain bin. A positive average

estimate does not mean that all estimates of persistence are positive; in fact at all marker distances we have found negative estimates. The percentage of negative estimates (reversed phases) increased with marker distance, may be as a consequence of recent events, reaching values across breeds of 29-30%, 36-38%, 40-43% and 43-48% for average marker intervals of 20 kb, 50 kb, 100 kb and 1 Mb.

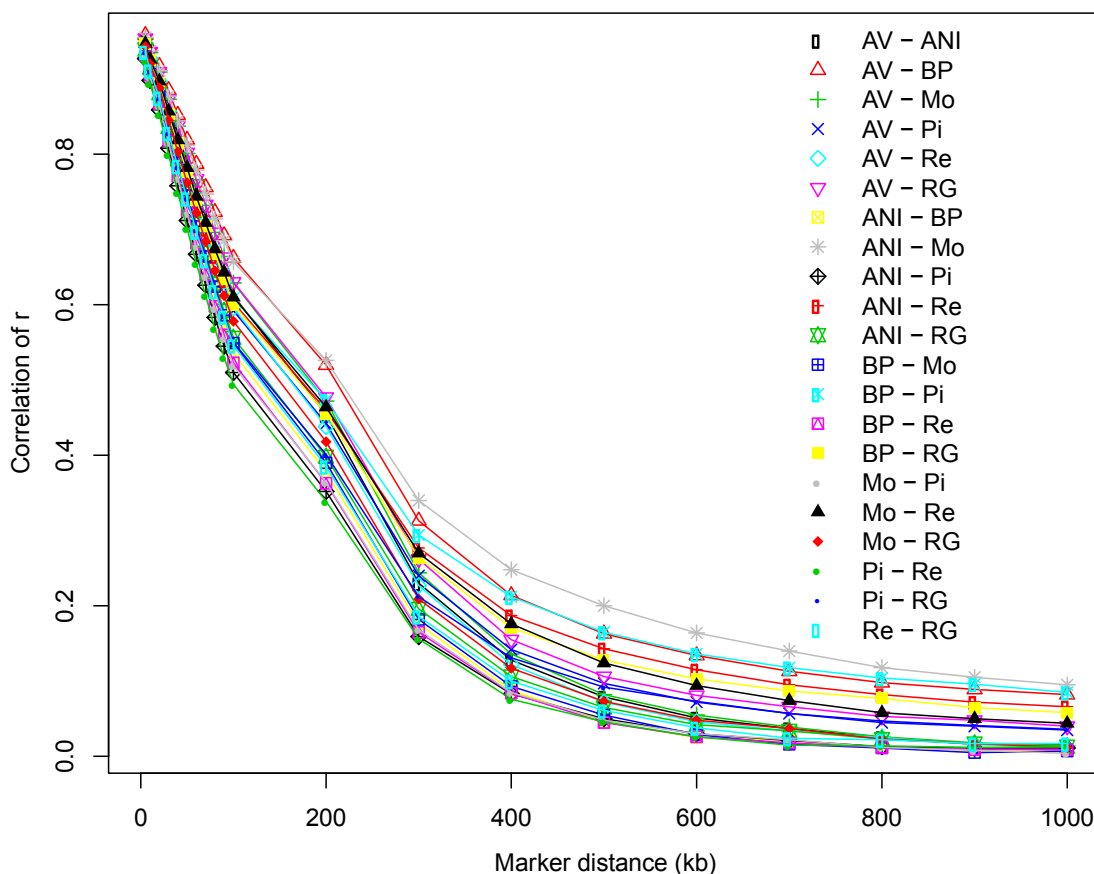


Figure 2 Evolution of pairwise persistence of phase across all autosomes in seven Spanish beef breeds.

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

The comparison of Heatmap graphics corresponding to two of the distances analyzed confirms that there is a lower persistence of phase (lighter color) for greater distances among markers (Figure 3). There are also differences between the pairwise persistence between breeds across distances. At the two distances shown, but also at distances not shown in figures, the group formed by ANI, Mo and Re breeds presents the greater values of persistence among them, indicating their genetic proximity, whereas the persistence is

the lowest between Pi and BP and the rest of the breeds. AV, on the contrary, has a similar pattern of persistence with the rest of the breeds, confirming its central position among the Spanish breeds seen by Cañas-Álvarez et al. (2015). The Heatmap also allowed us to present the grouping of breeds by a dendrogram, synthesizing the divergence among breeds. At the different distances, the same two clusters are clearly stated. The first one includes ANI, Mo and Re, whereas the second one includes the rest of the breeds. This second cluster, however, does not provide a consistent grouping with the relationships described in a previous study in these breeds (Cañas-Álvarez et al., 2014).

The estimates of persistence of phase in the Spanish beef breeds were in the upper range of those observed by (Gautier et al., 2007) at short marker distances using a 1536 SNP panel. At greater distances, up to 500 kb, the distances were in general lower. De Roos et al. (2008) found correlations of 0.97 for marker pairs 5 kb apart in closely related populations (Black and white Holstein-Friesian bulls and Red and white Holstein-Friesian Bulls), but greater than 0.80 for less related cattle (Australian Holstein and Australian Angus), using panels ranging from 1252 to 5237 SNPs. The correlation decreased with the distance in proportion to the degree of divergence of the breeds considered. In pigs, however, the persistence continued over larger distances, probably due to higher selection pressure and smaller effective population size (Harmegnies et al., 2006).

We have described that at short distances correlations of r between breeds were very high, although it diminished rapidly at longer distances. If we want to find markers that work consistently in two populations we need to set a minimum correlation of 0.8 (as in de Roos et al., 2008). Similar to some results of these authors, in the Spanish beef breeds, correlations greater than 0.8 are found for markers apart up to 30 to 50 kb, depending upon the pair of breeds considered (not shown in tables). Because of that, the number of markers equally distributed across the genome must be in the range of 50,000 to 83,000 depending upon the breed. This implies that for using SNP information across breeds in genomic selection, for some pairs of breeds we would need to use at least twice the number of markers needed for within-population genomic selection which was set previously in a minimum of 38,000 SNPs.

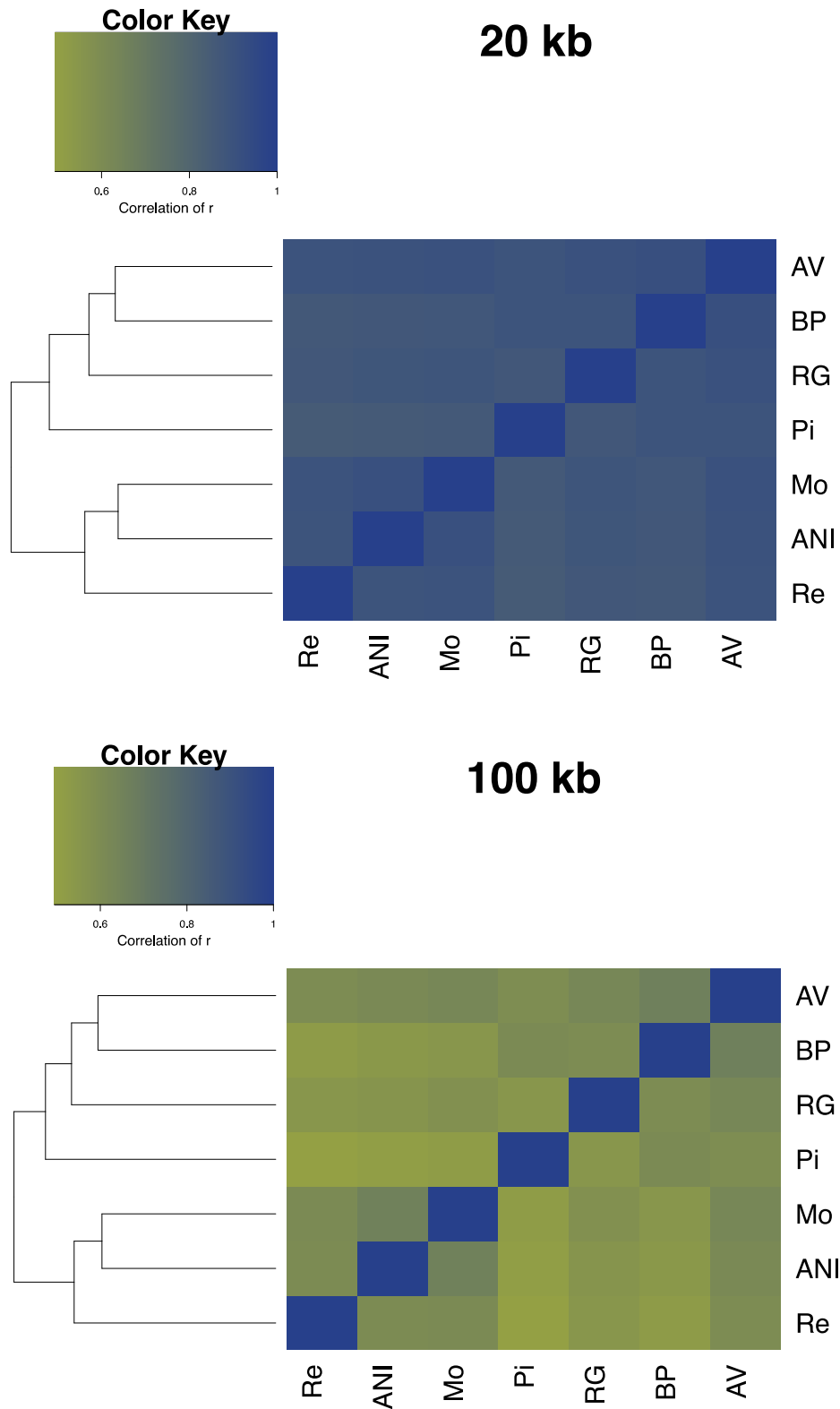


Figure 3 Persistence of phase matrices across all autosomes among seven Spanish beef breeds evaluated at 20 and 100 kb

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

Time since divergence between breeds

The slope of the persistence of phase, i.e., the correlation of r , has a strong implication in determining the time in which possibly the breeds diverged. Specifically, for two breeds, k and k' , we estimated the time of divergence (T) as from the approximation $r_{k,k'} = e^{-2cT}$, where $r_{k,k'}$ is the correlation of r , c is the recombination distance between markers and T is time since the breeds diverged in generations (de Roos et al., 2008). Because there are differences among chromosomes in recombination rates, Arias et al. (2009) demonstrated that the average genetic distance (cM) per Mb was equal to 1.25. Using this ratio, the breeds diverged between 129.4 (AV - Mo pair) and 206.8 generations ago (Pi - Re pair) (Additional file 3). Pi breed diverged first from the rest of breeds, whereas AV showed the lowest times of divergence with the other breeds. The results parallel the observed ones in the graphics of persistence of phase between populations and to the neighbor joining representation of the pairwise Nei's distances (Cañas-Álvarez et al., 2015). The range of divergence times found in our study was lower than that observed for the pair Australian Holstein and Australian Angus ($T = 364$), breeds specialized in dairy and meat production, respectively. However, our estimates were in the order of the divergence time between two dairy breeds: New Zealand Friesian and New Zealand Jersey ($T = 191$) (de Roos et al., 2008). These results would suggest that the divergence between dairy and beef breeds preceded the subsequent divergence among breeds of the same aptitude.

If we consider an approximate average interval of 6.5 yr for the Spanish beef breeds (Cañas-Álvarez et al., 2014), their divergence occurred between ~ 841 and 1344 years ago, which corresponds to the years of the first half of the Medieval period. This fact could be consistent with Spanish history at those times, when there was a political subdivision between Christian and Muslim kingdoms at the North and South of the country, respectively. The populations located in the ancient Muslim territory (Re-Mo-Av) were surely more influenced by the African populations (Decker et al., 2014), whereas the populations located in North Spain were more related with the rest of the European populations. However, after the political reunification of the Spanish kingdoms in 1492, the contact between populations were restored by the instauration of the transhumance, that implies a strong movement of animals, in special of the AV, explaining the central position of that population (Cañas-Álvarez et al., 2014).

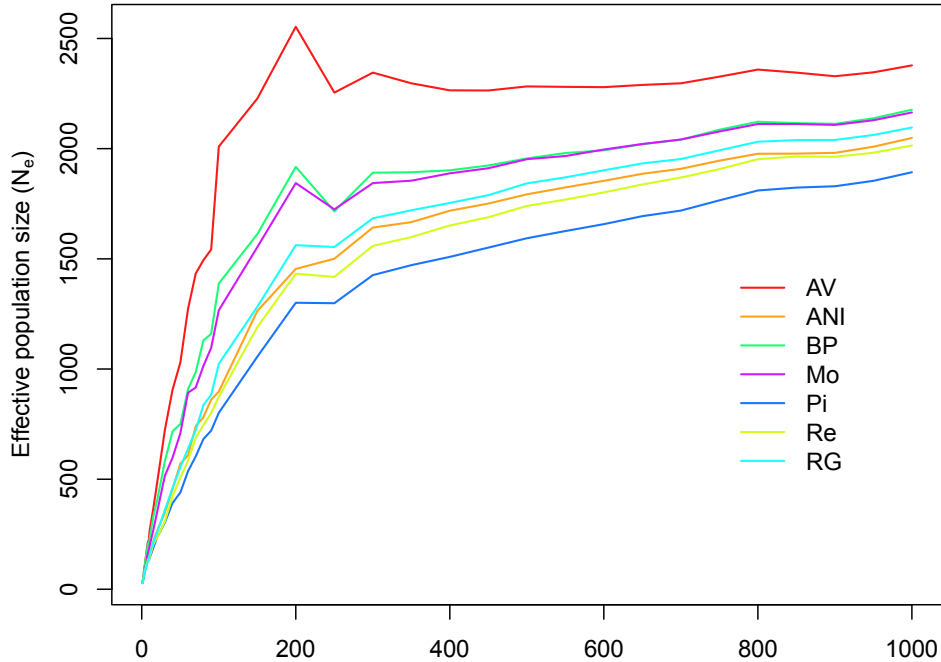
Past effective population size

We have estimated the effective population size, N_e , at different generations in the past from LD estimates using different approaches. The first one assumes the generalization widely used in the literature $1\text{Mb} = 1\text{cM}$ (M1). The second takes into account the specific variation in the rate of recombination among chromosomes as in Arias et al. (2009) (M2). Both estimates are based in the Sved's (1971) formula on the expectation of r^2 . The third approach is built on the Tenesa et al. (2007) formula and estimates iteratively the values of effective population size and the existence of mutation (a parameter) (M3). The use of a different recombination rate among chromosomes had a limited impact on the estimation of N_e . M2 tended to give similar values to those of M1, with inconsistent variations across breeds. As expected, the differences increased with the number of generations (Additional file 4). The N_e estimates in M2 oscillated between 2079.9 and 2487.8 for Pi and AV, respectively, at 1500 generations ago, whereas they were of 26.3 to 31.4 one generation ago for RG and Re, respectively. Model 3 (M3) gave in general N_e estimates lower than those of M1 and M2. Specifically, for the same generation and breeds, the estimates of N_e were 1753.1 and 2171.1 at 1500 generations ago, and 24.5 and 31 in RG and BP for the previous generation. Parameter a determines the value of expected r^2 when the distance between markers is effectively 0 (Corbin et al., 2010). We have found estimates of a ranging from 1.0 to 2.1, with an average of 1.4. This would support, as in Corbin et al. (2010) in Thoroughbred horses, the alternative version of Sved's (1971) equation derived by Tenesa et al. (2007).

