# BUSQUEDA DE ALELOS ALTERNATIVOS ENTRE CERDOS IBERICOS Y DUROC

Barragán, C., Benítez, R., Fernández, A. I., Fernández, A., García, F., Núñez, Y., Ovilo, C., Rodríguez, C. y Silió, L.
Departamento de Mejora Genética Animal, INIA, 28040 Madrid
E-mail: silio@inia.es

## INTRODUCCIÓN

Cerdos Ibéricos colorados procedentes de Portugal y España contribuyeron en el siglo XIX al origen de la raza Duroc–Jersey, posteriormente denominada Duroc. Hacia 1964 se comenzaron a importar animales de esta raza, cuyos híbridos con madres Ibéricas se utilizan mayoritariamente en la producción de cerdos de tipo Ibérico, especialmente en la basada en engorde intensivo. Diversas comparaciones entre cerdos Ibéricos puros y cruzados con Duroc han permitido cuantificar el menor crecimiento y rendimiento en piezas nobles así como la superior calidad de carne de los primeros (Ventanas et al., 2006; Serrano et al., 2008). El estudio del transcriptoma de *m. longissimus* ha permitido identificar decenas de genes diferencialmente expresados entre estos tipos genéticos integrados en redes génicas relacionadas con procesos metabólicos subyacentes a las diferencias fenotípicas observadas (Ovilo et al., 2014). En el presente trabajo se pretende continuar el análisis de las diferencias genéticas entre ambas razas, utilizando información de genotipado masivo de SNPs y de secuenciación masiva, con el objetivo de identificar posibles variantes génicas con alelos exclusivos de cada genoma para el diseño de un panel de marcadores genéticos utilizable en la autentificación del origen racial de animales y productos.

## MATERIAL Y MÉTODOS

**Animales, Genotipos y Secuencias.** Mediante el *Porcine SNP60 BeadChip* (Illumina, San Diego, CA, USA), se realizó el genotipado de un total de 169 cerdos: 32 de diversas líneas Duroc, 76 Ibéricos de diferentes orígenes genéticos y otros 61 de la línea Torbiscal. Tras pasar el control de calidad estándar, se descartaron aquellos SNPs monomórficos en todos los animales genotipados, los localizados en los cromosomas X e Y y los no mapeados según *Sscrofa* 10.2, siendo 43.451 los SNPs autosómicos utilizados. Muestras de ADN de 17 de estos animales, cinco Duroc y 12 Ibéricos (dos Torbiscal), fueron secuenciadas en grupos de dos por el Centro Nacional de Análisis Genómico (Barcelona).

**Analisis de datos.** La divergencia entre las frecuencias alélicas de Duroc y cada uno de los grupos Ibéricos se estimó mediante los estadísticos de Wright $F_{ST} = {(H_T - H_S)}/{H_T}$ y el propuesto por Hedrick (2005) corregido por su máximo alcanzable $F'_{ST} = {F_{ST}}/{F_{ST(max)}}$, en función de la heterocigosidad dentro de grupo $F_{ST(max)} = {(1 - H_S)}/{(1 + H_S)}$. De acuerdo con Akey et al. (2010), se calcularon los valores tipificados $d_i = \sum_{j \neq i} {\left(F_{ST}^{ij} - E\left[F_{ST}^{ij}\right]\right)}/{sd\left[F_{ST}^{ij}\right]}$ para cada SNP, siendo *i* = Duroc, *j* = Ibéricos o Torbiscal, $E[F_{ST}^{ij}]$ y $sdE[F_{ST}^{ij}]$ la media y desviación típica de $F_{ST}^{ij}$ (o en su caso de $F'^{ij}_{ST}$). Los valores $d_i$ (o $d'_i$) se promediaron por ventanas solapantes de diez SNPs sucesivos. Las regiones de máxima diferenciación genética se identificaron como aquellas que exceden los percentiles $\mathbf{P}_{99\cdot9}$ o $\mathbf{P}_{99\cdot5}$ de la distribución a lo largo del genoma de los valores promediados $d_i$. El contenido génico de estas regiones se identificó utilizando Biomart y las funciones de los mismos mediante la herramienta de notación funcional del Max Planck Institute for Molecular Genetics. El alineamiento de las secuencias obtenidas, y la identificación de variantes nucleotídicas (SNVs) y de los homocigotos se realizó siguiendo el protocolo GATK. Las variantes inicialmente detectadas se verificaron mediante el software SAMTools.

