ANÁLISIS DE POTENCIA EN LA DETECCIÓN DE QTL MEDIANTE GENOTIPADO SELECTIVO Y POOLES DE ADN

J. A. Baro, C. Carleos, * J. Cañón †
CIATA-Somió, Ppdo. Asturias; *Dept. de Estadística, Univ. Oviedo; †Dpto. de Producción Animal, UCM

INTRODUCCION

Los métodos para la detección y mapeo de QTLs se basan en el análisis de la asociación entre fenotipos y alelos de marcadores. Para obtener la máxima potencia en la detección, se recurre a información obtenida de grandes esquemas de hibridación entre grupos genéticos muy separados, pero recientemente se han desarrollado métodos que permiten el empleo de poblaciones existentes. Uno de estos métodos es el genotipado selectivo unido a la técnica de genotipado por lotes, muy adecuado para la exploración de poblaciones en que se haga uso extensivo de IA (Inseminación Artificial) y MOET (Ovulación Múltiple y Transferencia de Embriones). El genotipado selectivo en familias de medio hermanos (Lander y Botstein, 1989) consiste en tomar unicamente las colas de la distribución de valores fenotípicos de medio-hermanos, y el genotipado por lotes es un procedimiento laboratorial para la obtención —a partir de una muestra con material genético de varios individuos— las frecuencias alélicas de los marcadores asociadas a cada una de las colas.

Se analiza la potencia de varios estadísticos para la detección de QTLs con diversos escenarios y métodos, empleando para ello la simulación de un sencillo esquema de segregación con un QTL y un marcador con cinco alelos equifrecuentes.

MATERIAL Y METODOS

Simulación

Los datos se obtuvieron generando por simulación familias de medio-hermanos, derivados de un progenitor común (semental) doblemente heterozigoto. En el modelo de simulación se introdujo un efecto de ruido normalmente distribuido para recoger los efectos ambientales y del fondo genético poligénico, con varianza fijada en 1.0 en aras de la simplicidad. Para el tamaño de las familias, se hicieron asunciones realistas que reflejasen las que puede presentar un esquema de IA de escala regional: 200 a 500 medio hermanos contemporáneos por semental.

El efecto aditivo del QTL se modeló con tres valores: nulo, lo que permitiría comprobar la proporción de rechazos bajo la hipótesis nula (error de tipo I) de que no hay QTL segregando en la población, 0.25 unidades para el caso de QTL con efecto moderado y 0.5 unidades que correponderían a un gen de efecto grande.

El efecto de la dominancia del QTL se investigó en todo el rango de valores, de nulo a completo.

El efecto de la distancia de mapeo entre el QTL y el marcador se estudió para valores de distancia nula —el caso de un marcador pleiotrópico—, 10 y 25 cM, distancias máximas para un mapa con intervalos de 20 y 50 cM (Darvasi y Soller, 1994).

El progenitor no común (ej. hembras) se asume no emparentado y en equilibrio de ligamiento para marcador y QTL. Sus genotipos no se consideran, una asunción razonable en un esquema real de detección que recoja los datos en el momento del sacrificio de los animales genotipados.

La segregación de los alelos de origen paterno es la única que aporta información y, por consiguiente, se introduce un efecto s de alelo paterno, definido como la mitad de la diferencia fenotípica media entre los grupos de progenie que heredan cada uno de los alelos paternos.

Los alelos del QTL se asumieron equifrecuentes, caso particular para el que el efecto aditivo es el doble efecto s de alelo paterno, y 0.1. La implicaciones de la elección de estos dos valores se discutirán.

El genotipado selectivo se simuló tomando las colas de las distribución de los valores fenotípicos de los medio hermanos. Las citadas colas comprendían el 10, 20 y 50% de los animales, y el

100% en el caso, tambien contemplado, de no emplear esta técnica en cuyo caso cada una de las colas comprende la mitad de los individuos disponibles. Para que la estimación de la frecuencia de los alelos del marcador en las colas reflejase las implicaciones del empleo de pooles de ADN, se incorporó un error técnico. El error técnico puede atribuirse a la contribución desigual de los individuos a la muestra mezcla, y los errores cometidos en la estimación de las frecuencias alélicas de los marcadores, provocadas a su vez por imprecisión en la lectura de la densidad de las bandas electroforéticas. Para modelar el error técnico se introdujo en el modelo una variable aleatoria independiente $N(0, \sigma^2_t)$ que distorsionaba aditivamente la estimación de la frecuencia, siempre dentro de su espacio paramétrico.

Cálculo de la potencia

Todos los cálculos de la potencia se basaron en el parámetro π (Darvasi & Soller 1994), que se define como la frecuencia relativa de individuos que, habiendo heredado cierto alelo del marcador del semental, se encuentran en la cola superior de la distribución de fenotipos. Bajo determinadas asunciones (*ibidem*), toma el mismo valor para los individuos que heredan el otro alelo paterno y pertenecen a la cola inferior. Para la hipótesis nula (ausencia de asociación), π toma el valor de 0.5, la segregación del alelo paterno es independiente de la pertenencia a una u otra cola de la distribución de fenotipos.

Tests Normales. Se basan en las propiedades de la teoría de tests normales, como en la fórmula nº 5 de (Darvasi & Soller 1994) y derivan de una serie de aproximaciones a la distribución de los fenotipos de la progenie y del estadístico π, que se asumen normalmente distribuidos bajo las hipótesis nula y alternativa.

