

# Testaje del desequilibrio de ligamiento poblacional versus el desequilibrio dentro de familias utilizando familias de medios hermanos

Luis Gomez Raya

Centre UdL –IRTA, Area de Produccio Animal, Av. Alcalde Rovira Roure, 177.  
25198 Lleida. Spain

## Introducción

En los últimos años se ha intensificado la búsqueda de QTL (del inglés Quantitative Trait Loci) utilizando marcadores moleculares. El primer paso consiste generalmente en realizar un barrido genómico en el que se detectan áreas del genoma en las que están segregando QTL. A continuación se realiza mapeo fino en el que se aumenta la densidad de marcadores y de genes candidatos en dichas áreas cromosómicas y se lleva a cabo una localización mas precisa de los QTL. En diseños de medios hermanas, que son frecuentes en especies domesticas comerciales como el vacuno lechero, es posible investigar el desequilibrio de ligamiento dentro de familias testando la segregación dentro de machos. Normalmente no es posible identificar el gen si hay varios genes candidatos muy ligados debido a que el fuerte desequilibrio dentro de familias impide discriminar entre ellos por la falta de meiosis informativas. En este estudio proponemos detectar el desequilibrio a nivel poblacional utilizando el alelo procedente de la madre. Estos alelos son muestras tomadas al azar de la población de madres y pueden servir para testar el desequilibrio poblacional. Si el gen candidato es la mutación causal misma entonces un fuerte desequilibrio de ligamiento poblacional se espera. Los objetivos de este trabajo son desarrollar métodos de máxima verosimilitud para testar el desequilibrio dentro familias y poblacional y para comparar ambos desequilibrios. Este método proporciona información para comparar genes candidatos respecto al desequilibrio de ligamiento dentro de familias y poblacional en zonas cromosómicas donde se han detectado QTL.

## Teoría

Métodos de máxima verosimilitud pueden utilizarse para estimar los efectos asociados a los alelos heredados por vía paterna y materna. Por simplicidad asumimos que el carácter se distribuye normalmente, la varianza es la misma para todos los genotipos y que el gen es aditivo. También se asume que el gen es candidato y dialélico en el que uno de los alelos esta asociado a producción en un estudio previo en el que se detectaron QTL en esa zona cromosómica. En la Figura 1 se muestra un esquema de las contribuciones de machos y hembras a la descendencia así como las frecuencias esperadas de cada genotipo y su valor esperado. En el esquema, el primer alelo se hereda del padre (subíndice S) y el segundo de la madre (subíndice D). Los genotipos de los heterocigotos  $A_{SA} a_{Da}$  y  $a_{Sa} A_{DA}$  no pueden distinguirse sin información adicional que podría obtenerse con marcadores en la proximidad de dicho locus.

Para la descendencia j de padre i y con genotipo  $G_0 = Aa$ , la verosimilitud es:

$$\begin{aligned} L_{ij}(x_{ij}, \alpha_{SA}, \alpha_{DA}, \alpha_{Sa}, \alpha_{Da} | G_0 = Aa, G_s) &= p_{ij}(G_0 = A_{SA} a_{Da} | G_s, f_A) z_{ij1} + \\ & p_{ij}(G_0 = a_{SA} A_{Da} | G_s, f_A) z_{ij2} \\ &= \frac{1}{2} f_a z_{ij1} + \frac{1}{2} f_A z_{ij2} \end{aligned}$$

where

$$z_{ij1}(x, \mathbf{m}_{j1}, \mathbf{s}^2) = \frac{1}{\mathbf{s}\sqrt{2\mathbf{p}}} e^{-\frac{(x-\mathbf{m}_{j1})^2}{2\mathbf{s}^2}} \text{ and } z_{ij2}(x, \mathbf{m}_{j2}, \mathbf{s}^2) = \frac{1}{\mathbf{s}\sqrt{2\mathbf{p}}} e^{-\frac{(x-\mathbf{m}_{j2})^2}{2\mathbf{s}^2}}$$

con  $\mu_{j1} = \alpha_{SA} + \alpha_{Da}$  and  $\mu_{j2} = \alpha_{Sa} + \alpha_{DA}$ .

