

# IDENTIFICACIÓN DE REGIONES GENÓMICAS POR DETECCIÓN DE HUELLAS DE SELECCIÓN PARA GRASA INTRAMUSCULAR EN LÍNEAS DE CONEJO.

Sosa-Madrid<sup>1</sup>, B.S, Varona<sup>2</sup>, L., Hernandez<sup>1</sup>, P., Ibañez-Escriche<sup>1</sup>, N., Casto-Rebollo<sup>1</sup>, C. y Blasco, A.<sup>1</sup>

<sup>1</sup>Instituto de Ciencia y Tecnología Animal, Universitat Politècnica de València, 46022 Valencia, España.

<sup>2</sup>Instituto Agroalimentario de Aragón (IA2). Universidad de Zaragoza, 50013 Zaragoza, España.

bosomad@posgrado.upv.es

## INTRODUCCIÓN

La Universitat Politècnica de València llevó a cabo un experimento de selección divergente para grasa intramuscular (GIM). La respuesta a la selección, en la novena generación, fue de 3.1 desviaciones típicas (Martínez-Álvaro et al., 2016). Esta alta respuesta indicaría que la selección pudo generar modificaciones relevantes en regiones genómicas con alta importancia a nivel funcional del carácter. En este sentido, la información genómica procedente de poblaciones de selección divergente constituye un material de excepcional valor para el estudio de huellas de selección, ya que permite atribuir las señales detectadas a la variabilidad genética del carácter que fue objeto de selección. Hay numerosos métodos para detectar huellas de selección, que se clasifican en función de la definición de la hipótesis nula (ausencia de selección) (Oleksyk et al., 2010). Estos métodos pueden estar basados en el análisis de diversidad genética, la reducción de la variabilidad local, la modificación del espectro de frecuencias y la extensión del desequilibrio de ligamiento (Qanbari y Simianer, 2014). El propósito del actual trabajo es identificar regiones genómicas por tres métodos de detección de huella de selección basados en hipótesis distintas ( $F_{st}$  – diversidad genética, XP-CLR – modificación del espectro de frecuencias y XP-EHH – extensión del desequilibrio de ligamiento) con el objetivo final de generar una lista de genes candidatos asociados a la selección divergente por GIM.

## MATERIAL Y MÉTODOS

**Material Biológico:** Las muestras biológicas procedieron de conejos de la novena generación de selección divergente: 240 de la línea seleccionada para alta GIM y 237 de la línea para baja GIM. Además, se utilizó la información de los progenitores de la novena generación para calcular las fases haplotípicas (10 padres y 38 madres de cada línea). Estas muestras fueron genotipadas con el “Affymetrix Axiom OrcunSNP Array” (~200k SNPs). El control de calidad fue realizado utilizando el programa “*Axiom Analysis Suite v. 4.0*”. Los SNPs con un “call rate”  $\geq 95\%$ , menor frecuencia alélica  $\geq 5\%$ , y una posición conocida en el mapa genómico fueron utilizados en el estudio. Para la imputación y las fases haplotípicas se utilizó el programa *Flmpute* (Sargolzaei et al., 2014). Después del control de calidad, un total de 89,864 SNPs fueron utilizados en el análisis.

**Métodos:** El análisis consistió en el cálculo de tres estadísticos:

- Índice de fijación ( $F_{st}$ ): se calculó el parámetro  $F_{st}$  de Wright (1951). Los valores de  $F_{st}$  fueron promediados en ventanas deslizantes de 0,5 Mb en intervalos de un SNP. Además, los valores fueron normalizados basados en el número de SNPs dentro de cada ventana.
- Coeficiente de verosimilitud compuesto entre poblaciones (XP-CLR) (Chen et al., 2010): Los valores XP-CLR fueron calculados utilizando la línea baja como población de referencia y la línea alta como objetivo. El software *XP-CLR* ([http://genetics.med.harvard.edu/reich/Reich\\_Lab/Software.htm](http://genetics.med.harvard.edu/reich/Reich_Lab/Software.htm)) fue implementado para la obtención de los valores y se seleccionó el valor obtenido en la posición más cercana a cada SNP.
- Extensión de homocigosidad haplotípica entre poblaciones (XP-EHH) (Sabeti et al., 2007): La línea alta fue utilizada para calcular el valor del numerador y la línea baja para el denominador. El cálculo de los valores se realizó mediante el software *selscan* (Szpiech y Hernandez, 2014) haciendo una normalización para los valores de XP-EHH.

