

Análisis de estimadores basados en información molecular para inferir el parentesco genealógico

S.T. Rodríguez-Ramilo*, M.A. Toro*, A. Caballero**, J. Fernández*

* Departamento de Mejora Genética Animal. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). Ctra. A Coruña Km. 7,5. 28040 Madrid, Spain.

** Departamento de Bioquímica, Genética e Inmunología. Edificio de Ciencias Experimentales. Campus Universitario. Universidad de Vigo. 36310 Vigo, Spain.

Enviar correspondencia a: Silvia Teresa Rodríguez Ramilo. Departamento de Mejora Genética Animal. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA). Ctra. A Coruña Km. 7,5. 28040 Madrid (Spain). Teléfono: +34913473759. Fax: +34913572293. E-mail: ramilo@inia.es.

Resumen

Muchas poblaciones en cautividad y la mayoría de las poblaciones naturales no disponen de información genealógica. Debido a ello se han desarrollado numerosos estimadores del parentesco a partir de la información proporcionada por marcadores neutros moleculares. En este trabajo realizamos simulaciones para comprobar en qué situaciones dos estimadores del parentesco permiten obtener estimas precisas del parentesco genealógico. Para ello simulamos 10 o 100 marcadores moleculares neutros codominantes con seis posibles alelos, junto con 120 loci, aditivos y bialélicos, con efecto sobre un carácter cuantitativo. Se consideró un censo de 50 o 500 individuos en dos escenarios poblacionales: (1) la población no disponía de una estructura familiar, ya que los individuos se apareaban y contribuían a la siguiente generación de forma libre y aleatoria; (2) la población estaba estructurada en 5 o 50 familias con contribuciones iguales de cada familia a la siguiente generación (10 descendientes por familia), siendo la elección de los progenitores, o bien aleatoria, o bien dependiente de su valor fenotípico. Los resultados no aportan evidencias de un comportamiento satisfactorio de los estimadores de parentesco analizados, aunque se observa una mejora cuando se dispone de 100 marcadores. Los estimadores parecen ser más adecuados para inferir parentesco que consanguinidad, y su precisión depende de la varianza o dispersión de los valores del parentesco genealógico. Finalmente, se deduce que, a efectos de comparación y ordenación de los parentescos, el parentesco molecular resulta igualmente efectivo que cualquiera de los estimadores estudiados.

Palabras clave: parentesco, marcador molecular, consanguinidad

Summary

Analysis of estimators based on molecular information to infer genealogical coancestry

Many captive populations and most natural ones lack genealogical information. In this situation, several coancestry estimators have been proposed to infer coancestry from molecular markers. In this study we carried out simulations to test in which situations two analyzed coancestry *pairwise* estimators produce accurate estimates of the genealogical coancestry. We simulated 10 or 100 codominant neutral markers with six potential alleles, as well as 120 additive, biallelic loci controlling a quantitative trait. Simulations involved census sizes of 50 or 500 individuals in two different population scenarios: (1) populations with no family structure, where contributions of parents to progeny are free and mating is at random; (2) populations with a familiar structure, where 5 or 50 families are obtained through equal contributions of parents to the next generation (10 offspring per family), and parents of each family are chosen at random or depending on their phenotypic value. Our results provide no

evidence to suggest a satisfactory behaviour of the analyzed coancestry estimators, although it improves when 100 markers are available. The estimators are likely to be more suitable to infer coancestry than inbreeding, and their accuracy depends on the variance and dispersion of the genealogical coancestry values. Finally, we deduce that molecular coancestry is as effective as the analyzed estimators in order to compare and sort the coancestries.

Key words: coancestry, molecular marker, inbreeding

Introducción

El conocimiento de las relaciones de parentesco entre los individuos que integran una población es trascendental en numerosas áreas de la biología y de la genética de poblaciones. Por ejemplo, es necesario para estimar parámetros genéticos como heredabilidades y correlaciones genéticas (Ritland, 1996a y 2000) y valores mejorantes (Lynch y Walsh, 1998), sirve de criterio para el manejo de poblaciones con el objetivo de evitar la pérdida de diversidad y el incremento de la consanguinidad (Caballero y Toro, 2000 y 2002) y, finalmente, permite el estudio de sistemas de apareamiento, cuidado parental, dispersión y otros parámetros en ecología o biología del comportamiento (Avisé, 1994; Frankham *et al.*, 2002). Habitualmente, el parentesco se calcula a partir de la información genealógica. Sin embargo, la mayoría de las poblaciones naturales y muchas poblaciones en cautividad no disponen de dicha información.

Durante los últimos años, el desarrollo y aplicación de marcadores moleculares altamente polimórficos (especialmente los microsatélites) ha sido enormemente útil, particularmente en genética de poblaciones y conservación (Avisé, 1994; Frankham *et al.*, 2002), debido a su capacidad para inferir las relaciones entre individuos cuando parte o la totalidad de la genealogía poblacional no está disponible. Consecuentemente, se han desarrollado numerosos estimadores de parentesco para evaluar el grado de relación entre individuos empleando la infor-

mación molecular proporcionada por los distintos tipos de marcadores. Una descripción de la mayoría de estos estimadores puede encontrarse en Toro *et al.* (2002) y Fernández y Toro (2006), pudiendo clasificarse en aquéllos que estiman parentescos de parejas de individuos (métodos *pairwise*) y aquéllos basados en la reconstrucción explícita de la genealogía.

