

## Detección de regiones genómicas con elevado desequilibrio de ligamiento en poblaciones de vacuno de carne españolas con análisis de BovineHD BeadChip

Mouresan, E.F.<sup>1</sup>; González-Rodríguez, A.<sup>1</sup>; Munilla, S.<sup>1,2</sup>; Moreno, C.<sup>1,7</sup>; Altarriba, J.<sup>1,7</sup>; Díaz, C.<sup>3</sup>; Baro, J.A.<sup>4</sup>; Piedrafita, J.<sup>5</sup>; Molina, A.<sup>6</sup>; Cañas-Álvarez, J.J.<sup>5</sup> y Varona, L.<sup>1,7</sup>

<sup>1</sup>Facultad de Veterinaria. Universidad de Zaragoza. Unidad de Mejora Genética. Zaragoza. Spain.

<sup>2</sup>Facultad de Agronomía. Universidad de Buenos Aires. Producción Animal. CAB. Argentina.

<sup>3</sup>Instituto Nacional de Tecnología Agraria y Alimentaria - INIA, Mejora Genética Animal. Madrid. Spain.

<sup>4</sup>Universidad de Valladolid. Ciencias Agroforestales. Producción Animal. Valladolid. Spain.

<sup>5</sup>Universitat Autònoma de Barcelona. Ciència Animal y de los Alimentos. Barcelona. Spain.

<sup>6</sup>Universidad de Córdoba. Departamento de Genética. Córdoba. Spain.

<sup>7</sup>IA2: Instituto Agroalimentario de Aragón. Zaragoza. Spain.

### PALABRAS CLAVE ADICIONALES

Desequilibrio de ligamiento.  
Vacuno de carne.

### ADDITIONAL KEYWORDS

Linkage disequilibrium.  
Beef cattle.

### INFORMACIÓN

Cronología del artículo.

Recibido/Received: 30.05.2015

Aceptado/Accepted: 17.06.2016

On-line: 15.01.2017

Correspondencia a los autores/Contact e-mail:

[mouresan@unizar.es](mailto:mouresan@unizar.es)

### RESUMEN

El objetivo de este trabajo fue evaluar el patrón de desequilibrio de ligamiento a lo largo del genoma en siete poblaciones españolas autóctonas de vacuno de carne (*Asturiana de los Valles*, *Avileña Negra-Ibérica*, *Bruna dels Pirineus*, *Morucha*, *Pirenaica*, *Retinta* y *Rubia Gallega*). Para ello, se utilizó el BovineHD BeadChip con el que se genotiparon 171 tríos formados por individuo/padre/madre. Después del filtrado, se dispuso de 573.134 SNP. A partir de esta información se definió un parámetro que mide el desequilibrio medio del genoma por regiones de 1Mb en cada una de las poblaciones. Los resultados mostraron que el desequilibrio de ligamiento es muy heterogéneo a lo largo del genoma y que, además, esta heterogeneidad es consistente entre poblaciones. Las causas de esta heterogeneidad pueden ser, o bien estructurales y atribuibles a una menor tasa de mutación y/o recombinación, o bien consecuencia de procesos de selección estabilizadora.

### Detection of genomic regions with high linkage disequilibrium in Spanish Beef Cattle Populations by BovineHD BeadChip analysis

### SUMMARY

The objective of this study was to evaluate the pattern of linkage disequilibrium along the genome in seven autochthonous Spanish cattle beef populations (*Asturiana de los Valles*, *Avileña Negra-Ibérica*, *Bruna dels Pirineus*, *Morucha*, *Pirenaica*, *Retinta* and *Rubia Gallega*). The BovineHD BeadChip was used to genotype 171 trios of individual/sire/dam. 573,134 SNPs were available after filtering. With this information, a parameter that measures the mean disequilibrium of the genome in regions of 1 Mb in each population was defined. The results show that the linkage disequilibrium is very heterogeneous along the genome, and this heterogeneity is consistent among the considered populations. The causes of this heterogeneity could be structural, and attributed to a lower mutation rate and/or recombination rate, or a result of stabilizing selection.

