

ESTIMACIÓN DE LOS PARÁMETROS BIOMÉTRICOS DE LOS QTL EN POBLACIONES F_2 : UN NUEVO ENFOQUE

ESTIMATING QTL BIOMETRICS PARAMETERS IN F_2 POPULATIONS: A NEW APPROACH

J. Jesús Cerón-Rojas y Jaime Sahagún-Castellanos

Fitotecnía. Universidad Autónoma Chapingo. 56230. Chapingo, Estado de México
(jsahagun@correo.chapingo.mx).

RESUMEN

Los procedimientos estándar para construir mapas de loci de caracteres cuantitativos (QTL o Quantitative Trait Loci, por sus siglas en inglés) se basan en la estimación por máxima verosimilitud o mínimos cuadrados de los parámetros biométricos del QTL y generalmente requieren métodos numéricos y software especial, ya que no producen estimadores explícitos. El objetivo del presente trabajo fue derivar estimadores explícitos de los parámetros biométricos de un QTL en una población F_2 . La derivación se basó en tres combinaciones lineales de variables aleatorias con distribución normal mezclada y un marcador molecular. Para cada combinación lineal se determinaron su media y varianza y, con éstas, se derivaron estimadores del efecto aditivo y del efecto de dominancia del QTL, además de la frecuencia de recombinación entre éste y el marcador. Los estimadores explícitos aquí obtenidos fueron asintóticamente consistentes e insesgados y no requieren software especial para su evaluación.

Palabras clave: Caracteres cuantitativos, distribuciones mezcladas, frecuencia de recombinación, marcadores moleculares, perfil de verosimilitud.

INTRODUCCIÓN

Un método para identificar loci de caracteres cuantitativos (QTL, por sus siglas en inglés) es el método Marcador-QTL-Marcador (MQM) o mapeo por intervalos simples propuesto por Lander y Botstein (1989) y más tarde generalizado por Jasen (1993, 1996, 2003) y Zeng (1994) como “mapeo por intervalos compuestos (composite interval mapping o CIM, por sus siglas en inglés)”. MQM considera dos marcadores moleculares (MM) adyacentes al QTL y recurre al método de máxima verosimilitud. Desde entonces se han publicado varios artículos (Wright y Mowers, 1994; Whittaker *et al.*, 1996; Xu, 1998; Xu y Volg, 2000) que describen metodologías que pretenden facilitar la estimación de los parámetros biométricos: medias y varianzas del carácter cuantitativo (QT, por

ABSTRACT

The standard procedures for constructing maps of Quantitative Trait Loci (QTL) are based on the estimation by maximum likelihood or minimum squares of the QTL biometrics parameters and generally require numerical methods and special software, given that they do not produce explicit estimators. The objective of the present study was to derive explicit estimators of the biometrics parameters of a QTL in an F_2 population. The derivation was based on three linear combinations of random variables with mixed normal distribution and a molecular marker. For each linear combination, its mean and variance were determined, with which estimators were derived of the additive effect and dominance effect of the QTL, as well as the frequency of recombination between the QTL and the marker. The explicit estimators obtained here were asymptotically consistent and unbiased, and do not require special software for their evaluation.

Key words: Quantitative traits, mixed distributions, recombination frequency, molecular markers, likelihood profile.

INTRODUCTION

One method of identifying quantitative trait loci (QTL) is the Marker-QTL-Marker (MQM) method or simple intervals mapping proposed by Lander and Botstein (1989), later generalized by Jasen (1993, 1996, 2003) and Zeng (1994) as composite interval mapping or CIM. MQM considers two molecular markers (MM) adjacent to the QTL and recurs to the maximum likelihood method. Since then, various articles have been published (Wright and Mowers, 1994; Whittaker *et al.*, 1996; Xu, 1998; Xu and Volg, 2000) that describe methodologies which intend to facilitate the estimation of the biometric parameters: means and variances of the quantitative traits (QT), the additive effect and the dominance effect of the QTL, the frequencies of recombination between the QTL and the pair of MMs adjacent to it, along with the evaluation of the test statistic to identify linkage. MQM require numerical methods for the evaluation of the test statistic and do not provide explicit estimators of the biometrics parameters of the QTL.

Recibido: Mayo, 2005. Aprobado: Noviembre, 2006.

Publicado como ENSAYO en *Agrociencia* 41: 57-73. 2007.

sus siglas en inglés), el efecto aditivo y de dominancia del QTL, las frecuencias de recombinación entre éste y el par de MM's adyacentes a él, así como la evaluación del estadístico de prueba para identificar ligamiento. MQM requiere métodos numéricos para la evaluación del estadístico de prueba y no proporciona estimadores explícitos de los parámetros biométricos del QTL.

Haley y Knott (1992) y Martínez y Curnow (1992) propusieron utilizar la regresión lineal para identificar el ligamiento entre el QTL y los MM's, y el método de mínimos cuadrados para estimar los parámetros biométricos y la frecuencia de recombinación. Sin embargo, aunque es posible implementar el procedimiento en una computadora personal los residuos no tienen distribución normal, y, consecuentemente, se disminuye la potencia de la prueba (Xu, 1998). Kearsey y Hyne (1994) y Hyne y Kearsey (1995) desarrollaron un método ligeramente diferente al de Haley y Knott (1992) y Martínez y Curnow (1992): en vez de considerar las observaciones de los caracteres cuantitativos de manera individual, utilizaron las diferencias de las medias de las poblaciones definidas por los genotipos de los MM's para identificar y localizar el QTL. El procedimiento puede implementarse en una computadora personal sin dificultad alguna y es uno de los más potentes para la identificación de QTL. Desafortunadamente, no proporciona estimadores explícitos de los parámetros biométricos del QTL y de la frecuencia de recombinación entre éste y los MM's.

El método de regresión está disponible en el paquete SAS pero es inadecuado en el contexto del análisis de QTL ya que en él se confunde el efecto del QTL con la frecuencia de recombinación. Por ello Lander *et al.* (1987) desarrollaron MAPMAKER y Knapp *et al.* (1990) GENEMAP, aunque Darvasi *et al.* (1993) han utilizado el algoritmo Newton-Raphson, el cual está implementado en SAS, para estimar las frecuencias de recombinación y los parámetros biométricos del QTL. Varios investigadores han desarrollado software estadístico adicional para la identificación de QTL (ver Cuadro A.1 del Apéndice A).

Los procedimientos de análisis referidos están diseñados para poblaciones endogámicas. La ventaja de tales poblaciones radica en la posibilidad de conocer con certeza los genotipos parentales y la fase de ligamiento. Además, los datos de estas poblaciones pueden considerarse provenientes de una sola familia, ya que toda la progenie comparte los mismos genotipos parentales (Xu y Atchley, 1995). Sin embargo, las metodologías de análisis señaladas no permiten analizar los datos de poblaciones no endogámicas ya que en éstas los genotipos parentales generalmente son desconocidos y provienen de familias diferentes, lo que significa que las probabilidades para la identificación de

Haley and Knott (1992) and Martínez and Curnow (1992) proposed the utilization of linear regression to identify the linkage between the QTL and the MMs, and the minimum squares method to estimate the biometrics parameters and recombination frequency. However, although it is possible to implement the procedure in a personal computer, the residues do not have normal distribution, and consequently, the power of the test is reduced (Xu, 1998). Kearsey and Hyne (1994) and Hyne and Kearsey (1995) developed a method that was slightly different from that of Haley and Knott (1992) and Martínez and Curnow (1992): instead of considering the observations of the quantitative traits individually, they utilized the differences of the means of the populations defined by the genotypes of the MMs to identify and locate the QTL. The procedure can be implemented without difficulty in a personal computer and is one of the most powerful for the identification of the QTL. Unfortunately, it does not provide explicit estimators of the biometrics parameters of the QTL and of the frequency of recombination between it and the MMs.

The regression method is available in the SAS package, but it is inadequate in the context of the QTL analysis, because the effect of the QTL is confused with the recombination frequency. For this reason Lander *et al.* (1987) developed MAPMAKER and Knapp *et al.* (1990) GENEMAP, although Darvasi *et al.* (1993) have used the Newton-Raphson algorithm, which is implemented in SAS, to estimate the recombination frequencies and the QTL biometrics parameters. Various investigators have developed additional statistical software for the identification of the QTL (see Table A.1 of Appendix A).

The analysis procedures referred to are designed for endogamic populations. The advantage of such populations lies in the possibility of knowing with certainty the parental genotypes and the linkage phase. Furthermore, the data of these populations can be considered to come from a single family, given that all of the progeny share the same parental genotypes (Xu and Atchley, 1995). However, the above mentioned analysis methodologies do not permit the analysis of data of non-endogamic populations, because the parental genotypes in these populations are generally unknown and proceed from different families, which means that the probabilities for QTL identification, given the information of the markers, varies from one family to another. In addition, the linkage phase of the parents is difficult to know and the number of individuals which comprise a family is generally small, which hinders the estimation of the parameters of interest. For this reason, various investigators recur to the variance components model for the analysis of this type of data

QTL, dada la información de los marcadores, es diferente de familia en familia. Además, la fase de ligamiento de los progenitores es difícil de conocer y el número de individuos que conforman una familia es generalmente pequeño, lo que dificulta la estimación de los parámetros de interés. Por esto, varios investigadores recurren al modelo de componentes de varianza para el análisis de este tipo de datos (Xu y Atchley, 1995; Williams, 1998; Blangero *et al.*, 2001; Elston y Cordell, 2001). Adicionalmente, se ha aplicado la estadística bayesiana en la identificación de QTL en poblaciones endogámicas y no endogámicas (Hoeschele y van-Raden, 1993a, 1993b; George *et al.*, 1995; Sillanpää y Arjas, 1998, 1999; Hoeschele, 2003). Sin embargo, como sucede en el caso MQM, tanto la estadística bayesiana como las metodologías utilizadas en poblaciones no endogámicas requieren métodos numéricos y no proporcionan estimadores explícitos de los parámetros biométricos del QTL.

El objetivo del presente trabajo fue derivar un método para construir estimadores explícitos de los parámetros biométricos: medias del QT; del efecto aditivo y de dominancia de un QTL, y de la frecuencia de recombinación entre éste y un MM.

MÉTODOS Y MARCO TEÓRICO

En la derivación de los estimadores explícitos se hicieron los siguientes supuestos: 1) el QTL y los MM's son dialélicos; 2) los alelos de los MM's son codominantes, sus genotipos observables y no afectan el fenotipo del QT; 3) las poblaciones definidas a partir de los genotipos de los MM's son una mezcla de distribuciones normales de los fenotipos asociados a varios QTL (Weller, 1992; Zeng, 1994).

Para la población F₁ de la cruce de dos progenitores homocigotos P₁ y P₂ cuyos genotipos respectivos son A₁₁A₁₁Q₁Q₁ y A₁₂A₁₂Q₂Q₂, considérese un par de cromosomas homólogos con k MM's heterocigotos (A_{1i}A₁₂, i=1, 2,..., k) genotípicamente distinguibles y supóngase que la frecuencia de recombinación entre un QTL y el MM i-ésimo (M_i) es r. En la población F₂ los genotipos de M_i: A₁₁A₁₁, A₁₁A₁₂ y A₁₂A₁₂ serán distinguibles y observables; sin embargo, los genotipos correspondientes al QTL no se observarán y sólo será posible clasificar las observaciones fenotípicas del QT de acuerdo con el genotipo marcador al que están asociadas. Esto genera, para cada genotipo marcador, una mezcla de tres distribuciones fenotípicas normales, correspondientes a los tres genotipos del QTL (Cuadro 1).

