

ANÁLISIS BAYESIANO PARA LA DETECCIÓN DE HUELLAS DE SELECCIÓN EN POBLACIONES DE VACUNO DE CARNE ESPAÑOL

González-Rodríguez, A.¹, Munilla, S., Mouresan, E.F., Cañas-Álvarez, J. J., Baro, J. A., Molina, A., Díaz, C., Piedrafita, J., Altarriba, J. y Varona, L.

¹Grupo de Genética Cuantitativa y Mejora Animal. Facultad de Veterinaria, Universidad de Zaragoza. Calle Miguel Servet 177, 50013 Zaragoza, España. aldemar@unizar.es

INTRODUCCIÓN

Los procesos de selección y adaptación pueden dejar huellas en la estructura del genoma de las poblaciones. En la literatura, se han descrito un gran número de procedimientos para la detección de esta huella (Qanbari y Simianer, 2014), tanto basados en el espectro de frecuencias como en la longitud de los haplotipos. Otros procedimientos, como el F_{ST} , tienen su fundamento en la comparación entre las frecuencias alélicas de cada población. En este mismo sentido, Vitalis et al. (2014) ha propuesto un nuevo método (SelEstim), basado en un enfoque bayesiano, que permite distinguir entre los polimorfismos seleccionados y casi neutros, bajo un modelo genético que asume una población dividida en varias subpoblaciones entre las que puede existir migración.

El objetivo de este estudio es utilizar este procedimiento para localizar regiones genómicas asociadas a la selección en siete poblaciones de vacuno de carne Español (Asturiana de los Valles –AV–, Avileña-Negra Ibérica –ANI–, Bruna dels Pirineus –BP–, Morucha –Mo–, Pirenaica –Pi– and Retinta –Re–, Rubia Gallega –RG–).

MATERIAL Y MÉTODOS

Se dispuso de un total de 171 tríos (padre/madre/descendiente) de siete poblaciones de vacuno de carne español (AV, n=25; ANI, n=24; BP, n=25; Mo, n=25; Pi, n=24; Re, n=24; RG, n=24). Los padres fueron escogidos con el objetivo de capturar la mayor variabilidad existente entre las razas. Tras un control de calidad estándar se utilizaron 703.707 SNPs. Las fases de los cromosomas de los padres se establecieron por medio del software BEAGLE (Browning y Browning, 2009).

SelEstim: El procedimiento de Vitalis et al. (2014) asume un modelo Bayesiano jerárquico que permite distinguir el polimorfismo seleccionado desde un fondo neutral (o casi neutral) y estima la intensidad de selección para cada alelo y población. El modelo asume que los alelos cuentan en cada loci con dos estados (A y a), cuya distribución estadística es:

$$L(p_{ij}; n_{ij}) = \binom{n_{ij}}{x_{ij}} p_{ij}^{x_{ij}} (1-p_{ij})^{n_{ij}-x_{ij}}$$

Donde p_{ij} es la frecuencia alélica del estado A , x_{ij} es el número de alelos en estado A y n_{ij} es el número total de alelos para el $j^{\text{ésimo}}$ loci en la $i^{\text{ésima}}$ población. Además, asume una distribución a priori de p_{ij} a partir de una densidad estacionaria en un proceso de difusión (Vitalis et al., 2014):

$$\Psi(p_j; M_j, \sigma_{ij}, k_j, \pi_j) = C^{-1} \exp(\sigma_{ij} p_j) p_j^{M_j \pi_j - 1} (1 - p_j)^{M_j (1 - \pi_j) - 1}$$

Donde M_j es la escala de parámetros de migración ($M_j = 4N_j m_j$), π_j es la frecuencia de los alelos en toda la población, $\sigma_{ij} = 2N_j s_{ij}$ es la escala del coeficiente de selección en la población i en el locus j y k_j es un indicador variable que toma valores de 0 si el alelo A es seleccionado y 1 si el alelo a es seleccionado para el $j^{\text{ésimo}}$ locus en la $i^{\text{ésima}}$ población. La distribución a priori para k_j es una Bernoulli (Ber (0,5)), para π_j es una Beta (1,1) y para M es una log-uniforme (0,001 a 10.000). También se asume la siguiente distribución a priori: $f(\sigma_{ij} | \delta_j) \sim \exp(-\delta_j^{-1})$, siendo δ_j un hiperparámetro que representa el efecto medio de la selección en el locus j sobre todas las poblaciones. En una jerarquía anterior, se postula una distribución exponencial a priori para

$f(\delta_j|\Lambda) \sim \exp(\Lambda^{-1})$, y $f(\Lambda) \sim \exp(\Lambda^{-1})$ con $\Lambda=1,0$. Este modelo se implementó mediante un método de MCMC con el software de acceso público SELESTIM (<http://www1.montpellier.inra.fr/CBGP/software/selestim/>).

