

MODELOS DE EVALUACIÓN GENÓMICA CON DOMINANCIA DIRECCIONAL

Varona, L.¹, Legarra², A., Herring³, W. y Vitezica², Z.G.

¹ Unidad de Genética Cuantitativa y Mejora Animal. Universidad de Zaragoza. 50013 ZARAGOZA. ²INRA, UMR GenPhySE, CS52627, 31326 Castanet Tolosan, Francia. ³ PIC North America, 100 Bluegrass Commons Blvd., Suite 2200, Hendersonville, TN, 37075, USA
lvarona@unizar.es

INTRODUCCIÓN

La selección genómica (Meuwissen et al., 2001) se ha constituido como una herramienta de indudable interés en mejora genética animal. Las herramientas habituales para selección genómica consideran los efectos de sustitución de los marcadores SNP, que son capaces de capturar gran parte de los efectos de dominancia e interacciones de mayor orden (Hill et al., 2008). Sin embargo, la estimación de los efectos de dominancia es relevante porque permite diseñar estrategias de diseño de apareamientos (Toro y Varona, 2010).

Los procedimientos de evaluación genómica exigen un proceso de regularización de los efectos asociados a los marcadores, como por ejemplo la asunción de una distribución a priori gaussiana (Su et al., 2012; Vitezica et al., 2013). Esta distribución implica una distribución simétrica de las estimaciones posteriores de los efectos de dominancia. Sin embargo, la teoría clásica de la genética cuantitativa argumenta que la depresión endogámica y la heterosis se deben a la presencia de dominancia direccional (i.e. un mayor porcentaje de efectos de dominancia positivos que negativos).

Esta limitación puede solventarse al menos de dos modos: 1) bajo la asunción de un efecto de dominancia medio distinto de cero (Xiang et al., 2016) o 2) con la utilización de distribuciones asimétricas (Sahu et al., 2003) para la regularización de los efectos de dominancia. Por lo tanto, el objetivo de este trabajo es comparar los resultados de las dos aproximaciones en una base de datos de porcino.

MATERIAL Y MÉTODOS

Los datos para este trabajo fueron proporcionados por *Genus plc* (Hendersonville, TN, USA). Los genotipos para todas las cerdas fueron generados mediante el *Illumina PorcineSNP60 BeadChip* (Illumina, San Diego, CA). Después de un control de calidad, se utilizaron 37011 SNP en las poblaciones 1 y 2, respectivamente. El número de cerdas incluidas en el análisis fueron 2612 y con un total de 11581 registros del carácter nacidos vivos, respectivamente. El tamaño medio de camada fue de $12,40 \pm 3,00$.

En primer lugar, se define el modelo completo que incluye las dos aproximaciones para la inclusión de dominancia direccional. Posteriormente, fijando a cero alguno de los parámetros se generarán los modelos reducidos. Por lo tanto, el modelo completo (FULL) fue:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{f}b + \mathbf{X}\mathbf{t} + \mathbf{W}\mathbf{r} + \mathbf{Q}\mathbf{c} + \mathbf{Z}\mathbf{a} + \mathbf{K}\mathbf{d} + \mathbf{e}$$

donde \mathbf{y} es el vector de registros fenotípicos, μ es la media general, b es el parámetro de la depresión por consanguinidad, \mathbf{t} es un vector con el orden de parto, \mathbf{r} es un vector del efecto granja-año-estación (3163 –población 1- y 4293 –población 2- niveles), \mathbf{c} es el vector de efectos ambientales permanentes (3631 y 2612 niveles), \mathbf{a} y \mathbf{d} son los vectores de efectos aditivos y dominantes y \mathbf{e} es el vector de residuos. Además, \mathbf{f} es el vector de la homocigosidad media individual y \mathbf{X} , \mathbf{W} , \mathbf{Q} , \mathbf{Z} y \mathbf{K} son las matrices de incidencia.

