

## Análisis de agrupamiento de cerdos Ibéricos, Duroc y de sus cruces

S.T. Rodríguez-Ramilo<sup>\*,\*\*</sup>, J. Fernández<sup>\*\*</sup>, M.A. Toro<sup>\*\*</sup>

\* Departamento de Bioquímica, Genética e Inmunología. Facultad de Biología, Universidad de Vigo, 36310 Vigo, Spain

\*\* Departamento de Mejora Genética Animal. INIA. Crta. A Coruña Km. 7,5. 28040 Madrid, Spain  
E-mail: silviat@uvigo.es

### Resumen

En este estudio se utilizaron datos reales de individuos Ibéricos y Duroc, así como individuos simulados para generar una F1 y los retrocruces respectivos de Ibérico y de Duroc. Se evaluó la capacidad de tres métodos de agrupamiento para distinguir entre estas cinco poblaciones. Los resultados indican que los métodos de agrupamiento evaluados no detectan la partición óptima. En parte el problema se debe a la distinta configuración genética de diferentes variedades (estirpes) de la raza Ibérica que los métodos son capaces de detectar e interfieren en la separación entre individuos puros y cruzados.

**Palabras clave:** Cerdo Ibérico, Método Bayesiano, Simulated annealing

### Summary

#### Cluster analysis of Iberian, Duroc and crossed pigs

In this study real data from Iberian and Duroc individuals were used, as well as simulated F1, Iberian and Duroc backcrossed individuals. The ability of three clustering methods to separate the five populations was evaluated. Results indicate that the clustering methods do not detect the optimal partition. The problem is, in part, due to the different genetic configurations of the different strains of Iberian pigs that the methods can detect and, therefore, interfering in the separation between pure and crossed individuals.

**Key words:** Iberian pig, Bayesian method, simulated annealing

### Introducción

El incremento del censo así como los cambios significativos en la normativa sobre la selección y reproducción del ganado porcino Ibérico son algunas de las razones que han motivado la publicación del nuevo Libro Genealógico de la Raza Porcina Ibérica. Asimismo, la publicación de la Norma de Calidad para los productos Ibéricos obliga a

los ganaderos a que sus ejemplares estén inscritos en el Libro Genealógico o que estén certificados por una entidad de inspección que califica a los individuos. La calificación de los animales se hace en función de caracteres morfológicos, por lo que puede estar influenciada subjetivamente. Igualmente, la legislación española actual permite hasta un 50% de genoma de origen de raza Duroc en animales empleados para

originar productos Ibéricos. Sin embargo, la denominación de Ibérico puro se atribuye, únicamente, a individuos pertenecientes a la raza Ibérica (Patiño, 2007).

Los métodos de agrupamiento permiten separar un grupo de individuos en varias subpoblaciones si su origen genético es desconocido y/o estudiar la correspondencia entre los grupos inferidos y las poblaciones predefinidas. En los últimos años se han propuesto varios métodos Bayesianos para estimar la estructura genética poblacional subyacente a partir de la información que proporcionan los marcadores moleculares (Pritchard et al., 2000; Corander et al., 2004). Estos métodos Bayesianos actúan buscando el número de grupos inferido ( $K$ ) y la asignación de individuos a esos grupos que minimizan el desequilibrio de Hardy-Weinberg y de ligamiento dentro de grupos.

Existen otros métodos que no asumen ni equilibrio de Hardy-Weinberg (HWE) ni de ligamiento (LE) a la hora de detectar estructura genética poblacional subyacente, implementando algoritmos de *simulated annealing*. Uno de estos métodos es el propuesto por Rodríguez-Ramilo et al. (2008) que intenta maximizar la distancia genética promedio entre subpoblaciones.

En el presente artículo se evalúa la capacidad de tres métodos de agrupamiento para separar poblaciones puras y cruzadas. Para ello se emplean datos reales de individuos Ibéricos y Duroc y datos simulados de individuos pertenecientes a una F1 y retrocruces de Ibérico y de Duroc.

## Material y métodos

### Datos reales y simulados

Se utilizaron los genotipos para 36 microsatélites de 170 individuos Ibéricos con la siguiente

distribución: 31 Torbiscal, 32 Guadyerbas, 50 Retintos, 30 Lampiños y 27 Entrepelados así como de 64 individuos Duroc representando siete núcleos de selección españoles (Alves et al., 2006).

A partir de los datos genotípicos reales se obtuvieron 100 individuos simulados para generar la primera generación filial (F1). Estos individuos F1 se cruzaban con los 170 Ibéricos y los 64 Duroc respectivamente, para generar otros 100 individuos respectivamente de cada retrocruce.

### Métodos Bayesianos

Los análisis se realizaron con el STRUCTURE versión 2.1 (Pritchard et al., 2000) y con el BAPS versión 4.14 (Corander et al., 2004).

En el STRUCTURE se realizaron 10.000 burn-in y 50.000 MCMC (Markov chain Monte Carlo). Se eligió el modelo "admixture" y la opción de frecuencias alélicas correlacionadas. Para determinar  $K$  se implementó el criterio  $\Delta K$  propuesto por Evanno et al. (2005). En el BAPS el número máximo de grupos permitidos fue 20 y se empleó la opción de agrupamiento de individuos.

### Método de la matriz de parentesco molecular (MPM)

La idea que subyace a este método es que se espera que poblaciones distintas presenten una elevada distancia genética entre ellas. Esta distancia se puede calcular a partir de la información que nos proporcionan los marcadores moleculares sin asumir HWE ni LE. El método implementa un algoritmo de *simulated annealing* para encontrar la partición que maximiza la distancia mínima de Nei promedio entre subpoblaciones. Para determinar  $K$  se implementó el criterio  $\Delta K$  propuesto por Evanno et al. (2005) modificado para la distancia genética.

