

# Uso del parentesco molecular genómico para la depuración genética de una población

Carmen Amador<sup>1\*</sup>, Jesús Fernández<sup>1</sup>, Theo HE Meuwissen<sup>2</sup>

\* amador.carmen@inia.es

<sup>1</sup> Departamento de Mejora Genética Animal, INIA, 28040 Madrid, España

<sup>2</sup> Department of Animal and Aquacultural Sciences, UMB, 1432 Ås, Norway

## Resumen

Si el mantenimiento en pureza de una población se ve alterado por un aporte de material genético exógeno, dicha población ha de ser depurada. Para estudiar la eficacia del uso de genotipado masivo en la eliminación de información genética indeseada, se simuló dos poblaciones (nativa y exógena) y varios escenarios de introgresión en los que un número variable de individuos exógenos se mezclaban durante varias generaciones con la población nativa. Para eliminar la información exógena se utilizó como criterio de gestión el determinar las contribuciones a la siguiente generación de manera que se minimizase el parentesco molecular (calculado a partir de 50.000 SNP) entre los reproductores y los individuos exógenos introducidos en la población. Los resultados mostraron una notable recuperación del fondo genético nativo, mayor que cuando se minimiza el parentesco genealógico, pero asociado a un gran incremento en la consanguinidad debido a la reducción en el número de individuos que contribuyen a las siguientes generaciones.

## Introducción

El mantenimiento en pureza de muchas poblaciones naturales y domésticas puede ser deseable desde el punto de vista de la conservación y la mejora genética. En el primer caso podemos encontrar diversos ejemplos de especies salvajes que se encuentran amenazadas por sus equivalentes domésticos (Rhymer y Simberloff, 1996; Allendorf *et al.*, 2001). Por otra parte, algunas razas domésticas están estrechamente ligadas a beneficios económicos que dependen directamente de su fondo genético específico (Dalvit *et al.*, 2007). En ambos casos es recomendable evitar intercambios de material genético con otras poblaciones.

En un estudio anterior (Amador *et al.*, 2011) se analizó la eficiencia del uso de genealogías para eliminar información exógena previamente mezclada con una población nativa que queremos recuperar. Los resultados mostraron que la información de la genealogía permite recuperar gran parte de la pureza de la población, con un aumento en la consanguinidad consecuencia de la depuración, y una eficacia limitada en situaciones en las que la mezcla tuvo lugar durante muchas generaciones o en las que el porcentaje de introgresión fue elevado.

Cuando la genealogía no está disponible, otras fuentes de información pueden emplearse en su lugar. Varios estudios han analizado la eficiencia de sustituir el pedigrí por información molecular para varias tareas. Hayes *et al.* (2009) demostraron que reemplazando la matriz de relaciones aditivas calculada mediante genealogías por una matriz G calculada mediante genotipado de alta densidad en análisis BLUP, la precisión

de los valores mejorantes se ve incrementada. De Cara *et al.* (2011) comprobaron que con suficiente densidad de marcadores, la información molecular mantiene más diversidad genética que utilizando el pedigrí en el manejo de poblaciones en programas de conservación.

El objetivo del presente estudio es analizar (mediante simulación por ordenador) la eficacia de sustituir el parentesco genealógico por parentesco calculado mediante marcadores de genotipado masivo para eliminar la mayor información exógena posible de una población cuyo fondo genético queremos recuperar.

## Material y Métodos

Las simulaciones comprendieron tres partes:

### 1) Generación de las poblaciones nativa y exógena.

Se simularon dos poblaciones independientes de  $N = 100$  individuos. El genoma de cada individuo constaba de 20 cromosomas de 1M con dos tipos de loci bialélicos: 2.500 marcadores (empleados para gestión) y 25.000 no-marcadores (empleados para evaluación) en cada uno de los cromosomas. Todos los loci eran equidistantes y los marcadores estaban regularmente insertados entre los loci no-marcadores. Las frecuencias iniciales tanto en marcadores como en no-marcadores fueron 0.5/0.5 en la población base. Cada población (nativa y exógena) se mantuvo de forma independiente con contribuciones y apareamientos aleatorios durante 100 generaciones discretas y con censo constante para generar patrones de desequilibrio de ligamiento diferenciales en cada población.

### 2) Simulación del evento de introgresión.

Un número variable de exógenos (10-50) se introdujeron en la población nativa creando una población mezcla con un total de 100 individuos. Dicha población se mantuvo al azar durante 5 generaciones discretas para simular la introgresión de genes exógenos en el genoma nativo.