Las distribuciones a priori para b y cada nivel de \mathbf{t} son uniformes. Para cada nivel de \mathbf{r} , \mathbf{c} and \mathbf{a} son distribuciones normales univariantes y para cada nivel de \mathbf{d} son distribuciones normales asimétricas (Sahu et al., 2003):

$$p(\mathbf{d}|\sigma_d^2, \lambda_d) = \prod_{i=1}^{N_{SNP}} \frac{2}{\sqrt{\sigma_d^2 + \lambda_d^2}} \phi\left(\frac{d_i}{\sqrt{\sigma_d^2 + \lambda_d^2}}\right) \Phi\left(\frac{\lambda_d}{\sigma_d} \frac{d_i}{\sqrt{\sigma_d^2 + \lambda_d^2}}\right)$$

donde σ_d^2 y λ_d son la varianza de los efectos dominantes y el parámetro de asimetría, respectivamente. Finalmente, las distribuciones a priori de los componentes de varianza y del parámetro de asimetría fueron uniformes. Partiendo de este modelo completo, se describieron los siguientes modelos reducidos:

- Modelo SC: En este modelo el parámetro de asimetría se fijó a cero. Por lo tanto, el modelo es equivalente al propuesto por Xiang et al. (2016).
- Modelo AN: En este modelo el parámetro de la depresión por consanguinidad (b) se fijó a cero.
- Modelo SN: Se fijan a cero tanto el parámetro de asimetría como el de depresión por consanguinidad. Este modelo es equivalente al propuesto por Su et al. (2012).

Además de los parámetros incluidos en el modelo, se estimaron la depresión endogámica esperada para una consanguinidad de $F=0.10$ (I_D) y las varianzas aditivas y de dominancia y sus ratios sobre la varianza total (h^2 y d^2). Para ello se utilizaron los valores de los efectos aditivos y dominantes en cada iteración junto con las frecuencias alélicas brutas obtenidas en la población.

La implementación se realizó mediante muestreo de Gibbs (Gelfand y Smith, 1990) con un total de 100.000 iteraciones después de descartar las primeras 25.000. El muestro de los elementos de \mathbf{d} , exige la definición de una variable auxiliar (Varona et al., 2008) distribuida mediante una distribución normal truncada.

RESULTADOS Y DISCUSIÓN

Los resultados se presentan en la Tabla 1. En ella se puede observar como existe una evidencia clara de dominancia direccional, como se muestra en los parámetros de asimetría bajo el modelo AN y de depresión por consanguinidad bajo el modelo SC. Los resultados de estos dos parámetros bajo el modelo completo (FULL) mostraron mucha mayor dispersión, indicando cierto grado de confusión estadística entre ellos. De hecho, la correlación posterior entre las muestras del muestro de Gibbs fue de -0,90.

Pese a todo, la estimación de la potencial depresión endogámica fue similar bajo los modelos AN, SC y FULL, mientras que el modelo SN proporciono estimaciones inferiores, ya que no considera la dominancia direccional y por lo tanto infraestima la depresión endogámica.

Las estimaciones de los componentes de la varianza también fueron diferentes entre modelos. Si se toma el modelo SN como punto de partida, la inclusión de la covariada con la homocigosidad media redujo la varianza de los efectos dominantes, mientras que la asunción de una distribución asimétrica la incrementó. La causa de este fenómeno se debe a la naturaleza de las distribuciones asimétricas que se asumieron (Sahu et al., 2003). En este sentido, en un futuro se pretende explorar distribuciones asimétricas alternativas. Esta relación entre los estimadores de la varianza de los efectos de dominancia se contrarrestó con las diferencias en el componente de varianza permanente individual, que fue mayor con el modelo SC y más reducido con el modelo AN.