The results of historical N_e estimated with model M2 from 1,000 generations ago to the present time are shown graphically in Figure 4a. At the beginning of the interval of time studied, matching with the domestication period that started in the Neolithic, N_e estimates oscillated from ~1900 to ~2300 for AV and Pi, respectively. Except for AV, a slow decline was observed until 200 generations ago (about 1,300 years ago), when N_e have diminished to 1300 – 2500 depending upon the breed. From this period, corresponding with the time the breeds diverged, a sharp decline in population size has continued until today. AV and BP breeds showed the greatest values of N_e along all generations, particularly in the latest ten ones (Figure 4b). The N_e prior to the current generation stood at around 26-31 in all breeds studied in this work. These values are lower than the ones obtained from the analysis of pedigree records in the same breeds (Cañas-

Álvarez et al., 2014), what is quite reasonable because pedigree files allowed us to trace back a limited number of generations.

a)



b)

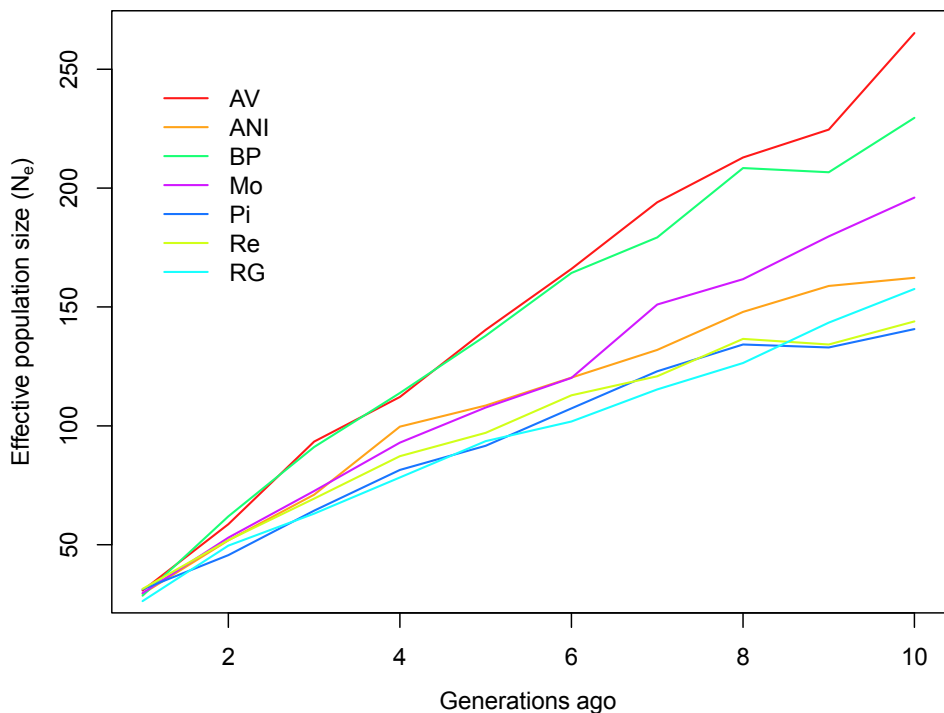


Figure 4 Effective population size estimated using specific recombination rates per chromosome (M2) in the periods 0 to 1000 generations ago (a) and 0 to 10 generations ago (b).

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

A way to test the performance of the three methods in the short term is to compare the N_e estimates with those found from average r^2 between nonsyntenic chromosomes to obtain an empirical null distribution. This represents an approximation of the LD that can be obtained by chance, assuming that the markers used have not undergone simultaneous selection. The mean of nonsyntenic r^2 values reflects both sampling of animals and genetic drift, and may be expected to lower down with increases in both sample size and N_e (Corbin et al., 2010). The estimates found in this study were very similar across breeds, in the range 0.010 to 0.011 (Additional table 5). All LD values between nonsyntenic markers are within the interval of LD at long distances in these breeds (Table 1). The values observed here are much greater than those observed in Australian Holstein-Friesian cattle (Khatkar et al., 2010) and in Thoroughbred horses (Corbin et al., 2010). The comparatively high values observed in our breeds suggest that some of the LD created by admixture during breed evolution (Cañas-Álvarez et al., 2015) has been kept until the present time. With respect to the effective size, assuming $c = 0.5$ for the nonsyntenic loci, N_e estimates ranged from 43 (Re) to 47.5 (AV and BP). This would suggest that all the three methods used above provide downwardly biased estimates of effective population size in the last generations or that estimates of N_e very much depend on how well the average LD estimates represent the whole LD patterns in the genome.

Estimates of effective population size lower than those presented in our work were obtained in European and West African breeds, where N_e estimates near to 500 for 50 generations ago and of 80 for five generations ago were found (Gautier et al., 2007; de Roos et al., 2008; Qanbari et al., 2010). The observed drop in N_e potentially represents a number of scenarios, including some founder event, the formation of different breeds and recently artificial reproduction techniques or a combination of any of them (Wang, 2005). To that decline, bottlenecks caused by the Great Famine of 1,315-1,322 in Europe could also have contributed, as reported in other breeds of cattle (Gautier et al., 2007; Villa-Angulo et al., 2009). Regarding the breeds studied, the different N_e found can be also the result of the initial confinement of the populations in different geographical areas, a drop of population censuses due to the partial replacement and un-controlled crossbreeding of autochthonous breeds with foreign breeds occurred in the middles of the XX century (García-Dory, 1986), and some genetic selection for specific characteristics.

Effective population size also determines the accuracy of genomic selection (Daetwyler et al., 2010), as that accuracy can be expressed as a function of the effective

number of loci (M_e) in a population (Goddard, 2009). M_e relates to the number of independent chromosome segments and equals the summation across chromosomes of $2N_eL/(\log 4N_eL)$, L being the genome size in Morgans for each specific chromosome. In our case, with chromosome lengths as in Arias et al. (2009), and the current effective sizes calculated from nonsyntenic LD estimates, M_e was in the range of 509 (Re) and 552 (AV and BP) (Additional file 5). From this value, using the deterministic formula of Daetwyler et al. (2010), we can estimate the size of the training population under different heritability scenarios.

Conclusions

We present the results of an overview of LD of Spanish beef cattle breeds using a high-density SNP panel. We observed that LD tended to decrease with increasing genetic distance in all studied populations, such that r^2 averages decreased from 0.58 to 0.07 in marker distances of 5 kb and 200 kb, respectively. Similarly, the persistence of phase was high at short distances and decreased at a moderate rate. A minimum of 38,000 and a maximum of 83,000 SNP markers would be needed for within breed and across-breed genomic evaluation, respectively. The time of divergence between the breeds analyzed ranged from 129 to 206 generations. Effective population sizes of the breeds were obtained from the expectation of LD estimates. We found a substantial decrease in the effective size of the Spanish native beef cattle populations starting 200 generations ago, and the N_e in the generation previous to the current one was estimated to be in the interval 26-47 depending upon the method of estimation and the breed involved. These estimates allow us to predict the size of the training population in different heritability scenarios.

Methods

Animal samples and SNP genotyping

A total of 504 animals belonging to 168 trios (sire/dam/son) were chosen as less unrelated as possible from seven Spanish beef cattle breeds, including Asturiana de los Valles (AV; N=75), Avileña-Negra Ibérica (ANI; N=72), Bruna dels Pirineus (BP; N=75), Morucha (Mo; N=75), Pirenaica (Pi; N=72), Retinta (Re; N=69) and Rubia Gallega (RG; N=66). 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 of Refinement (1993). Genomic DNA was extracted following the protocol described in Prefiler™

Forensic DNA Kit of Applied Biosystems, using automated equipment Mag-Max™ Express 96-Magnetic Particle Processor. High-density genotypes (777,962 SNPs) were obtained by using the BovineHD Genotyping BeadChip (Illumina, Inc 2012). The SNPs were mapped in the UMD3.1 assembly (Zimin et al., 2009). All the laboratory work was done by Xenética Fontao, Lugo, Spain. Markers kept for the study belonged to autosomal chromosomes and were not in repeated positions. Additional requirements were Mendel error rate < 0.05, individual call rate ≥ 0.95, SNP call rate ≥ 0.95, and MAF > 0.05. The quality control was made using PLINK software (Purcell et al., 2007).

Linkage disequilibrium decay

Sampling animals in trios yields improved accuracy of the estimated haplotypes (Marchini et al., 2006). The pair of haplotypes of the parental chromosomes within each breed was inferred by using the “trio” option of Beagle software v3.3.2 (Browning and Browning, 2009). To increase phasing accuracy we have run 100 iterations of the phasing algorithm and sampled 100 haplotype pairs for each individual during each iteration of the phasing algorithm. Linkage disequilibrium was computed from the phases of the parents’ trios across all autosomal chromosomes within each breed by using PLINK software. The LD measure adopted was the r^2 statistic (Hill and Robertson, 1968), which is generally accepted as the more robust and better interpretable LD parameter (Kruglyak, 1999). It is computed as:

$$r^2 = \frac{(p_{11} - p_1q_1)^2}{(p_1q_1p_2q_2)}$$

where p_1 and p_2 are the minor and major allele frequencies in the SNP1 respectively, q_1 and q_2 are the minor and major allele frequencies in the SNP2 respectively, and p_{11} corresponds to the frequency observed between both minor alleles in the same individual throughout the whole population. Marker pairs were grouped by their pairwise physical distance into bins of 5 kb, starting from 0 to 2 Mb. The average r^2 for SNP pairs in each bin was plotted using an R environment (R Core Team, 2015). The average r^2 between adjacent markers for different densities after thinning for a certain percentage of markers chosen at random (10%, 5% and 1%) was also estimated.

Persistence of LD phase

In this case we used as a measure of LD the signed square root of r^2 , $r = p_{11} - p_1q_1/\sqrt{p_1q_1p_2q_2}$, computed with the PLINK software. If r^2 between two markers is equal in two populations, but their corresponding r has opposite sign, the gametic phase is reversed. Persistence of phase quantifies the extent to which there is an excess of the same haplotypes in all breeds. It was estimated as the Pearson correlation coefficient between the LD between two specific markers in two different breeds as in Badke et al. (2012):

$$r_{k,k'} = \frac{\sum(r_{(k)} - \bar{r}_{(k)})(r_{(k')} - \bar{r}_{(k')})}{S_{(k)}S_{(k')}}}$$

where $r_{k,k'}$ is the correlation of phases between $r_{(k)}$ in population k and $r_{(k')}$ in population k' , $S_{(k)}$ and $S_{(k')}$ are the standard deviation of $r_{(k)}$ and $r_{(k')}$ respectively, and $\bar{r}_{(k)}$ and $\bar{r}_{(k')}$ are the average of r across all SNP within intervals mentioned above for population k and k' respectively. To represent the evolution of the persistence of phase with marker separation, the results were stacked in bins of 10 kb from 0 to 100 kb, and of 100 kb from 100 kb to 1 Mb of marker distances. In addition, to highlight some results we used a Heatmap graph to represent the ranges of correlation of r among populations. The calculus and subsequent plots were constructed using an R environment (R Core Team, 2015).

Past effective population size and time since divergence

To estimate the effective population size, N_e , a correction for sample size was applied for all r^2 values using the following equation (Nilsen et al., 2008):

$$r_{corrected}^2 = \frac{r_{computed}^2 - \frac{1}{n}}{1 - \frac{1}{n}}$$

where n represents the number of haplotypes in the sample. When the corrected r^2 was negative it was set to 0. The N_e was estimated applying the approximate expectation of r^2 in absence of mutation (Sved, 1971):

$$r^2 = 1/(1 + 4N_e c) \approx N_e = (1/4c) * (1/r^2 - 1)$$

where N_e is the effective population size, r^2 is the corrected linkage disequilibrium estimate, and c is the recombination frequency which is replaced by map distance in Morgans. N_e was estimated at different generations in the past (T), calculated by the

approximation $T = 1/2c$ (Hayes et al., 2003). To estimate N_e three models were employed. The first of these models (M1) assumes the generalization $1\text{Mb} = 1\text{cM}$; in the second model (M2), specific recombination rates for each chromosome in cattle were used (Arias et al., 2009), as suggested by Qanbari et al. (2010) and Corbin et al. (2010); the third model (M3) is a nonlinear least squares approach also using specific recombination rates for each chromosome:

$$r_i^2 = 1/(a + 4bd_i) + e_i \approx b = (1/4d_i) * (1/r_i^2 - a)$$

where r_i^2 is the estimate of r^2 for each i^{th} marker pair at a distance d_i in Morgans (Tenesa et al., 2007). The a parameter in the Sved's formula takes a value of 1, but Tenesa et al. (2007) showed that it takes a value of approximately 2 when mutation is taken into account.

Furthermore, we estimate the time since breed divergence (T). Assuming that the variance of r remains constant in both breeds, the expected correlation of $r_{k,k'}$ equals e^{-2cT} , where $r_{k,k'}$ is the persistence of phase, c is the recombination distance between markers (in Morgans) and T is the time since breed divergence in generations (de Roos et al., 2008). We calculated a linear regression for the natural logarithm of correlation of phase with pairwise distance between 5 kb and 400 kb, where the slope of this regression was an estimate of $-2T$. Consequently, the slope divided by -2 was the number of generations (T) since these two breeds have diverged.

Additional files provided with this submission

Additional file 1. Figures A1-A7. Frequency distribution of linkage disequilibrium for adjacent markers in 7 Spanish beef breeds.

Additional file 2. Figures B1-B7. Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in 7 Spanish beef breeds.

Additional file 3. Estimates of time since divergence between breeds from linear regression of the natural logarithm of the expected correlation of r .

Additional file 4. Past effective population size at various generations ago in seven Spanish beef breeds.

Additional file 5. Average of linkage disequilibrium between nonsyntenic SNPs, effective population sizes, and effective number of loci in seven Spanish beef breeds.

Abbreviations

AV, Asturiana de los Valles breed; ANI, Avileña-Negra Ibérica breed; BP, Bruna dels Pirineus breed; GEBV, Genomic estimated breeding value; LD, linkage disequilibrium; MAF, Minor allele frequency; Mo, Morucha breed; N_e , Effective population size; Pi, Pirenaica breed; QTL, Quantitative trait locus; Re, Retinta breed; RG, Rubia Gallega breed; SNP, Single-nucleotide polymorphism.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

LV, CD and JP designed the experiment. JA, CD, AM, JAB and JP identified trios for data collection. LV and CD coordinated the genotyping. JJCA and JP completed the statistical analysis and wrote a first version of the manuscript. All authors contributed to the discussion and approved the paper.

Acknowledgements

The research leading to these results has received funding of a Ministerio de Ciencia e Innovación AGL 2010-15903 grant from the Spanish government. The collaboration of Breed societies in collecting samples and the support of FEAGAS is also acknowledged. The COLCIENCIAS Francisco José de Caldas 497/2009 fellowship awarded to J. J. Cañas-Álvarez is also recognized.