En el presente artículo presentamos un estudio de simulación cuyo objetivo es evaluar el sesgo y precisión de dos estimadores de parentesco genealógico de parejas de individuos basados en información molecular.

Métodos y simulaciones

Parentescos calculados

Parentesco genealógico: Las relaciones genéticas entre individuos se pueden obtener mediante el coeficiente de parentesco de Malécot (1948) calculado a partir de las genealogías. El coeficiente de parentesco entre los individuos i y j ($f_{g,ij}$) se define como la probabilidad de que un alelo elegido al azar del individuo i sea idéntico por descendencia a uno elegido al azar del individuo j .

Parentesco molecular: De forma análoga, también es posible definir el parentesco molecular aplicando la definición de Malécot (1948) a loci marcadores. El parentesco molecular entre dos individuos i y j ($f_{M,ij}$) es la probabilidad de que dos alelos elegidos al azar, uno de cada individuo, sean iguales (idénti-

cos en estado). Esta probabilidad puede calcularse para cada alelo l del locus k ($f_{M,ijkl}$) y sumarse para todos los alelos $f_{M,ijk}$. Considerando k marcadores, el parentesco molecular estimado será la media aritmética de los valores obtenidos con cada uno de ellos.

Estimadores del parentesco: Con objeto de aproximar la información proporcionada mediante la identidad en estado a la identidad por descendencia, se han propuesto estimadores que intentan corregir el parentesco molecular mediante el uso de las frecuencias alélicas de la población actual. En este estudio se consideraron dos de estos estimadores:

1. Estimador de Ritland (1996b):

$$\hat{f}_{Rij} = \frac{\sum_k^L \sum_l^n \frac{f_{M,ijkl} - p_{kl}^2}{p_{kl}}}{\sum_k^L (n_k - 1)}$$

donde p_{kl} es la frecuencia del alelo l del locus k , n_k es el número de alelos del locus k y L es el número de loci marcadores. Este estimador asume que la correlación entre las estimas de diferentes loci es cero, por lo que el factor de ponderación de cada locus es $(n_k - 1)$.

Ritland (1996a y 2000) indicó que una manera más correcta de estimar la varianza de los parentescos era la varianza actual, la cual refiere a la varianza que queda una vez eliminada la varianza muestral. La varianza actual del parentesco se estima como

$$var(act.) = \frac{1}{M} \sum_j \left[\frac{\left(\sum_k w_k \hat{f}_{Rijk} \right)^2 - \sum_k w_k^2 \hat{f}_{Rijk}^2}{1 - \sum_k w_k^2} \right] - \hat{f}_R^2$$

donde \hat{f}_R es el parentesco estimado promedio de la población (sobre las M parejas de individuos posibles) y w_k es la ponderación específica del locus k , estimada como

$$w_k = \frac{n_k - 1}{\sum_k (n_k - 1)}$$

2. Estimador de Li y Horvitz (1953):

$$\hat{f}_{LHij} = \frac{\sum_k^L \frac{f_{M,ijk} - \sum_l^n p_{kl}^2}{1 - \sum_l^n p_{kl}^2}}{L}$$

Simulaciones

Marcadores: Se simularon 10 o 100 marcadores moleculares neutros y codominantes con seis alelos equiprobables en cada marcador, en equilibrio de Hardy-Weinberg y de ligamiento, permitiéndose recombinación libre entre los mismos.

Carácter: Se simuló un carácter cuantitativo controlado por 120 loci bialélicos, de manera que las frecuencias alélicas eran $p = q = 0,5$ y todos los loci se encontraban en equilibrio de Hardy-Weinberg y de ligamiento, con recombinación libre. Para cada locus, los valores genotípicos fueron 5,0, 3,5 y 2,0 para los genotipos AA, Aa y aa, respectivamente. El valor genotípico de un individuo se obtuvo como la suma de los valores genotípicos de todos los loci. El valor fenotípico de los individuos se obtenía a partir de la suma del valor genotípico y un valor ambiental muestreado al azar de una distribución normal con media cero y varianza tal que la heredabilidad inicial del carácter fuera $h^2 = 0,50$.

Estructura poblacional: Se consideraron dos censos poblacionales ($N = 50$ o 500) con el mismo número de machos y hembras, sin fluctuaciones durante 10 generaciones discretas y para dos escenarios posibles. La Figura 1 muestra el procedimiento seguido para desarrollar el esquema en todas las situaciones poblacionales consideradas. En el primer escenario, la población no presentaba estructura familiar, por lo que las contribuciones de los progenitores a la descendencia eran libres y el apareamiento aleatorio. En el segundo escenario, las poblaciones presenta-

ban una estructura familiar, compuesta por 5 o 50 familias de hermanos (dependiendo del censo poblacional) con 10 individuos por familia. Los progenitores de cada familia eran elegidos al azar o seleccionados los de mayor valor fenotípico. En el caso de poblaciones estructuradas, el apareamiento era aleatorio o asociativo según su parentesco genealógico, de manera que se apareaban los progenitores más relacionados. El objetivo de este procedimiento era forzar una mayor varianza entre los parentescos de los individuos de la población.