Las verosimilitudes de descendencia de padre i y con genotipo  $G_0 = AA$  y  $G_0 = aa$  son:

$$L_{ij}(x_{ij}, \alpha_{SA}, \alpha_{DA}, \alpha_{Sa}, \alpha_{Da} | G_0 = AA, G_s) = z_{ij3},$$

$$L_{ij}(x_{ij}, \alpha_{SA}, \alpha_{DA}, \alpha_{Sa}, \alpha_{Da} | G_0 = aa, G_s) = z_{ij4},$$

respectivamente. Las medias de estas distribuciones normales  $\mu_{j3} = \alpha_{SA} + \alpha_{DA}$  and  $\mu_{j4} = \alpha_{Sa} + \alpha_{Da}$ .

La verosimilitud conjunta para toda la descendencia n de cada uno de los nb machos es:

$$L(\mathbf{a}_{SA}, \mathbf{a}_{Sa}, \mathbf{a}_{DA}, \mathbf{a}_{Da}, \mathbf{s}^2) = \prod_{j=1}^{nb} \prod_{i=1}^n L_{ij}$$

El testaje del desequilibrio de ligamiento poblacional versus dentro de familias se puede llevar a cabo mediante LRT (del inglés Likelihood Ratio Statistics):  $LRT \sim -2 \ln(\max L_{ho} / \max L_{ha})$ , donde  $\max L_{ho}$  es la verosimilitud de la hipótesis nula ( $\alpha_{SA} = \alpha_{DA}$ ;  $\alpha_{Sa} = \alpha_{Da}$ ) y  $\max L_{ha}$  es la verosimilitud del modelo sin restricciones en los parámetros a estimar ( $\alpha_{SA} \neq \alpha_{DA}$ ;  $\alpha_{Sa} \neq \alpha_{Da}$ ). LRT se distribuye como una  $\chi^2$  con  $2nb$  grados de libertad.

## Discusión

En este estudio se propone comparar los desequilibrios poblacional y dentro de familias para genes candidatos en zonas cromosómicas donde se han detectado QTL. Si el test propuesto (LRT) es significativo entonces indicaría que los desequilibrios de ligamiento son diferentes. Lo cual implicaría que el gen candidato testado es un marcador pero no la mutación causal. Este test podría ser significativo también si existe imprinting sexual.

De la observación de las formulas para escribir la verosimilitud se deduce que si la frecuencia del gen candidato es 0.5 entonces el método no puede discriminar entre los dos desequilibrios. La potencia estadística se debe calcular para una variedad de escenarios con frecuencias alélicas diferentes.

En la práctica se puede incluir marcadores moleculares en la proximidad del gen candidato que se quiere testar con el fin de identificar el origen paterno o materno de alelos en los heterocigotos ( $A_{SA} a_{Da}$  y  $a_{Sa} A_{DA}$ ) y de esta manera incrementar la potencia estadística del método.

## Agradecimientos

Esta investigación se ha financiado por el MCYT de España a través del proyecto AGL2002-04271-C03-02

**Figura 1.** Esquema de las contribuciones de un macho heterocigoto y de las hembras a la descendencia en familias de medios hermanos. Prob. es la frecuencia del genotipo correspondiente. Valor esperado es la media de cada genotipo.  $f_i$  = frecuencia del alelo  $i$  en la población de hembras.

A/a	A, a		Prob	Valor esperado
-	↙			
$A_{SA}$	$A_{DA}$		$\frac{1}{2} f_A$	$\alpha_{SA} + \alpha_{DA}$
$A_{SA}$	$a_{Da}$		$\frac{1}{2} f_a$	$\alpha_{SA} + \alpha_{Da}$
$a_{Sa}$	$A_{DA}$		$\frac{1}{2} f_A$	$\alpha_{Sa} + \alpha_{DA}$
$a_{Sa}$	$a_{Da}$		$\frac{1}{2} f_a$	$\alpha_{Sa} + \alpha_{Da}$