ANALISIS DE QTLs EN CARACTERES LONGITUDINALES.

L. Varona, L. Gómez-Raya, W. Rauw, J. L. Noguera.
Area de Producció Animal. Centre UdL-IRTA.
Av/ Rovira Roure 177. 25198. LLEIDA.

INTRODUCCIÓN

Los caracteres expresados a lo largo del tiempo, como el crecimiento, la lactación o el consumo, están cobrando interés en los últimos años. Los procedimientos para su análisis se basan o bien en un modelo jerárquico de funciones de producción (Varona et al., 1997, 1998, Blasco y Varona, 1999, Toro et al., 1999, Rekaya et al., 2000), o en la utilización de modelos de regresión aleatoria (Jamrozik et al., 1997; Misztal et al., 2000). Por otra parte, el desarrollo de las técnicas de genética molecular ha permitido disponer de un mayor número de marcadores moleculares que se han utilizado para detectar genes de expresión cuantitativa (QTLs) en las distintas especies de interés zootécnico (Riquet et al., 1999; Rothschild, 2000). El objetivo de este trabajo es describir un procedimiento para la detección de QTL con efecto en caracteres longitudinales en el contexto del análisis jerárquico de funciones de producción. El procedimiento se comprobó mediante simulación.

SIMULACIÓN

Se han simulado datos de crecimiento procedentes de 400 individuos de una población F2, constituida por el cruce de dos líneas homocigotas tanto para el QTL (QQ *vs.* qq) como para dos marcadores flanqueantes (SS *vs.* ss y MM *vs.* mm). La fracción de recombinación entre los marcadores y el QTL fueron del 10%. Por cada individuo se han obtenido 5 registros de peso a los 80, 90, 100, 110 y 120 días de edad. Estos datos se simularon asumiendo una función lineal de crecimiento a partir de la siguiente expresión:

$$y_{ij} = a_i + b_i x_{ij} + r_{ij}$$

donde y_{ij} es el registro j del individuo i, a_i es el peso a 100 días de edad del individuos i, b_i es la velocidad de crecimiento del individuo i, x_{ij} es del día de control de la pesada j del individuo i, y r_{ij} es el residuo asociado, que se obtuvo de una distribución normal con media 0 y varianza 1.

Los parámetros de la función lineal de crecimiento de cada individuos se obtuvieron a partir de las siguientes expresiones:

$$a_i = \mu(a) + 1(G = QQ)ad(a) + 1(G = Qq)d(a) - 1(G = qq)ad(a) + e(a)_i$$
$$b_i = \mu(b) + 1(G = QQ)ad(b) + 1(G = Qq)d(b) - 1(G = qq)ad(b) + e(b)_i$$

Donde $\mu(a)$=100, $ad(a)$=0, $d(a)$=0, $\mu(b)$=0.6, $ad(b)$=-0.02, $d(b)$=-0.02. Las medias poblacionales del peso a los 100 dias y la velocidad de crecimiento son $\mu(a)$ y $\mu(b)$, $ad(a)$ y $ad(b)$ son los efectos aditivos, $d(a)$ y $d(b)$ son los efectos dominantes, respectivamente, y $e(a)$ y $e(b)$ son los residuos que se extrajeron de distribuciones normales univariantes con media cero y varianzas 25 y 9 x 10^{-4} respectivamente.

METODOLOGÍA

Se ha realizado un análisis bayesiano jerárquico a partir de la distribución conjunta de datos y parámetros. La verosimilitud de los datos de la función de producción fue:

$$f(\mathbf{y} \mid \mathbf{a}, \mathbf{b}, \sigma_r^2) \propto \prod_{i=1}^{nan} \prod_{j=1}^{ni} \frac{1}{\sigma_r} \exp \frac{-\left(y_{ij} - a_i - b_i x_{ij}\right)^2}{2\sigma_r^2}$$

Donde *nan* es el número de individuos con datos y n_i es el número de datos del individuo i.