Variables aleatorias y distribuciones mezcladas

Sean x_{1j} , x_{2j} y x_{3j} , j=1, 2,..., n_{it}, las j-ésimas expresiones fenotípicas de los genotipos Q₁Q₁, Q₁Q₂ y Q₂Q₂ del QTL (n_{it} es el número de observaciones del QT en la población definida por el t-ésimo genotipo del i-ésimo MM (M_i), t=1, 2, 3; t representa el

Cuadro 1. Población mezclada F₂ para cada genotipo del MM M_i ligado a un QTL con frecuencia de recombinación r. La distribución del valor fenotípico (X_q) asociada al q-ésimo genotipo del QTL es f_q(x) (q=1, 2, 3, representa a los genotipos Q₁Q₁, Q₁Q₂ y Q₂Q₂).

Table 1. Mixed F₂ population for each genotype of the MM M_i linked to a QTL with recombination frequency r. The distribution of the phenotypic value (X_q) associated to the q-th genotype of the QTL is f_q(x) (q=1,2,3, represents the genotypes Q₁Q₁, Q₁Q₂ and Q₂Q₂).

Genotipos del marcador (M _i)	Genotipos del QTL		
	Q ₁ Q ₁	Q ₁ Q ₂	Q ₂ Q ₂
A ₁₁ A ₁₁	(1-r) ² f ₁ (x)	2r(1-r) f ₂ (x)	r ² f ₃ (x)
A ₁₁ A ₁₂	r(1-r) f ₁ (x)	[1-2r(1-r)] f ₂ (x)	r(1-r) f ₃ (x)
A ₁₂ A ₁₂	r ² f ₁ (x)	2r(1-r) f ₂ (x)	(1-r) ² f ₃ (x)

(Xu and Atchley, 1995; Williams, 1998; Blangero *et al.*, 2001; Elston and Cordell, 2001). In addition, Bayesian statistics has been applied in the identification of the QTL in endogamic and non-endogamic populations (Hoeschele and van-Raden, 1993a, 1993b; George *et al.*, 1995; Sillanpää and Arjas, 1998, 1999; Hoeschele, 2003). However, as in the case of MQM, both the Bayesian statistic and the methodologies used in non-endogamic populations require numerical methods and do not provide explicit estimators of the QTL biometrics parameters.

The objective of the present study was to derive a method for constructing explicit estimators of the biometrics parameters: means of the QT; of the additive and the dominance effect of a QTL, and of the frequency of recombination between it and an MM.

METHODS AND THEORETICAL FRAMEWORK

In the derivation of the explicit estimators, the following assumptions were made: 1) the QTL and the MMs are diallelic; 2) the alleles of the MMs are co-dominant, these genotypes are observable and do not affect the phenotype of the QT; 3) the populations defined from the genotypes of the MMs are a mixture of normal distributions of the phenotypes associated to various QTL (Weller, 1992; Zeng, 1994).

For the F₁ population of the cross of two P₁ and P₂ homozygote parents whose respective genotypes are A₁₁A₁₁Q₁Q₁ and A₁₂A₁₂Q₂Q₂, consider a pair of homologous chromosomes with k MMs heterozygotes (A_{1i}A₁₂, i=1, 2,..., k) that are genotypically distinguishable and assume that the frequency of recombination between a QTL and the i-th MM (M_i) is r. In the F₂ population the genotypes of M_i: A₁₁A₁₁, A₁₁A₁₂ and A₁₂A₁₂ will be distinguishable and observable; however, the genotypes corresponding to the QTL will not be observed and it will only be possible to classify the phenotypic observations of the QT according to the marker genotype to which it they are associated. This generates, for each marker genotype, a mixture of three normal phenotypic distributions,

genotipo del MM. Así, $t=1$ indica el genotipo $A_{i1}A_{i1}$, $t=2$ el genotipo $A_{i1}A_{i2}$, etc.). Además, supóngase que las variables aleatorias X_1 , X_2 y X_3 [con medias $E(X_q)=\mu_q$ y varianzas $\text{Var}(X_q)=\sigma_q^2$, $q=1, 2, 3$] denotan los posibles valores fenotípicos asociados a los genotipos del QTL (Q_1Q_1 , Q_1Q_2 y Q_2Q_2 , respectivamente); son independientes y tienen distribución normal. Para el i -ésimo MM pueden definirse las tres variables aleatorias mezcladas (correspondientes a los tres genotipos del MM) siguientes:

$$\begin{aligned} Z_{i1} &= \theta_1 X_1 + 2\theta_2 X_2 + \theta_3 X_3, & Z_{i2} &= \theta_2 X_1 + (1-2\theta_2)X_2 + \theta_2 X_3 \\ y \\ Z_{i3} &= \theta_3 X_1 + 2\theta_2 X_2 + \theta_1 X_3 \end{aligned} \quad (1)$$

donde, $\theta_1=(1-r)^2$, $\theta_2=r(1-r)$ y $\theta_3=r^2$; r se definió anteriormente. Las distribuciones mezcladas respectivas de Z_{i1} , Z_{i2} y Z_{i3} son:

$$\begin{aligned} g_{i1}(z_i) &= \theta_1 f_1(x_j) + 2\theta_2 f_2(x_j) + \theta_3 f_3(x_j), \\ g_{i2}(z_i) &= \theta_2 f_1(x_j) + (1-2\theta_2) f_2(x_j) + \theta_2 f_3(x_j) \quad y \\ g_{i3}(z_i) &= \theta_3 f_1(x_j) + 2\theta_2 f_2(x_j) + \theta_1 f_3(x_j) \end{aligned} \quad (2)$$

mientras que $f_q(x)$, $q=1, 2, 3$, es igual a

$$f_q(x_j) = \frac{1}{\sigma_q \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_q^2} (x_j - \mu_q)^2 \right\}. \text{ Por las Ecuaciones 1 y}$$

2, las medias (m_{ii}) y varianzas (s_{ii}^2) de las Z_{ii} son:

$$\begin{aligned} m_{i1} &= \theta_1 \mu_1 + 2\theta_2 \mu_2 + \theta_3 \mu_3, \\ m_{i2} &= \theta_2 \mu_1 + (1-2\theta_2) \mu_2 + \theta_2 \mu_3 \quad y \\ m_{i3} &= \theta_3 \mu_1 + 2\theta_2 \mu_2 + \theta_1 \mu_3 \end{aligned} \quad (3)$$

$$\begin{aligned} s_{i1}^2 &= \theta_1^2 \sigma_1^2 + 4\theta_2^2 \sigma_2^2 + \theta_3^2 \sigma_3^2, \\ s_{i2}^2 &= \theta_2^2 \sigma_1^2 + (1-2\theta_2)^2 \sigma_2^2 + \theta_2^2 \sigma_3^2 \quad y \\ s_{i3}^2 &= \theta_3^2 \sigma_1^2 + 4\theta_2^2 \sigma_2^2 + \theta_1^2 \sigma_3^2 \end{aligned} \quad (4a)$$

Las covarianzas ($s_{iir'}$) entre las Z_{ii} 's descritas en (1) son

$$\begin{aligned} s_{i12} &= \theta_2 \theta_1 \sigma_1^2 + 2(1-2\theta_2) \sigma_2^2 + \theta_3 \sigma_3^2, \\ s_{i13} &= \theta_3 \theta_1 (\sigma_3^2 + \sigma_1^2) + 4 \theta_2^2 + \sigma_2^2 \quad y \\ s_{i23} &= \theta_2 \theta_3 \sigma_1^2 + 2(1-2\theta_2) \sigma_2^2 + \theta_1 \sigma_3^2 \end{aligned} \quad (4b)$$

Identificación de ligamiento

Como $\mu_1=\mu+\alpha$, $\mu_2=\mu+\delta$ y $\mu_3=\mu+\alpha$ son las medias de las variables fenotípicas X_1 , X_2 y X_3 correspondientes a los genotipos Q_1Q_1 , Q_1Q_2 y Q_2Q_2 del QTL, μ es la media poblacional del QT; $\alpha=0.5(\mu_1-\mu_3)$ y $\delta=0.5(2\mu_2-\mu_1-\mu_3)$ son el efecto aditivo y la desviación de dominancia del QTL (Liu, 1998a), es posible escribir las medias (m_{ii}) de las Z_{ii} ($i=1, 2, \dots, k$; $t=1, 2, 3$) de la manera siguiente:

$$\begin{aligned} m_{i1} &= \mu + \alpha(1-2r) + 2\delta r(1-r), \\ m_{i2} &= \mu + \delta(1-2r+2r^2) \quad y \\ m_{i3} &= \mu - \alpha(1-2r) + 2\delta r(1-r) \end{aligned} \quad (5)$$

corresponding to the three genotypes of the three genotypes of the QTL (Table 1).

Random variables and mixed distributions

Let x_{1j} , x_{2j} and x_{3j} , $j=1, 2, \dots, n_{it}$, be the j -th phenotypic expressions of the genotypes Q_1Q_1 , Q_1Q_2 , and Q_2Q_2 of the QTL (n_{it} is the number of observations of the QT in the population defined by the t -th genotype of the i -th MM (M_i), $t=1, 2, 3$; t represents the genotype of the MM. Thus, $t=1$ indicates the genotype $A_{i1}A_{i1}$, $t=2$ the genotype $A_{i1}A_{i2}$, etc.). Furthermore, assume that the random variables X_1 , X_2 and X_3 , and [with means $E(X_q)=\mu_q$ and variances $\text{Var}(X_q)=\sigma_q^2$, $q=1,2,3$] denote the possible phenotypic values associated to the genotypes of the QTL (Q_1Q_1 , Q_1Q_2 , and Q_2Q_2 , respectively); are independent and have normal distribution. For the i -th MM, the three mixed random variables (corresponding to the three genotypes of the MM) can be defined as follows:

$$\begin{aligned} Z_{i1} &= \theta_1 X_1 + 2\theta_2 X_2 + \theta_3 X_3, & Z_{i2} &= \theta_2 X_1 + (1-2\theta_2)X_2 + \theta_2 X_3 \\ \text{and} \\ Z_{i3} &= \theta_3 X_1 + 2\theta_2 X_2 + \theta_1 X_3 \end{aligned} \quad (1)$$

where, $\theta_1=(1-r)^2$, $\theta_2=r(1-r)$ and $\theta_3=r^2$; r was defined previously. The respective mixed distributions of Z_{i1} , Z_{i2} and Z_{i3} are as follows:

$$\begin{aligned} g_{i1}(z_i) &= \theta_1 f_1(x_j) + 2\theta_2 f_2(x_j) + \theta_3 f_3(x_j), \\ g_{i2}(z_i) &= \theta_2 f_1(x_j) + (1-2\theta_2) f_2(x_j) + \theta_2 f_3(x_j) \quad \text{and} \\ g_{i3}(z_i) &= \theta_3 f_1(x_j) + 2\theta_2 f_2(x_j) + \theta_1 f_3(x_j) \end{aligned} \quad (2)$$

whereas $f_q(x)$, $q=1, 2, 3$, is equal to

$$f_q(x_j) = \frac{1}{\sigma_q \sqrt{2\pi}} \exp \left\{ -\frac{1}{2\sigma_q^2} (x_j - \mu_q)^2 \right\}. \text{ For Equations 1 and}$$