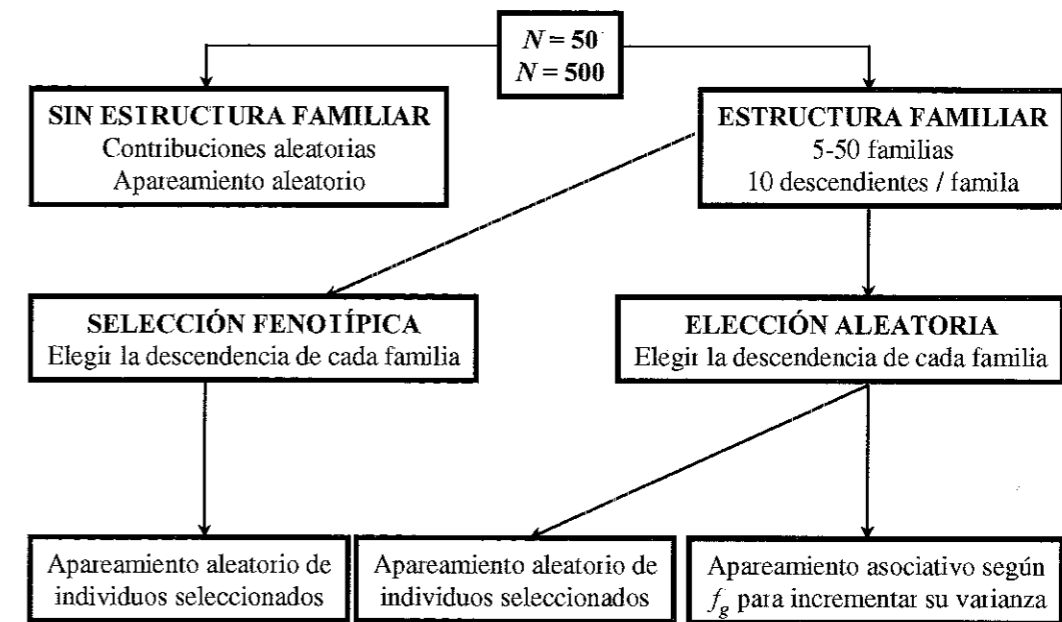


Figura 1. Procedimiento seguido para diseñar los distintos escenarios poblacionales en los que se evaluó el parentesco. f_g : parentesco genealógico.
Figure 1. Procedure followed to design the different population scenarios where the coancestry was evaluated. f_g : genealogical coancestry.

Parámetros evaluados: Para cada población analizada se calculó la media y la varianza del parentesco genealógico, el parentesco molecular y los estimadores de Ritland (1996b) y Li y Horvitz (1953), así como el promedio de la varianza actual descrita por

Ritland (1996a y 2000). La precisión de los estimadores se evaluó mediante los parámetros de la regresión y el coeficiente de determinación entre cada uno de los estimadores y el parentesco genealógico. Los cálculos se realizaron por separado para los

parentescos entre individuos (ij) y para los autoparentescos (ii), ya que estos últimos representan una estima de la consanguinidad de la población mediante la ecuación $F_i = 2 f_{ii} - 1$. Cada escenario se simuló 100 veces y los valores de los parámetros se promediaron entre simulaciones.

Resultados y discusión

En la tabla 1 se muestra, para cada población analizada, el promedio y la varianza de los parentescos genealógico, molecular y estimados mediante los métodos de Ritland (1996b) y Li y Horvitz (1953), así como el promedio de la varianza actual descrita por Ritland (1996a y 2000). Como es lógico, el parentesco molecular promedio presentó valores superiores al genealógico en todos los casos, ya que mide identidad en estado. El parentesco promedio de los dos estimadores analizados fue prácticamente nulo, tanto a corto como a largo plazo, lo que provoca un incremento en el sesgo de los estimadores a largo plazo debido al incremento continuado del parentesco genealógico promedio. Esta situación se debe al hecho de que las frecuencias iniciales se estimaron a partir de la propia población, provocando subestimas. Por esta razón se ha sugerido que los valores absolutos de las estimas de parentesco obtenidas únicamente pueden emplearse para clasificar el parentesco entre las parejas de individuos en términos relativos (Toro et al., 2002).

Con respecto a las varianzas de los parentescos, se observaron valores mayores para el parentesco molecular y los dos estimadores que para el parentesco genealógico. En el caso en el que se disponía de 100 marcadores la varianza media del parentesco molecular se redujo, adquiriendo valores similares a la varianza media del parentesco genealógico. Para los dos estimadores anali-

zados también se observó ligeramente esta reducción, siendo más acusada en poblaciones con censo $N = 500$ sin estructura familiar. Exceptuando el caso de poblaciones estructuradas y un censo reducido a largo plazo, es importante señalar la similitud entre la varianza genealógica media y el promedio de la varianza actual, lo cual tiene implicaciones a la hora de aplicar el estimador de heredabilidad propuesto por Ritland (1996a y 2000).