Las distribuciones a priori de los parámetros de la función de producción fueron:

$$f(\mathbf{a} \mid \mu(a), ad(a), d(a), \sigma_e^2(a))$$
$$\propto \prod_{i=1}^{nan} \frac{1}{\sigma_e(a)} \exp \frac{-\left(a_i - \mu(a) - pr(QQ)ad(a) - pr(Qq)d(a) + pr(qq)ad(a)\right)^2}{2\sigma_e^2(a)}$$

$$f(\mathbf{b} \mid \mu(b), ad(b), d(b), \sigma_e^2(b))$$
$$\propto \prod_{i=1}^{nan} \frac{1}{\sigma_e(b)} \exp \frac{-\left(b_i - \mu(b) - pr(QQ)ad(b) - pr(Qq)d(b) + pr(qq)ad(b)\right)^2}{2\sigma_e^2(b)}$$

donde $\mu(a)$ y $\mu(b)$ son las medias poblacionales, $pr(QQ)$, $pr(Qq)$ y $pr(qq)$ son las probabilidades de poseer el genotipo QQ, Qq o qq condicionado a los marcadores, las distancias entre éstos y la posición del genoma considerada. Finalmente, $\sigma^2_e(a)$ y $\sigma^2_e(b)$ son las varianzas residuales para ambos parámetros.

Las distribuciones a priori de los componentes de varianza y los efectos aditivo y dominante fueron consideradas planas. Por lo tanto la distribución conjunta de datos y parámetros es:

$$f(\mathbf{a}, \mathbf{b}, \sigma_r^2, \mu(a), ad(a), d(a), \sigma_e^2(a), \mu(a), ad(a), d(a), \sigma_e^2(a) \mid \mathbf{y}) \propto$$
$$f(\mathbf{y} \mid \mathbf{a}, \mathbf{b}, \sigma_r^2) f(\mathbf{a} \mid \mu(a), ad(a), d(a), \sigma_e^2(a)) f(\mathbf{b} \mid \mu(b), ad(b), d(b), \sigma_e^2(b))$$
$$f(\sigma_r^2) f(\mu(a)) f(ad(a)), f(d(a)) f(\sigma_e^2(a)) f(\mu(b)) f(ad(b)), f(d(b)) f(\sigma_e^2(b))$$

El análisis se ha realizado utilizando Muestro de Gibbs (Geman y Geman, 1984; Gelfand y Smith, 1990) combinado con Data Augmentation (Tanner y Wong, 1987). Se utilizó una cadena única de 110.000 iteraciones descartando las 10.000 primeras.

RESULTADOS

El procedimiento se aplicó a los datos simulados, obteniendo los resultados presentados en la figura 1, que, como se puede observar, son coherentes con los parámetros de simulación. Las distribuciones posteriores de los efectos aditivos y dominantes para peso incluyen al 0 en sus intervalos de mayor densidad posterior (HPD) al 95%. Por el contrario, para la velocidad de crecimiento la probabilidad posterior de ser mayor que 0, tanto para el efecto aditivo como para el dominante, fue prácticamente nula. Se ha presentado un caso extremadamente simple para comprobar la capacidad del procedimiento para recuperar los parámetros simulados. Pese a todo, es susceptible de ser adaptado a situaciones más realistas en varias vertientes:

