

Selección Genómica*

Cantet¹, R.J.C., Gualdrón Duarte¹, J.L. y Munilla Leguizamón¹, S.

Departamento de Producción Animal, Facultad de Agronomía, UBA
Consejo Nacional de Investigaciones Científicas y Técnicas

Introducción

La evaluación genética animal tiene como objeto la predicción del mérito genético individual o valor de cría (en inglés, *breeding value*, BV). En el modelo clásico de la genética cuantitativa, los BVs de los caracteres de importancia económica son controlados por un número muy grande (más formalmente, tendiendo a infinito) de genes: el *modelo infinitesimal*. Estos genes actúan en forma aditiva sin ligamiento entre ellos, cada uno con un efecto pequeño sobre la expresión fenotípica del carácter (Bulmer, 1985). En consecuencia, y hasta hoy, la selección en animales se realizó asumiendo una distribución normal para el BV resultante de una suma que tiende a infinito de variables aleatorias cada una de valor muy pequeño. En la actualidad cada uno de los genes o fracciones del genoma que afectan los caracteres productivos bajo un modelo infinitesimal es conocido como QTL (en inglés *Quantitative Trait Locus*). Si bien grande, el máximo número de genes en un genoma como por ejemplo el humano, se estima en el rango entre 20.000 a 25.000 (International Human Consortium, 2004), ciertamente un número finito. Los avances en genética molecular han mostrado que las características cuantitativas están afectadas por unos pocos genes con gran impacto y muchos con escaso, de modo tal que la distribución del efecto de los QTLs es aproximadamente exponencial (Otto y Jones, 2000).

Se han localizado un gran número de QTLs para caracteres productivos en especies animales y algunos de ellos han sido "mapeados" a distancias inferiores a un centi-

Morgan (1 cM). La información fue utilizada para la selección asistida por marcadores moleculares (MAS). Esta metodología consiste en la predicción del BV total sumando la contribución poligénica más los efectos individuales de los QTLs que se ubican en la cercanía de un marcador molecular (Fernando y Grossman, 1989). Un ejemplo de aplicación es el esquema francés de MAS en ganado lechero que se realiza desde 2001 (Boichard et al, 2003; Guillaume et al, 2008). El esquema MAS se efectúa en dos etapas: primero se hace una búsqueda a lo largo del genoma intentando detectar QTLs que afecten significativamente el carácter bajo estudio; el paso siguiente es producir las predicciones del BV sumando a los efectos de los QTLs marcados el BV poligénico restante debido a los QTLs no detectados en la primera etapa. Lamentablemente, solo una proporción limitada de la varianza aditiva total del carácter suele ser detectada por los marcadores. Alternativamente, Meuwissen et al. (2001) sugirieron seleccionar por todos los QTLs que afectan el carácter, prediciendo el BV mediante un gran número de marcadores moleculares distribuidos regularmente a lo largo del genoma. Esta metodología de predicción del BV es conocida como *selección genómica*. Para la detección de QTLs en MAS se realizan test de hipótesis altamente restrictivos que involucran correcciones por hipótesis múltiples dada la presencia de muchos marcadores o posiciones en el genoma, de modo de reducir al máximo el número de falsos positivos. Pero en la selección genómica estas pruebas de significancia son omitidas, estimándose simultáneamente

*Conferencia presentada durante el transcurso del 31° Congreso Argentino de Producción Animal, 15 al 17 de octubre de 2008, Potrero de los Funes, San Luis.

1. Dpto. Producción Animal, Fac.Agron., UBA. rcantet@mail.agro.uba.ar

el efecto de todos los marcadores o posiciones en los cromosomas (Meuwissen et al, 2001).

Muy recientemente se ha podido alcanzar la elevada densidad en el número de marcadores dispersos a lo largo del genoma que requiere la selección genómica debido al uso masivo de los marcadores SNP (siglas en inglés de *Single Nucleotide Polymorphism*). Estos consisten en una variación en la secuencia del ADN debida a una diferencia en un nucleótido (sea adenina, timina, guanina o citosina). En diciembre de 2007, la empresa Illumina, de origen estadounidense sacó al mercado un chip que contiene información de casi 58.000 SNPs, con un costo (a julio de 2008) de 245 u\$d/animal. Con este costo es posible predecir el BV del animal usando todos los marcadores disponibles mediante la selección genómica (Meuwissen et al, 2001).