### 3) Gestión.

Para eliminar la información exógena, la población se gestionó durante 10 generaciones minimizando el parentesco molecular, calculado a través de los marcadores, entre los individuos de la población actual y los exógenos padres en la población mezcla. El parentesco molecular se corrigió para eliminar la similitud entre los individuos para los marcadores debido a la identidad en estado (IBS, dependiente de las frecuencias iniciales) como se propone en Shepherd *et al.* (2010). Una vez calculados los parentescos genómicos estandarizados, se seleccionaron como padres de la siguiente generación (en cada generación de manejo) los 20 individuos (10 machos y 10 hembras) con menor valor del parentesco promedio. Cada uno de los individuos seleccionados contribuyó con diez hijos, por lo que el valor teórico de  $\Delta F$  es el correspondiente a  $N_e = 40$  (es decir, 0.025). Se fijó el número de reproductores efectivamente contribuyendo para evitar que todos los descendientes se generasen de una sola pareja, como ocurriría

si no se implementaran restricciones al ser imposible encontrar varios individuos con exactamente el mismo grado de parentesco molecular.

La genealogía se anotó durante todas las generaciones de la simulación incluyendo las fases de creación de las poblaciones, el periodo de introgresión y el de gestión.

Para evaluar los resultados, se emplearon los datos de los 500.000 loci no marcadores para calcular la proporción de genoma nativo presente en la población al final de la fase de gestión, así como los valores de  $\Delta F$  genealógico.

## Resultados y Discusión

### 1) Proporción de genoma nativo.

La proporción de genoma nativo recuperado después de 1 o 10 generaciones de manejo se muestra en la Figura 1 para diferentes proporciones de individuos exógenos introducidos, junto con los resultados para el manejo mediante genealogías obtenidos en Amador *et. al* (2011).

En la gráfica se observa como el parentesco molecular nos permitió obtener mejores resultados que el uso del pedigrí, recuperando un mayor porcentaje de genoma nativo tras 10 generaciones de manejo. Cuando el porcentaje de introgresión aumentaba, el método requería más generaciones para alcanzar la máxima depuración, aunque en la mayor parte de los casos estos valores máximos se obtuvieron tras 3 ó 4 generaciones de manejo (resultados no mostrados). Los beneficios del uso del parentesco molecular con respecto al genealógico fueron más apreciables en los escenarios con mayor número de exógenos en la mezcla.

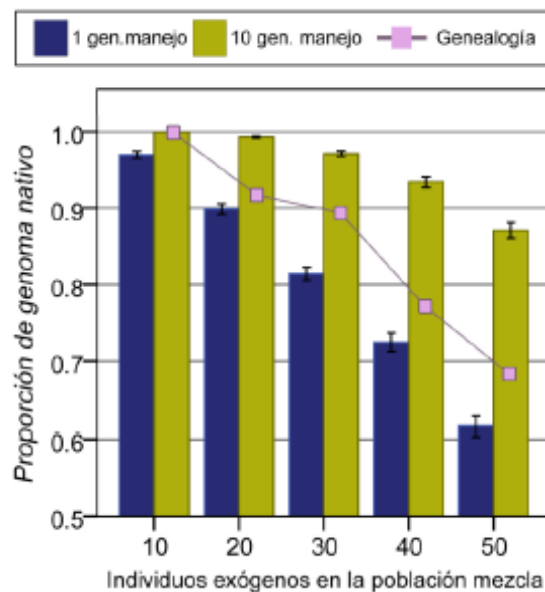


Figura 1. Proporción de genoma nativo recuperado tras 1 o 10 generaciones de manejo minimizando el parentesco molecular, y tras 10 generaciones minimizando el parentesco genealógico (Amador *et al.* 2011).

## 2) Consanguinidad.

Los niveles de  $\Delta F$  obtenidos en la desintrogresión mediante marcadores y en el caso de las genealogías se muestran en la Figura 2.

El incremento en consanguinidad general que se produce en los métodos de depuración es consecuencia directa de la gestión debido a la reducción en el número de individuos contribuyendo a las siguientes generaciones, tanto para el manejo con genealogías como el manejo con información molecular, respecto del tamaño poblacional.