2, the means (m_{ii}) and varianzas (s_{ii}^2) of the Z_{ii} are:

$$\begin{aligned} m_{i1} &= \theta_1 \mu_1 + 2\theta_2 \mu_2 + \theta_3 \mu_3, \\ m_{i2} &= \theta_2 \mu_1 + (1-2\theta_2) \mu_2 + \theta_2 \mu_3 \quad \text{and} \\ m_{i3} &= \theta_3 \mu_1 + 2\theta_2 \mu_2 + \theta_1 \mu_3 \end{aligned} \quad (3)$$

$$\begin{aligned} s_{i1}^2 &= \theta_1^2 \sigma_1^2 + 4\theta_2^2 \sigma_2^2 + \theta_3^2 \sigma_3^2, \\ s_{i2}^2 &= \theta_2^2 \sigma_1^2 + (1-2\theta_2)^2 \sigma_2^2 + \theta_2^2 \sigma_3^2 \quad \text{and} \\ s_{i3}^2 &= \theta_3^2 \sigma_1^2 + 4\theta_2^2 \sigma_2^2 + \theta_1^2 \sigma_3^2 \end{aligned} \quad (4a)$$

The covarianzas ($s_{iir'}$) among the Z_{ii} 's described in (1) are as follows:

$$\begin{aligned} s_{i12} &= \theta_2 \theta_1 \sigma_1^2 + 2(1-2\theta_2) \sigma_2^2 + \theta_3 \sigma_3^2, \\ s_{i13} &= \theta_3 \theta_1 (\sigma_3^2 + \sigma_1^2) + 4 \theta_2^2 + \sigma_2^2 \quad \text{and} \\ s_{i23} &= \theta_2 \theta_3 \sigma_1^2 + 2(1-2\theta_2) \sigma_2^2 + \theta_1 \sigma_3^2 \end{aligned} \quad (4b)$$

Identification of linkage

As $\mu_1=\mu+\alpha$, $\mu_2=\mu+\delta$ and $\mu_3=\mu+\alpha$ are the means of the phenotypic values X_1 , X_2 and X_3 corresponding to the genotypes

Los contrastes $m_{i1}-m_{i3}=2\alpha(1-2r)$ y $2m_{i2}-(m_{i3}+m_{i1})=2\delta(1-2r)^2$ pueden definirse a partir de las Ecuaciones 5. A éstos se asocian, respectivamente, las hipótesis nulas: $H_0: m_{i1}=m_{i3}$ y $H_0: 2m_{i2}=m_{i1}+m_{i3}$. Con la prueba de cualquiera de éstas es posible identificar ligamiento entre el QTL y el MM ya que si $H_0: m_{i1}=m_{i3}$ es verdadera, $r=0.5$; lo que se significa que el QTL y el MM están en cromosomas diferentes. La argumentación es válida ya que se trabaja bajo el supuesto: $\mu_1 > \mu_3$. Esto asegura que $\alpha > 0$, lo que se garantiza eligiendo líneas contrastantes respecto al QT bajo estudio (por ejemplo, si el QT es altura de planta, se eligen dos líneas con altura media diferente).

De acuerdo con el último párrafo, la identificación de ligamiento entre el QTL y el MM, M_i , se reduce a la comparación de las medias de dos poblaciones. Supóngase los tamaños de muestra (cuyos individuos tienen genotipos marcadores, $A_{i1}A_{i1}$, $A_{i1}A_{i2}$ y $A_{i2}A_{i2}$): n_{i1} , n_{i2} y n_{i3} ; y que $\hat{m}_{i1}$, $\hat{m}_{i2}$, $\hat{m}_{i3}$, $\hat{s}_{i1}^2$, $\hat{s}_{i2}^2$, y $\hat{s}_{i3}^2$ son estimaciones insesgadas de las medias y varianzas de Z_{i1j} , Z_{i2j} y Z_{i3j} (Ecuaciones 3 y 4a). Para probar la hipótesis $H_0: m_{i1}=m_{i3}$, un estadístico de prueba adecuado es (Liu, 1998a):

$$T_i = \frac{\hat{m}_{i1} - \hat{m}_{i3}}{\sqrt{\frac{n_{i1} + n_{i3}}{n_{i1}n_{i3}}} \hat{s}_{ip}^2} \quad (6)$$

donde, $\hat{s}_{ip}^2 = \frac{(n_{i1}-1)\hat{s}_{i1}^2 + (n_{i3}-1)\hat{s}_{i3}^2}{n_{i1} + n_{i3} - 2}$ es el estimador de la varianza de la diferencia; T_i se distribuye como la *t-Student* con $n_{i1}+n_{i3}-2$ grados de libertad. Si $|T_i| > t_{n_{i1}+n_{i3}-2, \lambda/2}$ ($t_{n_{i1}+n_{i3}-2, \lambda/2}$ es un valor de la distribución *t-Student* tal que $P\{|T_i| > t_{n_{i1}+n_{i3}-2, \lambda/2}\} = \lambda$; λ es el nivel de significancia de la prueba) existirá evidencia estadística de la presencia de, al menos, un QTL ligado al marcador.

Para ubicar el QTL sobre el cromosoma se construye un *perfil de verosimilitud* (Lander y Botstein, 1989) graficando el valor correspondiente de T_i vs la posición de cada marcador sobre el cromosoma bajo estudio. El máximo absoluto de la gráfica indica la posición más probable del QTL sobre el cromosoma.

Combinaciones lineales

Una vez identificado el QTL, el problema siguiente es estimar su posición sobre el cromosoma y la proporción con que contribuye a la variabilidad fenotípica total del QT. Para ello es necesario estimar las medias del QT: μ_1 , μ_2 , μ_3 y la frecuencia de recombinación, r , entre el marcador y el QTL. Esto se hará recurriendo a combinaciones lineales de las variables aleatorias mezcladas. Las medias y varianzas de tales combinaciones permiten encontrar estimadores de los parámetros biométricos indicados. Considérense las variables aleatorias del sistema de Ecuaciones 1 que permiten formular las combinaciones lineales siguientes:

$$U = Z_{i1} + 2Z_{i2} + Z_{i3} = X_1 + 2X_2 + X_3,$$

Q_1Q_1 , Q_1Q_2 and Q_2Q_2 of the QTL, μ is the population mean of the QT; $\alpha=0.5(\mu_1-\mu_3)$ and $\delta=0.5(2\mu_2-\mu_1-\mu_3)$ are the additive effect and the deviation of dominance of the QTL (Liu, 1998a), it is possible to write the means (m_{ii}) of the Z_{ii} ($i=1, 2, \dots, k$; $t=1, 2, 3$) as follows:

$$\begin{aligned} m_{i1} &= \mu + \alpha(1-2r) + 2\delta r(1-r), \\ m_{i2} &= \mu + \delta(1-2r+2r^2) \quad \text{and} \\ m_{i3} &= \mu - \alpha(1-2r) + 2\delta r(1-r) \end{aligned} \quad (5)$$

The contrasts $m_{i1}-m_{i3}=2\alpha(1-2r)$ and $2m_{i2}-(m_{i3}+m_{i1})=2\delta(1-2r)^2$ can be defined from Equations 5. Associated to these are, respectively, the null hypotheses: $H_0: m_{i1}=m_{i3}$ and $H_0: 2m_{i2}=m_{i1}+m_{i3}$. With the proof of any one of these it is possible to identify linkage between the QTL and the MM, given that if $H_0: m_{i1}=m_{i3}$ is true, then $r=0.5$; which means that the QTL and the MM are in different chromosomes. The argumentation is valid, given that it falls under the assumption: $\mu_1 > \mu_3$. This insures that $\alpha > 0$, which is guaranteed by selecting contrasting lines with respect to the QTL under study (for example, if the QTL is plant height, two lines with different mean height are selected).

According to the previous paragraph, the identification of linkage between the QTL and the MM, M_i , is reduced to the comparison of the means of the two populations. Suppose that the sample sizes (whose individuals have marker genotypes, $A_{i1}A_{i1}$, $A_{i1}A_{i2}$ and $A_{i2}A_{i2}$): n_{i1} , n_{i2} and n_{i3} , and that $\hat{m}_{i1}$, $\hat{m}_{i2}$, $\hat{m}_{i3}$, $\hat{s}_{i1}^2$, $\hat{s}_{i2}^2$, y $\hat{s}_{i3}^2$ are unbiased estimations of the means and variances of Z_{i1j} , Z_{i2j} and Z_{i3j} (Equations 3 and 4a). To test the hypothesis $H_0: m_{i1}=m_{i3}$, an adequate test statistic is (Liu, 1998a):

$$T_i = \frac{\hat{m}_{i1} - \hat{m}_{i3}}{\sqrt{\frac{n_{i1} + n_{i3}}{n_{i1}n_{i3}}} \hat{s}_{ip}^2} \quad (6)$$

where, $\hat{s}_{ip}^2 = \frac{(n_{i1}-1)\hat{s}_{i1}^2 + (n_{i3}-1)\hat{s}_{i3}^2}{n_{i1} + n_{i3} - 2}$ is the estimator of the variance of the difference; T_i is distributed as the *t-Student* with $n_{i1}+n_{i3}-2$ degrees of freedom. If $|T_i| > t_{n_{i1}+n_{i3}-2, \lambda/2}$ ($t_{n_{i1}+n_{i3}-2, \lambda/2}$ is a value of the distribution *t-Student* so that $P\{|T_i| > t_{n_{i1}+n_{i3}-2, \lambda/2}\} = \lambda$; λ is the level of significance of the test), there will be statistical evidence of the presence of at least one QTL linked to the marker.

To locate the QTL on the chromosome, a *likelihood profile* (Lander and Botstein, 1989) is constructed, graphing the corresponding value of T_i vs the position of each marker on the chromosome under study. The absolute maximum of the graph indicates the most likely position of the QTL on the chromosome.