La tabla 2 es equivalente a la tabla 1 pero referida a los valores de consanguinidad. De la misma manera que ocurría con el parentesco entre individuos, el coeficiente de consanguinidad molecular promedio presentó valores superiores al genealógico en todos los casos, mientras que la media para los dos estimadores analizados fue ligeramente inferior al valor real en todos los casos, excepto a corto plazo en poblaciones con $N = 500$. Empleando 100 marcadores se redujeron las varianzas medias del coeficiente de consanguinidad, sobre todo en el caso del coeficiente de consanguinidad molecular, pero sin alcanzar a la del genealógico. Esta leve reducción de varianza promedio en los dos estimadores analizados fue más acusada en poblaciones con censo $N = 500$ y sin estructura familiar.

En la tabla 3 se muestra el promedio del coeficiente de determinación, la pendiente de regresión y el intercepto, entre el parentesco genealógico y los estimadores. Los valores de pendiente de regresión e intercepto correspondientes a la situación ideal son $b = 1,00$ y $a = 0,00$, respectivamente. También se muestran los anteriores coeficientes para el parentesco molecular y el genealógico. En cuanto al coeficiente de determinación, en el caso de 100 marcadores se encontraron mayores valores que cuando se disponía únicamente de 10 marcadores. Si la población tenía una estructura familiar y se realizaba apareamiento aleatorio de los progenitores

seleccionados al azar, se obtenían mayores valores (tanto a corto como a largo plazo y con 10 y 100 marcadores) que si la población carecía de estructura familiar. Este hecho

probablemente sea debido a la menor varianza genealógica promedio en los casos en los que no existía una estructura familiar (véase la tabla 1).

Tabla 1. Media y varianza ($\times 10^2$) del parentesco genealógico (f_g), molecular (f_M), Ritland (f_R) (1996b), y Li y Horvitz (f_{LH}) (1953) y la varianza actual (act.) propuesta por Ritland (1996a y 2000) en las generaciones (t) 1 y 10; (1) y (2) población con $N = 50$; (3), (4), (5) y (6) $N = 500$; (1) y (3) población sin estructura familiar; (2), (4) y (5) población con estructura familiar y padres elegidos al azar; (6) población con estructura familiar y padres seleccionados según su valor fenotípico; (5) apareamiento asociativo de individuos según su parentesco genealógico

Table 1. Mean and variance ($\times 10^2$) of genealogical coancestry (f_g), molecular (f_M), Ritland (f_R) (1996b), and Li and Horvitz (f_{LH}) (1953) and the actual (act.) variance proposed by Ritland (1996a and 2000) at generations (t) 1 and 10; (1) and (2) population of $N = 50$; (3), (4), (5) and (6) $N = 500$; (1) and (3) population with no family structure; (2), (4) and (5) population with a family structure and parents selected at random; (6) population with a family structure and parents selected according to their phenotypic value; (5) assortative mating according to the genealogical coancestry

| Marcadores | Escenario | t | f_g | f_M | f_R | f_{LH} | $V(f_g)$ | $V(f_M)$ | $V(f_R)$ | $V(f_{LH})$ | $V(act)$ | |
|------------|-----------|-----|-------|-------|-------|----------|----------|----------|----------|-------------|----------|------|
| 10 | 1 | 1 | 0,01 | 0,18 | -0,01 | -0,01 | 0,12 | 0,44 | 0,56 | 0,66 | 0,11 | |
| | | 10 | 0,09 | 0,25 | -0,01 | -0,01 | 0,13 | 0,60 | 0,58 | 1,14 | 0,13 | |
| | 2 | 1 | 0,05 | 0,21 | -0,01 | -0,01 | 0,94 | 1,02 | 1,36 | 1,65 | 0,96 | |
| | | 10 | 0,34 | 0,44 | -0,01 | -0,01 | 0,40 | 0,88 | 1,73 | 1,92 | 0,87 | |
| | 3 | 1 | 0,00 | 0,17 | 0,00 | 0,00 | 0,01 | 0,36 | 0,51 | 0,51 | 0,01 | |
| | | 10 | 0,01 | 0,18 | 0,00 | 0,00 | 0,01 | 0,36 | 0,51 | 0,51 | 0,02 | |
| | 4 | 1 | 0,01 | 0,17 | 0,00 | 0,00 | 0,11 | 0,42 | 0,60 | 0,62 | 0,11 | |
| | | 10 | 0,04 | 0,20 | 0,00 | 0,00 | 0,13 | 0,51 | 0,62 | 0,81 | 0,14 | |
| | 5 | 1 | 0,01 | 0,17 | 0,00 | 0,00 | 0,11 | 0,42 | 0,60 | 0,62 | 0,11 | |
| | | 10 | 0,05 | 0,20 | 0,00 | 0,00 | 0,27 | 0,65 | 0,78 | 1,03 | 0,27 | |
| | 6 | 1 | 0,01 | 0,17 | 0,00 | 0,00 | 0,11 | 0,43 | 0,60 | 0,62 | 0,11 | |
| | | 10 | 0,07 | 0,22 | 0,00 | 0,00 | 0,14 | 0,56 | 0,63 | 0,94 | 0,14 | |
| | 100 | 1 | 1 | 0,01 | 0,18 | -0,01 | -0,01 | 0,12 | 0,12 | 0,15 | 0,18 | 0,11 |
| | | | 10 | 0,09 | 0,25 | -0,01 | -0,01 | 0,13 | 0,14 | 0,17 | 0,26 | 0,12 |
| | | 2 | 1 | 0,05 | 0,21 | -0,01 | -0,01 | 0,94 | 0,69 | 1,00 | 1,11 | 0,96 |
| | | | 10 | 0,34 | 0,45 | -0,01 | -0,01 | 0,39 | 0,34 | 0,93 | 0,98 | 0,84 |
| | | 3 | 1 | 0,00 | 0,17 | 0,00 | 0,00 | 0,01 | 0,04 | 0,06 | 0,06 | 0,01 |
| | | | 10 | 0,01 | 0,18 | 0,00 | 0,00 | 0,01 | 0,04 | 0,07 | 0,06 | 0,02 |
| 4 | | 1 | 0,01 | 0,17 | 0,00 | 0,00 | 0,11 | 0,11 | 0,16 | 0,16 | 0,11 | |
| | | 10 | 0,04 | 0,20 | 0,00 | 0,00 | 0,13 | 0,13 | 0,19 | 0,21 | 0,14 | |
| 5 | | 1 | 0,01 | 0,17 | 0,00 | 0,00 | 0,11 | 0,11 | 0,16 | 0,16 | 0,11 | |
| | | 10 | 0,05 | 0,20 | 0,00 | 0,00 | 0,26 | 0,23 | 0,31 | 0,37 | 0,26 | |
| 6 | | 1 | 0,01 | 0,17 | 0,00 | 0,00 | 0,11 | 0,11 | 0,16 | 0,16 | 0,11 | |
| | | 10 | 0,07 | 0,22 | 0,00 | 0,00 | 0,13 | 0,14 | 0,19 | 0,23 | 0,14 | |

