APLICACIÓN DE CORRELACIONES PARCIALES Y TEORÍA DE LA INFORMACIÓN A REDES DE ESTIMAS DE CONSANGUINIDAD EN CONEJOS

Rodríguez-Ramilo[1], S. T., Reverter[2], A., Sánchez[3], J. P., Velasco-Galilea[3], M., González[3], O. y Piles[3], M.
[1]INRA, Castanet-Tolosan, 31326, Francia. [2]CSIRO Agriculture & Food, Brisbane, QLD 4067, Australia. [3]IRTA, Caldes de Montbui, 68140, España.
miriam.piles@irta.es

INTRODUCCIÓN

Obtener estimas precisas de consanguinidad es importante para el manejo de poblaciones en programas de selección. Entre las consecuencias más importantes de la consanguinidad se encuentran la pérdida de variabilidad genética y la reducción de la media de caracteres con interés económico (Falconer y Mackay, 1996). Tradicionalmente, el coeficiente de consanguinidad se ha estimado a partir del pedigrí. Sin embargo, los chips de SNPs permiten obtener estimas genómicas de consanguinidad.

Es habitual encontrar en la literatura la correlación que existe entre distintas estimas de consanguinidad genómicas y las obtenidas a partir de la información de pedigrí (e. g. Rodríguez-Ramilo et al., 2019). Sin embargo, algunas de estas correlaciones podrían ser espurias si, por ejemplo, dadas tres estimas de consanguinidad (A, B y C) sucede que obtenemos una correlación moderada entre dos estimas independientes (A y B) debida a la confusión con la otra (C), cuando la correlaciones entre A y C y entre B y C son ambas elevadas. La cuestión es si la correlación moderada que existe entre A y B es significativamente independiente de las altas correlaciones que existen entre A y C y entre B y C. Para tratar de responder a esta cuestión hay algoritmos que emplean los coeficientes de correlación parcial de primer orden y la teoría de la información (PCIT) para identificar asociaciones significativas entre dos variables e independientes de una tercera variable (Reverter y Chan, 2008). En este trabajo se propone el uso del algoritmo PCIT para detectar asociaciones significativas entre distintas estimas de consanguinidad y fenotipos en una población seleccionada de conejos.

MATERIAL Y MÉTODOS

Datos

Los animales proceden de la población seleccionada del IRTA fundada en 1983 cruzando conejos de 5 líneas New Zealand White y una línea sintética de California x New Zealand. Esta población ha sido seleccionada para peso de camada y tasa individual de crecimiento durante unos años y, posteriormente para tasa individual de crecimiento. Después de unos años sin ser seleccionada, está siendo seleccionada actualmente para eficiencia alimentaria durante el engorde empleando distintas estrategias de selección. La selección se ha realizado mediante generaciones solapadas y evitando apareamientos entre individuos con abuelos en común (Piles et al., 2017).

Se genotiparon N = 490 animales con el chip de 200K de Affymetrix. El control de calidad se realizó con el programa PLINK v1.90 (Purcell et al., 2007). Después del filtrado los análisis se realizaron con 114.604 SNPs autosómicos.

Como información fenotípica se empleó la ganancia de peso diaria promedio (*ADG*) en dos regímenes alimenticios: a voluntad (*ADGRegV*, N = 228) y con restricción (*ADGRegR*, N = 207). Esta fue medida sólo en los gazapos (435 de los 490 individuos).

Estimas de consanguinidad basadas en información genealógica

Las estimas de consanguinidad basadas en pedigrí (*Fped*) se obtuvieron mediante el programa inbupgf90 (Aguilar y Misztal, 2008). Se dividió la genealogía en tres períodos de 20 años y, siguiendo a Ragab et al. (2015), se estimó la consanguinidad basada en pedigrí antigua (*FpedA*), intermedia (*FpedI*) y reciente (*FpedR*).