Linear combinations

Once the QTL has been identified, the next problem is to estimate its position on the chromosome and the proportion with which it

$$\begin{aligned} V &= Z_{i1} - 2Z_{i2} + Z_{i3} = (1-2r)^2 (X_1 - 2X_2 + X_3) \quad y \\ W &= Z_{i1} - Z_{i3} = (1-2r)(X_1 - X_3) \end{aligned} \quad (7)$$

Las medias de U , V y W son:

$$\begin{aligned} E(U) &= \mu_1 + 2\mu_2 + \mu_3 = m_u, \\ E(V) &= (1-2r)^2 (\mu_1 - 2\mu_2 + \mu_3) = m_v \quad y \\ E(W) &= (1-2r)(\mu_1 - \mu_3) = m_w \end{aligned} \quad (8)$$

En tanto que sus varianzas se expresan como:

$$\begin{aligned} s_u^2 &= \sigma_1^2 + 4\sigma_2^2 + \sigma_3^2 \\ s_v^2 &= (1-2r)^4 (\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2) \quad y \\ s_w^2 &= (1-2r)^2 (\sigma_1^2 + \sigma_3^2) \end{aligned} \quad (9)$$

y sus covarianzas son: $s_{uv} = (1-2r)^2 (\sigma_1^2 - 2\sigma_2^2 + \sigma_3^2)$, $s_{uw} = (1-2r)(\sigma_1^2 - \sigma_3^2)$ y $s_{vw} = (1-2r)^3 (\sigma_1^2 - \sigma_3^2)$. Como U , V y W son combinaciones lineales de variables aleatorias con distribución normal, su distribución también será normal. Además, bajo el supuesto: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$, U , V y W serán independientes.

RESULTADOS

Estimadores de los parámetros biométricos

Con las Ecuaciones 3, 4a, 4b, 5 y 9 se derivan estimadores para la frecuencia de recombinación, r , y para las medias: μ_1 , μ_2 y μ_3 . Éstos, a su vez, permiten estimar el efecto aditivo (α) y de dominancia (δ) del QTL. Un estimador insesgado de r se obtiene del co-

$$\text{cient} \quad \frac{s_v^2}{s_u^2} = \frac{(1-2r)^4 (\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2)}{\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2} = (1-2r)^4.$$

Este estimador (Apéndice B) es:

$$\hat{r} = \frac{1}{2} \left\{ 1 - \frac{8(n+2)}{8(n+2)+1} \sqrt{\frac{\hat{s}_v}{\hat{s}_u}} \right\} \quad (10)$$

$$\hat{s}_v = \sqrt{\hat{s}_{i1}^2 + 4\hat{s}_{i2}^2 + \hat{s}_{i3}^2 + 2|\hat{s}_{i13}| - (2|\hat{s}_{i12}| + 2|\hat{s}_{i23}|)}$$

y

$$\hat{s}_u = \sqrt{\hat{s}_{i1}^2 + 4\hat{s}_{i2}^2 + \hat{s}_{i3}^2 + 2(2|\hat{s}_{i12}| + 2|\hat{s}_{i13}| + 2|\hat{s}_{i23}|)}$$

En $\hat{s}_v$ y $\hat{s}_u$ el símbolo $|\cdot|$ se utiliza para expresar el valor absoluto de las estimaciones de las covarianzas. Esto asegura que $0 \leq \hat{r} \leq 1/2$.

contributes the total phenotypic variability of the QT. For this purpose it is necessary to estimate the means of the QT: μ_1 , μ_2 , μ_3 and the frequency of recombination, r , between the marker and the QTL. This will be done by recurring to linear combinations of the mixed random variables. The means and variances of these combinations make it possible to find estimators of the indicated biometrics parameters. Consider the random variables of the system of Equations 1 which permit the formulation of the following linear combinations:

$$\begin{aligned} U &= Z_{i1} + 2Z_{i2} + Z_{i3} = X_1 + 2X_2 + X_3, \\ V &= Z_{i1} - 2Z_{i2} + Z_{i3} = (1-2r)^2 (X_1 - 2X_2 + X_3) \quad \text{and} \\ W &= Z_{i1} - Z_{i3} = (1-2r)(X_1 - X_3) \end{aligned} \quad (7)$$

The means of U , V and W are as follows:

$$\begin{aligned} E(U) &= \mu_1 + 2\mu_2 + \mu_3 = m_u, \\ E(V) &= (1-2r)^2 (\mu_1 - 2\mu_2 + \mu_3) = m_v \quad \text{and} \\ E(W) &= (1-2r)(\mu_1 - \mu_3) = m_w \end{aligned} \quad (8)$$

Whereas their variances are expressed as:

$$\begin{aligned} s_u^2 &= \sigma_1^2 + 4\sigma_2^2 + \sigma_3^2 \\ s_v^2 &= (1-2r)^4 (\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2) \quad \text{and} \\ s_w^2 &= (1-2r)^2 (\sigma_1^2 + \sigma_3^2) \end{aligned} \quad (9)$$

and their covariances are: $s_{uv} = (1-2r)^2 (\sigma_1^2 - 2\sigma_2^2 + \sigma_3^2)$, $s_{uw} = (1-2r)(\sigma_1^2 - \sigma_3^2)$ and $s_{vw} = (1-2r)^3 (\sigma_1^2 - \sigma_3^2)$. Because U , V and W are linear combinations of random variables with normal distribution, their distribution will also be normal. In addition, under the assumption: $\sigma_1^2 = \sigma_2^2 = \sigma_3^2$, U , V and W will be independent.

RESULTS

Estimators of the biometrics parameters

With Equations 3, 4a, 4b, 5 and 9, estimators re derived for the frequency of recombination, r , and for the means: μ_1 , μ_2 and μ_3 . These in turn make it possible to estimate the additive effect (α) and the dominance effect (δ) of the QTL. An unbiased estimator of r is obtained from the quotient

$$\frac{s_v^2}{s_u^2} = \frac{(1-2r)^4 (\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2)}{\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2} = (1-2r)^4. \quad \text{This estimator (Appendix B) is:}$$

$$\hat{r} = \frac{1}{2} \left\{ 1 - \frac{8(n+2)}{8(n+2)+1} \sqrt{\frac{\hat{s}_v}{\hat{s}_u}} \right\} \quad (10)$$

Los estimadores de las medias de las distribuciones fenotípicas asociadas a los genotipos de un QTL se obtienen a partir de las Ecuaciones 8. De éstas resulta:

$$\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 1/4 & -1/4 & 0 \\ 1/4 & 1/4 & -1/2 \end{bmatrix} \begin{bmatrix} \hat{m}_u \\ \hat{m}_v / (1-2\hat{r})^2 \\ \hat{m}_w / (1-2\hat{r}) \end{bmatrix}$$

$$\begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 1/4 & -1/4 & 0 \\ 1/4 & 1/4 & -1/2 \end{bmatrix} \text{ es la matriz inversa de}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & -2 & 1 \\ 1 & 0 & -1 \end{bmatrix}$$

Lo anterior permite obtener los estimadores de μ_1 , μ_2 y μ_3 siguientes:

$$\begin{aligned} \hat{\mu}_1 &= \frac{\hat{m}_u}{4} + \frac{\hat{m}_v}{4(1-2\hat{r})^2} + \frac{\hat{m}_w}{2(1-2\hat{r})}, \\ \hat{\mu}_2 &= \frac{1}{4} \left[\hat{m}_u - \frac{1}{(1-2\hat{r})^2} \hat{m}_v \right] \quad \text{y} \\ \hat{\mu}_3 &= \frac{\hat{m}_u}{4} + \frac{\hat{m}_v}{4(1-2\hat{r})^2} - \frac{\hat{m}_w}{2(1-2\hat{r})} \end{aligned} \quad (11)$$

donde,

$$\begin{aligned} \hat{m}_u &= \hat{m}_{i1} + 2\hat{m}_{i2} + \hat{m}_{i3}, \quad \hat{m}_v = \hat{m}_{i1} - 2\hat{m}_{i2} + \hat{m}_{i3} \quad \text{y} \\ \hat{m}_w &= \hat{m}_{i1} - \hat{m}_{i3}. \end{aligned}$$

Con los estimadores $\hat{\mu}_1$, $\hat{\mu}_2$ y $\hat{\mu}_3$, el efecto aditivo (α), la desviación de dominancia (δ) y la varianza genotípica (σ_G^2) del QTL tienen, respectivamente, como estimadores:

$$\begin{aligned} \hat{\alpha} &= 0.5(\hat{\mu}_1 - \hat{\mu}_3), \quad \hat{\delta} = 0.5(2\hat{\mu}_2 - \hat{\mu}_1 - \hat{\mu}_3) \quad \text{y} \\ \hat{\sigma}_G^2 &= \frac{1}{2}\hat{\alpha}^2 + \frac{1}{4}\hat{\delta}^2 \end{aligned} \quad (12)$$

En el Apéndice B se exponen con detalle las propiedades muestrales de los estimadores:

$$\hat{\mu}_1, \hat{\mu}_2, \hat{\mu}_3, \hat{\alpha}, \hat{\delta} \text{ y } \hat{\sigma}_G^2 \text{ (Ecuaciones 11 y 12).}$$

$$\hat{s}_v = \sqrt{\hat{s}_{i1}^2 + 4\hat{s}_{i2}^2 + \hat{s}_{i3}^2 + 2|\hat{s}_{i13}| - (2|\hat{s}_{i12}| + 2|\hat{s}_{i23}|)}$$

and

$$\hat{s}_u = \sqrt{\hat{s}_{i1}^2 + 4\hat{s}_{i2}^2 + \hat{s}_{i3}^2 + 2(2|\hat{s}_{i12}| + 2|\hat{s}_{i13}| + 2|\hat{s}_{i23}|)}$$

In $\hat{s}_v$ and $\hat{s}_u$ the symbol $|\cdot|$ is used to express the absolute value of the estimations of the covariances. This insures that $0 \leq \hat{r} \leq 1/2$.

The estimators of the means of the phenotypic distributions associated to the genotypes of a QTL are obtained from Equations 8. From these the following is obtained:

$$\begin{bmatrix} \hat{\mu}_1 \\ \hat{\mu}_2 \\ \hat{\mu}_3 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 1/4 & -1/4 & 0 \\ 1/4 & 1/4 & -1/2 \end{bmatrix} \begin{bmatrix} \hat{m}_u \\ \hat{m}_v / (1-2\hat{r})^2 \\ \hat{m}_w / (1-2\hat{r}) \end{bmatrix}$$

$$\begin{bmatrix} 1/4 & 1/4 & 1/2 \\ 1/4 & -1/4 & 0 \\ 1/4 & 1/4 & -1/2 \end{bmatrix} \text{ is the inverse matrix of}$$

$$\begin{bmatrix} 1 & 2 & 1 \\ 1 & -2 & 1 \\ 1 & 0 & -1 \end{bmatrix}$$

The above makes it possible to obtain the following estimators of μ_1 , μ_2 and μ_3 :

$$\begin{aligned} \hat{\mu}_1 &= \frac{\hat{m}_u}{4} + \frac{\hat{m}_v}{4(1-2\hat{r})^2} + \frac{\hat{m}_w}{2(1-2\hat{r})}, \\ \hat{\mu}_2 &= \frac{1}{4} \left[\hat{m}_u - \frac{1}{(1-2\hat{r})^2} \hat{m}_v \right] \quad \text{and} \\ \hat{\mu}_3 &= \frac{\hat{m}_u}{4} + \frac{\hat{m}_v}{4(1-2\hat{r})^2} - \frac{\hat{m}_w}{2(1-2\hat{r})} \end{aligned} \quad (11)$$

where,

$$\begin{aligned} \hat{m}_u &= \hat{m}_{i1} + 2\hat{m}_{i2} + \hat{m}_{i3}, \quad \hat{m}_v = \hat{m}_{i1} - 2\hat{m}_{i2} + \hat{m}_{i3} \quad \text{and} \\ \hat{m}_w &= \hat{m}_{i1} - \hat{m}_{i3}. \end{aligned}$$

With the estimators $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$, the additive effect (α), the deviation of dominance (δ) and the genotypic variance (σ_G^2) of the QTL have respectively as estimators:

Construcción del perfil de verosimilitud

Kearsey y Hyne (1994) estudiaron el caso hipotético de un QTL ubicado a 50 centi-Morgans (cM) del primer marcador en un cromosoma de 100 cM con seis marcadores ubicados (de izquierda a derecha sobre el cromosoma) a 0, 20, 40, 60, 80 y 100 cM; supusieron los valores $\alpha=1.0$, $\delta=0.5$ y una variabilidad fenotípica total de 5.0. En el Cuadro 2 se muestran las estimaciones de las medias fenotípicas y de las posiciones de los marcadores obtenidos por estos autores. Para construir el perfil de verosimilitud que permita identificar el QTL sobre el cromosoma es necesario calcular el valor del estadístico de prueba (EP) para cada marcador.