Tabla 2. Media y varianza ($\times 10^2$) del coeficiente de consanguinidad genealógico (F_g), molecular (F_M), Ritland (F_R) (1996b), y Li y Horvitz (F_{LH}) (1953) en las generaciones (t) 1 y 10; (1) y (2) población con $N = 50$; (3), (4), (5) y (6) $N = 500$; (1) y (3) población sin estructura familiar; (2), (4) y (5) población con estructura familiar y padres elegidos al azar; (6) población con estructura familiar y padres seleccionados según su valor fenotípico; (5) apareamiento asociativo de individuos según su parentesco genealógico

Table 2. Mean and variance ($\times 10^2$) of the genealogical inbreeding coefficient (F_g), molecular (F_M), Ritland (F_R) (1996b), and Li and Horvitz (F_{LH}) (1953) at generations (t) 1 and 10; (1) and (2) population of $N = 50$; (3), (4), (5) and (6) $N = 500$; (1) and (3) population with no family structure; (2), (4) and (5) population with a family structure and parents selected at random; (6) population with a family structure and parents selected according to their phenotypic value; (5) assortative mating according to the genealogical coancestry

| Marcadores | Escenario | t | F_g | F_M | F_R | F_{LH} | $V(F_g)$ | $V(F_M)$ | $V(F_R)$ | $V(F_{LH})$ |
|------------|-----------|----|-------|-------|-------|----------|----------|----------|----------|-------------|
| 10 | 1 | 1 | 0,00 | 0,16 | -0,02 | -0,02 | 0,00 | 1,36 | 2,80 | 2,04 |
| | | 10 | 0,08 | 0,24 | -0,02 | -0,02 | 0,12 | 1,84 | 10,24 | 3,48 |
| | 2 | 1 | 0,00 | 0,16 | -0,06 | -0,06 | 0,00 | 1,28 | 3,52 | 2,08 |
| | | 10 | 0,30 | 0,40 | -0,08 | -0,08 | 0,12 | 2,08 | 10,04 | 3,56 |
| | 3 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 1,40 | 2,04 | 2,00 |
| | | 10 | 0,02 | 0,18 | 0,00 | 0,00 | 0,00 | 1,40 | 2,04 | 2,00 |
| | 4 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 1,36 | 2,12 | 2,00 |
| | | 10 | 0,04 | 0,20 | 0,00 | 0,00 | 0,04 | 1,56 | 5,04 | 2,48 |
| | 5 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 1,36 | 2,12 | 1,96 |
| | | 10 | 0,08 | 0,24 | 0,04 | 0,04 | 0,28 | 1,96 | 6,44 | 3,16 |
| | 6 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 1,40 | 2,16 | 2,00 |
| | | 10 | 0,06 | 0,22 | 0,00 | 0,00 | 0,04 | 1,72 | 7,76 | 2,88 |
| 100 | 1 | 1 | 0,00 | 0,16 | -0,02 | -0,02 | 0,00 | 0,12 | 0,44 | 0,20 |
| | | 10 | 0,08 | 0,24 | -0,02 | -0,02 | 0,12 | 0,24 | 1,60 | 0,48 |
| | 2 | 1 | 0,00 | 0,16 | -0,06 | -0,06 | 0,00 | 0,12 | 0,36 | 0,20 |
| | | 10 | 0,28 | 0,40 | -0,08 | -0,08 | 0,12 | 0,28 | 1,88 | 0,84 |
| | 3 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 0,12 | 0,20 | 0,20 |
| | | 10 | 0,02 | 0,18 | 0,00 | 0,00 | 0,00 | 0,12 | 0,20 | 0,20 |
| | 4 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 0,12 | 0,20 | 0,20 |
| | | 10 | 0,04 | 0,20 | 0,00 | 0,00 | 0,04 | 0,16 | 0,60 | 0,28 |
| | 5 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 0,12 | 0,20 | 0,20 |
| | | 10 | 0,08 | 0,24 | 0,04 | 0,04 | 0,28 | 0,36 | 1,12 | 0,56 |
| | 6 | 1 | 0,00 | 0,16 | 0,00 | 0,00 | 0,00 | 0,12 | 0,20 | 0,20 |
| | | 10 | 0,06 | 0,22 | 0,00 | -0,02 | 0,04 | 0,20 | 1,12 | 0,32 |