### INTRODUCCIÓN

En los últimos años, la disponibilidad de técnicas de genotipado masivo de marcadores SNP ha permitido el desarrollo de procedimientos de evaluación genómica (Meuwissen *et al.*, 2001),

que se ha constituido en una herramienta imprescindible en la mejora genética de vacuno de leche (Hayes *et al.*, 2009). Sin embargo, su utilización en vacuno de carne es mucho más limitada. Entre las causas de esta dificultad se pueden mencionar el menor censo de las poblaciones de

vacuno de carne y la menor implantación de la inseminación artificial (Berry *et al.*, 2016).

La evaluación genómica a partir de la información de varias poblaciones se ha planteado como una posible alternativa a este problema, pero los resultados obtenidos tanto mediante simulación (Mouresan *et al.*, 2014), como con datos experimentales (Kachman *et al.*, 2013) no son muy esperanzadores.

El éxito de la evaluación genómica a partir de varias poblaciones está ligado a la presencia de un elevado desequilibrio de ligamiento y a la persistencia de su fase entre las poblaciones. Varios autores han estudiado la persistencia de la fase entre poblaciones en relación a la distancia física entre los marcadores (de Roos *et al.*, 2008) y como una medida más de la diversidad genética (Cañas-Alvarez *et al.*, 2016), pero apenas se ha estudiado la arquitectura genética de esta persistencia.

Por lo tanto, el objetivo de este trabajo es realizar un estudio comparativo de las regiones del genoma donde aparezca un alto grado de desequilibrio de ligamiento entre siete poblaciones de ganado vacuno autóctono español (*Asturiana de los Valles*- AV-, *Avileña Negra-Ibérica*- ANI-, *Bruna dels Pirineus* -BP-, *Morucha* -Mo-, *Pirenaica* -Pi-, *Retinta* -Re- y *Rubia Gallega* -RG-).

## MATERIAL Y MÉTODOS

Se realizó un genotipado de SNP de alta densidad mediante el *BovineHD BeadChip* (Illumina Inc, USA) a un total de 171 tríos de padre/madre/hijo pertenecientes a 7 poblaciones españolas de vacuno de carne, que incluyeron la razas *Asturiana de los Valles* (AV, n=25), *Avileña-Negra Ibérica* (ANI, n=24), *Bruna dels Pirineus* (BP, n=25), *Morucha* (Mo, n=25), *Pirenaica* (Pi, n=24), *Retinta* (Re, n=24) y *Rubia Gallega* (RG, n=24). Los progenitores fueron seleccionados con la menor relación de parentesco posible entre sí con el objetivo de maximizar la representatividad de la muestra.

En primer lugar, se realizó un control de calidad mediante el software PLINK (Purcell *et al.*, 2007). En

concreto, se utilizaron los siguientes requisitos: 1) SNP localizados en los cromosomas autosómicos, 2) error mendeliano inferior al 0.05, 3) call rate superior al 0.95 a nivel de individuo y de SNP, 4) Frecuencia Alélica Menor (MAF) superior al 0.01 en cada una de las poblaciones. El resultado de este filtrado seleccionó 573.134 marcadores SNP, que cubrieron un total de 2.511.259 kb, y representan, en promedio, la presencia de un marcador cada 4,317 kb. Posteriormente, se establecieron los haplotipos parentales mediante el software BEAGLE (Browning y Browning, 2009).

A partir de esta información, se definieron ventanas de 1 Megabase (Mb) que cubrieron todo el genoma, y se calculó el desequilibrio de ligamiento (Falconer y Mackay, 1996) entre todos los marcadores presentes dentro de cada ventana como

$$r = D / \sqrt{(p_1q_1p_2q_2)}$$

donde

$$D = f_{11}f_{22} - f_{12}f_{21}$$

El número promedio de marcadores presentes en cada ventana fue 226 con una desviación típica de 61. Con el objetivo de evitar valores extremos generados por un número reducido de marcadores, se analizaron exclusivamente aquellas ventanas que presentaron al menos 100 SNP. Por lo tanto, en cada ventana de 1 Mb se calcularon, como mínimo, 4900 estimaciones de desequilibrio de ligamiento.