References

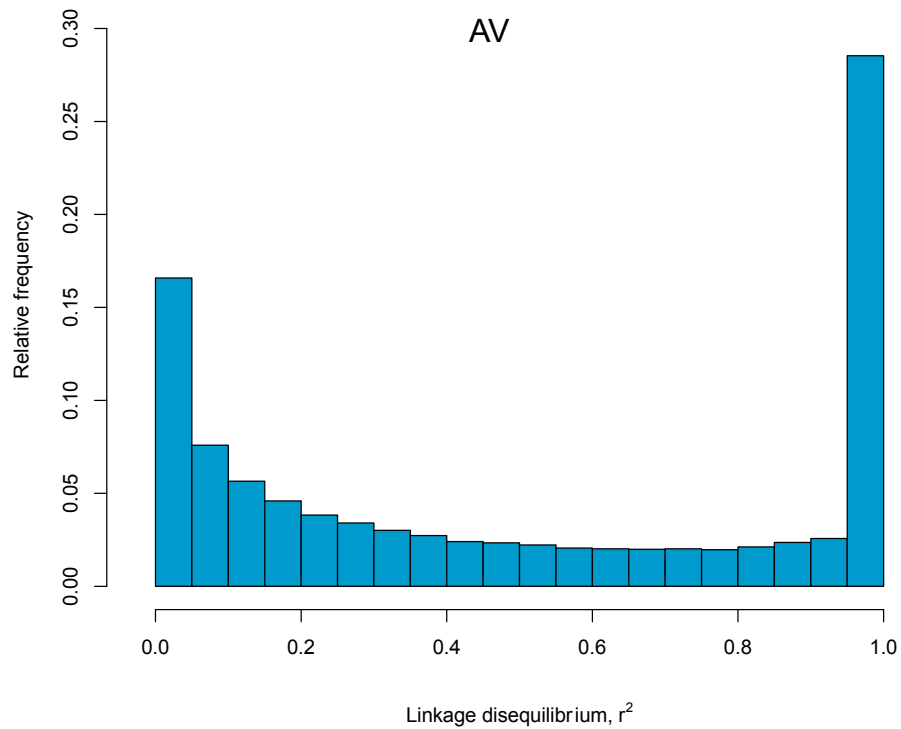
- Andreescu C, Avendano S, Brown SR, Hassen A, Lamont SJ, Dekkers JCM. Linkage disequilibrium in related breeding lines of chickens. *Genetics*. 2007;177:2161–2169.
- Ardlie K, Kruglyak L, Seielstad M. Patterns of linkage disequilibrium in the human genome. *Nature Rev Genet*. 2002;3:299–309.
- Arias JA, Keehan M, Fisher P, Coppieters W, Spelman R. A high density linkage map of the bovine genome. *BMC Genetics*. 2009;10:18.
- Badke YM, Bates RO, Ernst CW, Schwab C, Steibel JP. Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics*. 2012;13:24.
- Browning BL, Browning SR. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*. 2009;84:210–23.
- Cañas-Álvarez JJ, González-Rodríguez A, Martín-Collado D, Avilés C, Altarriba J, Baro JA, et al. Monitoring changes in the demographic and genealogical structure of the main Spanish local beef breeds. *J Anim Sci*. 2014;92:4364–74.
- Cañas-Álvarez JJ, González-Rodríguez A, Munilla S, Varona L, Díaz C, Baro JA, et al. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density chip. *J Anim Sci*. 2015;submitted.

- Corbin LJ, Blott SC, Swinburne JE, Vaudin M, Bishop SC, Woolliams JA. Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Anim Genet.* 2010;41:8–15.
- Corbin LJ, Liu AYH, Bishop SC, Woolliams JA. Estimation of historical effective population size using linkage disequilibria with marker data. *J Anim Breed Genet.* 2012;129:257–70.
- Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA. The impact of genetic architecture on genome-wide evaluation methods. *Genetics.* 2010;185:1021–31.
- Decker JE, McKay SD, Rolf MM, Kim, J, Molina-Alcala, A et al. Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *Plos Genet.* 10:e1004254.
- Dekkers JCM, Hospital F. The use of molecular genetics in the improvement of agricultural populations. *Nat Rev Genet.* 2002;3:22–32.
- Espigolan R, Baldi F, Boligon AA, Souza FR, Gordo DG, Tonussi RL, et al. Study of whole genome linkage disequilibrium in Nelore cattle. *BMC Genomics.* 2013;14:305.
- García-Dory MA. Las razas bovinas autóctonas de España están en vías de extinción. *Quercus.* 1986;16:4–9.
- Gautier M, Faraut T, Moazami-Goudarzi K, Navratil V, Foglio M, Grohs C, et al. Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics.* 2007;177:1059–70.
- Goddard ME, Hayes BJ, McPartlan H, Chamberlain AJ. Can the same genetic markers be used in multiple breeds? In: *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production.* Belo Horizonte, Brazil. 2006;16–22.
- Goddard M. Genomic selection: prediction of accuracy and maximization of long term response. *Genetica.* 2009;136:245–57.
- González-Rodríguez A, Toro MA, Varona L, Carabaño MJ, Cañas-Álvarez JJ, Altarriba J, et al. Genome-wide analysis of genetic diversity in autochthonous Spanish populations of beef cattle. In: *Proceedings of the 10th World Congress on Genetics Applied to Livestock Production.* Vancouver, Canada. 2014. https://asas.org/docs/default-source/wcgalp-proceedings-oral/255_paper_9607_manuscript_1524_0.pdf?sfvrsn=2. Accessed 19 Feb 2015.
- Gutiérrez JP, Altarriba J, Díaz C, Quintanilla R, Cañón J, Piedrafita J. Pedigree analysis of eight Spanish beef cattle breeds. *Genet Sel Evol.* 2003;35:43–63.
- Harmegnies N, Farnir F, Davin F, Buys N, Georges M, Coppieters W. Measuring the extent of linkage disequilibrium in commercial pig populations. *Anim Genet.* 2006;37:225–31.

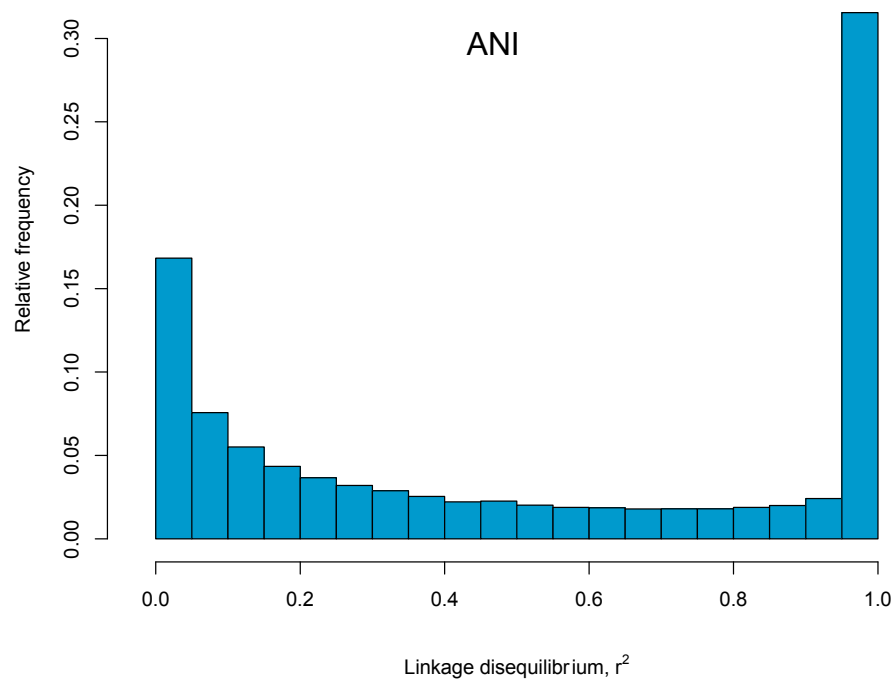
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME. Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Res.* 2003;13:635–43.
- Hill WG, Robertson A. Linkage disequilibrium in finite populations. *Theor Appl Genet.* 1968;38:226–31.
- Illumina Inc. BovineHD Genotyping BeadChip Datasheet. 2012. http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf. Accessed 26 Sept 2014.
- Joint Working Group on Refinement. Removal of blood from laboratory mammals and birds. First report of the BVA/FRAME/RSPCA/UFAW Joint Working Group on Refinement. *Lab Anim.* 1993;27:1–22.
- Khatkar M, Nicholas F, Collins A, Zenger K, Cavanagh J, Barris W, Schnabel R, Taylor J, Raadsma H. Extent of genome-wide linkage disequilibrium in Australian Holstein-Friesian cattle based on a high density SNP panel. *BMC Genomics.* 2008;9:187.
- Kruglyak L. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genet.* 1999;22:139–44.
- Lu D, Sargolzaei M, Kelly M, Li C, Vander Voort G., Wang Z, Plastow G, Moore S, Miller SP. Linkage disequilibrium in Angus, Charolais and crossbreed beef cattle. *Front Genet.* 2012;3(152):1-9.
- Marchini J, Cutler D, Patterson N, Stephens M, Eskin E, Halperin E, et al. A comparison of phasing algorithms for trios and unrelated individuals. *Am J Hum Genet.* 2006;78:437–50.
- McKay SD, Schnabel RD, Murdoch BM, Matukumally LK, Aerts J, Coppieters W, et al. Whole genome linkage disequilibrium maps in cattle. *BMC Genetics.* 2007;8:74.
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, et al. Development and characterization of a high density SNP genotyping assay for cattle. *PLoS One.* 2009;4:e5350.
- Meuwissen THE. Accuracy of breeding values of ‘unrelated’ individuals predicted by dense SNP genotyping. *Genet Sel Evol.* 2009;41:35.
- Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics.* 2001;157:1819–29.
- Nilsen H, Hayes B, Berg PR, Roseth A, Sundsaasen KK, Nilsen K, et al. Construction of a dense SNP map for bovine chromosome 6 to assist the assembly of the bovine genome sequence. *Anim Genet.* 2008;39:97–104.
- Perez-O’Brien AM, Utsunomiya YT, Mészáros G, Bickhart DM, Liu GE, et al. Assessing signatures of selection through variation in linkage disequilibrium between taurine and indicine cattle. *Genet Sel Evol.* 2014;46: 19.

- Porto-Neto L, Kijas JW, Reverter A. The extent of linkage disequilibrium in beef cattle breeds using high-density SNP genotypes. *Genet Sel Evol.* 2014;46:22.
- Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet.* 2001;69:1–14.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet.* 2007;81:559–75.
- Qanbari S, Pimentel ECG, Tetens J, Thaller G, Lichtner P, Sharifi AR, et al. The pattern of linkage disequilibrium in German Holstein cattle. *Anim Genet.* 2010;41:346–56.
- de Roos APW, Hayes BJ, Goddard ME. Reliability of genomic predictions across multiple populations. *Genetics.* 2009;183:1545–53.
- de Roos APW, Hayes BJ, Spelman RJ, Goddard ME. Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics.* 2008;179:1503–12.
- Salomon-Torres R, Matukumalli LK, Van Tassell CP, Villa-Angulo C, Gonzalez-Vizcarra VM, Villa-Angulo R. High density LD-based structural variations analysis in cattle genome. *PLoS One.* 2014;9:e103046.
- Sved JA. Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theor Popul Biol.* 1971;2:125–41.
- Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, et al. Recent human effective population size estimated from linkage disequilibrium. *Genome Res.* 2007;17:520–6.
- Toosi A, Fernando RL, Dekkers JCM. Genomic selection in admixed and crossbred populations. *J Anim Sci.* 2010;88:32–46.
- Villa-Angulo R, Matukumalli LK, Gill CA, Choi J, Van Tassell CP, Grefenstette JJ. High-resolution haplotype block structure in the cattle genome. *BMC Genet.* 2009;10:19.
- Wang J. Estimation of effective population sizes from data on genetic markers. *Philos Trans R Soc Lond B Biol Sci.* 2005;360:1395–409.
- Warnes GR, Bolker B, Bonebakker L, Gentleman R, Liaw WHA, Lumley T, et al. gplots: Various R programming tools for plotting data. R package version 2.16.0. 2015. <http://cran.r-project.org/web/packages/gplots/index.html>. Accessed 11 Feb 2015.
- Zimin AV, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, et al. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 2009;10:R42.

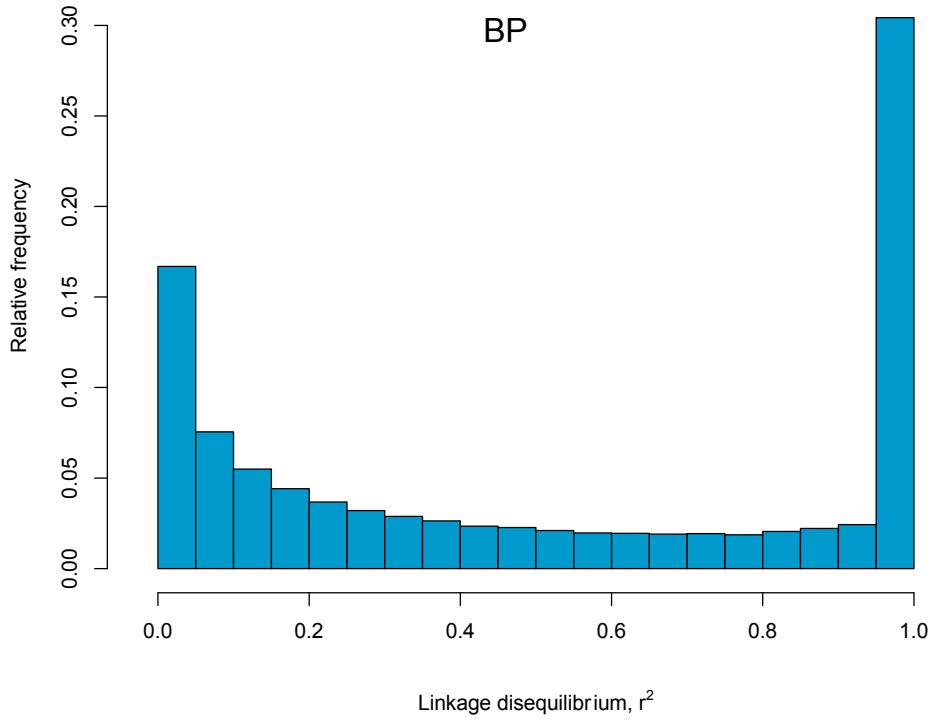
ADDITIONAL FILES



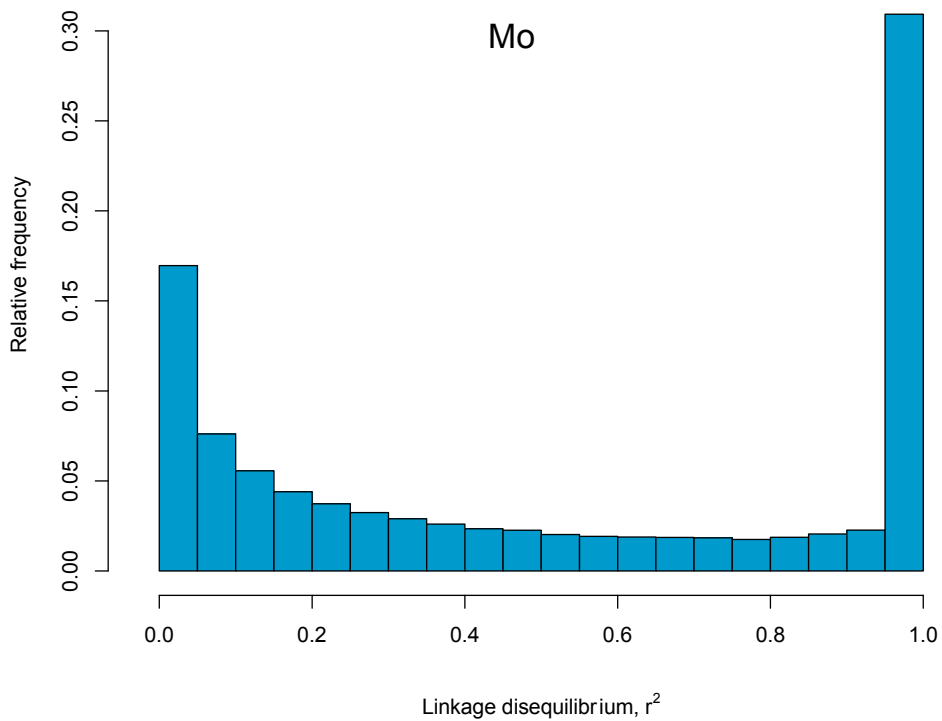
Additional file 1. Figure A1 Frequency distribution of linkage disequilibrium for adjacent markers in Asturias de los Valles breed