Las diferencias entre los coeficientes de determinación calculados fueron mínimas, apreciándose una ligera superioridad en el caso del estimador de Ritland (1996b). Los valores fueron exactamente iguales entre los coeficientes de determinación molecular y del estimador de Li y Horvitz (1953) en

todos los casos analizados excepto en el caso de poblaciones estructuradas de censo poblacional reducido a largo plazo, en las que el cambio en las frecuencias alélicas es más fácilmente detectable debido al censo efectivo reducido que presenta esta situación poblacional. Los coeficientes de deter-

minación permanecieron, en general, constantes a lo largo del tiempo excepto para los casos de poblaciones con censo reducido, con estructura familiar y apareamiento aleatorio de los individuos seleccionados al azar, en los que se redujo con el paso de las generaciones, y los de poblaciones con censo elevado, con estructura familiar y apareamiento atendiendo al parentesco

genealógico de los individuos seleccionados al azar, en los que se observó un aumento. La razón podría ser que, en el primer escenario poblacional se reduce la varianza promedio del parentesco genealógico debido al reducido censo. Por el contrario, en el segundo caso la varianza se incrementa con el paso de las generaciones (véase la tabla 1).

Tabla 3. Coeficiente de determinación (r^2), pendiente (b) e intercepto (a) de la regresión del parentesco genealógico dado el parentesco molecular y los estimadores de Ritland (1996b), y de Li y Horvitz (1953) en las generaciones (t) 1 y 10; (1) y (2) población con $N = 50$; (3), (4), (5) y (6) $N = 500$; (1) y (3) población sin estructura familiar; (2), (4) y (5) población con estructura familiar y padres elegidos al azar; (6) población con estructura familiar y padres seleccionados según su valor fenotípico; (5) apareamiento asociativo de individuos según su parentesco genealógico

Table 3. Coefficient of determination (r^2), coefficient (b) and intercept (a) of regression of the genealogical coancestry given the molecular coancestry, Ritland (1996b), and Li and Horvitz (1953) at generations (t) 1 and 10; (1) and (2) population of $N = 50$; (3), (4), (5) and (6) $N = 500$; (1) and (3) population with no family structure; (2), (4) and (5) population with a family structure and parents selected at random; (6) population with a family structure and parents selected according to their phenotypic value; (5) assortative mating according to the genealogical coancestry