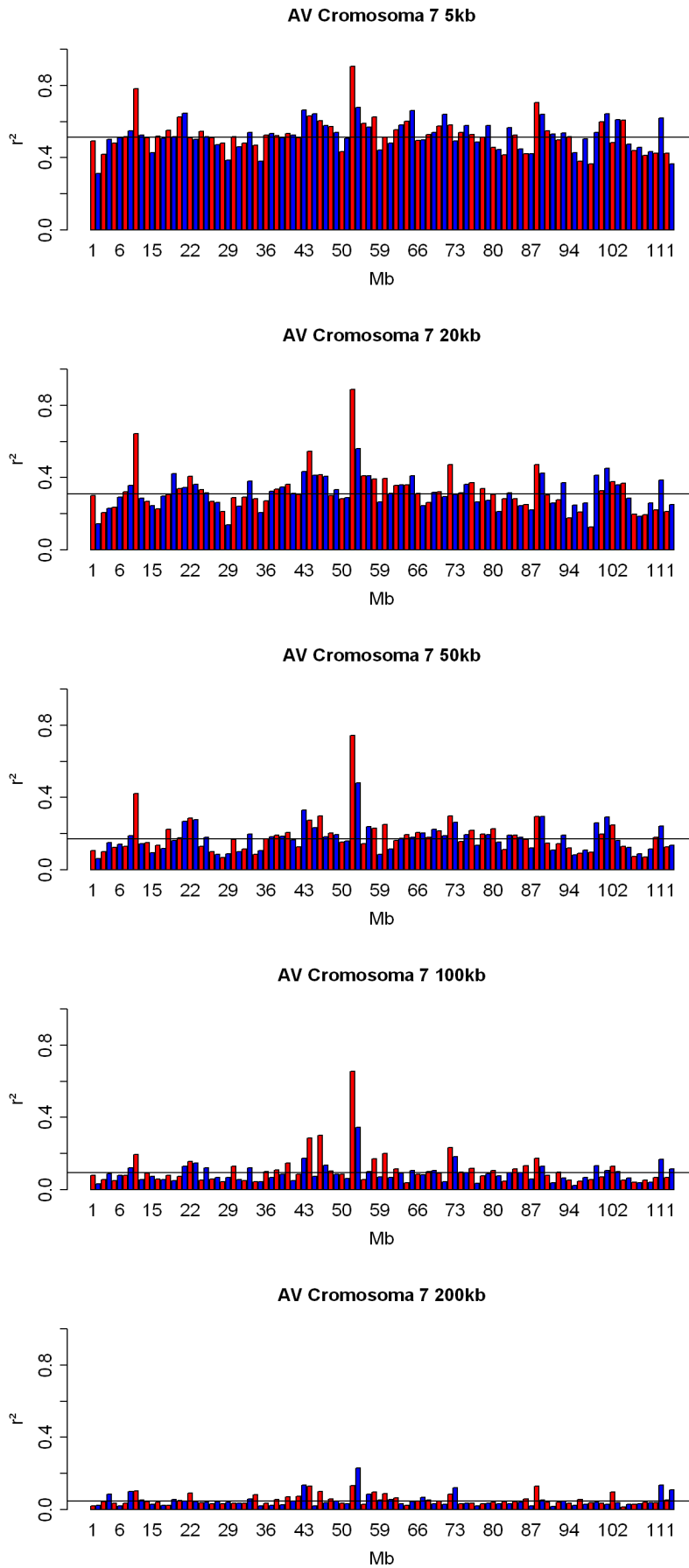
En primer lugar, se calculo el r<sup>2</sup> medio entre marcadores que se encuentran a distancias de 5, 20, 50, 100 y 200 kb (±2,5 kb) dentro de cada región de 1 Mb, y finalmente, con el objetivo de obtener una medida global del desequilibrio de ligamiento por región y su relación con la distancia de mapa, se definió un parámetro β, que se calculó como

$$\beta = \sum_{i=1}^n r_i^2 / \sum_{i=1}^n c_i$$

donde r<sup>2</sup> es el desequilibrio de ligamiento elevado al cuadrado, c es la distancia de mapa en Mb y n es el número de pares de SNP en cada región.

**Tabla I.** Valores medios (desviación típica) de r<sup>2</sup> entre marcadores con varias distancias entre ellos en las 7 poblaciones (Mean values (standard deviation) of r<sup>2</sup> between markers with various distances between them in the 7 populations).