Estimas de consanguinidad basadas en información genómica
Las estimas genómicas de consanguinidad basadas en segmentos de homocigotos (*Froh*) se obtuvieron con el programa PLINK v1.90 (Purcell et al., 2007). Los criterios para definir un segmento de homocigotos fueron: (1) tener un mínimo de 100 SNPs, (2) tener una densidad mínima de 1 SNP cada 50 Kb, (3) que la distancia máxima entre 2 SNPs homocigotos en un segmento fuese 1 Mb, (4) que hubiese un máximo de 5 genotipos faltantes, y (5) un máximo de un genotipo heterocigoto en un segmento en particular. La longitud mínima que constituyó un segmento de homocigotos se estableció en > 1,25 y < 2,5, > 2,5 y < 10, y > 10 Mb para reflejar consanguinidad antigua (*FrohA*), intermedia (*FrohI*) y reciente (*FrohR*), respectivamente (Curik et al., 2014).

Se obtuvieron estimas genómicas de consanguinidad a partir de la proposición de VanRaden (2008) (*Fvan*). También se calcularon la proporción de genotipos homocigotos (*Fsnp*) y la proporción de SNPs homocigotos para el alelo con menor frecuencia (*PHoMA*).

Finalmente, se calculó la eficiencia de compresión (*CE*) como *CE*=(*Sa*-*Sd*)/*Sa*, donde *Sa* y *Sd* representan el tamaño del fichero de genotipos en bytes antes y después de la compresión, respectivamente (Hudson et al., 2014).

Identificación de asociaciones significativas y reconstrucción de redes
Se emplearon los coeficientes de correlación parcial de primer orden combinados con una aproximación de la teoría de la información (Reverter y Chan, 2008) para identificar asociaciones significativas entre los distintos coeficientes de consanguinidad y fenotipos. Estas correlaciones se calcularon con el programa PCIT (Watson-Haigh et al., 2010). Una vez identificadas las asociaciones significativas, el análisis de redes complejas y su visualización se realizó mediante el programa Cytoscape 2.8.3 (Shannon et al., 2003).

RESULTADOS Y DISCUSIÓN

La Figura 1.a muestra las asociaciones significativas entre las distintas estimas de consanguinidad. *FrohR* es el nodo más importante ya que tiene 6 enlaces, de los cuales uno de ellos es negativo con *FrohA*. *CE* es un nodo balanceado porque tiene 2 conexiones positivas y 2 negativas. *Fvan* es la que está menos relacionada porque solamente está conectada con *PHoMA* indicando que Fvan tiene que ver con individuos que comparten alelos poco frecuentes. Finalmente, *FrohI* no aporta ninguna información útil que no den las otras variables ya que no presenta conexión con ninguna otra variable.

En la Figura 1.b se pueden observar las correlaciones significativas entre las distintas consanguinidades evaluadas y *ADGRegV*. Se sigue manteniendo que la consanguinidad más importante es *FrohR* porque sigue teniendo 6 conexiones, y sigue manteniendo la asociación negativa con *FrohA* indicando quizás el posible detrimento y beneficio que pudieran ejercer *FrohR* y *FrohA*, respectivamente, sobre los caracteres productivos. *FrohI* se correlaciona positivamente con *FpedI*, y ésta negativamente con *ADGRegV*. Sin embargo, tanto *FrohI* como *ADGRegV* tienen solamente una conexión, al igual que *Fvan*.

La Figura 1.c muestra las asociaciones significativas entre las distintas estimas de consanguinidad y *ADGRegR*. *Fvan* sigue mostrando solamente una conexión. *FrohR* sigue manteniendo el número máximo de conexiones, y la correlación negativa con *FrohA* Es importante resaltar que *FrohA* tiene una correlación negativa con *ADGRegR*. Además, *FpedI* tiene correlación positiva con *ADGRegR* y negativa con *ADGRegV* (Figura 1.b) pero ambas son muy bajas.

En resumen, los resultados indican que el algoritmo PCIT permite inferir conexiones significativas entre estimas de consanguinidad y fenotipos y enfatiza la importancia de *FrohR* frente a otras estimas.

