

NORMALIZACIÓN DE COEFICIENTES DE CONSANGUINIDAD BASADOS EN REGIONES DE HOMOCIGOSIDAD

Gomez-Raya, L., Rodríguez, C., Barragán, C. y Silió, L.

Departamento de Mejora Genética Animal

Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA)

Ctra. de La Coruña km 7, 28040 Madrid

INTRODUCCIÓN

Un coeficiente de consanguinidad es una estima puntual de la cantidad de apareamientos entre ancestros relacionados genéticamente de un individuo. Los coeficientes genealógicos de consanguinidad, definidos como la probabilidad de que los dos alelos de un locus de un individuo sean idénticos por descendencia (Crow y Kimura, 1970), han sido de uso generalizado. Los avances en tecnologías genómicas permiten un conocimiento más profundo de la autocigosidad, que podría proporcionar una mayor información de la magnitud de las relaciones genéticas entre los ancestros de un individuo. Existen diferentes medidas de autocigosidad, basadas en marcadores moleculares, que permiten estimar el nivel de consanguinidad de un individuo o población. El primero de los estadísticos basado en información molecular fue la proporción de posiciones monomórficas a nivel individual (Toro et al., 2009). El valor esperado de esta medida empírica de similaridad es una función de la consanguinidad genealógica, que es fácil de estimar utilizando información molecular (Van Raden, 2008; Toro et al., 2011). Sin embargo, estos estadísticos ignoran los SNPs ligados ubicados en el mismo cromosoma cuyos haplotipos se heredan como una unidad. El apareamiento entre parientes conlleva a la herencia de fragmentos de cromosomas o Regiones de Homocigosidad (ROH) que son idénticos por descendencia y no simples SNP individuales distribuidos a lo largo del genoma. Por lo tanto, se podrían utilizar ROH para estimar a nivel genómico la autocigosidad (Keller et al., 2011; Pemberton et al., 2012). La idea inicial fue separar ROH de acuerdo a su longitud, y asociar la proporción del genoma cubierta por los ROH mas largos a la consanguinidad reciente. Por ejemplo, Ferencakovic et al. (2011) utilizó fragmentos de entre 1-2 Mb a >16 Mb de longitud para buscar la relación entre tamaño de ROH y consanguinidad reciente en vacuno. Los umbrales utilizados por estos autores no utilizan toda la información disponible sobre los ROH. Recientemente, Gomez-Raya et al. (2015) han propuesto nuevos coeficientes basados en la función de probabilidad de la longitud de ROH de un individuo. Sin embargo, la normalización propuesta por estos autores sólo obliga a los coeficientes de consanguinidad a estar entre 0 y 1, lo cual complica la comparación de coeficientes de consanguinidad de diferentes grupos, poblaciones o individuos si los datos de la distribución de la longitud de ROH de todas esas fuentes no están disponibles. En este estudio, proponemos un nuevo método de normalización de coeficientes genómicos de consanguinidad que no requiere comparaciones de la distribución de la longitud de ROH entre varias fuentes. El método se ilustra con estimaciones de consanguinidad genómica de 217 cerdas ibéricas.

MATERIAL Y MÉTODOS

Gomez-Raya et al. (2015) propusieron la estimación de coeficientes de consanguinidad después de ajustar una distribución exponencial a la longitud de ROH de un individuo y a la distribución de todos los individuos de la población de referencia por

$$F_{ROH-E^m} = \frac{1}{\lambda_I-ROH} - \frac{1}{\lambda_A-ROH}$$

donde λ_{I-ROH} y λ_{A-ROH} son los parámetros después de ajustar una distribución exponencial a un individuo o a todos los individuos, respectivamente. Este método requiere el ajuste de las dos distribuciones a comparar: la del individuo y la de la población de referencia. Los nuevos coeficientes de consanguinidad normalizados utilizan la función de densidad acumulada de un individuo (o grupo) después de ajustar una distribución exponencial. De esta manera se puede calcular la probabilidad de que un ROH extraído al azar sea mayor o igual a un umbral predeterminado. Así, el coeficiente de consanguinidad normalizado se define como:

$$F_{E(T)} = F(T) = \int_T^{\infty} \lambda_{I-ROH} e^{-(\lambda_{I-ROH})x} d(x)$$

donde T es el umbral. Así pues, $F_{E(T)}$, estima el grado de autocigoidad de un individuo. El coeficiente normalizado de consanguinidad esta forzando a variar entre 0 y 1 debido a que esta definido como una probabilidad. El umbral es arbitrario pero comparaciones entre individuos (o grupos) es posible si se utiliza el mismo umbral. Los nuevos coeficientes fueron utilizados para estimar la consanguinidad de 217 cerdas de la estirpe Torbiscal de cerdo ibérico. Esta línea ha permanecido genéticamente cerrada entre 1963 y 2013 y su genealogía registrada se completa con la de sus ancestros desde 1944. Se comparó el coeficiente de consanguinidad genealógico (F_{ped}) con los nuevos coeficientes de consanguinidad utilizando un análisis de regresión.

