

SESGO EN LA ESTIMACIÓN DE EFECTOS DE QTLs EN
EXPERIMENTOS DE RETROCRUZAS.

ANA LAURA TORRES RAMÍREZ.

TESIS

Presentada como requisito parcial para obtener el grado de:
Maestro en Ciencias en
Estadística Experimental.

UNIVERSIDAD AUTONOMA AGRARIA
ANTONIO NARRO

Programa de Graduados

Buenavista, Saltillo, Coahuila. Octubre del 2004

UNIVERSIDAD AUTONOMA AGRARIA
ANTONIO NARRO

SUBDIRECCIÓN DE POSTGRADO

SESGO EN LA ESTIMACIÓN DE EFECTOS DE QTLs EN
EXPERIMENTOS DE RETROCRUZAS

TESIS

POR

ANA LAURA TORRES RAMÍREZ.

Presentada bajo la supervisión del comité particular de asesoría y
aprobada como requisito parcial para obtener el grado de:

MAESTRO EN CIENCIAS EN
ESTADÍSTICA EXPERIMENTAL

COMITÉ PARTICULAR

Asesor Principal

Dr. Manuel Humberto Reyes Valdés

Asesor

Dr. Octavio Martínez de la Vega

Asesor

Dr. Mario Cantú Sifuentes

Asesor

M.C. José Refugio Reyes Valdés.

Dr. Jerónimo Landeros Flores
Subdirector de Postgrado

Buenavista, Saltillo, Coahuila, México. Octubre de 2004

AGRADECIMIENTO

A Dios quien no ha permitido que me deje vencer por el desaliento, cansancio o desesperación. Por el contrario siempre ha sido una compañía constante y alentadora la cual ha logrado que llegue hasta este momento. **Gracias.**

Con amor y admiración a **mis padres** quienes me han enseñado no sólo con palabras sino con el ejemplo que siempre se debe dar lo mejor en cada tarea que se emprenda.

A **mis hermanos** que en todo momento me hacen saber que están conmigo y apoyan cada nuevo proyecto que emprendo.

A el **Dr. Humberto Reyes**, quien no limitó su apoyo sólo al aspecto académico, sino que lo llevó al plano humano y cordial.

A todos **los maestros** que durante mis estudios me ayudaron a alcanzar este objetivo. En especial a la **Maestra Nidia Pérez**

DEDICATORIA

El presente trabajo y logro alcanzado se lo dedico a todas las personas que de alguna u otra forma siempre me apoyaron. En especial quiero dedicarlo a:

Mis Padres:

Maria Juana Ramírez Olivo Por su constante amor, apoyo y preocupación

Constancio Torres González. Quién aunque ya no esté físicamente siempre me acompaña.

A Mis Hermanos:

Roberto
Gloria
Elisa
Dan

A Mis Sobrinos:

Tzolkin
Natalia
Pamela
El bebe

COMPENDIO

SESGO EN LA ESTIMACIÓN DE EFECTOS DE QTLs EN EXPERIMENTOS DE RETROCRUZA.

POR
ANA LAURA TORRES RAMÍREZ.

MAESTRIA EN
ESTADÍSTICA EXPERIMENTAL
UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO.
BUENAVISTA, SALTILLO, COAHILA, MÉXICO. Octubre 2004-09-22

Ph.D. Manuel Humberto Reyes Valdés. -----Asesor-----

Palabras Clave: QTLs, retrocruzas, sesgo, poder de detección.

Un problema asociado con la metodología en las investigaciones de QTLs es el sesgo en la estimación del número y efecto de estos. Esto se hizo evidente cuando fue reportado por Beavis (1994). El objetivo del presente estudio fue evaluar el sesgo en la estimación de los efectos de QTLs bajo diferentes condiciones experimentales en mapeo de regresión por intervalos en retrocruzas. Para realizar lo anterior se trabajó con tres factores experimentales: tamaño de muestra (100, 200 y 300), heredabilidad (10 y 5 por ciento) y número de marcadores (2, 4 y 6) . Se estableció un grupo de ligamiento de 100 cM donde se distribuyeron uniformemente cada una de las diferentes densidades

de marcadores y con la combinación entre este factor con heredabilidad y tamaño de muestra se obtuvieron 18 diferentes condiciones experimentales. Con la utilización de simulaciones se generaron los datos para la presente investigación. El análisis de los datos se llevo al cabo a través del programa SATORI (Reyes, 2003).

El análisis de los resultados indica que existe una tendencia a sobreestimar el valor absoluto de los efectos de los QTLs y que la magnitud del sesgo está estrechamente relacionada con las condiciones experimentales y la heredabilidad del QTL en cuestión. Desde el punto de vista puramente estadístico, el sesgo obedece a una correlación positiva entre el valor estimado del efecto y el valor del estadístico de prueba; esto acoplado a la regla de decisión que selecciona valores altos de dicho estadístico, a fin de declarar la presencia de un QTL.