	5 kb	20 kb	50 kb	100 kb	200 kb
AV	0,513 (0,087)	0,309 (0,097)	0,170 (0,081)	0,093 (0,061)	0,047 (0,036)
ANI	0,521 (0,087)	0,324 (0,097)	0,190 (0,082)	0,116 (0,063)	0,071 (0,042)
BP	0,517 (0,088)	0,316 (0,097)	0,181 (0,082)	0,107 (0,065)	0,062 (0,040)
Mo	0,515 (0,087)	0,316 (0,096)	0,182 (0,081)	0,107 (0,062)	0,062 (0,039)
Pi	0,528 (0,089)	0,334 (0,098)	0,202 (0,084)	0,129 (0,065)	0,084 (0,044)
Re	0,520 (0,087)	0,323 (0,096)	0,192 (0,082)	0,120 (0,063)	0,075 (0,042)
RG	0,519 (0,089)	0,321 (0,098)	0,187 (0,083)	0,114 (0,065)	0,070 (0,041)



**Figura 1.** Valores medios de  $r^2$  entre marcadores en varias distancias para el cromosoma 7 en la población AV (Mean values of  $r^2$  between markers in various distances on the chromosome 7 for the AV population).

**Tabla II.** Valores promedios y desviaciones típicas del parámetro  $\beta$  para cada población (Average values and standard deviations of the  $\beta$  parameter in each population).

Poblaciones	Promedio $\beta$	Desviación típica
AV	0,246	0,118
ANI	0,312	0,131
BP	0,284	0,128
Mo	0,284	0,126
Pi	0,350	0,138
Re	0,324	0,129
RG	0,311	0,129

## RESULTADOS Y DISCUSIÓN

Los resultados de desequilibrio de ligamiento medio a distancias de 5, 20, 50, 100 y 200 kb se presentan en la **tabla I**, presentando valores similares a los obtenidos por Cañas-Álvarez *et al.* (2016). Sin embargo, estos valores muestran heterogeneidad a lo largo del genoma. Así, en la **figura 1** se presentan los resultados obtenidos de este análisis en la población AV para el cromosoma 7. Se puede observar que los valores medios de  $r^2$  a una distancia específica son muy variables entre las regiones genómicas. Por este motivo, se optó por utilizar un parámetro  $\beta$  que cuantifica la cantidad de desequilibrio de ligamiento por unidad de distancia y permite comparar regiones del genoma mediante una única magnitud.

Los valores promedios y las desviaciones típicas del este parámetro a lo largo del genoma y para cada población se muestran en la **tabla II**. Los valores resultaron entre 0,246 (AV) y 0,350 (Pi) y, como puede observarse, la orden de las poblaciones es coherente con la presentada por Cañas-Álvarez *et al.* (2016) y con los resultados de la **tabla I**. Las diferencias entre los valores medios del parámetro  $\beta$  obtenidos para

**Tabla III.** Regiones genómicas que presentan un valor extremo en todas las poblaciones y los genes localizados en estas regiones (Genomic regions that present an extreme value in all populations and genes located in these regions).