RESULTADOS Y DISCUSIÓN

La densidad kernel de la distribución de la longitud de ROH de todos los individuos se utilizó para decidir el umbral que podría ser usado en la derivación de los coeficientes normalizados genómicos de consanguinidad (Figura 1). Se tomó 1Mb como el umbral que podría representar el punto a partir del cual la autocigoidad podría diferir entre individuos. La regresión de los coeficientes genómicos de consanguinidad normalizados sobre los coeficientes de consanguinidad genealógicos se presentan en la Figura 2. El valor de la correlación fue 0,691 (SD=0,0493). Los dos coeficientes están bien correlacionados pero el coeficiente normalizado de consanguinidad es superior por la gran cantidad de información genómica utilizada en su estimación. Por otra parte, el coeficiente de consanguinidad genealógico no es capaz de tener en cuenta la segregación mendeliana y la posibilidad de herencia de fragmentos de autocigoidad por azar y representa simplemente una estima puntual de la probabilidad promedio de autocigoidad en el genoma. Por ejemplo, estos coeficientes no pueden estimar el nivel de consanguinidad por cromosoma. Además, los coeficientes genealógicos de consanguinidad son incapaces de comparar dos individuos no relacionados genéticamente pero que pertenecen a poblaciones diferentes. Los nuevos coeficientes de consanguinidad genómicos si que permiten la comparación de los niveles de consanguinidad entre individuos pertenecientes a poblaciones diferentes.

La conclusión de este trabajo es que los nuevos coeficientes de consanguinidad genómicos normalizados (definidos como la probabilidad de que un ROH extraído al azar de un individuo es mayor o igual a 1Mb), podría ser superior a los coeficientes genealógicos de consanguinidad.

REFERENCIAS BIBLIOGRÁFICAS

- Crow, J.F. & Kimura, M. 1970 An Introduction to Population Genetics Theory. New York: Harper and Row.
- Gomez-Raya, L., Rodríguez, C, Barragán, C. & Silió, L. Molecular inbreeding coefficients based on the distribution of the length of runs of homozygosity in a closed line of Iberian pigs. (enviado). 2015.
- Ferenčaković, M., Hamzić, E., Gredler, B., Solberg, T.R., Klemetsdal, G., Curik, I. & Sölkner, J.. 2013. Estimates of autozygosity derived from runs of homozygosity: empirical evidence from selected cattle populations. J. Anim. Breed. Genet 130: 286-293.
- Keller, M.C., Visscher, P.M., & Goddard, M.E. 2011. Quantification of inbreeding due to distant ancestors and its detection using dense SNP data. Genetics, 189: 237-249.
- Pemberton, T., Asher, D., Feldman, M.W., Myers, R.M., Rosenberg, N.A. & Li, J.Z. 2012. Genomic patterns of homozygosity in worldwide human populations. Amer J. Hum. Genet. 91(2): 275-292.
- Toro, M., Caballero, A. & Fernández, J. 2009. Molecular characterization of breeds and its use in conservation. Livest. Sci. 120: 174–195.
- Toro, M., García-Cortés, L.A. & Legarra, A. 2011. A note on the rationale for estimating genealogical coancestry from molecular markers. Genet. Sel. Evol., 43: 27.
- Van Raden, P.M. 2008. Efficient methods to compute genomic predictions. J. Dairy Sci. 91: 4414-4423.

Agradecimientos: Esta investigación fue financiada con el proyecto RTA2011-00113. Agradecemos los esfuerzos de Jaime Rodríguez y toda la plantilla del 'Dehesón del

Encinar' por el mantenimiento de los registros de pedigrí de la estirpe Torbiscal y sus ancestros hasta el cierre reciente de esta granja.

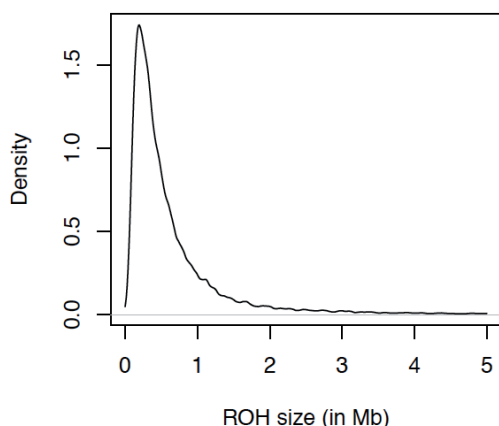


Figura 1. Kernel de densidad de la distribución de la longitud de ROH para todos los individuos (Numero de ROH = 299750, Bandwidth = 0,02935).

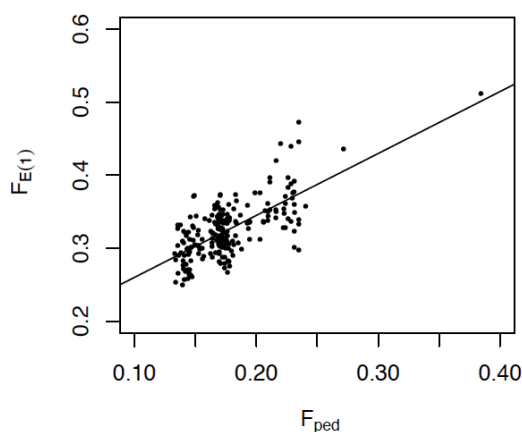


Figura 2. Regresión de los coeficientes de consanguinidad normalizados ($F_{E(1)}$) sobre los coeficientes de consanguinidad genealógicos (F_{ped}) para 217 cerdas ibéricas.

NORMALIZATION OF PARAMETRIC INBREEDING COEFFICIENTS BASED ON RUNS OF HOMOZYGOSITY

ABSTRACT: A new method for the estimation of normalized inbreeding coefficients based on the distribution of the length of Regions of Homozygosity (ROH) is proposed. The method is based on fitting an exponential distribution to the distribution of the length of ROH in order to estimate the probability that an ROH drawn at random from the distribution of that individual is greater or equal to a threshold (1Mb). The correlation between the normalized inbreeding coefficient and the pedigree inbreeding coefficient was 0.691 (0.0493). The novel normalized inbreeding coefficient could be superior to inbreeding coefficients based on pedigree information due to the use of all information from the distribution of ROH.

Keywords: Runs of homozygosity, Inbreeding, Iberian pigs, Swine.