**Análisis de anotaciones funcionales:** Los SNPs que presentaron valores por encima del percentil 0,01 en al menos dos métodos fueron considerados para este análisis. La búsqueda de los genes se realizó en regiones de +/- 250kb desde un SNP utilizando el programa

BIOMART (Ensembl Genes 95) (Aken et al., 2016). El análisis de las anotaciones funcionales fue realizado con el programa DAVID 6.8v. (Jiao et al., 2012)

## RESULTADOS Y DISCUSIÓN

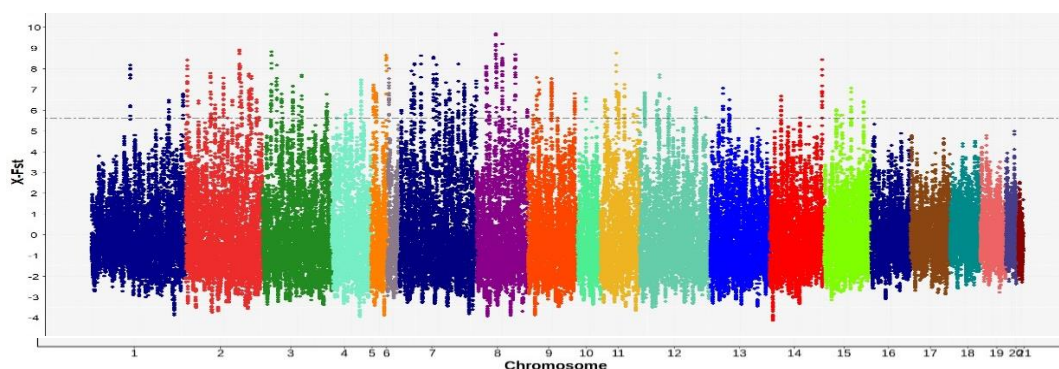
Las medias de los valores originales de  $F_{st}$  por ventanas, XP-CLR (no-normalizados) y XP-EHH (en valor absoluto) fueron 0,10, 1,55 y 0,68, respectivamente. El valor medio de  $F_{st}$  es superior a los encontrados en estudios de poblaciones de conejo europeo,  $F_{st}=0,08$  (Carneiro et al., 2014), y de poblaciones de selección divergente para capacidad uterina,  $F_{st} = 0,05$  (Sosa-Madrid et al., 2017). El experimento de selección divergente pudo haber generado una gran diferenciación genética entre las dos poblaciones a causa de la elevada heredabilidad de GIM (Martínez-Álvaro et al., 2016), o bien puede ser consecuencia del efecto de deriva genética. El efecto de deriva podría ser debido al reducido número de machos utilizados en las primeras generaciones (8 por línea), la estructura de apareamientos (parejas únicas no aleatorizadas) (Zomeño et al., 2013) y la duplicación de familias de macho en las últimas generaciones.

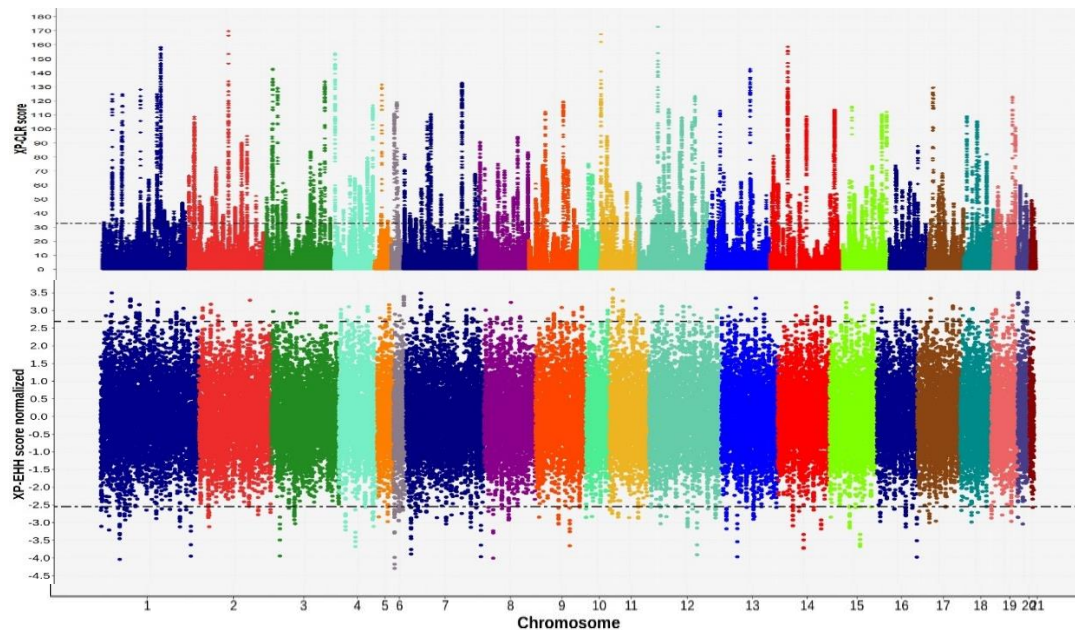
Los resultados presentados en forma de “manhattan plots” (Figura 1) sugieren que la selección por GIM no parece ser dominada por regiones genómicas específicas sino por múltiples variantes a lo largo del genoma, y además, la mayor parte de SNPs que superan el percentil 0,01 (899) no son coincidentes entre métodos. De hecho, un único SNP (OCU 13, 27,33 Mb) aparece entre los 899 SNPs seleccionados por los tres procedimientos, y solamente 79 están presentes en al menos dos de los métodos (ver Figura 2). Las regiones genómicas donde se sitúan estos 79 SNPs están localizadas en los OCU 1 (2 regiones), 2 (9), 3 (3), 5 (1), 7 (4), 8 (4), 9 (1), 11 (4), 12 (3), 13 (3), 16 (1) y 19 (2) y albergan 156 genes entre los que se encuentran algunos vinculados al metabolismo lipídico como *ADGRP5 -adhesion G protein-coupled receptor F5-* (OCU 12), *PRKAG2 -protein kinase AMP-activated non-catalytic subunit gamma 2-* (OCU 13), *P2RX1 – purinergic receptor P2X 1-*, *G6PC -glucose-6-phosphatase catalytic subunit -* y *BRCA1 -BRCA1, DNA repair associated-* (OCU 19). Por otro lado, cuando se realizó un análisis de anotación funcional con los 156 genes, se observó que las funciones con mayor p-valor fueron: unión de trifosfato de ribonucleósido de purina (GO:0035639), organización de orgánulos simple-organismo (GO:0006996) y ruta de señalización de AMPK (hsa04152).