Para poder evaluar el EP es necesario contar con estimaciones de dos medias, $\hat{m}_{i1}$ y $\hat{m}_{i3}$, y de dos varianzas muestrales: $\hat{s}_{i1}^2$ y $\hat{s}_{i3}^2$. Las primeras aparecen en el Cuadro 2. Los valores de $\hat{s}_{i1}^2$ y $\hat{s}_{i3}^2$ fueron estimados, en ambos casos (para los seis marcadores), por Kearsey y Hyne (1994) en 4.031; el tamaño de las muestras fue $n_{i1}=n_{i3}=75$. Para el primero y segundo marcador se tendrá (Ecuación 6):

$$T_1 = \frac{23.5669 - 22.5475}{\sqrt{\frac{150}{5625} \left[\frac{2(74)(4.031)}{148} \right]}} = 2.931$$

y

Cuadro 2. Estimaciones de las tres medias fenotípicas de las subpoblaciones de una F_2 correspondientes a cada uno de seis marcadores moleculares (M_i) ubicados a 0, 20, 40, 60, 80 y 100 centi-Morgans (cM). Se consideró un QTL ubicado a 50 cM de M_i y los valores: $\alpha=1.0$, $\delta=0.5$ y $h^2=0.1$ (Kearsey y Hyne, 1994).

Table 2. Estimations of the three phenotypic means of the subpopulations of an F_2 corresponding to each of the six molecular markers (M_i) located at 0, 20, 40, 60, 80 and 100 centi-Morgans (cM). A QTL was considered located at 50 cM from M_i and the values: $\alpha=1.0$, $\delta=0.5$ and $h^2=0.1$ (Kearsey and Hyne, 1994).

Marcador (M_i)	Posición del marcador M_i		Medias estimadas de las poblaciones mezcladas		
	asignada (cM)	estimada (cM)	$\hat{m}_{i1}$	$\hat{m}_{i2}$	$\hat{m}_{i3}$
1	0.0	0.0	23.566	22.883	22.547
2	20.0	20.6	23.961	22.865	22.363
3	40.0	45.2	23.899	23.077	21.675
4	60.0	64.5	24.028	23.086	21.676
5	80.0	82.5	23.617	23.038	22.121
6	100.0	103.7	23.166	23.027	22.591

$$\hat{\alpha} = 0.5(\hat{\mu}_1 - \hat{\mu}_3), \hat{\delta} = 0.5(2\hat{\mu}_2 - \hat{\mu}_1 - \hat{\mu}_3) \text{ y}$$

$$\hat{\sigma}_G^2 = \frac{1}{2}\hat{\alpha}^2 + \frac{1}{4}\hat{\delta}^2 \quad (12)$$

In Appendix B the sampling properties are exposed in detail of the estimators: (Equations 10, 11 and 12).

Construction of the likelihood profile

Kearsey and Hyne (1994) studied the hypothetical case of a QTL located at 50 centi-Morgans (cM) from the first marker in a chromosome of 100 cM with six markers located (from left to right on the chromosome) at 0, 20, 40, 60, 80 and 100 cM; assumed the values $\alpha=1.0$, $\delta=0.5$ and a total phenotypic variability of 5.0. The estimations of the phenotypic means and of the positions of the markers obtained by these authors are shown in Table 2. To construct the likelihood profile that permits the identification of the QTL on the chromosome, it is necessary to calculate the value of the test statistic (TS) for each marker.

In order to evaluate the TS, it is necessary to have estimations of two means, $\hat{m}_{i1}$ and $\hat{m}_{i3}$, and of two sampling variances: $\hat{s}_{i1}^2$ and $\hat{s}_{i3}^2$. The first appear in Table 2. The values of $\hat{s}_{i1}^2$ and $\hat{s}_{i3}^2$ were estimated, in both cases (for the six markers), by Kearsey and Hyne (1994) in 4.031; the size of the samples was $n_{i1}=n_{i3}=75$. For the first and second marker, we have the following (Equation 6):

$$T_1 = \frac{23.5669 - 22.5475}{\sqrt{\frac{150}{5625} \left[\frac{2(74)(4.031)}{148} \right]}} = 2.931$$

and

$$T_2 = \frac{23.9616 - 22.3639}{\sqrt{\frac{150}{5625} \left[\frac{2(74)(4.031)}{148} \right]}} = 4.842$$

The values of the TS for the remaining MMs were calculated in a similar way. This made it possible to construct the likelihood profile (Figure 1) where the maximum value indicates that, very probably (Lander and Botstein, 1989), the QTL is between marker 3 and 4.

Estimation of the biometrics parameters

The basic information of the four QTL located in different hypothetical chromosomes in a population

$$T_2 = \frac{23.9616 - 22.3639}{\sqrt{\frac{150}{5625} \left[\frac{2(74)(4.031)}{148} \right]}} = 4.842$$

De manera similar se calcularon los valores de los EP's para los MM's restantes. Esto permitió construir el perfil del verosimilitud (Figura 1) donde el valor máximo indica que, muy probablemente (Lander y Botstein, 1989), el QTL está entre el marcador 3 y 4.

Estimación de los parámetros biométricos

En el Cuadro 3 se muestra la información básica de cuatro QTL ubicados en diferentes cromosomas hipotéticos en un estudio de simulación para una población F₂ (Romero-Padilla *et al.*, 2004). En éste los parámetros del QTL se estimaron utilizando el paquete estadístico MAPMAKER (Lander *et al.*, 1987). Tales estimaciones se muestran en el Cuadro 4 junto con las estimaciones obtenidas con el método derivado en el presente trabajo. Excepto por el primer caso, no existen diferencias significativas entre los resultados obtenidos por ambos métodos. Sin embargo, en el trabajo de Romero-Padilla *et al.* (2004) se utilizó más información para la estimación, ya que se consideraron dos marcadores adyacentes al QTL de manera simultánea.

Podría pensarse que la diferencia observada en el primer caso se debe a que la distancia mapa entre el MM y el QTL es 23.45 cM mientras que en los restantes la distancia indicada es menor o igual a 10.0 cM. Esto parece una limitante, sin embargo no es así, ya que Darvasi *et al.* (1993) demostraron que para identificar

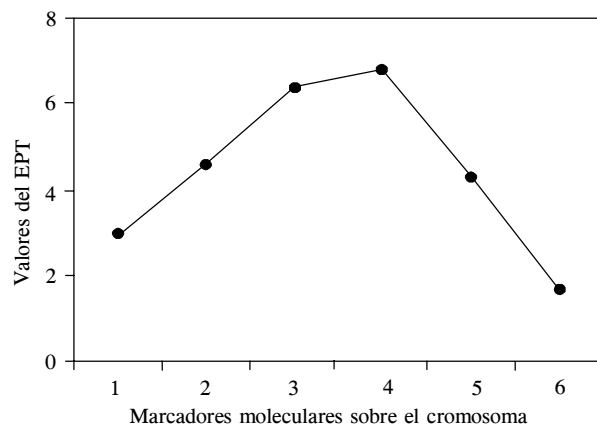


Figura 1. Perfil de verosimilitud para identificar un QTL ubicado entre los marcadores 3 y 4 en un cromosoma de 100 cM.

Figure 1. Profile of likelihood to identify a QTL located between markers 3 and 4 on a chromosome of 100 cM.

simulation study F₂ (Romero-Padilla *et al.*, 2004) is shown in Table 3. In this study, the parameters of the QTL were estimated using the MAPMAKER statistical package (Lander *et al.*, 1987). These estimations are shown in Table 4 along with the estimations obtained with the method derived in the present study. Except for the first case, there are no significant differences between the results obtained for both methods. However, in the work of Romero-Padilla *et al.* (2004), more information was used for the estimation, as they considered two markers adjacent to the QTL simultaneously.

It could be considered that the difference observed in the first case is due to the fact that the map distance between the MM and the QTL is 23.45 cM, whereas in the rest, the indicated distance is less than or equal to 10.0 c. This appears to be a limitation, however, this is not the case, given that Darvasi *et al.* (1993) demonstrated that to identify a QTL, the optimum distance between the MM and the QTL is precisely 10 cM.

Cuadro 3. Efectos asignados a cuatro QTL; marcadores entre los que se encuentra cada QTL y valores hipotéticos de la frecuencia de recombinación entre el primer MM (en este caso el izquierdo) y el QTL en una población F₂ (Romero-Padilla *et al.*, 2004).

Table 3. Effects assigned to four QTL; markers among which each QTL is found and hypothetical values of the recombination frequency between the first MM (in this case the left) and the QTL in an F₂ population (Romero-Padilla *et al.*, 2004).

QTL	Efecto del QTL	Marcadores entre los que está el QTL	Frecuencia de recombinación entre el MM y el QTL (r)
1	1.13	1-2	0.16
2	0.62	3-4	0.02
3	-0.98	5-6	0.03
5	0.76	9-10	0.03

Cuadro 4. Estimaciones de los parámetros biométricos de cuatro QTL (Cuadro 3) con MAPMAKER y con el procedimiento derivado en el presente trabajo.

Table 4. Estimations of the biometrics parameters of four QTL (Table 3) with MAPMAKER and with the procedure derived in the present study.

QTL	Con MAPMAKER			Con el método propuesto			
	$\hat{\alpha}$	$\hat{\delta}$	$\hat{\sigma}_g^2$	$\hat{\alpha}$	$\hat{\delta}$	$\hat{r}$	$\hat{\sigma}_g^2$
1	1.12	0.20	0.65	0.86	0.10	0.03	0.37
2	0.71	0.12	0.26	0.67	-0.13	0.03	0.23
3	-0.99	-0.60	0.60	-1.00	0.63	0.04	0.62
5	0.65	0.34	0.24	0.64	-0.40	0.02	0.24

un QTL la distancia óptima entre el MM y el QTL es, precisamente, 10 cM.