ABSTRACT

BIAS IN THE ESTIMATION QTLs EFFECTS IN BACKCROSS EXPERIMENTS

BY
ANA LAURA TORRES RAMIREZ

MAESTER OF SCIENCE
EXPERIMENTAL STATISTICS

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO
BUENAVISTA, SALTILLO, COAHUILA. OCTOBER 2004

Dr. Manuel Humberto Reyes Valdés ---Adviser---

Keywords: QTLs, Backcross, Bias, detection power.

A problem associated with the methodology on QTL research, is the bias on the estimation of numbers and effects. This was clearly evident when Beavis (1994) reported it. The goal of the present study was to evaluate the bias on the estimation of the QTL effects under different experimental conditions by regression interval mapping in backcross populations. Three factors were considered: sample size (100, 200 and 300), heritability (10 and 5 percent) and markers density (2, 4 and 6). A linkage group of 100 cM was established with markers uniformly distributed. The combination between this factor with

heritability and the sample size gave 18 different experimental conditions. Simulation was used to generate the data for this investigation. Data analysis was performed using the program SATORI (Reyes, 2003).

The data analysis indicates that there is a tendency to estimate over the absolute value of QTL effects and the bias magnitude is closely related with experimental conditions and heritability of the QTL. From a pure statistical point of view, the bias obeys to a positive correlation between the estimated value of the effect and the value of the test statistic coupled to the decision rule of selecting the high values of that statistic, with the object to establish the presence of a QTL.

INDICE DE CONTENIDO

	Página
INDICE DE CUADROS	<i>x</i>
INDICE DE FIGURAS	<i>xi</i>
I. INTRODUCCIÓN	1
OBJETIVO	2
II. REVISIÓN DE LITERATURA	4
Loci de Características Cuantitativas	4
Marcadores Genéticos	5
Fracción de Recombinación	8
Funciones Mapa	10
Métodos Estadísticos de Análisis	11
Máxima Verosimilitud	11
Mapeo por Intervalos	14
Mínimos Cuadrados	14
Prueba de Permutaciones	19
Sesgo en la Estimación	21
Factores que Influyen en la Estimación	23
Número de Marcadores	23
Heredabilidad	24
Tamaño de Muestra	26
Simulación	27
Mathematica	29
III. MATERIALES Y METODOS	31
IV. RESULTADOS Y DISCUSIÓN	45
V. CONCLUSIONES	68
VI. RESUMEN	69
VII. LITERATURA CITADA	71

INDICE DE CUADROS

Cuadro		Página
3.1	Valores de los parámetros (μ , σ) de una distribución Normal, para el caso de dos diferentes heredabilidades para los genotipos de QTL obtenidos en una retrocruza. -----	36
4.1	Concentrado de los resultados para cada una de las combinaciones obtenidas con los diferentes niveles de los tres factores experimentales estudiados. ----	46
4.2	Análisis de varianza de los valores de F Ajustada con respecto a la influencia de los tres factores experimentales. -----	49
4.3	Varianza para el efecto estimado en las diferentes condiciones experimentales. -----	57
4.4	Análisis de regresión para el sesgo condicional de cada uno de los tres factores experimentales y las combinaciones entre dos de ellos a la vez. -----	62

INDICE DE FIGURAS

Figura	Página
3.1 Cromosomas de 100 cM. Distribución de cada una de las densidades de marcadores -----	31
3.2 Pedigrí y valores genotípicos en una retrocruza 1 -----	32
3.3 Haplotipos presentes en el progenitor F1, así como los puntos donde es posible que se realice una recombinación. -----	37
4.1 Comportamiento del valor promedio de FA con respecto a cada uno de los tres factores experimentales estudiados. -----	51
4.2 Efecto estimado del QTL en todas las repeticiones, de acuerdo con el valor de F ajustada (FA).Combinando los niveles de tamaño de muestra y densidad de marcadores para una heredabilidad de 5%. -----	53
4.3 Efecto estimado del QTL en todas las repeticiones, de acuerdo con el valor de F ajustada (FA).Combinando los niveles de tamaño de muestra y densidad de marcadores para una heredabilidad de 10%. -----	54
4.4 Comportamiento medio de la dispersión de las estimaciones del efecto estimado en cada una de las 100 repeticiones de cada condición experimental. -----	56
4.5 Influencia de cada uno de los factores estudiados, sobre la precisión en la estimación del efecto. -----	58
4.6 Influencia de cada factor sobre el valor promedio del sesgo condicional. -----	61
4.7 Comportamiento de la precisión en la estimación de la ubicación del QTL, de acuerdo a los diferentes factores experimentales investigados. -----	65

I. INTRODUCCIÓN

Una de las herramientas fundamentales en las investigaciones biológicas son los mapas genéticos, los cuales son una representación lineal de loci polimórficos. El mapa permite la ubicación e identificación de sitios de expresión dentro del genoma, así como las mutaciones que en estas regiones se presentan. Constituye una útil herramienta en el estudio de la variación en las características de los seres vivos la cual puede ser clasificada como cualitativa y cuantitativa. Esta última tiene por lo general una base genética compleja.