## REFERENCIAS BIBLIOGRÁFICAS

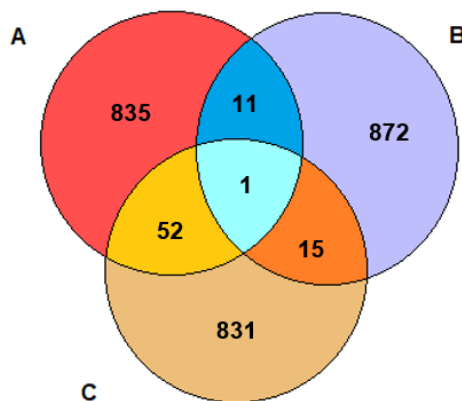
• Aken, B.L. 2016. Database 2016. • Carneiro, M. 2014. PLoS Genet. 10. • Chen, H. 2010. Genome Res. 20: 393–402. • Jiao, X. 2012. Bioinformatics 28: 1805–1806 • Martínez-Álvaro, M. 2016. J. Anim. Sci. 94: 4993–5003. • Oleksyk, T. K. 2010. Philos. Trans. R. Soc. B Biol. Sci. 365: 185–205. • Qanbari, S. 2014. Livest. Sci. 166: 133–143 • Sabeti, P.C. 2007. Nature 449: 913–918. • Sargolzaei, M. 2014. BMC Genomics 15: 478. • Sosa-Madrid, B.S. 2017. XVII Jorn. Prod. Anim: 558-560. • Szpiech, Z.A. 2014. Mol. Biol. Evol. 31: 2824–2827. • Wright, S. 1951. Ann. Eugen. 15:323-354. • Zomeño, C. 2013. J. Anim. Sci. 91: 4526-4531.

**Agradecimientos:** B. Samuel Sosa Madrid agradece su beca FPI al Ministerio de Economía y Competitividad de España (BES-2015-074194) del proyecto AGL2014-55921-C2-1-P.





**Figura 1.** Manhattan Plots para  $F_{st}$  (imagen superior), XP-CLR (medio) y XP-EHH (inferior). Las líneas punteadas indican el umbral del percentil 0,01



**Figura 2.** Diagrama de Venn para los SNPs superiores al percentil 0.01 utilizando los métodos: (A) XP-CLR, (B) XP-EHH y (C)  $F_{st}$ .

### IDENTIFICATION OF GENOMIC REGIONS BY DETECTING SIGNATURES OF SELECTION FOR INTRAMUSCULAR FAT IN RABBIT LINES

**ABSTRACT:** A divergent selection experiment for intramuscular fat (IMF) in rabbits with a response to selection of 3.1 standard deviations was performed. The aim of this study was to identify genomic regions related to IMF through three methods of detection of signatures of selection and to generate a list of candidate genes. The genomic data came from 477 rabbits of the high and low IMF lines at 9<sup>th</sup> generation. The methods implemented in this study were: Wright's fixation index ( $F_{st}$ ), cross population – composite likelihood ratio (XP-CLR) and cross population – extended haplotype homozygosity (XP-EHH). Only one SNP in rabbit chromosome (OCU) 13 showed values above the percentile 0.01 in all methods. Several genomic regions, harbouring 79 SNPs, were found in at least two methods (OCU 1, 2, 3, 5, 7, 8, 9, 11, 12, 13, 16 and 19). This suggests that either, IMF selection influenced multiple genomic regions, or the effect of genetic drift played an important role in these lines. The genes linked to fatty acid and lipid metabolic process, after pathways analysis, were *ADGRP5* (OCU 12), *PRKAG2* (OCU 13), *P2RX1*, *G6PC* and *BRCA1* (OCU 19).

**Keywords:** intramuscular fat, signature of selection, rabbit.