## Evaluación de los métodos

Para determinar el comportamiento de cada método se evaluó  $K$  y la proporción de individuos de cada población asignados a cada cluster. Para entender mejor el funcionamiento de los métodos también se evaluó esta proporción de individuos cuando se fuerza  $K = 2$ .

## Resultados y discusión

En cuanto al número de grupos inferido (tabla 1.a), ninguno de los tres métodos infiere las cinco subpoblaciones originales. El BAPS sobreestima el número de grupos real, y tanto el STRUCTURE como MPM lo subestiman. La mayor proporción de individuos Ibéricos se encuentra en el mismo grupo que su retrocruce en los tres métodos. Lo mismo sucede en los Duroc y su retrocruce. Tanto el STRUCTURE como MPM localizan la mayor proporción de F1 en el grupo de Duroc y su retrocruce e infieren un grupo adicional más para una proporción residual de asignaciones de distintas subpoblaciones (fundamentalmente individuos Guadyerbas y sus descendientes). El BAPS localiza la mayor proporción de individuos F1 en un tercer grupo e indica siete grupos adicionales para distintas proporciones residuales de las distintas subpoblaciones. En general, los tres métodos de agrupamiento detectan y separan distintas variedades de la raza Ibérica, por lo que la partición general queda distorsionada.

Si se implementa un procedimiento de aleatorización de alelos dentro de cada subpoblación se reestablece HWE y LE. En esta situación, los tres métodos infieren tres grupos (tabla 1.b). Tanto el BAPS como MPM dan una solución próxima a la ideal (separación de Ibéricos, cruzados con hasta el 50% de información Duroc y el resto). Si en el STRUCTURE se elige el modelo "No admixture" la solución se aproxima también a esta partición ideal (datos no mostrados).

Cuando se fuerzan sólo dos grupos (tabla 2) tanto el BAPS como MPM diferencian la totalidad de los individuos Ibéricos y Duroc en grupos separados. En la F1 se esperaría que los individuos estuvieran constituidos por un 50% de genoma de Ibérico y un 50% de genoma de Duroc, por lo que el STRUCTURE es el que presenta un comportamiento más adecuado (García *et al.*, 2006). Con respecto al retrocruce de Ibérico, se esperaría un 75% del genoma de Ibérico (y el 75% de genoma de Duroc en el retrocruce de Duroc). En este caso, también es el STRUCTURE el que presenta una mayor precisión. Sin embargo, el modelo "No admixture" muestra un comportamiento similar al BAPS y MPM (datos no mostrados). En general, la aleatorización no modifica sustancialmente los resultados.

En resumen, los métodos de agrupamiento evaluados no detectan la estructura poblacional subyacente, ni realizan una partición óptima desde el punto de vista práctico debido en parte a la diferenciación entre variedades dentro de la población de individuos Ibéricos.



Tabla 2. Proporción de cada población asignada a cada grupo inferido al considerar dos grupos antes (a) y después (b) de la aleatorización de los alelos

*Table 2. Proportion of each population assigned to each inferred cluster when two clusters are considered before (a) and after (b) the allele randomisation*

Método	Población	a. Grupos inferidos		b. Grupos inferidos	
		1	2	1	2
BAPS	Ibérico	1,00		1,00	
	Duroc		1,00		1,00
	F1	0,16	0,84	0,03	0,97
	Retrocruce Ibérico	0,95	0,05	0,92	0,08
	Retrocruce Duroc		1,00		1,00
STRUCTURE	Ibérico	0,95	0,05	0,96	0,04
	Duroc	0,03	0,97	0,03	0,97
	F1	0,48	0,52	0,49	0,51
	Retrocruce Ibérico	0,72	0,28	0,73	0,27
	Retrocruce Duroc	0,22	0,78	0,22	0,78
MPM	Ibérico	1,00		1,00	
	Duroc		1,00		1,00
	F1	0,36	0,64	0,19	0,81
	Retrocruce Ibérico	0,96	0,04	0,94	0,06
	Retrocruce Duroc	0,01	0,99		1,00

## Agradecimientos

Este trabajo ha sido financiado mediante proyectos del Plan Estratégico del INIA (CPE03-004-C2) y del Plan Nacional del MEC (CGL2006-13445-C02-01). Agradecemos al grupo de porcino del Departamento de Mejora Genética del INIA por permitirnos utilizar los datos moleculares.

## Bibliografía

Alves E, Fernández AI, Barragán C, Óvilo C, Rodríguez C, Silió L, 2006. Inference of hidden

population substructure of the Iberian pig breed using multilocus microsatellite data. *S.J.A.R.*, 4, 37-46.

Corander J, Waldmann P, Marttinen P, Sillanpaa MJ, 2004. BAPS 2: enhanced possibilities for the analysis of genetic population structure. *Bioinformatics*, 20, 2363-2369.

Evanno G, Regnaut S, Goudet J, 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol. Ecol.*, 14, 2611-2620.

García D, Martínez A, Dunner S, Vega-Pla JL, Fernández C, Delgado JV, Cañón J, 2006. Estimation of the genetic admixture composition of Iberian dry-cured ham samples using DNA multilocus genotypes. *Meat Sci.*, 72, 560-566.

Patiño E, 2007. El nuevo libro genealógico del Ibérico refuerza la selección de la raza. *Cárnica* 2000, 288, 79-80.

Pritchard JK, Stephens M, Donnelly P, 2000. Inference of population structure using multilocus genotype data. *Genetics*, 155, 945-959.

Rodríguez-Ramilo ST, Fernández J, Toro MA, 2008. Assessing population genetic structure via the molecular coancestry matrix. *Mol. Ecol.* (en preparación).

(Aceptado para publicación el 28 de abril de 2008)