## RESULTADOS Y DISCUSIÓN

**Diversidad y diferenciación genética**. Los valores de diversidad genética o heterocigosidad esperada en los grupos de animales Duroc, Ibéricos e Ibéricos Torbiscal calculados para el conjunto de SNPs fueron 0,323, 0,250 y 0,196, respectivamente. Los coeficientes de diferenciación entre Duroc e Ibéricos fueron $F_{ST}$ = 0,137 y $F'_{ST}$ = 0,247, para un máximo posible $F_{ST(max)}$ = 0,554, dado el valor de $H_S$ = 0,287. En la comparación entre Duroc y Torbiscal los valores correspondientes fueron $F_{ST}$ = 0,188, $F'_{ST}$ = 0,320, $F_{ST(max)}$ = 0,588 y $H_S$ = 0,260. El empleo de $F'_{ST}$ permite la comparación de medidas de diversidad basadas en diferente número de loci o en marcadores de diversos niveles de variación. Su utilización en este caso parece justificada dado que entre los SNPs incluidos en el *Porcine SNP60 BeadChip* figuran un buen número de monomórficos en Ibérico, raza que a diferencia de la Duroc no fue considerada en su diseño. El uso de uno u otro estadístico, más allá de las diferencias de escala, da lugar a algunas diferencias en la distribución de los promedios por ventana de $d_i$ o $d'_i$ calculados a lo largo de los 18 autosomas (Figuras 1 *a* y *b*). Ello se refleja en las ventanas y regiones de máxima diferenciación genética identificadas en cada caso. Considerando el percentil $\mathbf{P}_{99.9}$, el procedimiento basado en el estadístico $F_{ST}$ permite identificar las regiones cromosómicas con frecuencias alélicas más divergentes en los cromosomas 4, 5, 14 y 15 (con dos regiones de ventanas sucesivas), mientras que mediante el basado en $F'_{ST}$ las regiones de mayor divergencia se localizan en los cromosomas 4, 5, 9, 14 (3 regiones) y 15 (2 regiones). Teniendo en cuenta los antecedentes conocidos de la raza Duroc, estas regiones no sólo pueden interpretarse como huellas en el genoma de la selección efectuada en esta raza sino como resultantes de las contribuciones de las poblaciones americanas ancestrales que intervinieron en su formación. No puede descartarse asimismo que representen la introgresión de alelos de origen asiático experimentada por esta raza como por la práctica totalidad de las occidentales.

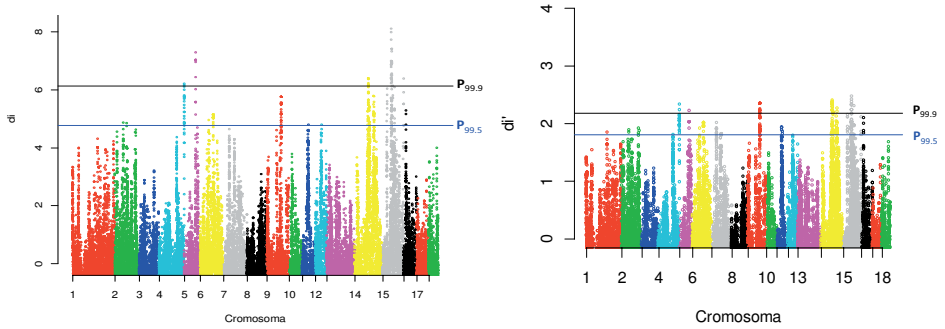



***Figura1 a).*** *Distribución de valores $d_i$ (Duroc vs. Ibericos + Torbiscal) promediados por ventanas deslizantes de 10 SNPs a lo largo de los autosomas* ***b)*** *Distribución equivalente de valores $d'_i$ basados en el estadístico $F'_{ST}$*

**Búsqueda de posibles variantes génicas privativas**. Definimos como variantes génicas (SNVs) privativas aquellos SNPs, inserciones o deleciones que presentan variantes alternativas en cada una de las razas analizadas. Nuestro interés en los SNVs poseedores de esta característica se debe a que aportan la máxima información por locus a la discriminación entre orígenes genéticos. Su búsqueda, a partir de las 17 secuencias disponibles, se ha centrado en las 31 regiones formadas por las 215 ventanas con frecuencias divergentes que exceden el percentil $\mathbf{P}_{99.5}$, con una longitud total de 28.378.218 pb distribuidas en once cromosomas. Estas regiones contienen en total 190 genes, de ellos sólo 123 conocidos, algunos duplicados (*7SK*, *CALCB*, *SNORA19*) o con hasta siete copias (*U6*). El análisis funcional preliminar de los 114 genes restantes ha mostrado el enriquecimiento de algunas funciones básicas, entre otras relacionadas con la cadena de proteinas MAPK (mitogen-activated protein kinases) que comunican señales en receptores de la membrana celular inductoras de transcripción del ADN nuclear.

