ITEA (1998), Vol. **94A N.° 3**, 265-270

DETECCIÓN DE QTLS MEDIANTE LA PARTICIÓN DE LA VARIANZA GENÉTICA EN FUNCIÓN DEL PARENTESCO ATRIBUIBLE A SEGMENTOS DEL GENOMA

L. Varona
M. Pérez-Enciso

Centre UDL-IRTA
Area de Producción Animal
25198 Lleida

España

RESUMEN

Se presenta un procedimiento de detección de segmentos cromosómicos que determinan la variación de un determinado carácter, lo que permite generalizar los métodos de análisis y detección de QTL. El procedimiento consiste en definir matrices de relaciones genéticas para un segmento particular de ADN. Estas matrices pueden ser calculadas tanto analítica como numéricamente. Se presenta una aplicación en una población simulada, donde se puede observar, por una parte, la capacidad de discriminar entre fuentes de variación de origen genético y, por otra, el incremento de precisión en la predicción del mérito genético de los individuos.

Palabras clave: QTL, Parentesco, ADN.

SUMMARY
QTL DETECCION USING GENETIC RELATIONSHIPS FOR GENOME SEGMENTS

A new procedure to detect genome segments that affect variation in a particular trait is presented, and thus allowing the detection of QTL. The procedure defines genetic relathionship matrices for any DNA segment. Those matrices can be calculated both analitically and numerically. An aplication of the procedure in a simulated population is presented. In this example, the ability of the procedure to discriminate between sources of variation is shown. Morever, the accuracy improvement in the prediction of genetic merit is also presented..

Key words: QTL, Genetic relationship, DNA.

Introducción

A pesar del gran número de métodos estadísticos disponibles para la identificación de genes que afectan a la expresión de caracteres cuantitativos (QTLS), todos ellos se basan, explícita o implícitamente, en asumir que existe uno o dos genes más un residuo genético infinitesimal. Este enfoque presenta una serie de problemas, a) la distribución de efectos génicos asumida

no es realista, el tránsito entre genes de efecto grande y de efecto muy pequeño es seguramente gradual, b) la elección del nivel de significación no es obvia y c) puede no ser un método robusto si las frecuencias alélicas no son las asumidas. Por ejemplo, en una F2, si no hay alelos fijados para el QTL en cada una de las poblaciones origen.

Un enfoque alternativo es identificar regiones que expliquen más variación genética que la esperada bajo un modelo infinitesimal. Esta posibilidad ha sido planteada por VISSCHER y HALEY (1996) y LIU y DEKKERS (1998), pero no desarrollaron la metodología necesaria. En este trabajo definimos la teoría y la ilustramos con un ejemplo sencillo.

Material y métodos

Teoría

Bajo un modelo estrictamente infinitesimal, la covarianza genética aditiva entre 2 individuos es proporcional al porcentaje de genoma compartido:

$$a(i,i') = \frac{1}{2G} \sum_{H=1}^{2} \sum_{H'=1}^{2} \int_{0}^{G} \delta_{H,H'}(x)dx =$$

$$= \frac{1}{2\sum_{1}^{n_c} G_i} \sum_{i=1}^{n_c} \sum_{H=1}^{2} \sum_{H'=1}^{2} \int_{0}^{G_i} \delta_{H,H'}(x)dx$$

donde H y H' son los haplotipos de cada individuo, G es la longitud del genoma, con

$$G = \sum_{i=1}^{n_c} G_i$$

n_c es el número de cromosomas y es una variable indicadora que toma el valor 1 si los haplotipos comparados son idénticos por descendencia en la posición x y 0 en caso contrario. En ausencia de marcadores, a(i,i') es el coeficiente de parentesco de Wright. Por el contrario, si se dispusiese de un numero infinito de marcadores completamente informativos, a(i,i') podría ser calculado con total precisión. En la práctica la situación es intermedia.

a) Marcadores completamente informativos.

Si los marcadores son completamente informativos, se conoce con certeza el origen gamético de cada uno de los marcadores y, por lo tanto, la probabilidad de origen de cualquier posición del genoma es sencillamente calculable.

Si dos haplotipos presentan los mismos alelos en ambos marcadores y en la misma fase que sus padres, el coeficiente de parentesco para el segmento de longitud λ del haplotipo j entre el individuo i y el haplotipo k del individuo i' es:

$$\delta(H_{ij}, H_{i'k}) =$$

$$= \frac{1}{\lambda} \int_{0}^{\lambda} \frac{\left[(1-r_\chi)^2(1-r_{\lambda-\chi})^2 + (r_\chi r_{\lambda-\chi})^2\right]}{(1-r_\lambda)^2} dx$$

donde r_i es la fracción de recombinación correspondiente a la distancia i.