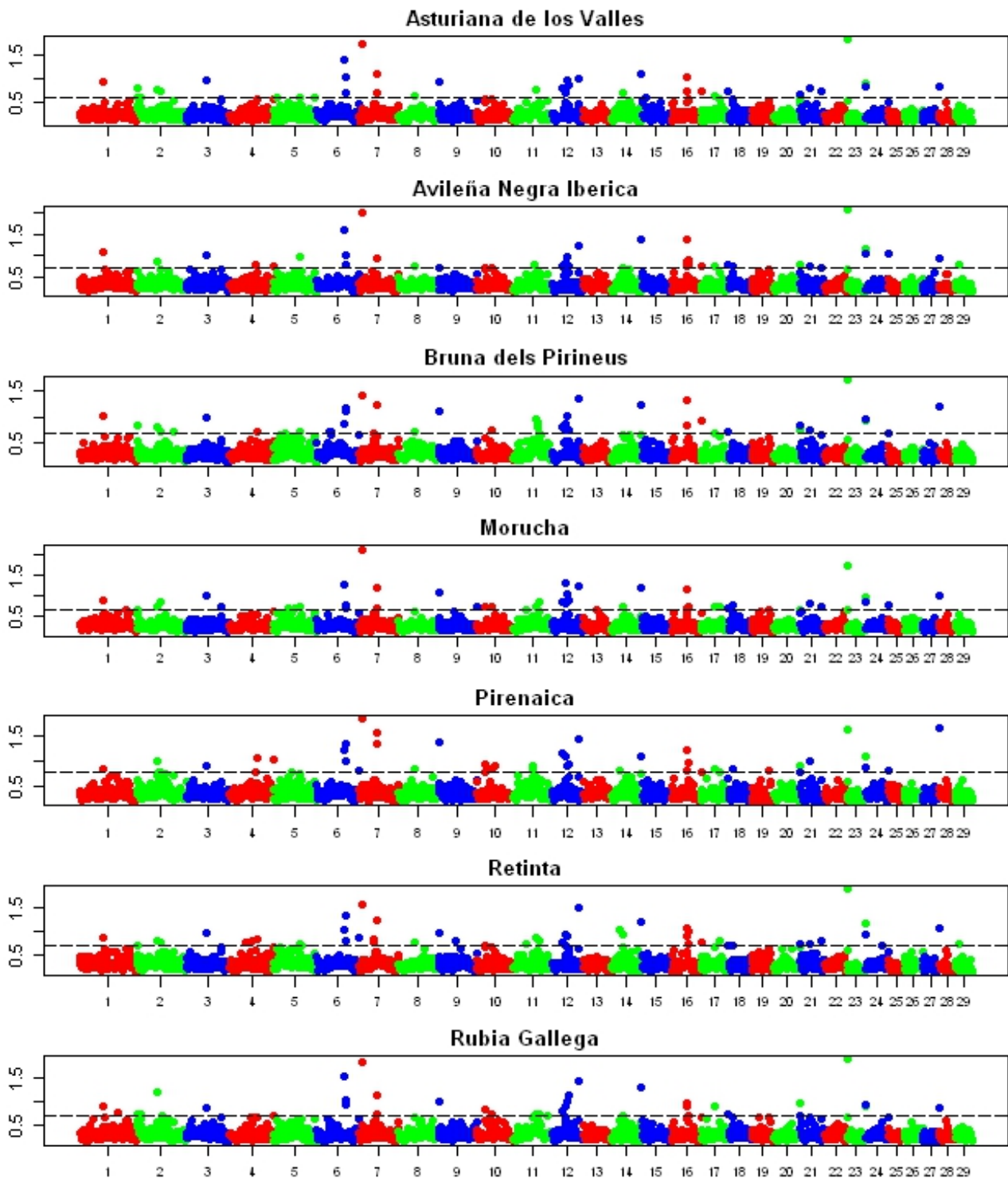
Cromosoma	Mb	Genes
6	78	-
7	10	ILVBL, SYDE1, CASP14, U6, CCDC105, SLC1A6, OR7C2, U6, Olfactory Receptor (x6)
	54	CYSTM1, PFDN1, HBEGF, SLC4A9, U6, SRA1, APBB3, SLC35A4, CD14, TMCO6, IK, WDR55, DND1, HARS, HARS2, ZMAT2, PCDHA5, PCDHA9, PCDHB13
12	77	CLDN10, DZIP1, DNAJC3
15	1	Olfactory Receptor (x7), OR9G1
23	1	KHDRBS2

cada una de las poblaciones pueden ser atribuibles a la presencia de mayor o menor grado de parentesco entre los individuos a causa de la variabilidad en el tamaño efectivo de las poblaciones.

La variabilidad a lo largo de genoma en cuanto al desequilibrio de ligamiento también se refleja en los resultados del valor  $\beta$  medio en las regiones genómicas de 1 Mb en cada población (**figura 2**). Pese a todo, es muy relevante que el patrón de desequilibrio de ligamiento muestre un alto grado de similitud entre las poblaciones. Como prueba, si se seleccionan aquellas regiones que presentan un valor de  $\beta$  superior a la media más 3 desviaciones típicas (0,600 (AV) – 0,764 (Pi)) en cada población, se encuentra que de las 84 regiones genómicas presentes, 20 de ellas aparecen seleccionadas en todas las poblaciones. Dos pueden ser las causas potenciales de esta heterogeneidad. Por una parte, es posible que reflejen la variabilidad en las tasas de mutación y recombinación presentes en el genoma bovino, como sugirió Arias *et al.* (2009). Por otra parte, es plausible que estas regiones hayan sido afectadas por procesos de selección estabilizadora, que permite sospechar que alberguen genes o regiones genómicas de especial relevancia, cuya estructura se mantiene evolutivamente en mayor medida que en otras regiones.