REFERENCIAS BIBLIOGRÁFICAS

- Gelfand, A. E. & Smith, A. F. M. 1990. Journal of American Statistical Association 85:398-409.
- Hill, W. G., Goddard, M. E. & Visscher, P. M. 2008. Plos Genetics 4:e1000008
- Meuwissen, T. H. E., Hayes, B. J. & Goddard, M. E. 2001. Genetics 157:1819-1829
- Sahu, S. K., Dey, D. K. & Branco, M. D. 2003. The Canadian Journal of Statistics 31:129-150
- Su, G., Christensen, O. F., Ostersen, T., Henryon, M. & Lund, M. S. 2012. Plos ONE 7:e45293
- Toro, M. A. & Varona, L. 2010. Genet. Sel. Evol. 43:33
- Varona, L., Ibanez-Escriche, N., Quintanilla, R., Noguera, J. L. & Casellas, J. 2008. Genetics Research 90:179-190
- Vitezica, Z. G., Varona, L. & Legarra A. 2013. Genetics 195:1223-1230.
- Xiang, T., Christensen, O. F., Vitezica, Z. G. & Legarra, A. 2016. Genet. Sel. Evol. 48:92

Agradecimientos: Este trabajo ha sido financiado por el proyecto *INRA-SELGEN EpiSel* (AL y ZV) y el proyecto CGL2016-80155 del Ministerio de Economía y Competitividad (LV). Los autores agradecen a Genus plc (Hendersonville, TN, USA) la disponibilidad de los datos fenotípicos y genotípicos.

Tabla 1. Estimaciones posteriores de los componentes de varianza, parámetro de asimetría, depresión endogámica y ratios de varianza aditiva y de dominancia.

	MODELO			
	SN	AN	SC	FULL
σ_a^2 ($\times 10^{-4}$)	0,653 (0,079)	0,601 (0,082)	0,718 (0,090)	0,635 (0,087)
σ_d^2 ($\times 10^{-4}$)	0,488 (0,151)	0,922 (0,174)	0,254 (0,156)	0,857 (0,357)
σ_c^2	0,294 (0,058)	0,295 (0,060)	0,298 (0,063)	0,296 (0,060)
σ_r^2	0,560 (0,115)	0,366 (0,110)	0,654 (0,119)	0,390 (0,158)
b	-	-	6,540 (2,200)	0,551 (6,311)
λ_d ($\times 10^{-3}$)	-	0,263 (0,094)	-	0,226 (0,224)
σ_e^2	6,635 (0,109)	6,636 (0,109)	6,629 (0,111)	6,634 (0,109)
I_D	0,087 (0,047)	0,299 (0,078)	0,247 (0,074)	0,277 (0,082)
V_A	1,108 (0,121)	1,171 (0,168)	1,159 (0,124)	1,189 (0,166)
V_D	0,759 (0,234)	1,456 (0,278)	0,452 (0,244)	1,400 (0,489)
h^2	0,116 (0,012)	0,118 (0,015)	0,126 (0,012)	0,121 (0,014)
d^2	0,081 (0,023)	0,146 (0,024)	0,049 (0,025)	0,134 (0,043)

σ_a^2 y σ_d^2 son las varianzas de los efectos aditivos y dominantes de los SNP, σ_c^2 es la varianza de los efectos granja-año-estación, σ_r^2 es la varianza de los efectos individuales permanentes, σ_e^2 es la varianza residual, b es la covariada con la heterocigosidad individual, λ_d es el parámetro de asimetría para los efectos dominantes, I_D es la depresión endogámica, V_A y V_D son la varianzas aditivas y de dominancia, y h^2 y d^2 son la heredabilidad y el cociente de varianza de dominancia.

GENOMIC EVALUATION WITH DIRECTIONAL DOMINANCE

ABSTRACT: Most of the statistical procedures for genomic evaluation are associated with the statistical problem of large p, small n, because the number of parameters to estimate is larger than the number of data. The most frequent approach to solve this problem is the use to some kind of regularization of SNP markers effects. Most of these procedures involve symmetric distributions, which is a reasonable prior for the additive genetic effects. However, symmetric prior distributions are in conflict with the traditional approaches for heterosis and inbreeding depression, when they are assumed for dominant effects. Two possible alternatives can be outlined: 1) the assumption of an a priori mean of dominance effects different from zero and 2) the use skewed distributions for regularization of dominance effects. Here we used both alternatives with a pig database. Both approaches provide clear evidence of directional dominance and lead to similar estimates of inbreeding depression. However, the distribution of the variance between the dominance effects and the permanent environmental effects were clearly different between models.

Keywords: Genomic selection, directional dominance, pig, litter size