Si comparten ambos marcadores en fase distinta a sus padres:

$$\delta(H_{ij},H_{i'k}) =$$

$$= \frac{1}{\lambda} \int_{0}^{\lambda} \frac{\left[(1-r_\chi)^2 r_{\lambda-\chi}{}^2 + (1-r_{\lambda-\chi})^2 r_\chi{}^2\right]}{r_\lambda{}^2} dx$$

Si no comparten ninguno de los alelos en los marcadores y los haplotipos se encuentran en la misma fase que sus padres, el parentesco es:

$$\delta(H_{ij},H_{i'k}) = \frac{1}{\lambda}\int_0^{\lambda} \frac{\left[2(1-r_\chi)\,(1-r_{\lambda-\chi})\,r_\chi\,r_{\lambda-\chi}\right]}{(1-r_\lambda)^2}\,dx$$

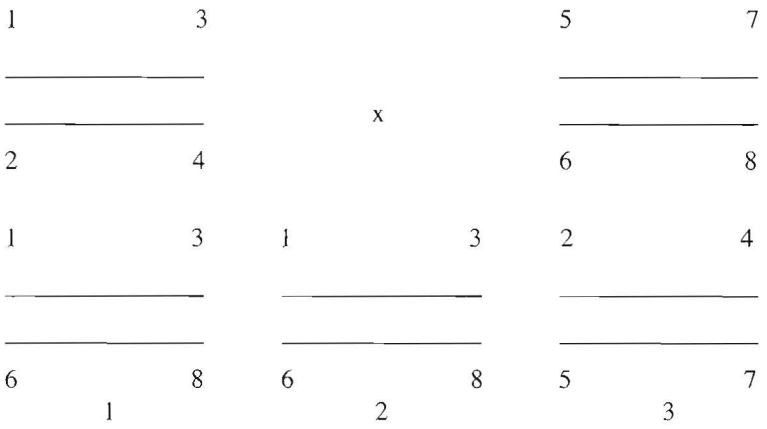
Si no comparten ninguno de los marcadores y los haplotipos se encuentran en distinta fase que los padres:

$$\delta(H_{ij},H_{i'k}) = \frac{1}{\lambda}\int_0^{\lambda} \frac{\left[2(1-r_\chi)\,r_{\lambda-\chi}\,(1-r_{\lambda-\chi})r_\chi\right]}{r_\lambda^{\,2}}\,dx$$

Si los haplotipos sólo comparten uno de los marcadores:

$$\delta(H_{ij},H_{i'k}) = 0{,}5$$

Supongamos dos marcadores polimórficos separados por una distancia de 20cM y el siguiente pedigrí:

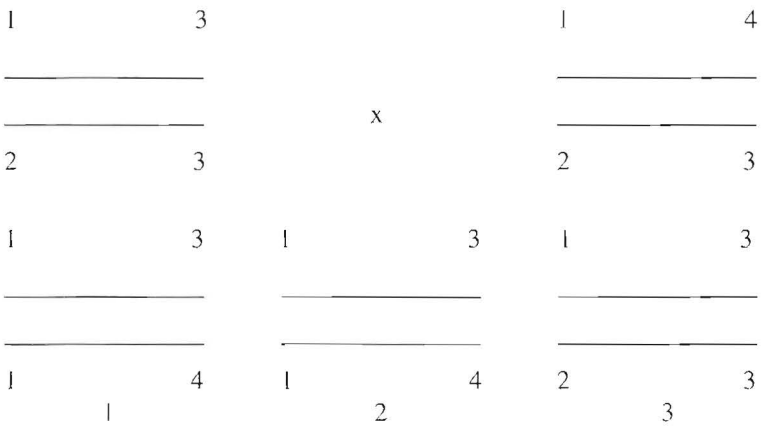


$$a(1,2) = \frac{1}{2} \times \left(\sum_{i=1}^{2}\sum_{j=1}^{2} \delta(H_{ij},H_{i'j})\right) = \frac{0{,}987+0{,}000+0{,}000+0{,}987}{2} = 0{,}987$$

$$a(1,3) = \frac{1}{2} \times \left(\sum_{i=1}^{2}\sum_{j=1}^{2} \delta(H_{ij},H_{i'j})\right) = \frac{0{,}013+0{,}000+0{,}000+0{,}013}{2} = 0{,}013$$

– b) Marcadores parcialmente informativos y/o fase desconocida.

Cuando los marcadores son parcialmente informativos y se desconoce la fase de los mismos, el origen genético no puede ser definido completamente y es necesario ponderar todas las posibles combinaciones. Supongamos la misma genealogía con marcadores bialélicos:



En este caso, y si de nuevo $\lambda = 0,2$

$$a(1,2) = 0,86$$

$$a(1,3) = 0,38$$

para el segmento considerado.

Modelo

Supongamos:

$$y = X\beta + Zu + e$$

donde **y** es el vector de datos fenotípicos, **X** y **Z** son las matrices de incidencia, β es el vector de efectos sistemáticos, **u** es el vector de efectos genéticos aditivos y **e** es el vector de residuos.

$$Var(u) = A^M \sigma^2_a$$

donde $\mathbf{A}^M$ es la matriz de relaciones aditivas calculadas utilizando la información molecular tal y como se ha descrito. En principio, la estimación de h^2 y de los valores genéticos aditivos será más precisa que utilizando **A** (calculada exclusivamente con la información aportada por el pedigrí). Sin embargo, el incremento de precisión no es muy elevado si la longitud del genoma es suficientemente grande.