DISCUSIÓN

El procedimiento para la identificación de QTL ligados a MM's en el que se analiza un MM y un QTL a la vez (MQTL) proporciona toda la información necesaria en la investigación de QTL. Sus desventajas son: a) el método no permite determinar si el marcador está ligado a uno o más QTL; b) requiere un tamaño de muestra relativamente grande para que las pruebas sean estadísticamente significativas; c) el criterio de decisión con el nivel de significancia estándar (0.05) pasa por alto el hecho de que varios marcadores se incluyen en la prueba; así, mientras la probabilidad de cometer un error tipo uno (ETI) en cualquier marcador es sólo 5%, la probabilidad de que al menos un ETI ocurra en una de las varias hipótesis que se prueban sobre el genoma completo es mucho mayor (Lander y Botstein, 1989); d) confunde la fracción de recombinación con el efecto del QTL, lo que conduce a subestimar el efecto del QTL. Sin embargo, el argumento del punto (c) no es privativo del método MQTL, sino que es un problema asociado a la búsqueda de QTL sin importar el método utilizado (Van Ooijen, 1999; Cheverud, 2001). Respecto al punto (d) el argumento es válido sólo en el contexto de la regresión lineal. En el presente trabajo se ha demostrado (Ecuaciones 10, 11 y 12) que el efecto del QTL y el de la frecuencia de recombinación pueden estimarse por separado.

El método MQM y mapeo por intervalos compuestos (CIM) tienen también problemas asociados; algunos de ellos son: 1) la identificación de ligamiento no es una prueba de intervalo, *i.e.*, una prueba que permita distinguir si existe o no un QTL en el intervalo definido por los MM's y que no se vea afectada por QTL fuera de él; 2) si varios QTL sobre el mismo cromosoma afectan al QT de interés, los estimadores de las posiciones de los QTL y sus efectos están sesgados (Whittaker *et al.*, 1996); 3) la estimación de los parámetros biométricos, y de las frecuencias de recombinación, sólo es posible por métodos numéricos; 4) no se conoce la distribución asintótica del estadístico de prueba para identificar ligamiento, lo que dificulta la determinación del nivel de significancia de la prueba. Esto último se debe a que las hipótesis no están apropiadamente anidadas (Churchill y Doerge, 1995), *i.e.*, las hipótesis bajo prueba comparan una distribución mezclada bajo la hipótesis alternativa con una distribución no mezclada bajo la hipótesis nula. Haley y Knott (1992) estudiaron las propiedades del estadístico de MQM y concluyeron que la aproximación

DISCUSSION

The procedure for the identification of QTLs linked to MMs in which an MM and a QTL are analyzed at the same time (MQTL) provides all of the necessary information in the investigation of a QTL. Its disadvantages are: a) the method does not allow the determination of whether the marker is linked to one or more QTL; b) it requires a relatively large size sample for the tests to be statistically significant; c) the criteria of decision with the standard level of significance (0.05) overlooks the fact that various markers are included in the test; thus, whereas the likelihood of committing a type one error (ETI) in any marker is only 5%, the likelihood of at least one ETI occurring in one of the various hypotheses that are being tested on the complete genome is much higher (Lander and Botstein, 1989); d) the fraction of recombination is confused with the effect of the QTL, which leads to a sub-estimation of the effect of the QTL. However, the argument of point (c) is not restricted to the MQTL method, but rather, it is a problem associated to the search of QTL regardless of the method used (Van Ooijen, 1999; Cheverud, 2001). With respect to point (d), the argument is valid only in the context of the linear regression. In the present study, it has been demonstrated (Equations 10, 11 and 12) that the effect of the QTL and that of the recombination frequency can be estimated separately.

The MQM method and composed intervals mapping (CIM) also have associated problems; some of them are: 1) the identification of linkage is not a proof of interval, *i.e.*, a test that makes it possible to distinguish whether or not a QTL exists in the interval defined by the MMs and that is not affected by QTL outside of it; 2) if various QTL on the same chromosome affect the QT of interest, the estimators of the positions of the QTL and their effects are biased (Whittaker *et al.*, 1996); 3) the estimation of the biometrics parameters, and of the recombination frequencies, is only possible by numerical methods; 4) the asymptotic distribution is not known of the test statistic for identifying linkage, which makes it difficult to determine the significance level of the test. This is due to the fact that the hypotheses are not appropriately clustered (Churchill and Doerge, 1995), *i.e.*, the hypotheses being tested compare a mixed distribution under the alternative hypothesis with an unmixed distribution under the null hypothesis. Haley and Knott (1992) studied the properties of the MQM statistic and concluded that the approximation to the chi-squared as an asymptotic distribution is not reliable in many cases and is at least questionable in others.

a la chi-cuadrada como una distribución asintótica no es confiable en muchos casos y es al menos cuestionable en algunos.

CONCLUSIONES

Las estimaciones derivadas en este trabajo con base en las combinaciones lineales de variables aleatorias con distribución normal mezclada para estimar los efectos aditivos y de dominancia del QTL y la frecuencia de recombinación entre éste y el MM resultaron consistentes e insesgadas. Para un caso reportado en la literatura, sus resultados se compararon con los que se obtienen utilizando MAPMAKER y se mostró que son muy similares. Sin embargo, el procedimiento aquí derivado tiene la ventaja de la simplicidad y sencillez algebraica, lo que facilita la comprensión de las ideas fundamentales subyacentes en la construcción de mapas de QTL's.

LITERATURA CITADA

Blangero, J., J. T. Williams, and L. Almasy. 2001. Variance component methods for detecting complex trait loci. *Adv. Genet.* 42: 151-181.

Cheverud, J. M. 2001. A simple correction for multiple comparisons in interval mapping genome scans. *Heredity* 87: 52-58.

Churchill, G. A., and R. W. Doerge. 1995. Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963-971.

Darvasi, A., A. Weinreb, V. Minke, J. I. Weller, and M. Soller. 1993. Detecting marker QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* 134: 943-951.

Elston R. C., and H. J. Cordell. 2001. Overview of model-free methods for linkage analysis. *Adv. Genet.* 42: 135-150.

George, M., D. Nielsen, M. Mackinnon, A. Mishra, R. Okimoto, A. T. Pasquino, L. S. Sargeant, A. Sorensen, M. R. Steel, X. Zhao, J. E. Womack, and I. Hoeschele. 1995. Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 139: 907-920.

Haley, C. S., and S. A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315-324.

Hoeschele, I., and P. M. Van Raden. 1993a. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. *Theor. Appl. Genet.* 85: 953-960.

Hoeschele, I., and P. M. Van Raden. 1993b. Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. *Theor. Appl. Genet.* 85: 946-952.

Hoeschele, I. 2003. Mapping quantitative trait loci in outbred pedigrees. In: *Handbook of Statistical Genetics*, 2nd Edition. Balding, D. Y., M. Bishop, and C. Cannings (eds). John Wiley and Sons, England. 1: 477-525.

Hyne, V., and M. J. Kearsey. 1995. QTL analysis: further uses of "maker regression". *Theor. Appl. Genet.* 91: 471-476.

Jansen, R. C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205-211.

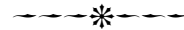
Jansen, R. C. 1996. A general Monte Carlo Method for mapping multiple quantitative trait loci. *Genetics* 142: 305-311.

Jansen, R. C. 2003. Quantitative trait loci in inbred lines. In: *Handbook of Statistical Genetics*, 2nd Edition. Balding D. J.,

CONCLUSIONS

The estimations derived in this study based on the linear combinations of random variables with mixed normal distribution to estimate the additive effects and those of dominance of the QTL and the recombination frequency between it and the MM, were consistent and unbiased. For a case reported in the literature, its results were compared with those that were obtained using MAPMAKER and it was shown that they are very similar. However, the procedure derived here has the advantage of algebraic simplicity, which facilitates the comprehension of the underlying fundamental ideas in the construction of maps of QTLs.

—End of the English version—



M. Bishop, and C. Cannings (eds). John Wiley and Sons, England, 1: 445-476.

Kearsey M. J., and V. Hyne. 1994. QTL analysis: A simple "marker-regression" approach. *Theor. Appl. Genet.* 89: 698-702.

Knapp, S. J., W. C. Bridges, and D. Birkes. 1990. Mapping quantitative trait loci using molecular marker linkage maps. *Theor. Appl. Genet.* 79: 583-592.

Lander, E. S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP's linkage maps. *Genetics* 121: 185-199.

Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daley, S. E. Lincol, and L. Newburg. 1987. MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1: 174-181.

Liu, B. H. 1998a. Statistical Genomics: Linkage, Mapping and QTL Analysis. CRC Pres., Boca Raton, New York. 611 p.

Liu, B. H. 1998b. Computational tools for study of complex trait: Compilation and distribution of data on complex traits. In: *Molecular Dissection of complex traits*, Paterson, A. H. (ed). USA, CRC Press, New York. pp: 43-79.

Lynch, M., and B. Walsh. 1998. *Genetics and Analysis of Quantitative Traits*. Sinauer Associates, Inc. Publisher Sunderland, Massachusetts, USA. 980 p.

Martínez, O., and R. N. Curnow. 1992. Estimating the localization and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* 85: 480-485.

Romero-Padilla, J. M., J. Sahagún-Castellanos, G. Ramírez-Valverde y G. Rendón-Sánchez. 2004. Índice de selección genotípica apoyada en marcadores moleculares ligados. *Agrociencia* 38: 293-303.

Sillanpää, M. J., and E. Arjas. 1998. Bayesian mapping of multiple Quantitative Trait Loci from incomplete inbred line cross data. *Genetics* 148: 1373-1388.

Sillanpää, M. J., and E. Arjas. 1999. Bayesian mapping of multiple Quantitative Trait Loci from incomplete outbreed offspring data. *Genetics* 151: 1605-1619.

Van Ooijen, J. W. 1999. LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* 83: 613-624.

Weller, J. I. 1992. Statistical Methodologies for Mapping and Analysis of Quantitative Trait Loci. In: *Plant Genomes: Methods for Genetic and Physical Mapping*. Beckmann, J. S., and T. C.

- (eds). Kluwer Academic Publishers, The Netherlands. pp: 181-207.
- Whittaker J. C., R. Thompson, and P. M. Visscher. 1996. On the mapping of QTL by regression of phenotype on marker-type. *Heredity* 77: 23-32.
- Williams, C. G. 1998. QTL mapping in outbreed pedigrees. *In: Molecular Dissection of Complex Traits*, Paterson, A. H. (ed). USA, CRC Press, New York, pp: 81-95.
- Wright, A. J., and R. P. Mowers. 1994. Multiple regression for molecular marker, quantitative trait data from large F2 populations. *Theor. Appl. Genet.* 89: 305-312.
- Xu, S., and W. R. Atchley. 1995. A random model approach to interval mapping of QTL. *Genetics* 141: 1189-1197.
- Xu, S. 1998. Further investigation on the regression method of mapping quantitative trait loci. *Heredity* 80: 364-373.
- Xu, S., and C. Volg. 2000. Maximum likelihood analysis of quantitative trait loci under selective genotyping. *Heredity* 84: 525-537.
- Zeng, Z. B. 1994. Precision mapping of quantitative trait loci. *Genetics* 136: 1457-1468.