| Marcadores | Escenario | t | r^2_{gM} | r^2_{gR} | r^2_{gLH} | b_{gM} | b_{gr} | b_{gLH} | a_{gM} | a_{gR} | a_{gLH} |
|------------|-----------|----|------------|------------|-------------|----------|----------|-----------|----------|----------|-----------|
| 10 | 1 | 1 | 0,19 | 0,19 | 0,19 | 0,23 | 0,20 | 0,19 | -0,03 | 0,01 | 0,01 |
| | | 10 | 0,16 | 0,21 | 0,15 | 0,19 | 0,22 | 0,13 | 0,05 | 0,10 | 0,10 |
| | 2 | 1 | 0,64 | 0,72 | 0,64 | 0,77 | 0,71 | 0,61 | -0,11 | 0,05 | 0,05 |
| | | 10 | 0,34 | 0,48 | 0,33 | 0,39 | 0,33 | 0,31 | 0,17 | 0,34 | 0,33 |
| | 3 | 1 | 0,02 | 0,02 | 0,02 | 0,03 | 0,02 | 0,02 | 0,00 | 0,00 | 0,00 |
| | | 10 | 0,03 | 0,03 | 0,03 | 0,04 | 0,03 | 0,03 | 0,00 | 0,01 | 0,01 |
| | 4 | 1 | 0,18 | 0,19 | 0,18 | 0,22 | 0,19 | 0,18 | -0,03 | 0,00 | 0,00 |
| | | 10 | 0,18 | 0,22 | 0,18 | 0,21 | 0,22 | 0,17 | 0,00 | 0,04 | 0,04 |
| | 5 | 1 | 0,18 | 0,19 | 0,18 | 0,22 | 0,19 | 0,18 | -0,03 | 0,00 | 0,00 |
| | | 10 | 0,29 | 0,32 | 0,29 | 0,35 | 0,34 | 0,27 | -0,03 | 0,05 | 0,05 |
| | 6 | 1 | 0,18 | 0,19 | 0,18 | 0,22 | 0,19 | 0,18 | -0,03 | 0,00 | 0,00 |
| | | 10 | 0,17 | 0,22 | 0,17 | 0,21 | 0,22 | 0,16 | 0,02 | 0,07 | 0,07 |
| 100 | 1 | 1 | 0,70 | 0,69 | 0,70 | 0,84 | 0,73 | 0,69 | -0,14 | 0,02 | 0,02 |
| | | 10 | 0,64 | 0,69 | 0,63 | 0,78 | 0,73 | 0,57 | -0,10 | 0,10 | 0,10 |
| | 2 | 1 | 0,95 | 0,96 | 0,95 | 1,14 | 0,95 | 0,90 | -0,19 | 0,06 | 0,05 |
| | | 10 | 0,82 | 0,86 | 0,81 | 0,98 | 0,60 | 0,57 | -0,10 | 0,34 | 0,33 |
| | 3 | 1 | 0,20 | 0,20 | 0,20 | 0,24 | 0,20 | 0,20 | -0,04 | 0,00 | 0,00 |
| | | 10 | 0,24 | 0,25 | 0,24 | 0,28 | 0,25 | 0,23 | -0,04 | 0,01 | 0,01 |
| | 4 | 1 | 0,69 | 0,70 | 0,69 | 0,83 | 0,69 | 0,68 | -0,14 | 0,01 | 0,01 |
| | | 10 | 0,69 | 0,73 | 0,68 | 0,82 | 0,72 | 0,65 | -0,12 | 0,04 | 0,04 |
| | 5 | 1 | 0,69 | 0,70 | 0,69 | 0,82 | 0,69 | 0,68 | -0,14 | 0,01 | 0,01 |
| | | 10 | 0,80 | 0,81 | 0,80 | 0,96 | 0,83 | 0,76 | -0,15 | 0,05 | 0,05 |
| | 6 | 1 | 0,69 | 0,70 | 0,69 | 0,83 | 0,69 | 0,68 | -0,14 | 0,01 | 0,01 |
| | | 10 | 0,67 | 0,71 | 0,67 | 0,81 | 0,71 | 0,62 | -0,11 | 0,07 | 0,07 |

El comportamiento de la pendiente de regresión, en general, fue muy similar al del coeficiente de determinación. En cuanto al intercepto, nótese que el parentesco molecular produjo valores inferiores, que se hacen más importantes al incrementar el número de marcadores, indicando el exceso que supone la identidad en estado. Sin embargo, los estimadores de parentesco presentan valores de intercepto próximos a cero indicando una mayor precisión, excepto en el escenario poblacional de una población con censo reducido, estructura familiar y apareamiento aleatorio de los individuos elegidos al azar, probablemente debido a la reducción en la varianza promedio del parentesco genealógico (véase la tabla 1).

La tabla 4 es equivalente a la tabla 3 pero referida a los valores de consanguinidad.

Tabla 4. Coeficiente de determinación (r^2), pendiente (b) e intercepto (a) de la regresión del coeficiente de consanguinidad genealógico dada la consanguinidad molecular y los estimadores de Ritland (1996b), y Li y Horvitz (1953) en la generación (t) 10; (1) y (2) población con $N = 50$; (3), (4), (5) y (6) $N = 500$; (1) y (3) población sin estructura familiar; (2), (4) y (5) población con estructura familiar y padres elegidos al azar; (6) población con estructura familiar y padres seleccionados según su valor fenotípico; (5) apareamiento asociativo de individuos según su parentesco genealógico

Table 4. Coefficient of determination (r^2), coefficient (b) and intercept (a) of regression of the genealogical inbreeding coefficient given the molecular inbreeding, Ritland (1996b), and Li and Horvitz (1953) at generation (t) 10; (1) and (2) population of $N = 50$; (3), (4), (5) and (6) $N = 500$; (1) and (3) population with no family structure; (2), (4) and (5) population with a family structure and parents selected at random; (6) population with a family structure and parents selected according to their phenotypic value; (5) assortative mating according to the genealogical coancestry

| Marcadores | Escenario | t | r^2_{gM} | r^2_{gR} | r^2_{gLH} | b_{gM} | b_{gR} | b_{gLH} | a_{gM} | a_{gR} | a_{gLH} |
|------------|-----------|-----|------------|------------|-------------|----------|----------|-----------|----------|----------|-----------|
| 10 | 1 | 10 | 0,08 | 0,03 | 0,08 | 0,06 | 0,01 | 0,05 | 0,06 | 0,09 | 0,09 |
| | 2 | 10 | 0,11 | 0,09 | 0,11 | 0,04 | 0,00 | 0,03 | 0,28 | 0,30 | 0,29 |
| | 3 | 10 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 | 0,01 |
| | 4 | 10 | 0,02 | 0,01 | 0,02 | 0,02 | 0,00 | 0,01 | 0,04 | 0,04 | 0,03 |
| | 5 | 10 | 0,10 | 0,05 | 0,10 | 0,11 | 0,04 | 0,09 | 0,05 | 0,08 | 0,09 |
| | 6 | 10 | 0,03 | 0,02 | 0,03 | 0,03 | 0,00 | 0,02 | 0,05 | 0,06 | 0,06 |
| 100 | 1 | 10 | 0,33 | 0,06 | 0,32 | 0,40 | 0,06 | 0,29 | -0,02 | 0,08 | 0,09 |
| | 2 | 10 | 0,27 | 0,22 | 0,27 | 0,28 | 0,02 | 0,13 | 0,16 | 0,30 | 0,03 |
| | 3 | 10 | 0,07 | 0,06 | 0,07 | 0,08 | 0,06 | 0,07 | 0,00 | 0,00 | 0,01 |
| | 4 | 10 | 0,13 | 0,04 | 0,13 | 0,15 | 0,03 | 0,12 | 0,01 | 0,03 | 0,04 |
| | 5 | 10 | 0,49 | 0,19 | 0,49 | 0,59 | 0,20 | 0,47 | -0,07 | 0,08 | 0,07 |
| | 6 | 10 | 0,16 | 0,02 | 0,16 | 0,19 | 0,00 | 0,15 | 0,01 | 0,06 | 0,07 |