Umbrales Teóricos. Los umbrales de rechazo se establecen como en el caso anterior para el caso de la hipótesis nula, pero para la hipótesis alternativa se genera una distribución empírica. Los umbrales de rechazo son en este caso $\frac{1}{2} \pm \sigma_{\hat{\pi}} Z_{1-\alpha/2}$ donde $\sigma_{\hat{\pi}}$ es la desviación estándar de $\hat{\pi}$, y $Z_{1-\alpha/2}$ es la ordenada de la distribución normal estándar que verifica $\int_{-\infty}^{Z_{1-\alpha/2}} \frac{1}{\sqrt{2\pi}} \exp\left(-x^2\right) dx = 1 - \frac{\alpha}{2}$

Umbrales Empíricos. Los umbrales de rechazo se obtienen de los cuartiles $\pi/2$ and $1-\pi/2$ de la distribución empírica de $\hat{\pi}$, simulada bajo la hipótesis nula (π=0.5). Los umbrales de rechazo resultantes se aplican a la distribución empírica de $\hat{\pi}$ derivada de las asunciones que se estén estudiando en cada caso.

RESULTADOS Y DISCUSION

Efecto aditivo del QTL

La potencia de detección de QTL es más elevada cuanto mayor es su efecto aditivo (Tabla 1). Asumiendo un efecto aditivo (*a*) de 0.25, y fijando la frecuencia del alelo favorable en la poblatción de madres (*t*) en 0.5, la potencia obtenida con 500 medio-hermanos es de 0.74 bajo el enfoque de Tests Normales de Darvasi y Soller. Cuando se recurre a umbrales teóricos, la potencia que se obtendría sería de 0.66, y si los umbrales se extraen de la distribución empírica del estadístico $\hat{\pi}$ bajo la hipótesis nula, la potencia baja a 0.55: una pérdida da casi el 20% .

Tabla 1. Potencia (%) con α=5%

efecto aditivo	a=0	a=0.25	a=0.5
Tests Normales	2.5	74	100
Umbrales Teóricos	10	66	100
Umbrales Empíricos	5	55	99

Con valores mas elevados, a=0.5, la potencia se ve incrementada de forma apreciable (casi 0.75) pero debe notarse que en este caso no se cumplen las asunciones de de Darvasi y Soller de

normalidad del estadístico $\hat{\pi}$ bajo la hipótesis alternativa. La potencia de detección de QTL es casi completa para efectos aditivos superiores a 0.5, pero el verdadero interés del genotipado selectivo reside en su capacidad para detectar QTL con efecto pequeño (Spellman, PhD Thesis, 1998).

Tabla 2. Rechazo empírico con distintos Errores de Tipo I

Error de Tipo I (α)	Rechazo empírico (Umbrales Teóricos)
0.01	0.03
0.05	0.10
0.10	0.17

Se estudió asimismo la distribución de $\hat{\pi}$ bajo la hipótesis nula de ausencia de QTL (a=0). Nuestros resultados sugieren que, bajo el enfoque de los Tests Normales, los umbrales que se asumen resultan en unos errores de tipo I muy superiores a los valores que se les asume. La Tabla 2 recoge valores de rechazo empírico con Umbrales Teóricos para diversos Errores de Tipo I.

Fondo poligénico, Dominancia y Frecuencia alélica del QTL

La dominancia afecta muy poco a la potencia: 3% de pérdida entre un esquema con alelo del QTL completamente dominante y otro con un QTL solo aditivo. El fondo poligénico reduce la potencia del esquema de deteccioón ya que oculta el efecto del QTL. La potencia decrece 5-15% cuando la varianza debida a los efectos poligénicos pasa de 0 a 0.5. En la Tabla 3 se recogen las potencias correpondientes a varias combinaciones de estos parámetros.

Tabla 3. Potencia (%) para α=5%

frecuencia	t = 0 á 1	t = 0.5	t = 0.1	t = 0.5
dominancia	d = 0	d = 0 á 1	d = 1	d = 0 á 1
e. aditivo	a = 0.25	a = 0.25	a = 0.25	a = 0.5
Umbrales Teóricos	66	66	100	100
Umbrales Empíricos	55	55	96	99

REFERENCIAS BIBLIOGRAFICAS

Darvasi A. y M. Soller. 1994. Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. Genetics 138:1365-1373.

Hill W. G. 1998. A note on the theory of artificial selection in finite populations and application to QTL detection by bulk segregant analysis. Genetical Research 72:55-58

Lipkin E., M. O. Mosig, A. Darvasi, E. Ezra, A Shalom, A. Friedmann, M. Soller. 1998. Quantitative trait locus mapping in dairy cattle by means of selective milk DNA pooling using dinucleotide microsatellite markers: analysis of milk protein percentage. Genetics 149: 1557-1567.

Muranty, H., B. Goffinet, F. Santi, 1997. Multitrait and multipopulation QTL search using selective genotiping

Martinez, M. L., N. Vukasinovic, A. C. Freeman y R. L. Fernando, 1998. Mapping QTL in outbred populations using selected samples. Genetic, selection, evolution 30:453-468.

Lander y Botstein, 1989