APÉNDICE A

Cuadro A.1. Direcciones de internet donde hay software estadístico para el análisis de QTL (Liu, 1998b).**Table A.1. Internet addresses where there is statistical software for QTL analysis (Liu, 1998b).**

Nombre del programa	Dirección en internet	Nombre de quien creó el programa
MAPMAKER/QTL	mapmaker @genome.wi.mit.edu	Eric Lander
QTLSTAT	sknapp @ helix.css.orst.edu	Steve Knapp
PGRI	benliu @ unity.ncsu.edu	Ben Liu
QTL Cartographer	basten @ esssjp.stat.ncsu.edu	Christopher Basten
MAPQTL	J.W.vanOOIJEN @ cpro.dlo.nl	Johan Van Ooijen
Map Manager QTL	kmanly @ mcbio.med.bufflo.edu	Kenneth Manly
QGENE	jcn 5 @ cornell.edu	James C. Nelson

APÉNDICE B

Análisis de los estimadores

Las propiedades de los estimadores (Ecuaciones 10, 11 y 12) se determinaron considerando el resultado siguiente respecto a la esperanza y a la varianza de un cociente de variables aleatorias (Lynch y Walsh, 1998): si U y V son variables aleatorias independientes (U y V no necesariamente son las definidas en las Ecuaciones 7), entonces, la esperanza y varianza de U/V , son, aproximadamente:

$$E\left(\frac{U}{V}\right) \cong \frac{E(U)}{E(V)} \left\{ 1 + \frac{Var(V)}{[E(V)]^2} \right\} \quad (B.1)$$

$$Var\left(\frac{U}{V}\right) \cong \left[\frac{E(U)}{E(V)} \right]^2 \left\{ \frac{Var(U)}{[E(U)]^2} + \frac{Var(V)}{[E(V)]^2} \right\} \quad (B.2)$$

Considérese el estimador $\hat{r}$ (Ecuación 10a); se demostrará que éste es asintóticamente insesgado. Supóngase que $n=n_u=n_v$ [n_u y n_v son, respectivamente, los tamaños de muestra de U y V (Ecuaciones 7)] y que $\hat{s}_v$ y $\hat{s}_u$ son independientes. La esperanza de $\hat{r}$ es

$$E(\hat{r}) = \frac{1}{2} \left[1 - \frac{8(n+2)}{8(n+2)+1} E\left(\frac{\sqrt{\hat{s}_v}}{\sqrt{\hat{s}_u}}\right) \right]$$

Por la Ecuación B.1,

$$E\left(\frac{\sqrt{\hat{s}_v}}{\sqrt{\hat{s}_u}}\right) \cong \frac{E(\sqrt{\hat{s}_v})}{E(\sqrt{\hat{s}_u})} \left\{ 1 + \frac{Var(\sqrt{\hat{s}_u})}{[E(\sqrt{\hat{s}_u})]^2} \right\} \quad (B.3)$$

APPENDIX B

Analysis of the estimators

The properties of the estimators (Equations 10, 11 and 12) were determined considering the following result with respect to the expectation and the variance of a quotient of random variables (Lynch and Walsh, 1998): if U and V they are independent random variables (U and V not necessarily those defined in Equations 7), then the expectation and variance of U/V are approximately:

$$E\left(\frac{U}{V}\right) \cong \frac{E(U)}{E(V)} \left\{ 1 + \frac{Var(V)}{[E(V)]^2} \right\} \quad (B.1)$$

$$Var\left(\frac{U}{V}\right) \cong \left[\frac{E(U)}{E(V)} \right]^2 \left\{ \frac{Var(U)}{[E(U)]^2} + \frac{Var(V)}{[E(V)]^2} \right\} \quad (B.2)$$

Consider the $\hat{r}$ estimator (Equation 10a); it will be demonstrated that it is asymptotically unbiased. Suppose that $n=n_u=n_v$ [n_u and n_v are, respectively, the sample sizes of U and V (Equations 7)] and that $\hat{s}_v$ and $\hat{s}_u$ are independent. The expectation of $\hat{r}$ is:

$$E(\hat{r}) = \frac{1}{2} \left[1 - \frac{8(n+2)}{8(n+2)+1} E\left(\frac{\sqrt{\hat{s}_v}}{\sqrt{\hat{s}_u}}\right) \right]$$

For Equation B.1,

$$E\left(\frac{\sqrt{\hat{s}_v}}{\sqrt{\hat{s}_u}}\right) \cong \frac{E(\sqrt{\hat{s}_v})}{E(\sqrt{\hat{s}_u})} \left\{ 1 + \frac{Var(\sqrt{\hat{s}_u})}{[E(\sqrt{\hat{s}_u})]^2} \right\} \quad (B.3)$$

Para determinar las esperanzas: $E(\sqrt{\hat{s}_v})$ y $E(\sqrt{\hat{s}_u})$, y la varianza $Var(\sqrt{\hat{s}_u})$, se utiliza el método delta (Lynch y Walsh, 1998). Éste se basa en la esperanza y varianza de la expansión de una serie de Taylor alrededor de una constante arbitraria, c . Así, supóngase que f es una función de x , entonces,

$$f = f(c) + (x - c) \frac{\partial f(c)}{\partial x} + (x - c)^2 \frac{\partial^2 f(c)}{2\partial x^2} + (x - c)^3 \frac{\partial^3 f(c)}{(3)(2)\partial x^3} + \dots$$

donde $f(c)$ es la función, f , evaluada en $x=c$. La primera y segunda derivada de f son suficientes para obtener resultados muy aproximados a los esperados (Lynch y Walsh, 1998). Supóngase que $f = \sqrt{\hat{s}_v}$; la primera y segunda derivadas de f respecto a $\hat{s}_v$ serán,

$$\text{respectivamente, } \frac{\partial f}{\partial \hat{s}_v} = \frac{1}{2\sqrt{\hat{s}_v}} \text{ y } \frac{\partial^2 f}{\partial \hat{s}_v^2} = -\frac{1}{4(\hat{s}_v)^{3/2}}.$$

Con $c = E(\hat{s}_v)$,

$$f = \sqrt{E(\hat{s}_v)} + \frac{[\hat{s}_v - E(\hat{s}_v)]}{2\sqrt{E(\hat{s}_v)}} - \frac{[\hat{s}_v - E(\hat{s}_v)]^2}{8[E(\hat{s}_v)]^{3/2}}$$

Como la esperanza y varianza de la desviación estándar muestral ($\hat{s}$) y la esperanza y varianza de la varianza muestral ($\hat{s}^2$) de variables con distribución normal, desviación estándar, σ , y varianza, σ^2 , son, respectivamente (Lynch y Walsh, 1998): $E(\hat{s}) = \sigma$,

$$Var(\hat{s}) = \frac{\sigma^2}{2(n+2)}, E(\hat{s}^2) = \sigma^2 \text{ y } Var(\hat{s}^2) = \frac{2(\sigma^2)^2}{n+2};$$

y, puesto que U y V tienen distribución normal,

$$\begin{aligned} E(f) &= E \left\{ \sqrt{E(\hat{s}_v)} + \frac{[\hat{s}_v - E(\hat{s}_v)]}{2\sqrt{E(\hat{s}_v)}} - \frac{[\hat{s}_v - E(\hat{s}_v)]^2}{8[E(\hat{s}_v)]^{3/2}} \right\} \\ &\equiv \sqrt{s_v} - \frac{s_v^2}{16(n+2)(s_v)^{3/2}} = \sqrt{s_v} \left(\frac{16(n+2)-1}{16(n+2)} \right) \\ &\equiv \sqrt{s_v} \end{aligned}$$

(B.4)

To determine the expectations: $E(\sqrt{\hat{s}_v})$, and $E(\sqrt{\hat{s}_u})$ the variance $Var(\sqrt{\hat{s}_u})$, the delta method is used (Lynch and Walsh, 1998). This is based on the expectation and variance of the expansion of a Taylor series around an arbitrary constant, c . Thus, supposing that f is a function of x , then,

$$f = f(c) + (x - c) \frac{\partial f(c)}{\partial x} + (x - c)^2 \frac{\partial^2 f(c)}{2\partial x^2} + (x - c)^3 \frac{\partial^3 f(c)}{(3)(2)\partial x^3} + \dots$$

where $f(c)$ is the function, f , evaluated in $x=c$. The first and second derivate of f are sufficient to obtain results that are very close to those expected (Lynch and Walsh, 1998). Suppose that $f = \sqrt{\hat{s}_v}$; the first and second derivatives of f with respect to $\hat{s}_v$ will be,

$$\text{respectively, } \frac{\partial f}{\partial \hat{s}_v} = \frac{1}{2\sqrt{\hat{s}_v}} \text{ y } \frac{\partial^2 f}{\partial \hat{s}_v^2} = -\frac{1}{4(\hat{s}_v)^{3/2}}.$$

With $c = E(\hat{s}_v)$,

$$f = \sqrt{E(\hat{s}_v)} + \frac{[\hat{s}_v - E(\hat{s}_v)]}{2\sqrt{E(\hat{s}_v)}} - \frac{[\hat{s}_v - E(\hat{s}_v)]^2}{8[E(\hat{s}_v)]^{3/2}}$$

As the expectation and variance of the sample standard deviation ($\hat{s}$) and the expectation and variance of the sample variance ($\hat{s}^2$) of variables with normal distribution, standard deviation, σ , and variance, σ^2 , are, respectively (Lynch and Walsh, 1998): $E(\hat{s}) = \sigma$;

$$Var(\hat{s}) = \frac{\sigma^2}{2(n+2)}, E(\hat{s}^2) = \sigma^2 \text{ and } Var(\hat{s}^2) = \frac{2(\sigma^2)^2}{n+2};$$

and, given that U and V have normal distribution,

$$\begin{aligned} E(f) &= E \left\{ \sqrt{E(\hat{s}_v)} + \frac{[\hat{s}_v - E(\hat{s}_v)]}{2\sqrt{E(\hat{s}_v)}} - \frac{[\hat{s}_v - E(\hat{s}_v)]^2}{8[E(\hat{s}_v)]^{3/2}} \right\} \\ &\equiv \sqrt{s_v} - \frac{s_v^2}{16(n+2)(s_v)^{3/2}} = \sqrt{s_v} \left(\frac{16(n+2)-1}{16(n+2)} \right) \\ &\equiv \sqrt{s_v} \end{aligned}$$

(B.4)

Un procedimiento similar al anterior permite encontrar la varianza de $f = \sqrt{\hat{s}_v}$. Ésta es

$$Var(f) = E \left\{ \left[\sqrt{E(\hat{s}_v)} + \frac{[\hat{s}_v - E(\hat{s}_v)]}{2\sqrt{E(\hat{s}_v)}} - \frac{[\hat{s}_v - E(\hat{s}_v)]^2}{8[E(\hat{s}_v)]^{3/2}} - E(f) \right]^2 \right\} \equiv \frac{s_v}{8(n+2)} \quad (B.5)$$

Por los resultados (B.4) y (B.5) la Ecuación B.3 es:

$$E \left(\frac{\sqrt{\hat{s}_v}}{\sqrt{\hat{s}_u}} \right) \equiv \sqrt{\frac{s_v}{s_u} \left[\frac{8(n+2)+1}{8(n+2)} \right]}, \text{ de donde es evidente}$$

que el estimador de $\hat{r}$ (Ecuación 10) es insesgado.