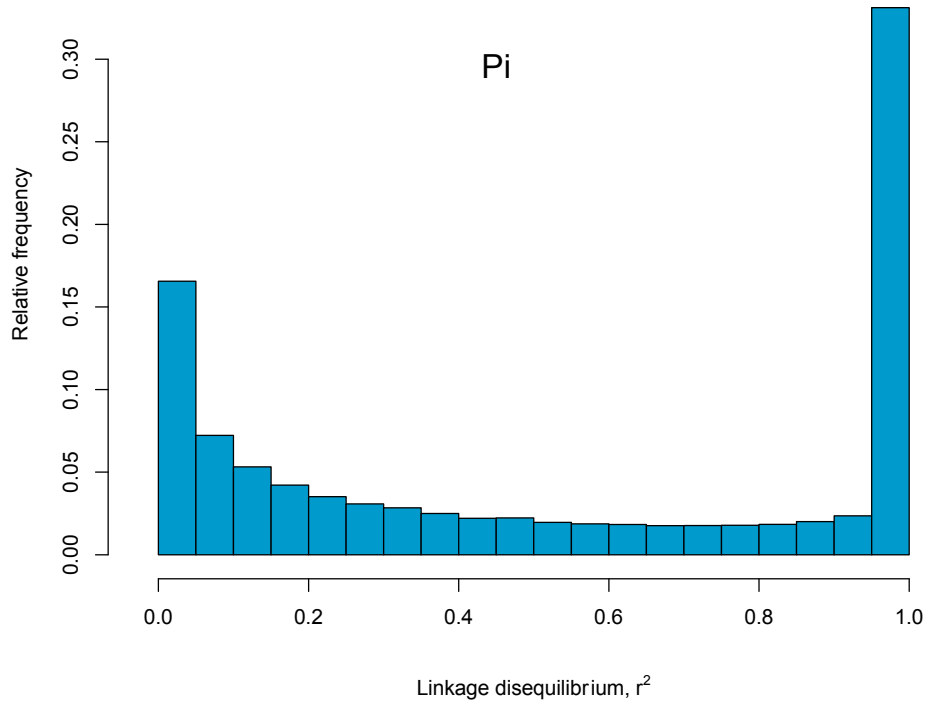
Additional file 1. Figure A2 Frequency distribution of linkage disequilibrium for adjacent markers in Avileña-Negra Ibérica breed



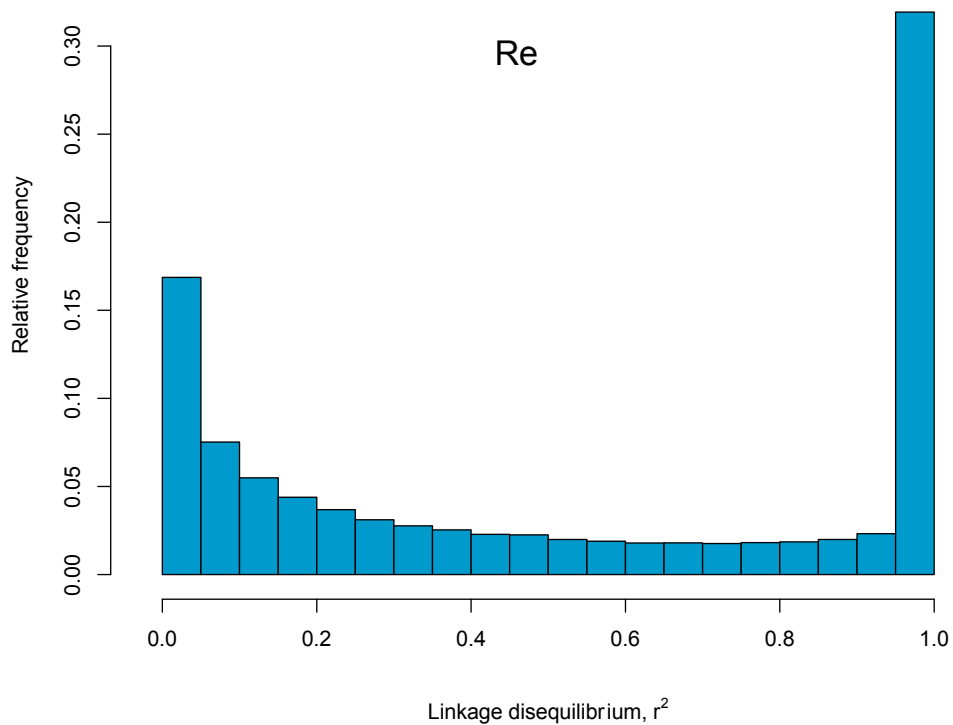
Additional file 1. Figure A3 Frequency distribution of linkage disequilibrium for adjacent markers in Bruna dels Pirineus breed



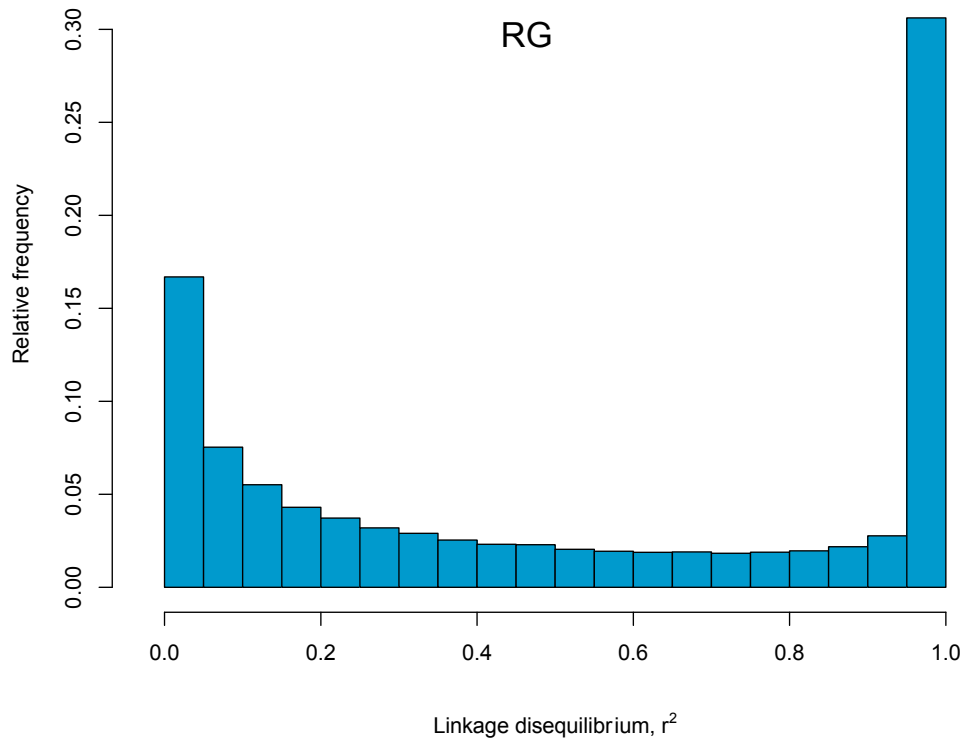
Additional file 1. Figure A4 Frequency distribution of linkage disequilibrium for adjacent markers in Morucha breed



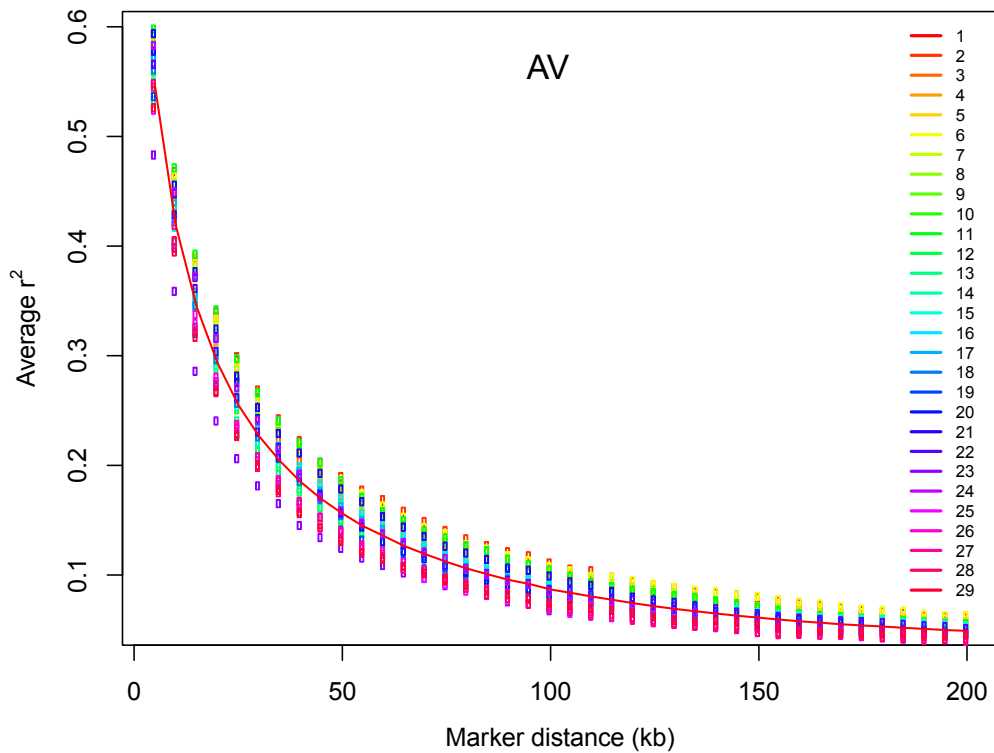
Additional file 1. Figure A5 Frequency distribution of linkage disequilibrium for adjacent markers in Pirenaica breed



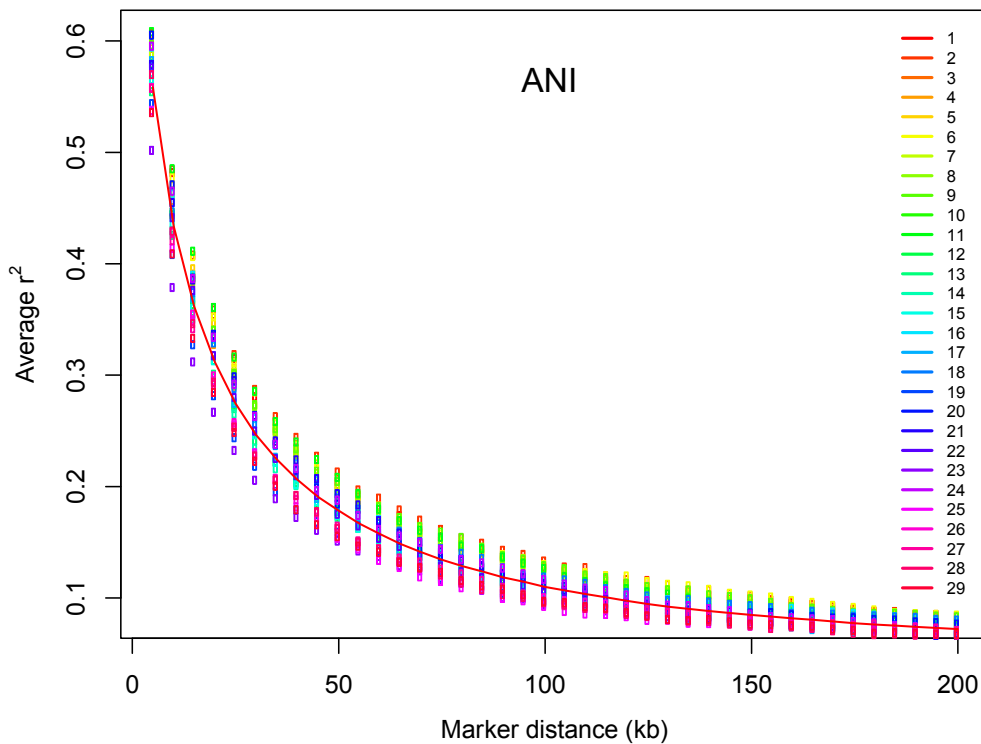
Additional file 1. Figure A6 Frequency distribution of linkage disequilibrium for adjacent markers in Retinta breed



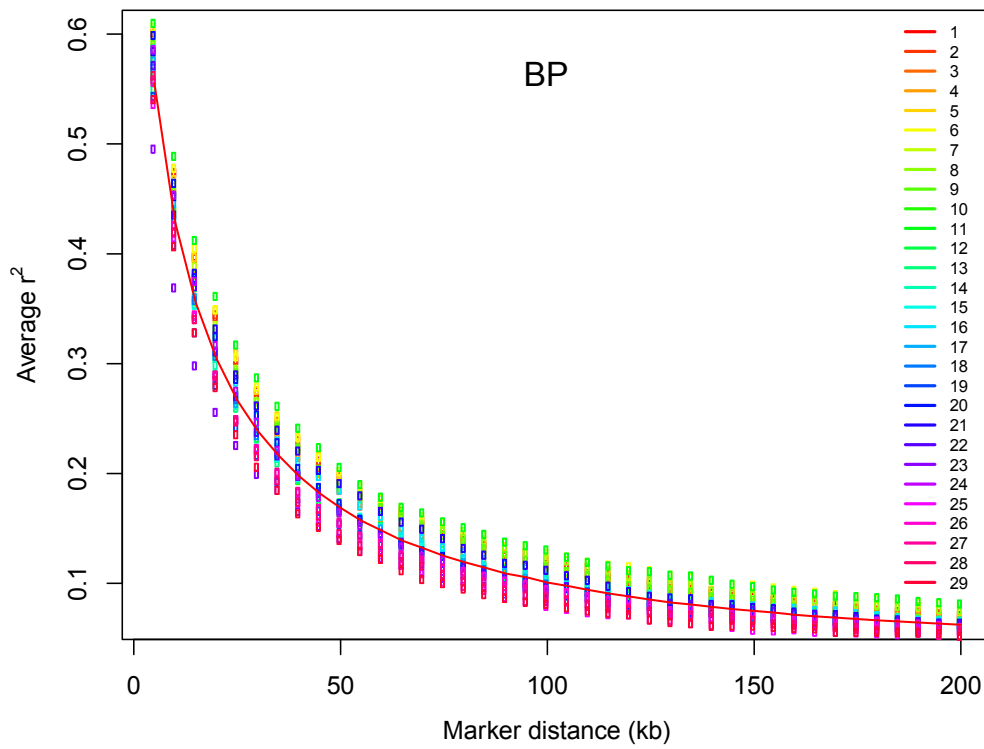
Additional file 1. Figure A7 Frequency distribution of linkage disequilibrium for adjacent markers in Rubia Gallega breed