Solamente se presentan los valores obtenidos en la generación 10, debido a que en la generación 1 no había individuos consanguíneos ($F = 0,00$). Los valores obtenidos del coeficiente de determinación fueron inferiores a los obtenidos con el parentesco (véase la tabla 3) y muy bajos incluso cuando se dispone de 100 marcadores. Los valores de r^2 fueron especialmente reducidos en el caso del estimador de Ritland (1996b). Las estimas de la pendiente de regresión presentaron valores muy reducidos, incluso próximos a cero, aunque mejoraron ligeramente en el caso de disponer de 100 marcadores. Las estimas del intercepto presentaron valores muy próximos a cero, probablemente debido a la reducida variación que se genera en los parentescos genealógicos (especialmente con $N = 500$).

Conclusiones

En este estudio hemos analizado el comportamiento de dos estimadores de parentesco genealógico en situaciones en las que se dispone de 10 o 100 marcadores. Sin embargo, en el desarrollo de estudios experimentales es habitual que el número de marcadores analizados sea próximo a 10 y, en cualquier caso, muy alejado de la situación hipotética de 100 marcadores, aunque es posible que este número se vea incrementado a medida que se abaraten las técnicas de obtención de información molecular. Nuestros resultados no proporcionan evidencias de un comportamiento adecuado de los estimadores de parentesco genealógico analizados, aunque mejora en el caso hipotético de que se disponga de un número considerable de marcadores. El factor crítico del que depende la precisión de los estimadores es la varianza o dispersión de los valores del parentesco genealógico. Por ello el comportamiento de los estimadores mejora en poblaciones que poseen un elevado censo poblacional y una estructura familiar definida. Los estimadores analizados serían más adecuados para inferir parentesco (10 marcadores r^2 0,02 – 0,72; 100 marcadores r^2 0,20 – 0,96) que consanguinidad (10 marcadores r^2 0,01 – 0,11; 100 marcadores r^2 0,07 – 0,49) y, concretamente, nuestros resultados descartan al estimador propuesto por Ritland (1996b) para inferir consanguinidad, debido al reducido coeficiente de determinación que presenta. Finalmente, hay que resaltar que, a efectos de comparación y ordenación de los parentescos, los resultados indican que resulta igualmente efectivo el parentesco molecular que cualquiera de los estimadores estudiados, debido al análogo coeficiente de determinación que presentan.

Agradecimientos

Este trabajo ha sido financiado mediante el proyecto Plan Estratégico del INIA (CPE03-004-C2).

Bibliografía

- Awise JC, 1994. Molecular Markers, Natural History and Evolution. Chapman and Hall, New York, USA.
- Caballero A, Toro MA, 2000. Interrelations between effective population size and other pedigree tools for the management of conserved populations. Genet. Res. 75: 331-343.
- Caballero A, Toro MA, 2002. Analysis of genetic diversity for the management of conserved subdivided populations. Conserv. Genet. 3: 289-299.
- Fernández J, Toro MA, 2006. A new method to estimate relatedness from molecular markers. Mol. Ecol. 15: 1657-1667.
- Frankham R, Ballou JD, Briscoe DA, 2002. Introduction to Conservation Genetics. Cambridge University Press, Cambridge, Reino Unido.
- Li CC, Horvitz DG, 1953. Some methods of estimating the inbreeding coefficient. Am. J. Hum. Genet. 5: 107-117.
- Lynch M, Walsh B, 1998. Genetics and Analysis of Quantitative Traits. Sinauer Associates. Sunderland, Massachusetts, USA.
- Malécot G, 1948. Les Mathématiques de l'hérédité. Masson et. Cie., Paris, Francia.
- Ritland K, 1996a. A marker-based method for inferences about quantitative inheritance in natural populations. Evolution 50: 1062-1073.
- Ritland K, 1996b. Estimators for pairwise relatedness and individual inbreeding coefficients. Genet. Res. 67: 175-186.
- Ritland K, 2000. Marker-inferred relatedness as a tool for detecting heritability in nature. Mol. Ecol. 9: 1195-1204.

Toro MA, Barragán C, Óvilo C, Rodrigáñez J, Rodríguez C, Silió L, 2002. Estimation of coancestry in Iberian pigs using molecular markers. *Conserv. Genet.* 3: 309-320.

(Aceptado para publicación el 2 de mayo de 2006)