Para verificar esta hipótesis, se calculó el valor medio de  $\beta$  para cada región de 1 Mb entre todas las poblaciones (**figura 3**) y se seleccionaron aquellas regiones que mostraron un valor extremo por encima de la media + 7 desviaciones típicas (1,151). Este proceso seleccionó 6 regiones genómicas (**tabla III**). El contenido de estas regiones se analizó mediante la base de datos BIOMART ([www.biomart.org](http://www.biomart.org), Smedley *et al.* 2015) y se localizaron 48 genes (**tabla III**) que codifican componentes de la membrana celular o participan en procesos del sistema nervioso y de la transducción de señales (**figura 4**). Un buen ejemplo de genes conservados evolutivamente se encuentra en el cromosoma 7 y Mb 54. Son los genes PCDHA5 (*Protocadherin Alpha 5*), PCDHA9 (*Protocadherin Alpha 9*) y PCDHB13 (*Protocadherin Beta 13*) que codifican proteínas del grupo de protocadherinas. Estos genes se encuentran agrupados en 2 o 3 regiones del genoma y tienen un papel crítico en el establecimiento y la función de las conexiones de las células del sistema nervioso (Sano *et al.*, 1993).

Al margen de la causa que provoque esta heterogeneidad, se constata de manera muy evidente la presencia de variabilidad en el desequilibrio de ligamiento entre regiones del genoma. Sin embargo, los procedimientos de selección genómica (Meuwissen *et al.*, 2001; Aguilar *et al.*, 2010) asignan a-priori la misma capacidad predictiva a todos los marcadores SNP, independientemente de su localización en regiones con alto o bajo desequilibrio de ligamiento. En futuros estudios se pretende estudiar la posibilidad de utilizar esta información para ponderar la importancia de las



**Figura 2.** Distribución de los valores del parámetro  $\beta$  a lo largo del genoma en las 7 poblaciones (Distribution of the values of the  $\beta$  parameter along the genome for the 7 populations).

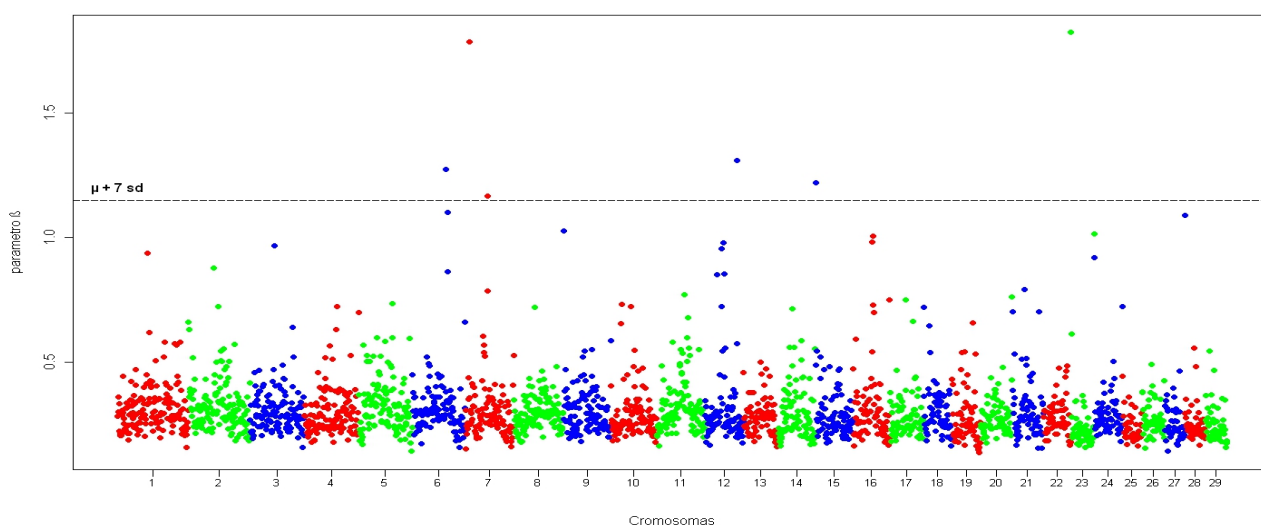
regiones del genoma en función de la magnitud del desequilibrio de ligamiento presente en el entorno de cada marcador SNP.