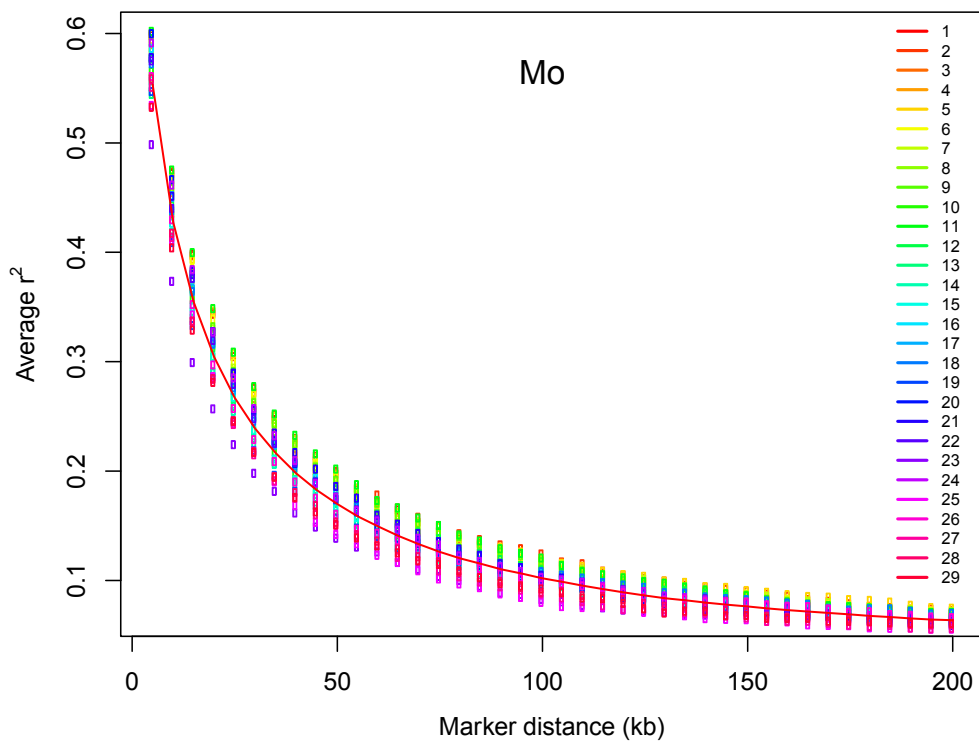
Additional file 2. Figure B1 Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in Asturiana de los Valles breed



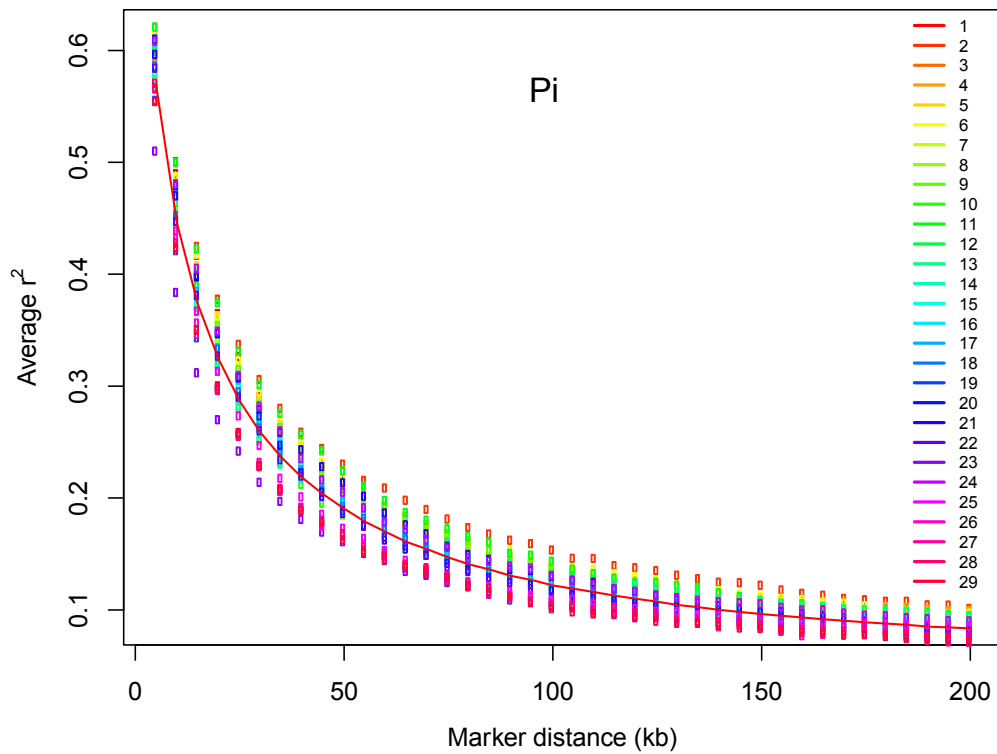
Additional file 2. Figure B2 Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in Avileña-Negra Ibérica breed



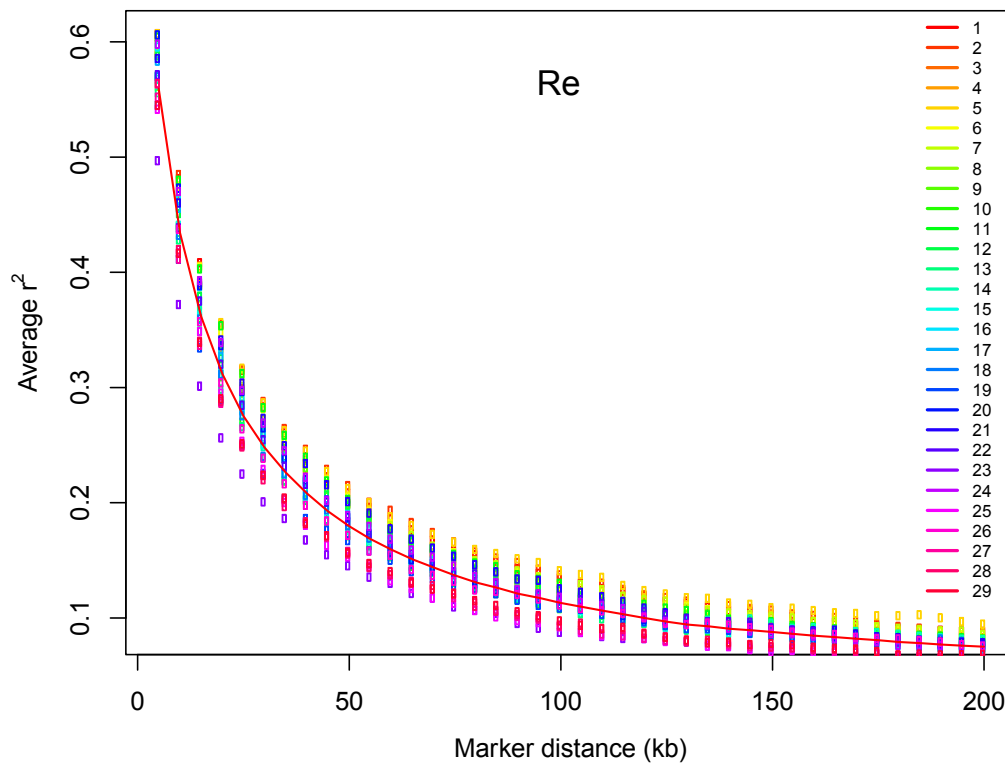
Additional file 2. Figure B3 Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in Bruna dels Pirineus breed



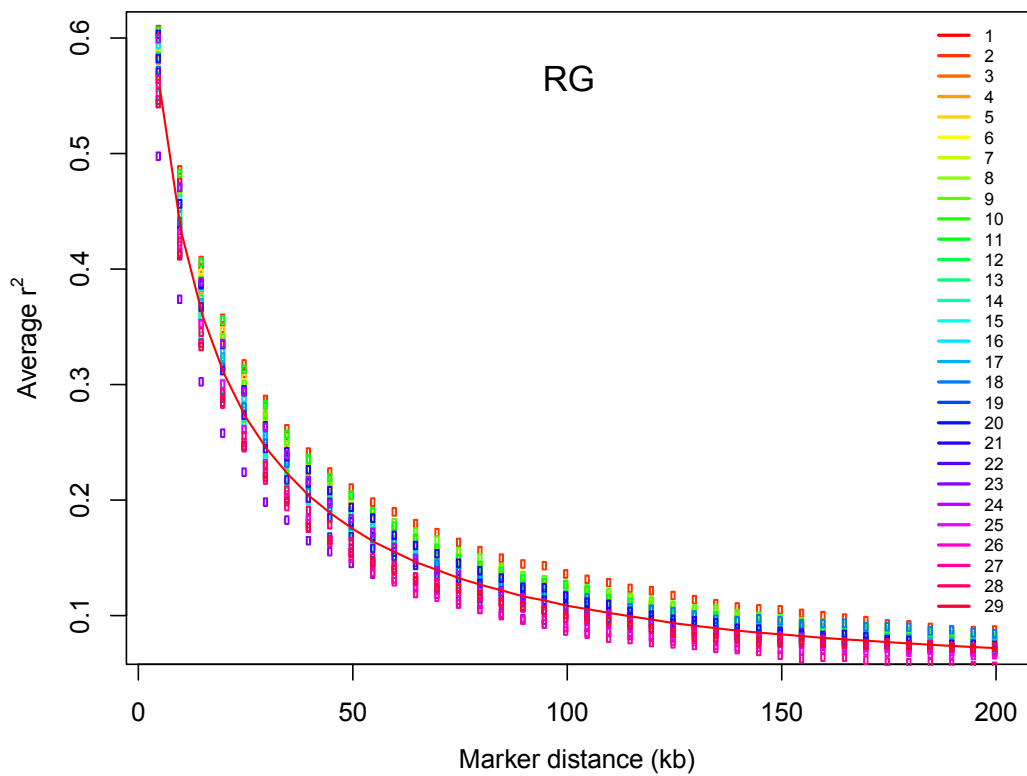
Additional file 2. Figure B4 Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in Morucha breed



Additional file 2. Figure B5 Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in Pirenaica breed



Additional file 2. Figure B6 Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in Retinta breed



Additional file 2. Figure B7 Evolution of average r^2 from 0 to 200 kb apart for each of the 29 autosomes in Rubia Gallega breed

Additional file 3 Estimates of time since divergence between breeds (T) from linear regression of the natural logarithm of the expected correlation of r

Breed ¹	AV	ANI	BP	Mo	Pi	Re
ANI	141.0					
BP	130.8	178.6				
Mo	129.4	137.8	172.2			
Pi	157.5	203.3	165.6	191.3		
Re	146.8	164.3	191.4	156.6	206.8	
RG	135.6	174.5	162.0	164.3	190.7	181.5

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

Additional file 4 Average effective population size (N_e) at various generations ago in seven Spanish beef breeds

Breeds ¹	Generations	Model ²		
		M1	M2	M3
AV	1	31.8	30.7	29.6
	5	139.2	140.4	133.5
	10	253.4	265.2	235.8
	50	1002.3	1029.6	991.0
	100	1762.5	2009.8	1655.0
	200	2550.6	2553.5	2212.5
	800	2742.1	2358.8	2194.2
	1500	2893.0	2487.8	2171.1
ANI	1	30.5	29.3	28.6
	5	115.6	108.5	105.7
	10	185.9	162.2	162.0
	50	603.6	570.5	531.9
	100	1018.3	897.3	920.3
	200	1575.2	1454.3	1344.9
	800	2271.7	1977.0	1814.9
	1500	2555.9	2227.2	1899.6
BP	1	32.2	28.6	31.0
	5	136.2	137.9	129.6
	10	239.4	229.6	230.5
	50	787.5	749.9	749.3
	100	1267.1	1388.1	1179.3
	200	1855.1	1917.2	1649.4
	800	2444.8	2122.4	1958.8
	1500	2678.4	2321.9	1997.7
Mo	1	28.1	29.7	25.7
	5	110.3	107.7	103.3
	10	188.7	196.0	180.9
	50	703.4	708.6	687.0
	100	1202.2	1266.4	1121.3
	200	1826.5	1843.7	1603.5
	800	2424.1	2112.2	1949.4
	1500	2681.2	2327.5	2003.2
Pi	1	30.1	31.2	27.3
	5	92.4	91.6	86.4
	10	142.7	140.7	135.1
	50	471.7	439.5	457.3

	100	812.0	800.9	794.8
	200	1284.8	1300.8	1163.3
	800	2054.0	1810.9	1650.2
	1500	2381.5	2079.9	1753.1
Re	1	28.3	31.4	25.5
	5	97.2	97.0	87.2
	10	154.6	143.9	134.7
	50	520.9	506.4	478.4
	100	904.3	873.9	775.6
	200	1445.3	1432.3	1239.6
	800	2216.7	1952.6	1792.7
	1500	2519.0	2199.2	1870.4
RG	1	26.1	26.3	24.5
	5	86.6	93.6	78.5
	10	141.1	157.6	130.3
	50	537.2	554.9	521.7
	100	961.2	1023.6	885.8
	200	1564.1	1561.9	1363.4
	800	2328.0	2031.7	1870.6
	1500	2607.7	2266.8	1944.2

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

²M1: Generalization of 1Mb = 1cM for the recombination distance is used; M2: Specific recombination rates for each chromosome; M3: Nonlinear least squares approach with specific recombination distance.

Additional file 5 Average LD (r^2) between nonsyntenic SNPs, effective population sizes (N_e), and effective number of loci (M_e) in seven Spanish beef breeds

Breed	Average r^2	N_e	M_e
AV	0.0104	47.5	551.5
ANI	0.0110	45.1	528.8
BP	0.0104	47.5	551.5
Mo	0.0111	44.6	524.1
Pi	0.0109	45.5	532.6
Re	0.0115	43.0	508.8
RG	0.0111	44.6	524.1

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

Capítulo 7

DISCUSIÓN

7. Discusión General

Los resultados obtenidos en la presente tesis doctoral han permitido describir y caracterizar la estructura genética de las principales razas de ganado de carne autóctono español. Esta descripción fue abordada desde dos puntos de vista. El primero de ellos evaluó los parámetros demográficos a partir de los pedigríes registrados en los libros genealógicos, y el segundo buscó dar una explicación de la diversidad genética y la divergencia partiendo de datos genómicos obtenidos a partir de un chip denso de marcadores SNP. Hasta ahora las poblaciones autóctonas de ganado vacuno de carne español no habían sido caracterizadas combinando tanto información genealógica como información genómica. Esta tesis permitió realizar esta caracterización, al mismo tiempo que evaluó algunas herramientas fundamentales en la exploración de la base genética aplicables en una futura implementación de la selección genómica.