Los valores de  $\Delta F$  obtenidos son superiores a los teóricos esperados durante las primeras generaciones cuando la eliminación de genoma exógeno es mayor (ver valores para una generación de manejo en la Figura 2). Esto se debe a que los 20 individuos seleccionados estarán más relacionados entre sí que si fueran elegidos al azar (como considera la estimación del  $\Delta F$  teórico). Dichos valores se reducen a partir de la generación 3-4 de manejo, cuando el método no puede eliminar más información exógena (se uniformizaban las relaciones de todos los individuos de la población) ajustándose al 0.025 esperado (resultados no mostrados).

En este estudio se estableció un control de la consanguinidad basado en imponer el número mínimo de reproductores. Otra estrategia posible sería la de poner una restricción en la tasa de consanguinidad máxima deseada de manera que el número de reproductores efectivo se fuese adaptando a la estructura de la población en cada generación.

De cualquier manera, el incremento en  $F$  que estos métodos conllevan y la rapidez a la hora de la depuración (esta ocurre en las primeras generaciones) sugieren que el número de generaciones que el método se utilice ha de reducirse lo máximo posible, para continuar el manejo con una estrategia que minimice la tasa de consanguinidad, como el método de contribuciones óptimas (Caballero y Toro, 2000; Sonesson y Meuwissen, 2001).

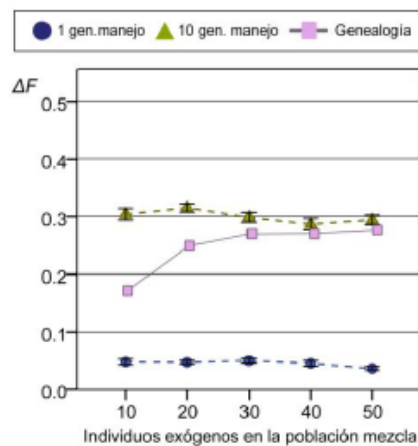


Figura 2. Valores de  $\Delta F$  obtenidos tras 1 o 10 generaciones de manejo minimizando el parentesco molecular y tras 10 generaciones de manejo minimizando el parentesco genealógico (Amador *et al.* 2011).

## Conclusiones

El parentesco molecular, calculado a partir de un número grande de marcadores tipo SNP, puede ser una herramienta eficiente para recuperar la pureza genética de una población introgresada. El fondo genético exógeno se elimina casi totalmente en casos con introgresión intermedia (10-30%) incluso tras 5 generaciones de mezcla. El incremento en la consanguinidad que se deriva del manejo es una consecuencia que hay que controlar/equilibrar cuando nos encontramos en esta situación. No obstante, usando el método el menor número de generaciones posibles, podemos alcanzar resultados positivos con tasas de consanguinidad aceptables.

## Agradecimientos

Este trabajo ha sido financiado por CGL2009-13278-C02-02 y fondos FEDER. Carmen Amador disfruta una beca FPI-INIA (INIA/Ministerio de Economía)

## Referencias

- Allendorf, F.W, Leary, R.F., Spruell, P. y Wenburg, J.K.: The problems with hybrids: setting conservation guidelines. *Trends In Ecology & Evolution* 16: 613-622. (2001)
- Amador, C., Toro, M.A. y Fernández, J.: Removing exogenous information using pedigree data. *Conservation genetics* 12: 1565–1573. (2011)
- Caballero, A. y Toro, M.A.: Interrelations between effective population size and other pedigree tools for the management of conserved populations. *Genetical Research* 75: 331-343. (2000)
- de Cara, M.A.R., Fernández, J., Toro, M.A. y Villanueva, B.: Using genome-wide information to minimise the loss of diversity in conservation programmes. *Journal of Animal Breeding and Genetics*. (2011)
- Dalvit, C., de Marchi, M. y Cassandro, M. 2007 Genetic traceability of livestock products: a review. *Meat Science* 77: 437-449. (2007)
- Hayes, B.J., Visscher, P.M. y Goddard, M.E.: Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research* 91: 47-60. (2009)
- Rhymer, J. M., y Simberloff, D.: Extinction by hybridization and introgression. *Annual Review Of Ecology And Systematics* 27: 83-109. (1996)
- Shepherd, R.K., Meuwissen, T.H.E. y Wooliams, J.A. : Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. *BMC Bioinformatics* 11. (2010)
- Sonesson, A.K., y Meuwissen, T.H.E.: Minimization of rate of inbreeding for small populations with overlapping generations. *Genetical Research* 77: 285-292. (2001)