#### CONCLUSIONES

En las poblaciones autóctonas de vacuno de carne se observa un patrón de desequilibrio de ligamiento heterogéneo a lo largo del genoma. Las posibles causas

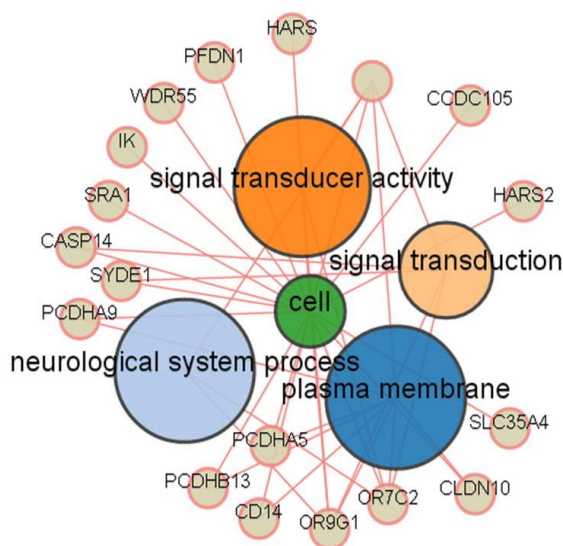
de esta heterogeneidad incluyen la variabilidad en la tasa de recombinación y mutación a lo largo del genoma. Sin embargo, las regiones genómicas asociadas a un mayor desequilibrio de ligamiento son consistentes entre poblaciones indicando la presencia de selección estabilizadora.

#### AGRADECIMIENTOS



**Figura 3.** Distribución de los valores medios del parametro  $\beta$  de cada region a lo largo del genoma (Distribution of the average values of the  $\beta$  parameter of each region along the genome).

Description	P-Value	Corrected P-Value
plasma membrane	1.08748e-08	1.54423e-06
neurological system process	3.23796e-08	4.5979e-06
signal transducer activity	1.96449e-05	0.00278958
signal transduction	0.000280588	0.0398436
cell	0.000422832	0.0600422
cellular_component	0.00741363	1
tRNA metabolic process	0.0306504	1
cell adhesion	0.0378676	1



**Figura 4.** Principales rutas biológicas en las que participan los genes encontrados (Main biological pathways in which the located genes participate).

Esta investigación ha sido posible gracias al soporte económico del proyecto AGL2010-15903. Aldemar González-Rodríguez agradece el apoyo financiero aportado por la beca universitaria BES-2011-045434. Se agradece el apoyo de las asociaciones españolas de vacuno de carne y de la Federación Española de Asociaciones de Ganado Selecto (FEAGAS).

**BIBLIOGRAFÍA**

Aguilar, I; Misztal, I; Johnson, D. L; Legarra, A; Tsuruta, S. and Lawlor, T.J. 2010. Hot topic: a unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci*, 93: 743–52.

Arias, J.A; Keehan, M; Fisher, P; Coppieters, W. and Spelman, R. 2009. A high density linkage map of the bovine genome. *BMC Genetics*. 10: 18.

Berry, D.P; Garcia, J.F and Garrick, D.J. 2016. Development and implementation of genomic predictions in beef cattle. *Animal Frontiers*, 6: 32.

Browning, B.L and Browning, S.R. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am J Hum Genet*, 84: 210–23.