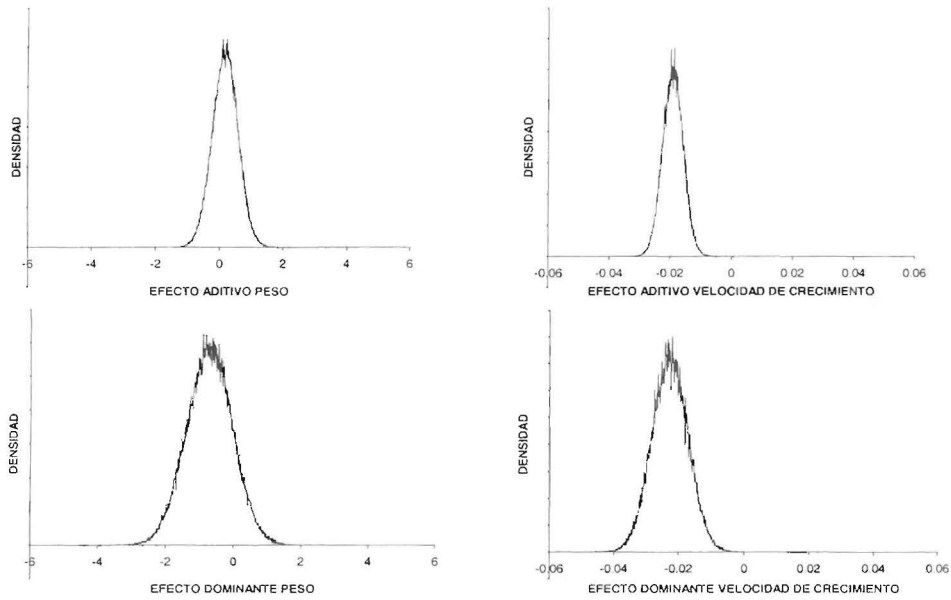
a) Se puede modificar la verosimilitud para introducir una función de crecimiento que describa más adecuadamente el crecimiento de los individuos (Gomperz, Von Bertalanffy, etc).

b) Se pueden modelizar los residuos mediante una función que dependa de la edad del individuo, y, además, se pueden introducir parámetros de autocorrelación entre residuos.

c) Las recombinaciones pueden ser consideradas como variables estocásticas, en lugar de asumir como conocidas las probabilidades condicionales de las configuraciones genéticas. En diseños sencillos, como el simulado, el resultado será equivalente, pero esta modificación hará factible el análisis de poblaciones más complejas.

d) Es posible incluir efectos atribuibles al efecto genético poligénico y al QTL en las distribuciones a priori de los parámetros de las funciones de crecimiento, permitiendo el análisis de poblaciones abiertas en el contexto de un modelo de componentes de varianza.

Figura 1. Distribuciones posteriores marginales de efectos aditivos y dominantes.



REFERENCIAS

Blasco A., Varona L. 1999. *ITEA*. 95:131-142.

Gelfand A., Smith A. F. M. 1990. *J.A.S.A.* 85:398-409.

Geman S., Geman D. 1984. *IEEE Trans. Pattn. Anal. Mach. Intell.* 6:721-741.

Jamrokzik J., Schaeffer L.R. 1997. *J. Dairy Sci.* 80:762-770.

Mignon-Grasteau S., Piles M., Varona L., de Rochambeau H., Poivey J.P., Blasco A., Beaumont C. *J. Anim. Sci.* 2515-2524.

Misztal I., Strabel T., Jamrozik J., Mantysaari E.A., Meuwissen T.H.E. 2000. *J. Dairy Sci.* 83:1125-1134.

Rekaya R., Carabaño M.J., Toro M.A. 2000. *J. Dairy Sci.* 2691-2701.

Riquet J., Coppieters W., Cambisano N., Arranz J.J., Berzi P., Davis S.K., Grisart B., Farnir F., Karim L., Mni M., Simon P., Taylor J. F., Vanmasshoven P., Wagenaar D., Womanck J. E., Georges M. 1999. *Proc. Natl. Acad. Sci. USA*. 84:2363-2367.

Rothschild, M. 2000. *ITEA* (en prensa).

Tanner M. A., Wong, W. H. 1987. *J.A.S.A.* 82:528-550.

Varona L., Moreno C., García-Cortes L.A., Altarriba J. 1997. *Livest. Prod. Sci.* 47:201-209

Varona L., Moreno C., García-Cortes L.A., Altarriba J. 1998. *J. Dairy Sci.* 1469-1478.