Predicción del valor de cría genómico

El cálculo directo de los efectos de los QTLs marcados genera información (Goddard, 2008) que está presente sólo en el residuo de segregación mendeliana del modelo aditivo para el BV (Bulmer, 1985). Cada padre pasa la mitad de sus genes a cada miembro de su progenie, si bien no transmite la misma mitad a cada hijo debido a la segregación aleatoria de segmentos cromosómicos durante la meiosis. Esta segregación es detectada por marcadores moleculares en alta densidad cercanos a cada QTL (Goddard, 2008). El uso de información molecular para controlar la información del residuo mendeliano queda en evidencia al comparar las *relaciones aditivas* incondicionales con aquellas que son obtenidas condicionales a los marcadores moleculares presentes (Fernando y Grossman, 1989; Abney, 2008). Las relaciones aditivas incondicionales son los valores esperados del número de pares de genes *idénticos por descendencia* (IBD) entre dos individuos. A nivel poblacional, las relaciones aditivas cuantifican adecuadamente la probabilidad de que dos individuos emparentados compartan regiones genómicas responsables de la variabilidad genética de un carácter. Por

ejemplo, Gagnon et al (2005) calcularon para 498 pares de hermanos enteros, empleando 1.522 marcadores microsatélites espaciados a una distancia promedio de 2.3 cM un valor promedio de genoma compartido de 0.4994 ± 0.0395 , valor cercano al 0.5 esperado. Sin embargo, la proporción de genes IBD comparados entre un individuo y su hermano puede ser inferior o superior a la esperada debido a la variabilidad en la transmisión del genoma de un padre a cada hijo. Esta información sólo es detectable mediante el uso de marcadores en alta densidad a lo largo de todo el genoma.

Meuwissen et al (2001) propusieron distintos métodos de estimación de los efectos de los QTLs, si bien el más promisorio resultó ser el denominado Bayes B. En esencia, es un método bayesiano que asume aleatorios a los efectos de los QTLs y asigna una varianza distinta a cada uno de ellos. La característica especial es la distribución *a priori* de dichas varianzas, asignando una probabilidad elevada al valor 0 del efectos del QTL; el resto de esta distribución mezcla es una χ^2 (chi-cuadrado invertida). La lógica de esta distribución es que la mayoría de los marcadores tienen una gran probabilidad de no estar asociados a QTL alguno.

Comparación entre la selección genómica y la convencional

VanRaden (2008) comparó la exactitud al pre-seleccionar toros lecheros para la prueba de progenie empleando el promedio de los valores de cría paterno y materno predichos mediante la evaluación convencional (PA), con el valor de cría genómico predicho (GEBV). La población empleada para estimar los efectos de los QTLs consistió en 3.576 toros Holstein nacidos entre 1952 y 1998. La población de prueba fueron 1.759 toros jóvenes nacidos entre 1999 y 2002. Los marcadores de ADN se obtuvieron a partir del Illumina Bovine-50K SNP Chip que contiene 58.000 SNPs. Sin embargo, dado que hubo casos de falta de resolución o incompatibilidad de la información entre padres e hijos, sólo se usaron 38.416 SNPs. Se predijeron los des-

víos de producción de las hijas (DYD) de los toros jóvenes en 2008. Los DYD consisten en el promedio de los datos de producción de las hijas de un toro, previamente corregidos por los efectos ambientales sistemáticos. Los PA se obtuvieron de la evaluación 2003 para 27 caracteres: 5 de producción, 5 de salud, 16 de conformación y el índice económico denominado Mérito Neto (*Net merit*). La exactitud de la prueba fue definida como la correlación entre el valor de cría verdadero (aproximado por el DYD 2008) y el predicho, sea mediante PA 2003 o por los GEBVs calculados con las estimaciones de los efectos de los SNPs obtenidos en la población de entrenamiento y proyectados en base a los genotipos de los toros jóvenes. Los resultados se muestran en el Cuadro 1.

Nótese que los GEBVs constituyeron un criterio de selección mucho más exacto que los valores PA. Debe indicarse que estos últimos no contienen información sobre el residuo de segregación mendeliano de los toros jóvenes, hecho que si ocurre con los GEBVs. La predicción genómica para los 26 caracteres fue significativamente mayor ($p < 0,0001$) que para el PA, además la ganancia en exactitud equivalió en promedio a que un toro joven tuviese 11 hijas con información fenotípica. Este aumento en precisión de selección permite disminuir el intervalo generacional dado que los animales pueden seleccionarse a una edad significativamente inferior

a la que actualmente se lo hace (5-6 años para la selección inicial). Asimismo, el hecho de tener que evaluar en la prueba de progenie unos pocos machos (20, según Chesnais et al, 2008) por camada, comparado con los 100 a 500 toros por año - dependiendo del poderío económico del centro de inseminación artificial – evaluados en un esquema convencional, conlleva una fuerte disminución de los costos de mantenimiento de toros y del esquema de selección en general.