Cañas-Alvarez, J; Mouresan, E; Varona, L; Diaz, C; Molina, A; Baro, J.A; Altarriba, J; Carabano, M; Casellas, J. and Piedrafita, J. 2016. Linkage disequilibrium, persistence of phase and effective population size in Spanish local beef cattle breeds assessed through a SNP high density chip. *J Anim Sci* (in press)

De Roos, A.P.W; Hayes, B.J; Spelman, R.J. and Goddard, M.E. 2008. Linkage Disequilibrium and Persistence of Phase in Holstein–Friesian, Jersey and Angus Cattle. *Genetics*, 179: 1503-1512.

Falconer and Mackay. 1996. Introducción a la genética cuantitativa. Addison Wesley Longman Limited. Essex, England

Hayes, B.J; Bowman, P.J; Chamberlain, A.J. and Goddard, M.E. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J Dairy Sci*, 92: 433-443.

Illumina Inc. 2012. BovineHD Genotyping BeadChip Datasheet. [http://www.illumina.com/documents/products/datasheets/datasheet\\_bovineHD.pdf](http://www.illumina.com/documents/products/datasheets/datasheet_bovineHD.pdf). (20/02/2014).

- Kachman, S.D; Spanger, M.L; Bennett, G.L; Hanford, K.J; Kuehn, L.A; Snelling, W.M; Thallman, R.M; Saatchi, M; Garrick, D.J; Schnabel, R.D; Taylor, J.F. and Pollak, E.J. 2013. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. *Genetics, Selection, Evolution: GSE*, 45: 30
- Meuwissen, T.H.E; Hayes, B.J. and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 157: 1819–29.
- Mouresan, E.F; Munilla, S; Diaz, C; Gonzalez-Rodriguez, A; Piedrafita, J; Avilés, C; Baro, J.A; Moreno, C. and Varona, L. 2014. Evaluation of the use of a Meta-population for Genomic Selection in Spanish Beef Cattle Breeds, 10 WCGALP, 17-22 August 2014, Vancouver, BC, Canada.
- 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. and Sham, P.C. 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet*, 81: 559–575.
- Sano, K; Tanihara, H; Heimark, R.L; Obata, S; Davidson, M; St John, T; Taketani, S. and Suzuki, S. 1993. Protocadherins: a large family of cadherin-related molecules in central nervous system. *EMBO J*, 12: 2249–56.
- Smedley, D; Haider, S; Durinck, S; Pandini, L; Provero, P; Allen, J; Arnaiz, O; Awedh, M.H; Baldock, R; Barbiera, G; Bardou, P; Beck, T; Blake, A; Bonierbale, M; Brookes, A.J; Bucci, G; Buetti, I; Burge, S; Cabau, C; Carlson, J.W; Chelala, C; Chrysostomou, C; Cittaro, D; Collin, O; Cordova, R; Cutts, R.J; Dassi, E; Di Genova, A; Djari, A; Esposito, A; Estrella, H; Eyra, E; Fernandez-Banet, J; Forbes, S; Free, R.C; Fujisawa, T; Gadaleta, E; Garcia-Manteiga, M.J; Goodstein, D; Gray, K; Guerra-Assunção, J.A; Haggarty, B; Han, D.J; Han, B.W; Harris, T; Harshbarger, J; Hastings, R.K; Hayes, R.D; Hoede, C; Hu, S; Hu, Z.L; Hutchins, L; Kan, Z; Kawaji, H; Keliet, A; Kerhornou, A; Kim, S; Kinsella, R; Klopp, C; Kong, L; Lawson, D; Lazarevic, D; Lee, J.H; Letellier, T; Li, C.Y; Lio, P; Liu, C.J; Luo, J; Maass, A; Mariette, J; Maurel, T; Merella, S; Mohamed, A.M; Moreews, F; Nabihoudine, I; Ndegwa, N; Noirot, C; Perez-Llamas, C; Primig, M; Quattrone, A; Quesneville, H; Rambaldi, D; Reecy, J; Riba, M; Rosanoff, S; Saddiq, A.A; Salas, E; Sallou, O; Shepherd, R; Simon, R; Sperling, L; Spooner, W; Staines, D.M; Steinbach, D; Stone, K; Stupka, E; Teague, J.W; Dayem Ullah, A.Z; Wang, J; Ware, D; Wong-Erasmus, M; Youens-Clark, K; Zadissa, A; Zhang, S.J. and Kasprzyk, A. 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res*, 43: 589–98.