Estructura demográfica de las poblaciones autóctonas. La importancia de los censos poblacionales

Durante muchos siglos las poblaciones bovinas autóctonas han jugado un papel esencial en la ganadería española. La evolución de los censos poblacionales a lo largo del tiempo se ha dado dependiendo de las circunstancias que les afectan y de la coyuntura social de su entorno. Estos censos han sufrido numerosos cambios durante muchos años, incluso comprometiendo, en algunos casos, la continuidad de algunas razas (Sánchez-Belda, 2002; Tisdell, 2003). Sin embargo, éstas han sabido sobreponerse pese a la continua inclusión de razas foráneas con mejores índices productivos. La alta adaptabilidad de las razas de las razas autóctonas (Hoffmann, 2010), de la mano de políticas de fomento de las comunidades autónomas y el interés particular de algunos productores, han permitido mantener e incluso aumentar los censos de estas poblaciones en las últimas décadas. En los últimos 11 años, por ejemplo, los censos de las razas Avileña-Negra Ibérica, Morucha, Pirenaica y Retinta han incrementado entre un 16 y 41%, y en las razas Asturiana de los Valles, Bruna dels Pirineus y Rubia Gallega este incremento ha sido superior, llegando incluso a valores de entre un 158 y 381%. Otra característica a destacar fue el bajo número de hembras por toro encontrado en la mayoría de las razas evaluadas. Este es un indicativo

del poco uso de la inseminación artificial en la mayoría de los rebaños, donde la monta natural facilita el manejo de los animales explotados bajo sistemas extensivos. Sin embargo, la excepción a este caso fue la raza Rubia Gallega, la cual se caracteriza por el amplio uso de la inseminación artificial, debido principalmente a la gran producción de dosis seminales de esta raza que provee Xenética Fontao, el mayor centro de inseminación artificial en España (ACRUGA, 2011). Este hecho se vio reflejado en el alto número de hembras por toro y en el alto número de crías por toro encontrado en los tamaños familiares.

Según la pirámide estructural presentada en el capítulo 4, las razas autóctonas se caracterizaron por el alto intercambio de toros reproductores entre los diferentes rebaños, lo cual es una práctica habitual en la mayoría de las razas manejadas bajo sistemas extensivos. Particularmente la raza Morucha alcanzó a tener más de la mitad de las granjas dedicadas a comprar y vender reproductores. Este intercambio de reproductores generó un alto número de granjas multiplicadoras donde el proceso de selección genética no se concentra en unas pocas granjas, sino por el contrario está repartido entre ellas, lo cual es vital en el mantenimiento de la variabilidad genética de las poblaciones. Sin embargo, esta clasificación no fue la misma en todas las razas. Por ejemplo, en las razas Avileña-Negra Ibérica, Retinta y Rubia Gallega más del 70% de los rebaños fueron clasificados como comerciales, dedicados exclusivamente a producir animales para sacrificio sin aportar reproductores a la población. Específicamente, la raza Rubia Gallega obtuvo los más altos valores de granjas clasificadas como comerciales, confirmando que la amplia utilización de la inseminación artificial provoca que pocos rebaños aporten toros reproductores en la variabilidad genética de esta raza.

La endogamia y el tamaño efectivo. Parámetros claves en la variabilidad genética

La profundidad del pedigrí es una característica muy importante en la estimación de los parámetros poblacionales obtenidos a partir de los estudios genealógicos (Boichard et al., 1997). Por esta razón es importante, antes de hablar propiamente de la endogamia, hacer énfasis en como la información genealógica existente puede restringir algunos de los resultados obtenidos. La subestimación de los valores de endogamia puede ser un claro ejemplo de la falta de completación del pedigrí (Cassell et al., 2003; Lutaaya et al., 1999).

Este hecho se evidenció en las razas Avileña-Negra Ibérica, Pirenaica y Retinta, donde la elevada profundidad del pedigrí coincidió con elevados valores de consanguinidad promedio. De la misma manera, este suceso pudo ser observado al comparar nuestras estimaciones con las obtenidas por Gutiérrez et al. (2003), realizadas también en razas autóctonas españolas. Nuestros valores mostraron un incremento en los niveles de profundidad del pedigrí de la mano con incrementos medios de la endogamia. El mayor conocimiento del pedigrí permitió obtener valores de endogamia acumulada muchas más generaciones atrás. De esta manera, el poco conocimiento de las genealogías es un hecho que se presenta habitualmente debido a que en muchas ocasiones la información de los padres no se conoce, o si se conoce no se completa en su totalidad hasta la máxima generación, ya sea por la dificultad en la obtención de los datos o por simple falta de interés. Sin embargo, en la actualidad, la realización continua de las evaluaciones genéticas en las distintas razas autóctonas ha generado una alta preocupación por actualizar y mejorar continuamente los libros genealógicos propios de cada raza. Tanto así, que desde el año 1995 hasta el año 2009, el incremento en la profundidad del pedigrí ha permitido llegar a valores promedios cercanos al 92% una generación atrás y valores promedios del 61% seis generaciones atrás.

Las tasas o incrementos de endogamia estimados en la presente tesis proporcionaron información relevante sobre el avance de la endogamia en el tiempo, y permitió además, vigilar el descenso de la variabilidad dentro de cada raza. Generalmente, el aumento de la endogamia y la consecuente reducción de la heterocigosis, resulta en reducción de sobrevivencia, salud, vigor y eficiencia reproductiva (Charlesworth y Willis, 2009). A esto se le conoce comúnmente como depresión endogámica, la cual se define como la disminución de los valores fenotípicos medios de los caracteres relacionados con la capacidad reproductora o la eficiencia fisiológica ocasionados por la endogamia (Thompson et al., 2000). Si observamos los valores de endogamia a lo largo de los 11 años evaluados en la presente tesis (1998-2009) se observan incrementos moderados de endogamia en todas las razas. Generalmente, en la mayoría de las poblaciones domésticas, la movilidad de los individuos es restringida a un contexto reducido, por lo tanto, los animales tienden a aparearse con individuos cercanos genéticamente (Falconer y Mackay, 1996). En nuestro caso, al tratarse de poblaciones con censos reducidos, el apareamiento entre todos los animales de una población no se realiza completamente al azar y menos aún cuando se tratan de poblaciones bajo selección. Todo esto conlleva a incrementos de los

valores de endogamia. Particularmente, las razas Avileña-Negra Ibérica, Morucha y Retinta presentaron las mayores tasas o incrementos de endogamia bajo las diferentes metodologías estudiadas, lo cual se vio reflejado en ser las razas con los menores valores de tamaño efectivo de la población.

En la presente tesis los tamaños efectivos poblacionales fueron estimados por medio de diferentes metodologías basadas en el incremento de la endogamia; los resultados obtenidos en cada una de ellas fueron altamente variantes. El tamaño efectivo estimado a partir de los incrementos de endogamia y coascendencia presentaron unos puntos débiles relacionados con la fluctuación de sus valores a lo largo del tiempo; de hecho, en algunos casos éstos disminuyeron ocasionando valores no estimables del tamaño efectivo. Este efecto también ha sido observado en diferentes estudios realizados en diferentes especies (Groeneveld et al., 2009; Leroy et al., 2009). En algunos casos para evitar que esto suceda, se ha optado por tomar periodos de tiempo relativamente cortos (dos generaciones) para estimar el tamaño efectivo, lo cual permite reducir el número de casos con valores negativos de tamaño efectivo (Leroy et al., 2013). Los tamaños efectivos obtenidos mediante los incrementos de coascendencia fueron muy superiores a los valores obtenidos mediante los incrementos de endogamia. Bajo condiciones de apareamiento al azar se espera que la coascendencia sea mayor que la endogamia. Este no fue el caso para la mayoría de las razas analizadas, donde las tasas de coascendencia fueron bajas comparadas con las tasas de endogamia. Este hecho puede ser explicado ya sea por la existencia de subpoblaciones o debido a las prácticas de cría particulares tales como una alta frecuencia de apareamiento entre individuos emparentados (Leroy et al., 2013). Por otro lado, la metodología basada en regresión para estimar el tamaño efectivo, depende en gran medida de los cambios en la política de apareamientos producidos durante el periodo de tiempo contemplado (Cervantes et al., 2008). De este modo, si tradicionalmente se han gestionado los apareamientos evitando los cruces entre parientes cercanos, luego de haber descuidado los apareamientos en el pasado, la pendiente de la recta de regresión puede llegar a ser muy baja. En nuestro caso, este hecho se vio reflejado en valores altos o no estimables del tamaño efectivo de la población. Por su parte, los valores de tamaño efectivo estimados a partir del incremento de endogamia individual permitieron obtener valores más aproximados y menos sesgados. Este método supera los problemas obtenidos con el cálculo del tamaño efectivo a partir de las tasas negativas de endogamia y coascendencia debido a que este representa una estandarización del coeficiente de

endogamia con relación a la profundidad del pedigrí de cada individuo (González-Recio et al., 2007). De esta forma, los tamaños efectivos no dependen de la profundidad del pedigrí y reflejan todos los fenómenos ocurridos en cada individuo de la población, como el solapamiento generacional, la selección y la subdivisión que pueda existir (Gutiérrez, et al., 2008).

Según Mc Parland et al. (2007), el tamaño efectivo de la población puede ser definido como un indicador de la pérdida de diversidad genética durante un periodo de tiempo determinado. En nuestro trabajo, el tamaño efectivo estimado a partir del incremento de endogamia individual en generaciones equivalentes osciló entre 19 y 90, colocando de manifiesto el riesgo de pérdida de variabilidad genética en las poblaciones. Pese a que el tamaño efectivo generalmente es bajo en la mayoría de las razas de ganado a nivel mundial (Boichard et al., 1997; Cleveland et al., 2005; Mc Parland et al., 2007), es necesario tener cuidado con los valores obtenidos en esta tesis, ya que pueden comprometer la diversidad genética de las poblaciones. Lo más importante es que en la actualidad las diferentes asociaciones de las razas vienen aumentando continuamente los censos de los animales, lo cual debe de ir de la mano también con la implementación de estrategias que permitan controlar las tasas de endogamia y los censos efectivos. En este sentido, cualquier programa de mejora genética deberá continuar fomentando un sistema de apareamientos que permita el intercambio de reproductores entre todos los rebaños dentro de cada raza.

Estructura y diversidad genética en los recursos genéticos locales

La información sobre la diversidad genética y la estructura de la población entre las razas de ganado es esencial para el mejoramiento genético, la comprensión de la adaptación del medio ambiente, así como la utilización y la conservación de las razas de ganado (Groeneveld et al., 2010). En el capítulo 5 empleamos el chip bovino de 778K SNP de Illumina para describir la diversidad genética de las poblaciones de bovinos autóctonos españoles de carne. Sin embargo, antes de dar inicio a la descripción de la diversidad, es importante mencionar que hay implícito un sesgo de verificación con respecto a los SNP analizados (*ascertainment bias*). El chip de 778K SNP de Illumina fue diseñado con SNP obtenidos de 27 razas distintas de las especies *Bos taurus taurus*, *Bos taurus indicus* e híbridos de ambas especies (Illumina Inc., 2012). Por lo tanto, los SNP utilizados en los

análisis genéticos poblacionales de las siete razas analizadas en esta tesis, se determinaron entre los más comunes del genoma bovino, con lo cual se asume que existe un rango lo suficientemente amplio para realizar los análisis de diversidad genética requeridos. El uso de la matriz de datos proporcionada por este chip ha sido particularmente útil para caracterizar numerosas razas de ganado vacuno (Salomon-Torres et al., 2014), inclusive en chip bovinos de menor cantidad de SNP ha sido posible llevar a cabo análisis de diversidad (The Bovine HapMap Consortium, 2009; Villa-Angulo et al., 2009). En cada uno de estos trabajos se han encontrado variaciones en los niveles de polimorfismo, indicando que el sesgo afecta a todas las razas y que ésta variación se relaciona más con el diseño del chip que con las razas que participan en el diseño. En nuestro caso, al determinar el grado de polimorfismo de las razas autóctonas españolas con respecto al chip de Illumina, se encontraron altos valores de marcadores genotipados (97% - 98%) en al menos el 95 % de las muestras empleadas. Adicionalmente el porcentaje de loci monomórfico varió entre el 11% y el 15%.

El parámetro más ampliamente utilizado para medir la diversidad dentro de las poblaciones es la heterocigosis esperada (Toro et al., 2009). Por su parte, la diversidad o riqueza alélica, también conocida como el número medio de alelos (MNA, por sus siglas en inglés), es un criterio alternativo que permite medir la diversidad genética. Este último parámetro es de importancia fundamental en los programas de conservación debido a que las razas con un bajo MNA tienen variación genética baja, debido posiblemente al aislamiento genético, a cuellos de botella poblacionales históricos o a efectos fundadores. Por el contrario, un alto MNA implica gran diversidad alélica, que podría haber sido influenciado por cruzamiento o mezcla (Keller, et al., 2001; Leberg, 2002). Nuestros resultados reflejaron una diversidad genética similar a la de la mayoría de las razas taurinas (Gautier et al., 2007). Sin embargo, en un estudio realizado en 19 razas de diferentes zonas geográficas (The Bovine HapMap Consortium, 2009), se encontró que existe una menor diversidad genética dentro del ganado taurino moderno respecto al ganado índico (cebuinos). Esta diferencia en la diversidad se debió principalmente a la poca diversidad de la población ancestral antes de la domesticación, y a efectos de cuellos de botella posterior a la domesticación, en lugar de las diferencias en la intensidad de selección natural o artificial entre *B. taurus* y *B. indicus* (The Bovine HapMap Consortium, 2009).

La distribución de la diversidad entre las diferentes poblaciones fue explicada a través de diferentes metodologías basadas en el análisis de las frecuencias alélicas. Estos

métodos se centraron en estimar la proporción de variación genética empleando el análisis de varianza molecular (Excoffier et al., 1992), donde se incluyeron también los estadísticos F (Wright, 1951, 1965). Bajo el análisis de varianza molecular se destacó un alto grado de variabilidad genética individual pero bajos niveles de variabilidad entre razas y entre individuos dentro de razas, lo cual puede ser un indicador de la poca estratificación o diferenciación genética entre razas. Es así, como esta amplia variabilidad genética individual ha permitido obtener los altos valores de diversidad genética mencionados anteriormente; además confirma el alto flujo de reproductores entre los distintos rebaños mencionado en el capítulo 4, debido a que la alta diversidad genética es facilitada por el intercambio continuo y mayor uso de diferentes reproductores entre los rebaños (Oyama et al., 2007).

En este mismo sentido, se dice que un grupo de poblaciones posee una alta diferenciación genética cuando se pueden detectar fuertes diferencias en las frecuencias alélicas entre estas poblaciones. En cambio, frecuencias alélicas similares indican una baja diferenciación genética (Hartl y Clark, 2007). En nuestro caso, el bajo valor de F_{ST} estimado en el capítulo 5, el cual representa la subdivisión entre poblaciones, indicó una baja diferenciación genética entre todas las razas, muestra de frecuencias alélicas similares. Esta poca diferenciación posiblemente se deba a la falta de presión de selección por tratarse de razas manejadas bajo sistemas extensivos, o también posiblemente se deba a niveles de migración o flujo genético entre las distintas razas (Slatkin, 1994; Wright, 1965).