Otro aspecto importante de la selección genómica es como impacta el número de SNPs a la exactitud en el carácter (Cuadro 2). Obsérvese que a mayor número de SNPs considerados en el análisis, mayor es la exactitud de predicción obtenida. Nótese además que los aumentos en exactitud dependieron del carácter analizado. Aparentemente al aumentar la densidad de SNPs, se acortan los intervalos entre los marcadores y los QTLs y el desequilibrio gamético (más conocido como *desequilibrio de ligamiento*, LD) es mayor. Al mismo tiempo, el incremento en el número de marcadores lleva a mayor número de parámetros en el modelo, muchos de ellos no ligados a genes. Esto provoca una disminución de la información por parámetro y la reducción en la exactitud (Cantet, 1997). Ambos procesos con efectos antagónicos interactúan, debiendo existir un número óptimo de SNPs a incluir en el modelo por encima del cuál la exactitud de predicción del GEBV disminuye.

Cuadro 1: Correlaciones entre DYD y PA o GEBV

	Tradicional: PA	Genómico: GEBV	Diferencia G - T
Mérito neto	0,3	0,53	0,23
Leche	0,35	0,58	0,23
Grasa	0,35	0,68	0,33
Proteína	0,35	0,57	0,22
Grasa %	0,35	0,78	0,43
Proteína %	0,35	0,69	0,34

Cuadro 2: Efecto del número de SNPs sobre la exactitud. 9.604(10K), 19.208 (20K), y 38.416(40K)

Carácter	Exactitud de PA	Exactitud de GEBV		
		10K	20K	40K
Merito Neto \$	0,3	0,48	0,5	0,53
Producción Leche	0,35	0,53	0,56	0,58
Producción Grasa,	0,35	0,64	0,66	0,68
Producción Proteína	0,35	0,54	0,56	0,57
Vida Útil	0,27	0,38	0,41	0,45
PCS (Mastitis)	0,3	0,45	0,47	0,51
T. Preñez Hijas	0,25	0,37	0,39	0,41

Comentarios finales

La selección genómica es la herramienta de mayor impacto sobre la eficiencia de la selección animal que el mejoramiento genético ha producido en muchos años, particularmente para el bovino de leche. Esto se debe a la disminución en el costo de evaluación de genotipos SNPs en alta densidad. La ganancia en la respuesta a la selección se debe a 1) la mayor exactitud de evaluación a edad temprana, y 2) al menor intervalo generacional, al poder seleccionar toros previamente a la prueba de progenie. Aún subsisten interrogantes en cuanto a la modelación de la varianza aditiva en el modelo del BV genómico, o a la respuesta a la selección en el largo plazo. Sin embargo, las posibilidades que abre la selección genómica impactarán la forma en que se realiza la selección de animales.

Bibliografía

- Abney, M. 2008. Identity by descent estimation and mapping of qualitative traits in large, complex pedigrees. *Genetics*. 179:1577-1590.
- Boichard, D., Grosh, C., Bourgeois, F., Cerqueira, F., Faugeras, R., Neau, A., Rupp, R., Amigues, Y., Boscher, M. and Levéziel, H. 2003. Detection of genes influencing economic traits in three French dairy cattle breeds. *Genetics Selection Evolution*. 35:77-101.
- Bulmer, M.G. 1985. The mathematical theory of quantitative genetics. Clarendon Press, Oxford, United Kingdom.
- Cantet, R.J.C. 1997. Effects of excluding a set of random effects on prediction error variance of breeding value. *J. Anim. Breed. Genet.* 114:267-274.
- Chesnais, J., Schenkel, F. and Caron, N. 2008. The next step in genomic selection: An industry perspective. ASAS-ADSA Joint meeting, Indianapolis, IN, Julio 9, 2008.
- Fernando, R. and Grossman, M. 1989. Marker assisted selection using best linear unbiased prediction. *Genetics Selection Evolution*. 21:467-477.
- Gagnon, A., Beise, J. and Vaupel, J. 2005. Genome wide identity by descent sharing among CEPH siblings. *Genet. Epidemiol.* 29:215-224.
- Goddard, M. 2008. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* (en prensa).
- Guillaume, F., Fritz, S., Boichard, D. and Druet, T. 2008. Estimation by simulation of the efficiency of the French marker-assisted selection program in dairy cattle. *Genetics Selection Evolution*. 40:91-102.
- International Human Consortium. 2004. Finishing the euchromatic sequence of the human genome. *Nature* 431:931-45.
- Meuwissen, T.H.E., Hayes, B.J. and Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157: 1819-1829.
- Otto, S. and Jones, C. 2000. Detecting the undetected: Estimating the total number of loci underlying a quantitative trait. *Genetics*. 156: 2093-2107.
- VanRaden, P.M. 2008. Reliability of genomic predictions for North American dairy bulls. ASAS-ADSA Joint meeting, Indianapolis, IN, Julio 9, 2008.