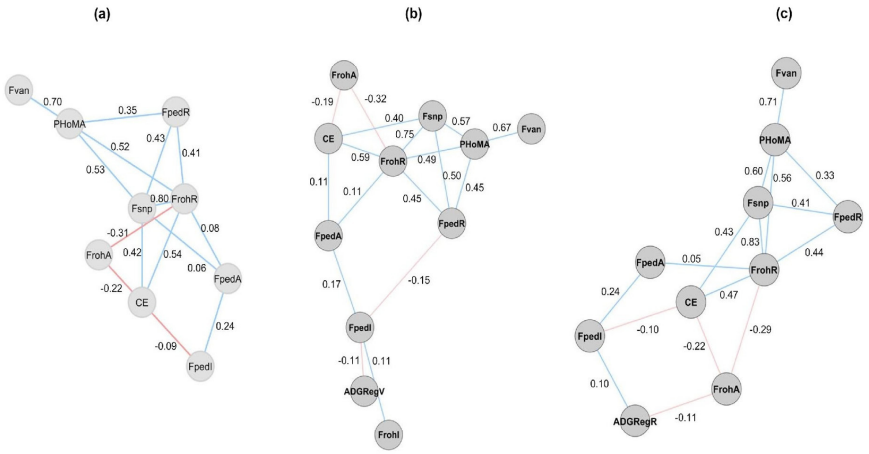


Figura 1. *Correlaciones parciales entre las distintas estimas de consanguinidad (FpedA, FpedI, FpedR, FrohA, FrohI, FrohR, Fvan, Fsnp, PHoMA y CE)* ***(a)****, incluyendo la ganancia de peso diaria promedio en los conejos alimentados a voluntad (ADGRegV)* ***(b)****, y con restricción (ADGRegR)* ***(c)****. Los enlaces en azul muestran las correlaciones positivas y en rojo las negativas.*

REFERENCIAS BIBLIOGRÁFICAS

• Aguilar, I. & Misztal, I. 2008. J. Dairy Sci. 91: 1669-1672. • Curik, I. et al. 2014. Livest. Sci. 166: 26-34. • Falconer, D.S. & Mackay, T.F.C. 1996. Harlow: Longman Scientific & Technical.• Hudson, N.J. et al. 2014. BMC Bioinformatics 15: 66.• Piles, M. et al. 2017. Genet. Sel. Evol. 49: 58. • Purcell, S. et al. 2007. Am. J. Hum. Genet. 81(3):559-575. • Ragab, M. et al. 2015. J. Anim. Breed. Genet. 132: 68-73.• Reverter, A. & Chan, E.K.F. 2008. Bioinformatics 24: 2491-2497. • Rodríguez-Ramilo, S.T. et al. 2019. J. Dairy Sci. En prensa. • Shannon, P. et al. 2003. Genome Res. 13(11): 2498-2504. • VanRaden, P.M. 2008. J. Dairy Sci. 91:4414-4423. • Watson-Haigh, N.S. et al. 2010. Bioinformatics 26: 411-413.

AGRADECIMIENTOS

Este trabajo ha sido financiado por los proyectos Feed-a-Gene (EU 633531), ARDI-POCTEFA (EFA 208/16), SMARTER (EU 772787), INRA (GDivSelGen MP-P10396) y el INIA (RTA2011-00064-00-00).

APLICATION OF PARTIAL CORRELATIONS AND INFORMATION THEORY TO ESTMATES OF INBREEDING NETWORKS IN RABBITS

ABSTRACT: This study aims to implement partial correlations and information theory to assess the relation between different estimators of inbreeding and phenotypes using a selected population of rabbits. Pedigree data and genomic information from a 200K chip were available. After filtering, the data set constituted 490 animals genotyped for 114,604 autosomal SNPs. The evaluated methodology allows inferring meaningful associations between estimates of inbreeding appropriately.

Keywords: inbreeding, average genetic gain, rabbit, partial correlation, network.