Más interesante es discernir cuál es la contribución de cada segmento del genoma a la variación genética total. Por ejemplo, se puede describir un modelo que atribuya una parte de la variación de origen genético a cada cromosoma.

$$u = \sum_{i=1}^{n_C} u_i \qquad Var(u_i) = Diag \left\{ A_i \sigma^2_{ai} \right\}$$

Ejemplo. Estimación del porcentaje de variación atribuido a cada cromosoma

Simulación

Se ha simulado una población consistente en 100 machos y 500 hembras, en un diseño jerárquico con 5 hembras cruzadas con cada macho y 5 descendientes por hembra, generándose familias de 25 medio hermanos en grupos de 5 hermanos completos. Para cada individuo se han generado 4 cromosomas con 20, 15, 10 y 5 loci infinitamente polimórficos repartidos homogéneamente en cada cromosoma. Cada loci explica el 1% de la variación, por lo tanto

CUADRO I
HEREDABILIDAD ESTIMADA (MEDIA ± DESVIACIÓN TÍPICA POSTERIORES)

Caso	Infinitesimal	Crom. I	Crom. II	Crom. III	Crom. IV
I	0,475 ± 0,035	–	–	–	–
II	0,023 ± 0,017	0,184 ± 0,032	0,175 ± 0,030	0,110 ± 0,027	0,035 ± 0,020
III	0,045 ± 0,034	0,.213 ± 0,066	0,096 ± 0,060	0,092 ± ,043	0,058 ± 0,045

la heredabilidad simulada es 0,50. En cada cromosoma se han simulado 2 marcadores infinitamente polimórficos (Caso II) y Bialélicos (Caso III).

Estimación

Se ha estimado la heredabilidad utilizando un modelo animal que no hace uso de la información de marcadores moleculares (Caso I), además se han estimado la heredabilidad atribuida a cada uno de los cromosomas utilizando 2 marcadores infinitamente polimórficos por cromosoma (Caso II) y dos marcadores bialélicos (Caso III), cuya frecuencia génica era 0,5. Los estimadores se han obtenido mediante inferencia bayesiana con información a priori vaga.

La heredabilidad total es estimada con precisión por los tres procedimientos. Sin la utilización de marcadores moleculares no es posible discriminar entre el origen cromosómico de la variación. Cuando se utilizan marcadores moleculares sí es posible discriminar entre los orígenes de la variación, aunque la precisión que se obtiene depende de la cantidad de información disponible en los marcadores. En el Caso II, la cantidad de información polimórfica (PIC) es 1, mientras que en el Caso III, su PIC es 0,375. Este hecho se refleja en la desviación de la distribución posterior marginal del porcentaje de variación explicada entre los casos II y III (Por ejemplo, 0,032 vs. 0,066 para el primer cromosoma).

La información molecular permite, además de atribuir las causas de la variación de origen genético a distintos segmentos de ADN, refinar la matriz de parentesco y, por lo tanto, mejorar la capacidad predictiva del mérito genético global, que puede ser usado como criterio de selección. En el cuadro 2, se presenta la correlación entre el mérito genético simulado y estimado para los tres casos considerados. Como se observa, con un marcador molecular con información polimórfica baja (Caso III), el aumento de la precisión es despreciable (1,6%), por el contrario, con un marcador completamente informativo se consigue un incremento del 6,8% en la precisión (Caso II).

CUADRO II
CORRELACIÓN ENTRE VALORES GENÉTICOS SIMULADOS Y PREDICHOS

Caso	r2
I	0.617
II	0.659
III	0.627

Conclusiones y perspectivas futuras

Los marcadores moleculares permiten refinar el parecido genético entre individuos para cada segmento de ADN, lo que permite calcular los porcentajes de variación atribuibles a dicho segmento, así como mejorar la capacidad predictiva de los modelos de evaluación. Incluso cuando no existen loci de gran efecto, pero los loci de efecto pequeño están repartidos desigualmente por el genoma, la información molecular se manifiesta de utilidad. Pese a todo, es necesario disponer de marcadores muy informativos para que estas ventajas se hagan evidentes. El procedimiento ha resultado eficiente en la estimación de variación genética a segmentos de ADN, por lo que puede ser utilizado para la detección de loci de variación cuantitativa (QTL). Con respecto a otras aproximaciones tiene la ventaja de que es fácilmente generalizable al análisis de varios caracteres simultáneamente y que, en principio, será más robusto que el enfoque clásico si existe pleiotropía y ligamiento.

Agradecimientos

Trabajo financiado por los proyectos UE PL96-2243 y CICYT AGF96-2510.

Bibliografía

VISSCHER P.M. y HALEY C., 1996. Detection of putative quantitative trait loci in line crosses under infinitesimal genetic models. Theoretical Applied Genetics. 93, 691-702.

LIU Z. y DEKKERS, J.C.M., 1998. Least squares interval mapping of quantitative trait loci under the infinitesimal genetic model in outbred populations. Genetics. 148, 495-505.