Por su parte, los bajos y no significativos valores de F_{IS} , permitieron confirmar que las frecuencias alélicas de todas las razas se encontraron dentro de las proporciones del equilibrio de Hardy-Weinberg. Sin embargo, es importante mencionar que las razas que obtuvieron los valores de F_{IS} positivos coincidieron justamente con las mismas razas que tienen las tasas más altas de endogamia obtenidas en el capítulo 4. Estas razas fueron Avileña-Negra Ibérica, Morucha y Retinta. De esta manera los índices de fijación se relacionan con el coeficiente de endogamia. Los valores mayores a cero de F_{IS} indican excesos de homocigotos en la población, indicativo inequívoco de incrementos de endogamia (Hartl y Clark, 2007). Inclusive el F_{IS} se interpreta muchas veces como el índice de fijación o de endogamia, F (Hedrick, 2010).

Diferenciación y divergencia entre poblaciones. ¿Qué tan cercanas son las razas autóctonas españolas?

Una manera de evaluar las relaciones genéticas entre poblaciones es mediante el cálculo de las distancias genéticas (Nei, 1972). Los valores obtenidos en las distancias genéticas entre pares de razas, tanto a partir de los valores de F_{ST} así como la distancia de Nei, mostraron claramente una estrecha relación entre las poblaciones de ganado de carne autóctono español y la cercanía genética que éstas presentan. Esta menor diferenciación era esperada debido a la ubicación de las razas estudiadas en un área geográfica bastante específica y a la mixtura entre poblaciones dada por la migración. Inclusive, Decker et al. (2014), quienes analizaron numerosas razas junto a razas autóctonas españolas, ubican a estas últimas como razas muy cercanas genéticamente. Sin embargo, pese a que las razas son muy cercanas, los tiempos de divergencia entre pares de razas ponen de manifiesto que las razas divergieron en la primera mitad del periodo medieval. Esto significa que en un principio el ganado ibérico tuvo una introgresión del ganado de África (Cymbron et al., 1999; Cymbron et al., 2005), posteriormente la división política provocó una divergencia entre las razas del norte y del sur del país, y más recientemente, la trashumancia restauró el contacto entre las poblaciones, lo cual explica en parte su cercanía genética. Las menores distancias genéticas entre las razas Avileña-Negra Ibérica, Morucha y Retinta, confirman esta cercanía entre las razas del sur y oeste de la península.

La reconstrucción de la filogenia realizada en la presente tesis mostró una distribución de las razas cercana a la clasificación del ganado español en tres troncos morfológicos (Sánchez-Belda, 1984). La excepción fue la raza Retinta, la cual según nuestro estudio se ubicó más cerca del tronco Ibérico, contrario a lo mencionado por Sánchez-Belda (1984), quien la ubicó en el tronco Turdetano. Por su parte, Jordana et al. (1991) situaron a la raza Retinta en una posición más intermedia entre ambos grupos, Turdetano e Ibérico. No obstante, es importante tener en cuenta que la reconstrucción de árboles filogenéticos presupone que la evolución de los linajes pueden divergir pero nunca ser resultado de cruces entre linajes. Esta presunción en algunas ocasiones no es aplicable al ganado, ya que a menudo las nuevas razas se originan por cruce entre dos o más razas ancestrales (FAO, 2010). Por esta razón, la interpretación del árbol filogenético realizado en este trabajo se basó más en determinar cercanías entre razas que en definir propiamente los orígenes de las mismas. El comportamiento de las diferentes medidas de las distancias

genéticas en el contexto del ganado ha sido revisado por Laval et al. (2002). Llegan a la conclusión de que todas las distancias dependen en gran medida del número de generaciones desde la divergencia y sobre el tamaño efectivo de la población de las razas y, por tanto, no se puede inferir la filogenia del árbol en el caso de las razas estrechamente relacionadas que exhiben diferentes tamaños efectivos. Por esta razón, generalmente se asume que en el trato con razas de animales de granja, la interpretación de los árboles en términos de filogenia puede ser engañoso (Felsenstein, 1982). Sin embargo, Barker (1999) ha argumentado que la diversidad filogenética ofrece el mejor criterio objetivo para la toma de decisiones de conservación (es decir, las razas que son taxonómicamente distintas deben ser favorecidas para la conservación).

El análisis de componentes principales nos permitió obtener una visión diferente, ya que en este caso las relaciones se detallan particularmente entre individuos y no entre poblaciones. Los resultados obtenidos mostraron que el análisis multivariante fue bastante aproximado a la clasificación realizada por el árbol filogenético. Adicionalmente, el análisis de componentes principales, al igual que el análisis de clúster, revelaron niveles de mixtura entre todas las razas del estudio, lo cual es apoyado por el bajo valor de F_{ST} mencionado anteriormente. Si observamos el gráfico del análisis de clúster en $K=4$ (capítulo 5), los individuos de la raza Avileña-Negra Ibérica se agrupan junto con los individuos de la raza Morucha, separándose luego en $K=6$. De esta manera, esta agrupación inicial indica de cierto modo ascendencia entre ambas razas. En general, la raza Asturiana de los Valles se caracterizó por ser la población con niveles de mixtura más amplios de todas las estudiadas. En los diferentes valores de K , los individuos de ésta raza siempre se caracterizaron por tener una combinación de las demás poblaciones. Esto es explicado por su amplia distribución geográfica en todo el territorio español. Igualmente, las razas Pirenaica y Bruna dels Pirineus mostraron niveles importantes de mixtura, generado principalmente por su cercanía en las montañas del Pirineo, confirmando igualmente su cercanía en el árbol filogenético explicado anteriormente. De esta manera, el alto porcentaje de trashumancia que se dio en la mayoría de las razas autóctonas, permitió su cercanía y el consecuente intercambio de genes. Otro gran ejemplo en la zona central del país fueron las razas Avileña-Negra Ibérica y Morucha, que en épocas de trashumancia debieron realizar intercambios genéticos entre ellas.

El futuro de la implementación de la selección genómica. Desequilibrio de ligamiento, persistencia de fases y tamaño efectivo ancestral

Sin lugar a dudas la selección genómica se ha convertido en una herramienta valiosa para mejorar genéticamente las especies de ganado y plantas. El uso directo de información de marcadores SNP en todo el genoma permite una precisa selección de los animales superiores a temprana edad, incluso si el fenotipo bajo la selección es difícil de medir (Daetwyler et al., 2013). Particularmente, los resultados en ganado lechero muestran que la aplicación de la selección genómica puede acortar los intervalos generacionales al tiempo que aumenta la tasa de ganancia genética (VanRaden et al., 2009). En un intento por mejorar la precisión de los estudios de selección genómica, las estimaciones del desequilibrio de ligamiento entre pares de SNP juegan un papel importante a la hora de encontrar marcadores que estén en desequilibrio de ligamiento con un QTL (de Roos et al., 2008).

En los resultados presentados en el capítulo 6, se muestra la estimación del desequilibrio de ligamiento en razas autóctonas de ganado de carne español empleando un panel de SNP de alta densidad. Nuestras estimaciones en promedio se encontraron dentro de los rangos de algunas otras estimaciones realizadas en ganado de carne y leche alrededor del mundo (Qanbari et al., 2010; Salomon-Torres et al., 2014) pero fueron menores a los valores estimados en caballos (Corbin et al., 2010) y en cerdos (Badke et al., 2012). Estas diferencias se deben a varios elementos que afectan las estimaciones del desequilibrio de ligamiento a lo largo de la distancia física entre loci, y que se encuentran relacionados entre sí. Entre estos elementos tenemos principalmente: la deriva génica, la endogamia y el flujo de genes (Slatkin, 2008). Generalmente, los bajos tamaños efectivos y la deriva génica interactúan con la selección afectando las frecuencias alélicas, y por ende generan mayores valores de desequilibrio de ligamiento entre algunos pares de loci (Hill y Robertson, 1966). Por su parte, los valores medios de endogamia retrasan levemente el descenso del desequilibrio de ligamiento debido a que la endogamia aumenta la covarianza entre alelos en diferentes loci (Golding y Strobeck, 1980; Weir y Cockerham, 1969). En nuestro caso, y como indicativo de un tamaño relativamente pequeño del tamaño efectivo de la población, se encontraron valores altos de desequilibrio de ligamiento a distancias cortas, pero un descenso progresivo a medida que se incrementaron las distancias. Debido a estas estimaciones de desequilibrio, la aplicación de la selección genómica en las razas

autóctonas vacunas tiene el potencial de lograr buenas exactitudes de las predicciones genómicas (GEBV, *genomic estimated breeding values*). Las densidades de marcadores obtenidas, de alrededor de 38.000 SNP, fueron cercanas a las encontradas en poblaciones de ganado lechero (de Roos et al., 2008). No obstante, es importante tener en cuenta que las predicciones genómicas no solo dependen del desequilibrio de ligamiento; la estructura familiar y la heredabilidad del carácter son otras características muy importantes a la hora de implementar un programa de selección genómica. De esta manera, y partiendo de la densidad mencionada anteriormente, se sugieren otros estudios adicionales que permitan comparar la capacidad predictiva de los valores mejorantes en función del número de individuos genotipados, la heredabilidad de los caracteres, número de individuos con fenotipos y los programas de selección actuales. Además, es igualmente importante identificar regiones particulares que pueden presentar diferentes grados de LD y evaluar su potencial uso para la detección de QTL y huellas de selección.

Generalmente, la implementación de la selección genómica requiere una gran muestra de individuos de referencia (*training individuals*) que posea un alto número de datos fenotípicos donde se pueda estimar los efectos de los SNP (Dekkers y Hospital, 2002). Por este motivo, se ha propuesto que la población de referencia sea tomada a través de varias poblaciones (Ibañez-Escriche et al., 2009; Toosi et al., 2010), basándose en el hecho de que el alcance y la fase del desequilibrio de ligamiento, y el efecto del QTL se conservan a través de las poblaciones (de Roos et al., 2009). Es de esta manera como la persistencia de fases del desequilibrio de ligamiento, medida como la correlación de fases, permite obtener una muestra de individuos que resuman un panel de SNP preciso para llevar a cabo un esquema de selección genómica (Goddard et al., 2006). Los niveles de persistencia de fases encontrados en las razas autóctonas son prometedores, ya que correlaciones mayores que 0,8 fueron halladas a una densidad de marcadores de 83.000 SNP distanciados entre 30 kb y 50 kb. De esta manera, hemos determinado cual es la distancia media entre marcadores que proveerá suficiente persistencia de fases a través de las poblaciones para llevar a cabo tanto estudios de asociación como de selección genómica (Meuwissen et al., 2001). Sin embargo estos resultados no son del todo concluyentes. Para complementar estos resultados se necesitarán estudios de simulación, que combinando haplotipos de referencia de animales de distintas poblaciones que estén estrechamente relacionados, determinen el efecto de su utilización en la predicción de GEBV en todas las poblaciones (de Roos et al., 2009). Además, es igualmente importante

determinar si estas precisiones justificarían el uso de una población de entrenamiento desde el punto de vista económico.

Las estimaciones del tamaño efectivo, obtenidas a partir del desequilibrio de ligamiento, permitieron complementar los tamaños efectivos calculados en el capítulo 4. De esta manera fue posible calcular los tamaños efectivos ancestrales muchas generaciones atrás, con lo cual se pudo hacer un seguimiento de los tamaños a través del tiempo (Hayes et al., 2003; Tenesa et al., 2007). El desequilibrio de ligamiento, es decir la correlación esperada de frecuencias alélicas en loci ligados, está dado en función del tamaño efectivo y de la tasa de recombinación (Tenesa et al., 2007). A efectos de calcular el tamaño efectivo de la población mediante la fórmula de Sved (1971) se puede asumir la generalización $1\text{Mb} = 1\text{cM}$, o bien tener en cuenta la variación específica de la tasa de recombinación entre los cromosomas obtenida a partir de una aproximación derivada de las diferencias entre el mapa físico y el mapa de ligamiento en cada cromosoma (Arias et al., 2009). Las mayores diferencias entre ambos métodos fueron reveladas a medida que se incrementaron el número de generaciones en el pasado, donde los modelos que implementaron la generalización $1\text{Mb} = 1\text{cM}$ presentaron, en general, mayores valores de tamaño efectivo ancestral. Esto es debido a que los mayores efectos de la recombinación se ven a pequeñas distancias genéticas entre marcadores, que coinciden con mayores generaciones en el pasado.

El progreso del tamaño efectivo de la población es un indicador crucial para evaluar el estado de peligro de una determinada población (FAO, 2010). Los resultados del tamaño efectivo ancestral mostraron que el cuello de botella más llamativo ocurrió hace alrededor de unas 200 generaciones (equivalente a 1300 años, asumiendo un intervalo generacional de 6,5 años). Esta inflexión correspondió a varios elementos relacionados con la divergencia y con el posterior aislamiento de las poblaciones generado por los procesos de selección. Sin embargo, pese a que los procesos de selección de algunas razas especializadas se dio a partir de generaciones más recientes (Gautier et al., 2007), en la actualidad estos valores de tamaño efectivo son cercanos a los obtenidos en la presente tesis en razas autóctonas. Esto indica que la intensidad de selección de razas como la Holstein, caracterizada por el uso continuo de la inseminación artificial, produjo disminuciones de la población efectiva en poco tiempo. En las razas autóctonas, pese a tener procesos de selección menos intensos, el bajo tamaño censal genera un tamaño efectivo mucho menor. Por otra parte, los tamaños efectivos de la población una

generación atrás obtenidos a partir del desequilibrio de ligamiento, comparados con los obtenidos a partir de las estimaciones del pedigrí, mostraron valores inferiores en todas las razas. Este hecho posiblemente es explicado por las sobreestimaciones causadas en los tamaños efectivos debido al poco conocimiento de la profundidad del pedigrí que se tiene actualmente.

Finalmente, se confirma la fuerte relación inversa entre los tamaños efectivos y el descenso del desequilibrio de ligamiento, al igual que la relación positiva entre endogamia y desequilibrio de ligamiento. Estos tres factores jugaron un papel muy importante en el comportamiento de las poblaciones acá evaluadas.