Las regiones del genoma que controlan a los caracteres cuantitativos se denominan QTLs (loci de características cuantitativas). Su naturaleza compleja se determina a partir de que su variación se debe a la segregación de varios genes polimórficos para cada uno de los cuales los efectos de las diferencias alélicas en el fenotipo son usualmente pequeñas. Esto, aunado a la variación ambiental, provoca que no exista una clara discontinuidad en la expresión del carácter, presentando así la característica una distribución de tipo continuo. La estimación del número, ubicación y medida de los efectos de los QTLs no es un proceso sencillo, pero con el surgimiento de la tecnología de marcadores y el empleo de métodos estadísticos, se ha logrado alcanzar dicho propósito. Para

un estudio de QTL lo ideal es contar con un mapa de marcadores y una población segregante para la variable de interés (Peterson *et al.*, 1988).

Uno de los primeros métodos propuestos para el análisis de los datos obtenidos a partir de los mapas de marcadores y que ha llegado a ser un procedimiento estándar, es el propuesto por Lander y Botstein (1989) denominado “mapeo por intervalos”. Este procedimiento se basó inicialmente en el enfoque estadístico de máxima verosimilitud, la cual constituye una herramienta muy poderosa al proveer de procedimientos para la estimación y prueba de hipótesis, como lo es la prueba generalizada de razón de verosimilitudes (Wilk, 1938). No obstante el empleo de la verosimilitud en el mapeo por intervalos presenta algunas limitantes, particularmente en distinguir efectos cuando se presenta más de un QTL en el cromosoma, pudiendo sesgarse seriamente la estimación de la posición del gen bajo estudio y llegar a conclusiones erróneas. (Haley y Knott, 1992; Martínez y Curnow, 1992). Se debe señalar que a medida que el número de parámetros por estimar aumenta, la solución por máxima verosimilitud se hace más complicada y es necesario recurrir al empleo de métodos numéricos para su solución.

A partir de las investigaciones de Haley y Knott(1992) y Martínez y Curnow (1992) se establece que el uso del método de regresión proporciona estimaciones muy aproximadas a las obtenidas por máxima verosimilitud, con la ventaja adicional de que la regresión es de fácil manejo e interpretación. Las variaciones ambientales pueden ser eliminadas por la inclusión de efectos fijos

en el modelo y pueden agregarse algunos otros factores o covariables sin tener un gran incremento en la dificultad del modelo. Además el método de regresión llega a ser superior computacionalmente a la máxima verosimilitud, en lo que se refiere a un análisis múltiple de datos y uso de la prueba de permutaciones (Churchill y Doerge, 1994) y el bootstrap para la construcción de intervalos de confianza (Visscher et al, 1996).

Al examinar un cromosoma para detectar la posible presencia de QTLs, se realizan numerosas pruebas correlacionadas y por lo tanto la distribución de F no puede ser empleada para obtener el umbral significativo, por lo que se debe realizar una prueba de permutaciones con los datos observados y así obtener el umbral significativo para trabajar (Churchill y Doerge, 1994).

El estudio de QTLs en una población segregante por medio de marcadores moleculares, se encuentra limitado por las características del experimento y la magnitud de los efectos individuales de cada gen. Se detectan los loci con mayores efectos, lo cual conduce a una sobreestimación en los efectos de los QTLs hasta ahora reportados en la literatura, en lo que se conoce como “efecto Beavis” (Beavis, 1994, Utz y Melchinger, 1994).

El objetivo del presente estudio fue evaluar el sesgo en la estimación de los efectos de QTLs bajo diferentes condiciones experimentales en poblaciones de retrocruza.

II. REVISIÓN DE LITERATURA.

Loci de Características Cuantitativas (QTL)

La variación en los caracteres cuantitativos se explica por la segregación de varios genes (herencia compleja) que se encuentran distribuidos en el genoma y cada uno de ellos contribuye con cierto porcentaje a la variación del carácter que controlan. Además, esta variación tiene un componente ambiental. El primero que estudió este tipo de características desde el punto de vista genético fue Francis Galton(Galton,1869; Galton, 1889), considerado además fundador de la escuela Biométrica a partir de la cual se traza la estadística moderna(Stingler, 1986; Crow, 1993).