Identificación de regiones genómicas: Se identificaron como señales significativas las regiones que presentaron más de 10 SNPs con una señal superior al 0,01% para el parámetro asociado a selección (δ) y dentro de una ventana de 1 Mb. Posteriormente, se utilizó la base de datos ENSEMBL (Flicek, et al., 2013) para localizar genes candidatos.

RESULTADOS Y DISCUSIÓN

El parámetro de migración (M) tiene una media posterior (Desviación típica) de 61,42 (0,21) para la raza AV, mientras que para el resto de las poblaciones tuvo un rango entre 16,46 (Pi) y 28,17 (Mo). Este resultado nos indica una mayor relación de la población AV con el resto. Este resultado concuerda con el obtenido por Cañas-Álvarez et al. (2014) mediante análisis de clúster con ADMIXTURE.

En la Figura 1 se presentan los resultados del parámetro δ_j , a lo largo de los autosomas bovinos y en la Tabla 1 se presenta la lista de regiones genómicas que presentan señales relevantes para al menos 10 SNPs y un δ_j mayor a 14,99 dentro de una región de 1Mb, además, en esta tabla se presentan los valores de σ_{ij} , y k_{ij} para el marcador de máxima señal en cada una de las poblaciones.

Los resultados de este procedimiento son similares a los obtenidos por González-Rodríguez et al. (2014), mediante el estadístico F_{ST} . Las regiones que involucran un mayor número de marcadores están localizadas en BTA2 (5,2-10,6 Mb), BTA5 (17,3-19,8 Mb), BTA6 (37,9-41,4 Mb), BTA11 (67,0-67,5 Mb) y BTA18 (13,4-16,5 Mb), y además, también presentan una señal de mayor intensidad de selección (32,066 - BTA2 -, 29,985 - BTA5 -, 29,968 - BTA6 -, 19,443 - BTA11 -, 25,403 - BTA18 -).

Algunas de estas regiones tienen asociado un gen candidato evidente como en BTA2 (*MTSN-Myostatin*), BTA5 (*KITLG-KITLigand*) y BTA18 (*MC1R-Melanocortin1 receptor*). Por el contrario, las regiones localizadas en los cromosomas BTA6 y BTA11 pueden estar asociadas a varios genes de interés. La región genómica del BTA6 ha sido asociada a multitud de caracteres de interés (Saatchi et al., 2014) y en ella se encuentran los genes *LAP3* (*Leucine aminopeptidase3*), *NCAPG* (*Non-SMC condensing I complex, subunit G*), *LOCRL* (*Ligand dependent nuclear receptor corepressor-like*) *ABCG2* (*ATP-binding cassette sub-family G member 2*) y *PPARG1A* (*Peroxisome proliferator-activated receptor gamma, coactivator 1 alpha*). Por otra parte, en la región del BTA11, se localizan los genes *PROKR1* (*Prokineticin receptor 1*), *GFPT1* (*Glutamine-fructose-6-phosphate transaminase 1*), *GMCL1* (*Germ cell-less spermatogenesis associated 1*), *PCBP1* (*Poly (rC) binding protein 1*) y *EHD3* (*EH-domain containing 3*), asociados a fertilidad.

REFERENCIAS BIBLIOGRÁFICAS

- Qanbari, S. & Simianer, H. 2014. *Livestock Science*, 116:133–143.
- Vitalis, R. et al. 2014. *Genetics*, 196:799–817.
- Browning, B. L., & Browning, S. R. 2009. *Am. J. Hum. Genet.* 84:210-223.
- Cañas-Álvarez, J. et al. 2014. 10th WCGALP.
- González-Rodríguez, A. et al. 2014. 10th WCGALP.
- Flicek, P., Ahmed, I., Amode, M. R. et al. 2013. *Nucleic Acid. Res.* 41(D1): D48-D55.
- Saatchi, M., et al. 2014. *BMC Genomics* 15:442.

Agradecimientos: Este estudio ha sido posible gracias al apoyo financiero de AGL2010-15903. González-Rodríguez A y Cañas J. J. disfrutaron de las becas BES-2011-045434 y COLCIENCIAS Francisco José de Caldas 497/2009, respectivamente.