Mediante la comparación de secuencias de individuos de ambas razas, se han detectado 255 SNVs, de ellos 91 con localización intragénica, para los que los animales secuenciados

son homocigotos con variantes alternativas en cada raza (Tabla 1). Estos resultados están fuertemente condicionados por la cobertura mínima exigida a las secuencias. Debe subrayarse asimismo que el limitado número de secuencias hace necesaria una posterior validación mediante el genotipado de aquellos más idóneos en un número mayor de animales. Entre los genes con SNVs privativos en sus secuencias destacan por el elevado número de éstos los genes *KHDRBS2* (SSC7) con 14 SNVs privativos, *WRN* y *TNKS* (SSC15) con 40 y 20 SNVs privativos respectivamente. El interés de estos genes se extiende a sus características funcionales relacionadas con las diferencias fenotípicas entre las razas comparadas. El gen *KHDRBS2* se ha relacionado en ratón con el desarrollo de la masa muscular (Karst et al., 2011). Los genes *WRN* y *TNKS* se relacionan con la regulación de la telomerasa, una de las funciones que presentan enriquecimiento en las regiones divergentes. Para Steibel et al. (2011), que han identificado en *WRN* un *e*QTL coincidente con un *p*QTL con efectos sobre área de lomo y terneza de la carne de cerdo, las funciones de esta proteina relacionadas con el crecimiento celular y la transcripción sugieren su posible efecto sobre el desarrollo muscular y la calidad de carne.

***Tabla 1***. *Regiones del genoma con frecuencias alélicas divergentes en las que se han detectado potenciales SNVs privativos, distribución de los 255 identificados y de los 91 con localización intragénica.*

| Cromosoma | Región (*pb*) | Nº de Genes | Nº SNVs privativos | Genes con SNVs privativos | SNVs privativos intragénicos |
|---|---|---|---|---|---|
| 2 | 138.702 | 3 | 2 | 1 | 2 |
| 4 | 1.286.528 | 8 | 26 | 4 | 1;3;1;1 |
| 7 | 738.747 | 4 | 17 | 1 | 14 |
| 7 | 339.534 | 3 | 1 | 1 | 1 |
| 15 | 1.538.921 | 11 | 113 | 2 | 40; 1 |
| 15 | 1.302.509 | 6 | 90 | 3 | 2; 1; 20 |
| 15 | 1.810.768 | 7 | 4 | 1 | 4 |
| 15 | 563.092 | 3 | 2 | 0 | 0 |
| Total | | 45 | 255 | 13 | 91 |

## REFERENCIAS BIBLIOGRÁFICAS

• Akey J.M. 2010. PNAS 107 (3): 1160-1165. • Hedrick P.W. 2005. Evolution 59: 1633-1638. • Karst S. et al. 2011. Mamm. Genome. 22: 530-543. • Óvilo C. et al. 2014 BMC Genomics 15: 413. • Serrano M.P. et al. 2008. Meat. Sci. 78: 420–428. • Steibel J.P. et al. 2011. PLoS ONE. 6 (2 ) e16766. • Ventanas et al. 2006. Meat Sci 73: 651–659.

**Agradecimientos:** Trabajo financiado por el Proyecto INIA RZ2012-00006-00-00.

## SEARCHING PRIVATIVE BREED ALLELES BETWEEN DUROC AND IBERIAN PIGS

**ABSTRACT**: The final goal of this study is the design of a panel of genetic markers with breed specific alleles suitable to authenticate purebred Iberian and Durocx Iberian crossbred pigs and their products with flexible and cheap lab methods. We report the results of a first research based on genome data from 169 representative pigs (32 Duroc and 137 Iberian) typed for over 43,000 autosomal SNPs, and DNA sequences of five Duroc and 12 Iberian. Using $F_{ST}$ and $F'_{ST}$ outlier tests we identified 31 genome regions with divergent allelic frequencies which include 114 known genes. The analysis of these regions in the available sequences allowed the identification of a 255 SNVs with potential breed specific alleles, 91 out of those with intragenic location. The functions of some of the differentiated genes indicate their possible role related to some of the main phenotypic differences between both breeds.

**Keywords:** Iberian pig, Duroc, breed specific alleles, authentication