Bibliografía

- ACRUGA. (2011). Programa de mejora de la raza bovina Rubia Gallega. Departamento de Anatomía, Embriología y Genética. Facultad de Veterinaria. Universidad de Zaragoza.
- Arias, J.A., Keehan, M., Fisher, P., Coppieters, W. y Spelman, R. (2009). A high density linkage map of the bovine genome. *BMC Genetics*, 10, 18.
- Badke, Y.M., Bates, R.O., Ernst, C.W., Schwab, C. y Steibel, J.P. (2012). Estimation of linkage disequilibrium in four US pig breeds. *BMC Genomics*, 13, 24.
- Barker, J.S.F. (1999). Conservation of livestock breed diversity. *AGRI*, 25, 33–43.
- Boichard, D., Maignel, L. y Verrier, É. (1997). The value of using probabilities of gene origin to measure genetic variability in a population. *Genetics Selection Evolution*, 29, 5–23.
- Cassell, B.G., Adamec, V. y Pearson, R.E. (2003). Effect of incomplete pedigrees on estimates of inbreeding and inbreeding depression for days to first service and summit milk yield in Holsteins and Jerseys. *Journal of Dairy Science*, 86, 2967–2976.
- Cervantes, I., Goyache, F., Molina, A., Valera, M. y Gutiérrez, J.P. (2008). Application of individual increase in inbreeding to estimate realized effective sizes from real pedigrees. *Journal of Animal Breeding and Genetics*, 125, 301-310.
- Charlesworth, D. y Willis, J.H. (2009). The genetics of inbreeding depression. *Nature Reviews. Genetics*, 10, 783–796.
- Cleveland, M.A., Blackburn, H.D., Enns, R.M. y Garrick, D.J. (2005). Changes in inbreeding of U.S. Herefords during the twentieth century. *Journal of Animal Science*, 83, 992–1001.

- Cymbron, T., Loftus, R.T., Malheiro, M.I. y Bradley, D.G. (1999). Mitochondrial sequence variation suggests an African influence in Portuguese cattle. *Proceedings of the Royal Society B*, 266, 597–603.
- Cymbron, T., Freeman, A.R., Isabel Malheiro, M., Vigne, J.D. y Bradley, D.G. (2005). Microsatellite diversity suggests different histories for Mediterranean and Northern European cattle populations. *Proceedings of the Royal Society B*, 272, 1837–1843.
- Corbin, L.J., Blott, S.C., Swinburne, J.E., Vaudin, M., Bishop, S.C. y Woolliams, J.A. (2010). Linkage disequilibrium and historical effective population size in the Thoroughbred horse. *Animal Genetics*, 41(Suppl. 2), 8–15.
- Daetwyler, H.D., Calus, M.P.L., Pong-Wong, R., de Los Campos, G. y Hickey, J.M. (2013). Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics*, 193, 347–365.
- de Roos, A.P.W., Hayes, B.J., Spelman, R.J. y Goddard, M.E. (2008). Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle. *Genetics*, 179, 1503–1512.
- de Roos, A.P.W., Hayes, B.J. y Goddard, M.E. (2009). Reliability of genomic predictions across multiple populations. *Genetics*, 183, 1545–1553.
- Decker, J.E., McKay, S., Rolf, M.M., Kim, J., Molina-Alcala, A., Sonstegard, T.S., Hanotte, O., Götherström, A., Seabury, C.M., Praharani, L., Babar, M.E., Regitano, L.C.A., Yildiz, M.A., Heaton, M.P., Lui W., Lei, C, Reency, J.M., Saif-Ur-Rehman, M., Schnabel, R.D. y Taylor, J.F. (2014). Worldwide patterns of ancestry, divergence, and admixture in domesticated cattle. *Plos Genetics*, 10, e1004254.
- Dekkers, J.C.M., y Hospital, F. (2002). The use of molecular genetics in the improvement of agricultural populations. *Nature Reviews. Genetics*, 3, 22–32.
- Excoffier, L., Smouse, P.E. y Quattro, J.M. (1992). Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics*, 131, 479–491.
- Falconer, D.S. y Mackay, T.F.C. (1996). *Introduction to quantitative genetics* (4th ed.). Harlow: Addison Wesley Longman Limited.
- FAO (2010). Marcadores moleculares: una herramienta para explorar la diversidad genética. En *La situación de los recursos zogenéticos mundiales para la alimentación y la agricultura* (pp. 393-341). Comisión de recursos genéticos para la alimentación y la agricultura. Roma.
- Felsenstein, J. (1982). How can we infer geographically and history from gene frequencies. *Journal of Theoretical Biology*, 96, 9–20.
- Gautier, M., Faraut, T., Moazami-Goudarzi, K., Navratil, V., Foglio, M., Grohs, C., Boland, A., Garnier, J.G., Boichard, D., Lathrop, G.M., Gut, I.G. y Eggen, A. (2007). Genetic and haplotypic structure in 14 European and African cattle breeds. *Genetics*, 177, 1059–1070.

- Goddard, M.E., Hayes, B.J., McPartlan, H. y Chamberlain, A.J. (2006). Can the same genetic markers be used in multiple breeds? En *Proceedings of the 8th World Congress on Genetics Applied to Livestock Production*. Belo Horizonte, Brazil, 16–22.
- Golding, G.B. y Strobeck, C. (1980). Linkage disequilibrium in a finite population that is partially selfing. *Genetics*, 94, 777–789.
- González-Recio, O., López de Maturana, E. y Gutiérrez, J.P. (2007). Inbreeding depression on female fertility and calving ease in Spanish dairy cattle. *Journal of Dairy Science*, 90, 5744–5752.
- Groeneveld, E., Westhuizen, B.D., Maiwashe, A., Voordewind, F. y Ferraz, J.B.S. (2009). POPREP: a generic report for population management. *Genetics and Molecular Research*, 8, 1158–1178.
- Groeneveld, L.F., Lenstra, J.A., Eding, H., Toro, M.A., Scherf, B., Pilling, D., Negrini, R., Finlay, E.K., Jianlin, H., Groeneveld, E., Weigend, S. y The GLOBALDIV Consortium. (2010). Genetic diversity in farm animals – a review. *Animal Genetics*, 41(Suppl. 1), 6–31.
- Gutiérrez, J.P., Altarriba, J., Díaz, C., Quintanilla, R., Cañón, J. y Piedrafita, J. (2003). Pedigree analysis of eight Spanish beef cattle breeds. *Genetics Selection Evolution*, 35, 43–63.
- Gutiérrez, J. P., Cervantes, I., Molina, A., Valera, M. y Goyache, F. (2008). Individual increase in inbreeding allows estimating effective sizes from pedigrees. *Genetics Selection Evolution*, 40, 359–378.
- Hartl, D.L. y Clark, A.G. (2007). *Principles of population genetics* (4th ed.). Sunderland: Sinauer Associates.
- Hayes, B.J., Visscher, P.M., McPartlan, H.C. y Goddard, M.E. (2003). Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research*, 13, 635–643.
- Hedrick, P.W. (2010). *Genetics of populations* (4th ed.). Tempe, Arizona: Jones and Bartlett Publishers.
- Hill, W.G. y Robertson, A. (1966). The effect of linkage on limits to artificial selection. *Genetical Research*, 8, 269–294.
- Hoffmann, I. (2010). Climate change and the characterization, breeding and conservation of animal genetic resources. *Animal Genetics*, 41(Suppl. 1), 32–46.
- Ibañez-Escriche, N., Fernando, R.L., Toosi, A., Dekkers, J.C. (2009). Genomic selection of purebreds for crossbred performance. *Genetics Selection Evolution*, 41, 12.
- Illumina Inc. (2012). BovineHD Genotyping BeadChip Datasheet. Recuperado de http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf

- Jordana, J., Pelegrín, M. y Piedrafita, J. (1991). Relaciones genéticas en bovinos españoles obtenidas a partir del estudio de caracteres morfológicos. *Información Técnica Económica Agraria*, 87A, 50–64.
- Keller, L.F., Jeffery, K.J., Arcese, P., Beaumont, M.A., Hochachka, W.M., Smith, J.N. y Bruford, M.W. (2001). Immigration and the ephemerality of a natural population bottleneck: evidence from molecular markers. *Proceedings of the Royal Society B*, 268, 1387–1394.
- Laval, G., SanCristobal, M. y Chevalet, C. (2002). Measuring genetic distances between breeds: use of some distances in various short term evolution models. *Genetics Selection Evolution*, 34, 481–507.
- Leberg, P.L. (2002). Estimating allelic richness: effects of sample size and bottlenecks. *Molecular Ecology*, 11, 2445–2449.
- Leroy, G., Verrier, E., Meriaux, J.C. y Rognon, X. (2009). Genetic diversity of dog breeds: within-breed diversity comparing genealogical and molecular data. *Animal Genetics*, 40, 323–332.
- Leroy, G., Mary-Huard, T., Verrier, E., Danvy, S., Charvolin, E. y Danchin-Burge, C. (2013). Methods to estimate effective population size using pedigree data: Examples in dog, sheep, cattle and horse. *Genetics Selection Evolution*, 45, 1
- Lutaaya, E., Misztal, I., Bertrand, J.K. y Mabry, J.W. (1999). Inbreeding in populations with incomplete pedigrees. *Journal of Animal Breeding and Genetics*, 116, 475–480.
- Mc Parland, S., Kearney, J.F., Rath, M. y Berry, D.P. (2007). Inbreeding effects on milk production, calving performance, fertility, and conformation in Irish Holstein-Friesians. *Journal of Dairy Science*, 90, 4411–4419.
- Meuwissen, T.H.E., Hayes, B.J. y Goddard, M.E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157, 1819–1829.
- Nei, M. (1972). Genetic distance between populations. *American Naturalist*, 106, 283–292.
- Oyama, K., Nojima, M., Shojo, M., Fukushima, M., Anada, K. y Mukai, F. (2007). Effect of sire mating patterns on future genetic merit and inbreeding in a closed beef cattle population. *Journal of Animal Breeding and Genetic*, 124, 73–80.
- Qanbari, S., Pimentel, E.C.G., Tetens, J., Thaller, G., Lichtner, P., Sharifi, A.R. y Simianer, H. (2010). The pattern of linkage disequilibrium in German Holstein cattle. *Animal Genetics*, 41, 346–356.
- Salomon-Torres, R., Matukumalli, L.K., Van Tassell, C.P., Villa-Angulo, C., Gonzalez-Vizcarra, V.M. y Villa-Angulo, R. (2014). High density LD-based structural variations analysis in cattle genome. *PLoS ONE*, 9(7), e103046.
- Sánchez-Belda, A. (1984). *Razas bovinas españolas*. Madrid: Publicaciones de Extensión Agraria.

- Sánchez-Belda, A. (2002). *Razas ganaderas españolas. I. Bovinas*. Madrid: FEAGAS y MAPA.
- Slatkin, M. (1994). Gene flow and population structure. En L.A. Real (Ed.), *Ecological Genetics* (pp. 3–17). Princeton: Princeton University Press.
- Slatkin, M. (2008). Linkage disequilibrium – understanding the evolutionary past and mapping the medical future. *Nature Reviews Genetics*, 9, 477–485.
- Sved, J.A. (1971). Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology*, 2, 125–141.
- Tenesa, A., Navarro, P., Hayes, B.J., Duffy, D.L., Clarke, G.M., Goddard, M.E. y Visscher, P.M. (2007). Recent human effective population size estimated from linkage disequilibrium. *Genome Research*, 17, 520–526.
- The Bovine HapMap Consortium. (2009). Genome-wide survey of SNP variation uncovers the genetic structure of cattle breeds. *Science*, 324, 528–532.
- Thompson, J.R., Everett, R.W. y Wolfe, C.W. (2000). Effects of inbreeding on production and survival in Jerseys. *Journal of Dairy Science*, 83, 2131–2138.
- Tisdell, C. (2003). Socioeconomic causes of loss of animal genetic diversity: analysis and assessment. *Ecological Economics*, 45, 365–376.
- Toosi, A., Fernando, R.L. y Dekkers, J.C.M. (2010). Genomic selection in admixed and crossbred populations. *Journal of Animal Science*, 88, 32–46.
- Toro, M.A., Fernández J. y Caballero A. (2009). Molecular characterization of breeds and its use in conservation. *Livestock Science*, 120, 174–195.
- VanRaden, P.M., Van Tassell, C.P., Wiggans, G.R., Sonstegard, T.S., Schnabel, R.D., Taylor, J.F. y Schenkel, F.S. (2009). Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science*, 92, 16–24.
- Villa-Angulo, R., Matukumalli, L.K., Gill, C.A., Choi, J., Van Tassell, C.P. y Grefenstette, J.J. (2009). High-resolution haplotype block structure in the cattle genome. *BMC Genetics*, 10, 19.
- Weir, B.S. y Cockerham, C.C. (1969). Group inbreeding with two linked loci. *Genetics*, 63, 711–742.
- Wright, S. (1951). The genetical structure of populations. *Annals of Eugenics*, 15, 323–354.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19, 395–420.

Capítulo 8

CONCLUSIONES

8. Conclusiones

1. Según la información registrada en los libros genealógicos, las razas de ganado vacuno de carne españolas se caracterizaron por presentar censos crecientes, un amplio intercambio de reproductores entre los diferentes rebaños dentro de cada raza y por un uso limitado de la inseminación artificial, con excepción de la raza Rubia Gallega.
2. La profundidad del pedigrí fue altamente variable entre razas y a lo largo de las generaciones. Esto llevó a posibles subestimaciones de los coeficientes de endogamia en algunas razas.
3. Las tasas de endogamia fueron mayores a las tasas de coascendencia en todas las razas con excepción de la raza Rubia Gallega. Esto evidencia una alta estructura intrapoblacional que restringe los apareamientos en manadas o grupos de vacas.
4. El número efectivo de antepasados fue menor al número efectivo de fundadores, lo cual es un indicativo de que la mayoría de las razas ha experimentado algún cuello de botella durante su vida productiva.
5. En algunas razas, los tamaños efectivos poblacionales fueron más bajos que el mínimo recomendado para el mantenimiento de la variabilidad genética, lo cual puede tener fuertes implicaciones en la subsistencia de la biodiversidad local a mediano y corto plazo.
6. Los resultados obtenidos a partir del panel bovino de alta densidad indicaron un alto grado de variabilidad dentro de individuos pero baja diferenciación entre poblaciones, condición que fue confirmada por el bajo valor de F_{ST} y por las bajas distancias genéticas halladas entre poblaciones.

7. Los análisis de componentes principales y el análisis de clústeres permitieron detectar señales de mixtura entre todas las razas. La raza Asturiana de los Valles fue la que más altos niveles de mezcla presentó con las demás razas.
8. El árbol filogenético mostró una distribución de las razas cercana a la clasificación del ganado español en tres troncos morfológicos: Turdetano, Cantábrico e Ibérico. La excepción fue la raza Retinta, la cual se ubicó más cerca del tronco Ibérico que del Turdetano.
9. El desequilibrio de ligamiento y la persistencia de fases disminuyeron progresivamente a medida que se incrementaron las distancias genéticas entre marcadores.
10. Las estimaciones del desequilibrio de ligamiento en marcadores adyacentes determinaron que un panel de alrededor de 38.000 SNP sería suficiente para realizar programas de selección genómica dentro de cada raza.
11. La persistencia de fases permitió deducir que un panel de entre 50.000 y 83.000 SNP es un buen punto de partida para determinar la viabilidad de establecer una población de referencia combinando las distintas razas.
12. El tiempo de divergencia estimado a partir de la persistencia de fases varió entre 129 y 206 generaciones atrás, lo que nos llevaría a que la divergencia se produjo en la primera mitad de la Edad Media.
13. Los tamaños efectivos ancestrales de las razas autóctonas estimados a partir del desequilibrio de ligamiento comenzaron a disminuir de forma sustancial hace 200 generaciones, prolongándose de forma continuada hasta la actualidad. Estas estimaciones nos permiten estimar el número independiente de segmentos cromosómicos que se situó por encima de 500 y a partir de ello predecir el tamaño de la población de referencia bajo diferentes escenarios de heredabilidad.