Tabla 1. Valores de $\sigma_{ij}(k_{ij})$ para marcadores de máxima señal en regiones genómicas con un δ mayor a 14,99 y un número superior a 10 SNPs.

Cr	Posición (Mb)	Nº Marcadores	Delta Max. (δ)	AV	ANI	BP	Mo	Pi	Re	RG
2	5,2-10,6	222	32,066	136,36 (0,026)	92,17 (0,995)	52,15 (0,970)	92,32 (0,994)	49,07 (0,985)	69,81 (0,991)	25,49 (0,227)
5	17,2-19,8	63	29,985	59,84 (0,018)	46,17 (0,890)	216,22 (0,000)	35,03 (0,809)	38,97 (0,001)	45,48 (0,881)	24,60 (0,050)
5	47,0-47,1	18	18,683	57,38 (0,998)	15,08 (0,338)	68,34 (1,000)	26,53 (0,188)	9,35 (0,632)	24,08 (0,182)	18,03 (0,894)
5	70,1-70,2	11	16,010	14,37 (0,326)	8,57 (0,624)	42,49 (0,004)	68,67 (0,995)	6,79 (0,552)	13,39 (0,837)	19,14 (0,077)
6	31,9-31,9	12	18,534	15,12 (0,420)	51,48 (1,000)	32,25 (0,106)	76,82 (1,000)	9,21 (0,425)	16,61 (0,934)	15,23 (0,264)
6	36,5-41,3	100	29,968	29,01 (0,721)	77,17 (0,009)	148,26 (0,999)	61,81 (0,029)	21,85 (0,925)	76,64 (0,006)	47,81 (0,985)
7	47,2-47,7	17	18,874	14,27 (0,462)	16,29 (0,899)	58,86 (0,999)	41,16 (0,995)	22,07 (0,078)	60,65 (0,007)	9,44 (0,445)
11	66,9-70,8	51	19,443	25,56 (0,328)	14,99 (0,077)	82,31 (1,000)	43,73 (0,105)	43,12 (0,886)	10,46 (0,336)	14,09 (0,468)
13	57,5-57,7	11	18,442	16,61 (0,307)	9,08 (0,589)	43,10 (0,007)	19,70 (0,878)	10,65 (0,190)	83,75 (0,999)	33,08 (0,018)
18	13,3-16,3	111	25,403	18,48 (0,565)	81,47 (1,000)	43,15 (1,000)	96,51 (1,000)	25,02 (0,104)	53,94 (0,024)	35,10 (0,074)

Cr (Cromosoma), AV (Asturiana de los Valles), ANI (Avileña-Negra Ibérica), BP (Bruna dels Pirineus), Mo (Morucha), Pi (Pirenaica), Re (Retinta), RG (Rubia Gallega).

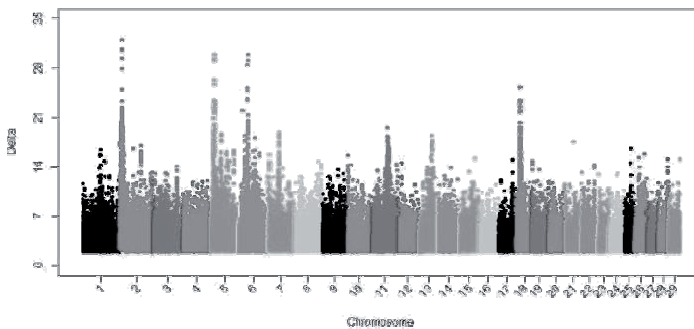


Figura 1. Resultados del efecto medio de la selección (delta)

BAYESIAN ANALYSIS TO DETECTING SELECTION FOOTPRINTS IN THE SPANISH POPULATIONS OF BEEF CATTLE

ABSTRACT: The BovineHD 770K BeadChip was used on 171 triplets (sire/dam/offspring) from seven Spanish local beef cattle breeds (Asturiana de los Valles, Avileña-Negra Ibérica, Morucha, Bruna dels Pirineus, Pirenaica, Retinta and Rubia Gallega). The parents were chosen to be as unrelated as possible. We use the software SELESTIM to identify loci targeted by a selection process and also to estimate the strength and the allele of selection within each subgroup. This approach showed five genomic regions with maximum signal, some of these have a clear candidate gene as BTA2 (*MTSN-Myostatin*), BTA5 (*KITLG-KITLigand*) and BTA18 (*MC1R-Melanocortin1 receptor*); while the located in BTA6 and BTA11 can be associated with several genes.

Keywords: Beef cattle, Selection footprint, Genomic regions, Bayesian analysis.