Similarmente, con el método delta se derivaron los resultados (asintóticos) siguientes respecto a $\hat{r}$, $(1-2\hat{r})$ y $(1-2\hat{r})^2$:

$$Var(\hat{r}) = \frac{(1-2r)^2}{16(n+2)}, \quad Var(1-2\hat{r}) = \frac{(1-2r)^2}{4(n+2)}$$

y

$$Var[(1-2\hat{r})^2] = \frac{2(1-2r)^4}{(n+2)} \quad (B.6)$$

$$(1-2\hat{r}) = \frac{8(n+2)}{8(n+2)+1} \sqrt{\frac{\hat{s}_v}{\hat{s}_u}} \text{ es un estimador insesgado de } (1-2r); \quad (B.7)$$

$$(1-2\hat{r})^2 = \frac{2(n+2)+1}{2(n+2)} \frac{\hat{s}_v}{\hat{s}_u}, \text{ es un estimador insesgado de } (1-2r)^2 \quad (B.8)$$

Con el método delta y las Ecuaciones B.6 a B.8 se demostró que los estimadores: $\hat{\mu}_1$, $\hat{\mu}_2$ y $\hat{\mu}_3$ (Ecuaciones 11) son consistentes e insesgados asintóticamente. Por ejemplo, considérese el sistema de Ecuaciones 11, la esperanza de $\hat{\mu}_1$ es:

$$E(\hat{\mu}_1) = E \left\{ \frac{\hat{m}_u}{4} \right\} + E \left\{ \frac{\hat{m}_v}{4(1-2\hat{r})^2} \right\} + E \left\{ \frac{\hat{m}_w}{2(1-2\hat{r})} \right\}$$

A procedure similar to the above makes it possible to find the variance of $f = \sqrt{\hat{s}_v}$. This is:

$$Var(f) = E \left\{ \left[\sqrt{E(\hat{s}_v)} + \frac{[\hat{s}_v - E(\hat{s}_v)]}{2\sqrt{E(\hat{s}_v)}} - \frac{[\hat{s}_v - E(\hat{s}_v)]^2}{8[E(\hat{s}_v)]^{3/2}} - E(f) \right]^2 \right\} \equiv \frac{s_v}{8(n+2)} \quad (B.5)$$

From the results (B.4) and (B.5), Equation B.3 is:

$$E \left(\frac{\sqrt{\hat{s}_v}}{\sqrt{\hat{s}_u}} \right) \equiv \sqrt{\frac{s_v}{s_u} \left[\frac{8(n+2)+1}{8(n+2)} \right]}, \text{ from which it is}$$

evident that the estimator of $\hat{r}$ (Equation 10) is unbiased.

Similarly, with the delta method, the following results (asymptotic) were derived with respect to $\hat{r}$, $(1-2\hat{r})$ and $(1-2\hat{r})^2$:

$$Var(\hat{r}) = \frac{(1-2r)^2}{16(n+2)}, \quad Var(1-2\hat{r}) = \frac{(1-2r)^2}{4(n+2)}$$

and

$$Var[(1-2\hat{r})^2] = \frac{2(1-2r)^4}{(n+2)} \quad (B.6)$$

$$(1-2\hat{r}) = \frac{8(n+2)}{8(n+2)+1} \sqrt{\frac{\hat{s}_v}{\hat{s}_u}} \text{ is an unbiased estimator of } (1-2r); \quad (B.7)$$

$$(1-2\hat{r})^2 = \frac{2(n+2)+1}{2(n+2)} \frac{\hat{s}_v}{\hat{s}_u}, \text{ is an unbiased estimator of } (1-2r)^2 \quad (B.8)$$

With the delta method and Equations B.6 to B.8, it was demonstrated that the estimators: $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$ (Equations 11) are consistent and asymptotically unbiased. For example, consider the system of Equations 11, the expectation of $\hat{\mu}_1$ is:

$$E(\hat{\mu}_1) = E \left\{ \frac{\hat{m}_u}{4} \right\} + E \left\{ \frac{\hat{m}_v}{4(1-2\hat{r})^2} \right\} + E \left\{ \frac{\hat{m}_w}{2(1-2\hat{r})} \right\}$$

De la Ecuación B.1 resulta que:

$$E(\hat{\mu}_1) = \frac{1}{4}m_u + \frac{1}{4}\left[\frac{n+3}{n+2}\right]\frac{m_v}{(1-2r)^2} \\ + \frac{1}{2}\left[\frac{4(n+2)+1}{4(n+2)}\right]\frac{m_w}{(1-2r)}$$

Las Ecuaciones 8 permiten escribir $E(\hat{\mu}_1)$ como:

$$E(\hat{\mu}_1) \cong \frac{1}{4}(\mu_1 + 2\mu_2 + \mu_3) + \frac{1}{4}(\mu_1 - 2\mu_2 - \mu_3) \\ + \frac{1}{2}(\mu_1 - \mu_3) = \mu_1$$

De manera similar a la anterior se encontró la varianza de $\hat{\mu}_1$:

$$Var(\hat{\mu}_1) = E\left\{\left[\frac{\hat{m}_u}{4} + \frac{\hat{m}_v}{4(1-2\hat{r})^2} + \frac{\hat{m}_w}{2(1-2\hat{r})} - E(\hat{\mu}_1)\right]^2\right\} \\ \cong \frac{\sigma_2^2}{2n} + \frac{3(\sigma_1^2 + \sigma_3^2)}{8n} + \frac{(\mu_1 - 2\mu_2 + \mu_3)^2}{8(n+2)} \\ + \frac{(\mu_1 - \mu_3)^2}{4(n+2)}$$

Las esperanzas y varianzas de $\hat{\mu}_2$ y $\hat{\mu}_3$ se obtienen de manera similar y de ello resulta que $\hat{\mu}_2$ y $\hat{\mu}_3$ son estimadores insesgados y consistentes.

Con el método delta se determinaron, también, las propiedades de los estimadores $\hat{\alpha}$, $\hat{\delta}$ y $\hat{\sigma}_G^2$ (Ecuaciones 12). Éstas son, respectivamente:

$$E(\hat{\alpha}) = 0.5E(\hat{\mu}_1 - \hat{\mu}_3) \cong \alpha \\ Var(\hat{\alpha}) \cong \frac{\sigma_2^2}{4n} + \frac{3(\sigma_1^2 + \sigma_3^2)}{16n} + \frac{(\mu_1 - 2\mu_2 + \mu_3)^2}{16(n+2)} \\ + \frac{(\mu_1 - \mu_3)^2}{8(n+2)}$$

$$E(\hat{\delta}) = 0.25E(2\hat{\mu}_2 - \hat{\mu}_1 - \hat{\mu}_3) \cong \delta$$

From Equation B.1, the result is that:

$$E(\hat{\mu}_1) = \frac{1}{4}m_u + \frac{1}{4}\left[\frac{n+3}{n+2}\right]\frac{m_v}{(1-2r)^2} \\ + \frac{1}{2}\left[\frac{4(n+2)+1}{4(n+2)}\right]\frac{m_w}{(1-2r)}$$

Equations 8 make it possible to write $E(\hat{\mu}_1)$ as follows:

$$E(\hat{\mu}_1) \cong \frac{1}{4}(\mu_1 + 2\mu_2 + \mu_3) + \frac{1}{4}(\mu_1 - 2\mu_2 - \mu_3) \\ + \frac{1}{2}(\mu_1 - \mu_3) = \mu_1$$

In a way similar to the above, the variance of $\hat{\mu}_1$ was found:

$$Var(\hat{\mu}_1) = E\left\{\left[\frac{\hat{m}_u}{4} + \frac{\hat{m}_v}{4(1-2\hat{r})^2} + \frac{\hat{m}_w}{2(1-2\hat{r})} - E(\hat{\mu}_1)\right]^2\right\} \\ \cong \frac{\sigma_2^2}{2n} + \frac{3(\sigma_1^2 + \sigma_3^2)}{8n} + \frac{(\mu_1 - 2\mu_2 + \mu_3)^2}{8(n+2)} \\ + \frac{(\mu_1 - \mu_3)^2}{4(n+2)}$$

The expectations and variances of $\hat{\mu}_2$ and $\hat{\mu}_3$ are obtained in a similar way, and as a result, $\hat{\mu}_2$ and $\hat{\mu}_3$ are unbiased and consistent estimators.

With the delta method, the properties of the estimators $\hat{\alpha}$, $\hat{\delta}$ and $\hat{\sigma}_G^2$ (Equations 12) were also determined. These are, respectively,

$$E(\hat{\alpha}) = 0.5E(\hat{\mu}_1 - \hat{\mu}_3) \cong \alpha \\ Var(\hat{\alpha}) \cong \frac{\sigma_2^2}{4n} + \frac{3(\sigma_1^2 + \sigma_3^2)}{16n} + \frac{(\mu_1 - 2\mu_2 + \mu_3)^2}{16(n+2)} \\ + \frac{(\mu_1 - \mu_3)^2}{8(n+2)}$$

$$E(\hat{\delta}) = 0.25E(2\hat{\mu}_2 - \hat{\mu}_1 - \hat{\mu}_3) \cong \delta$$

$$\begin{aligned} Var(\hat{\delta}) \cong & Var(\hat{\alpha}) + \frac{2(\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2)}{n} \\ & + \frac{2(\mu_1 - 2\mu_2 + \mu_3)^2}{(n+2)} \end{aligned}$$

$$\begin{aligned} Var(\hat{\delta}) \cong & Var(\hat{\alpha}) + \frac{2(\sigma_1^2 + 4\sigma_2^2 + \sigma_3^2)}{n} \\ & + \frac{2(\mu_1 - 2\mu_2 + \mu_3)^2}{(n+2)} \end{aligned}$$

$$\begin{aligned} E(\hat{\sigma}_G^2) &\cong 0.5\alpha^2 + 0.25\delta^2 + 0.5Var(\hat{\alpha}) + 0.25Var(\hat{\delta}) \\ Var(\hat{\sigma}_G^2) &\cong \alpha^2Var(\hat{\alpha}) + 0.25\delta^2Var(\hat{\delta}) \end{aligned}$$

$$\begin{aligned} E(\hat{\sigma}_G^2) &\cong 0.5\alpha^2 + 0.25\delta^2 + 0.5Var(\hat{\alpha}) + 0.25Var(\hat{\delta}) \\ Var(\hat{\sigma}_G^2) &\cong \alpha^2Var(\hat{\alpha}) + 0.25\delta^2Var(\hat{\delta}) \end{aligned}$$

Los seis últimos resultados se obtuvieron bajo el supuesto de que $\hat{\alpha}$, $\hat{\delta}$, $\hat{\sigma}_G^2$, $\hat{\mu}_1$, $\hat{\mu}_2$ y $\hat{\mu}_3$ son independientes.

The last six results were obtained under the assumption that $\hat{\alpha}$, $\hat{\delta}$, $\hat{\sigma}_G^2$, $\hat{\mu}_1$, $\hat{\mu}_2$ and $\hat{\mu}_3$ are independent.