En 1918 Ronald Fisher estableció las bases teóricas para explicar en términos genéticos la relación entre genes que controlan características cuantitativas y la correlación fenotípica. En las décadas de los 40s y 50s Mather y colaboradores demostraron por medio de una serie de trabajos con *Drosophila* que los genes que controlan las características cuantitativas se encuentran distribuidos en el genoma, introduciendo el término de poligenes (Kersey y Pooni, 1996). Gelderman (1975) sugirió el término de QTL (Locus de Carácter Cuantitativo) para remplazar el término de poligenes, lo que es

una de las razones principales para el estudio de QTLs, es que definen características relevantes de importancia económica en plantas y animales, así como caracteres que determinan el comportamiento social e intelectual del hombre además de enfermedades de origen genético. Un estudio de QTLs involucra el conocer número, posición y efecto de los genes que controlan el carácter bajo investigación. Sin embargo este tipo de estudios no es fácil debido a diferentes causas: La determinación del genotipo a partir del fenotipo no es factible en la mayoría de estudios con caracteres cuantitativos. No todos los genes involucrados contribuyen en la misma proporción para explicar la característica; unos pocos lo hacen en una proporción grande, mientras la mayoría aporta porcentajes pequeños y en ocasiones insignificantes. Además en ocasiones los diferentes loci presentan interacciones (epistasia) lo que dificulta aun más su estimación. Por lo tanto no es posible conducir la investigación bajo el análisis convencional de ligamiento solamente, sino que requieren métodos sofisticados asociados con el uso de marcadores genéticos.

Marcadores Genéticos

Un marcador es un carácter que debe presentar las siguientes cualidades: estar potencialmente asociado con la característica de interés, ser altamente heredable y de tipo cualitativo.

Inicialmente los marcadores empleados fueron caracteres fenotípicos, y el primer reporte de ligamiento entre un gen mayor y una característica

cuantitativa fue dado a conocer por Sax (1923) de un estudio con frijol (*Phaseolus vulgaris*) donde el marcador fue el color de semilla y el carácter cuantitativo fue el peso de ésta (Kersey y Pooni, 1996). Los estudios con QTLs durante los siguientes 60 años se realizaron con el empleo de genes mayores mutantes como marcadores de cromosomas que afectaban a características cuantitativas. Estas investigaciones fueron factibles en organismos donde se tenía la disponibilidad de líneas con un cromosoma que presentaba un QTL además de un gen mayor mutante, utilizado como marcador, tal fue el caso de *Drosophila* y trigo. Pero las anteriores condiciones no son el común para la mayoría de los organismos. Además se tiene los siguientes inconvenientes cuando se emplean como marcadores genes mayores mutantes.

- a) Baja frecuencia y número.
- b) En la gran mayoría de los casos provocan segregación distorsionada.
- c) Efectos pleiotrópicos del marcador sobre el QTL.
- d) Por lo general están limitados a ser usados bajo condiciones de laboratorio

Todo lo antes mencionado dificultó en gran medida el estudio de características cuantitativas hasta fines de la década de los 70s. Sin embargo el descubrimiento de la presencia natural de polimorfismos en el DNA a partir de 1980, introdujo la utilización de una nueva clase de marcadores: los marcadores de DNA. Presentan grandes ventajas para el análisis de QTLs por ejemplo:

- a) Altamente heredables
- b) Usualmente no provocan segregación distorsionada

- c) Usualmente no tienen ningún efecto sobre el carácter bajo estudio.
- d) Muchos de ellos son de tipo codominante.

Un gran número de marcadores moleculares se basa en el empleo de enzimas de restricción y el concepto es la mutación en el sitio de restricción. La propiedad de conservación del DNA entre los distintos organismos a lo largo de la evolución, ha permitido que la técnica de marcadores pueda ser elaborada para un organismo y posteriormente empleada en otros.

Los primeros marcadores moleculares empleados fueron: los polimorfismos en longitud de fragmentos de restricción (RFLPs) (Beckman y Soller, 1983; Lander y Botstein, 1989). Posteriormente surgieron nuevos tipos: DNA polimórfico amplificado aleatoriamente (RAPDs) (Williams et al, 1990), polimorfismo en longitud de fragmentos amplificados (AFLPs) (Zebau y Vos, 1993) y microsatélites (SSLPs) (Weissenbach J.A., 1993, Beattie CW., 1994) entre otros. Dichos marcadores moleculares permiten construir el mapa de ligamiento en base a cual los QTLs pueden ser localizados.

En la construcción del mapa de marcadores moleculares, en primer lugar se procede a detectar la presencia de marcadores polimorficos en distintas regiones del genoma o de un cromosoma de interés. A continuación se establece un orden de la posible ubicación de los marcadores en la o las regiones del genoma donde se detectaron, y se estiman las distancias entre los mismos.

Para construir un mapa se debe contar con una población en desequilibrio de ligamiento. En la práctica hay un gran número de posibles poblaciones que se podrían trabajar, pero por conveniencia y simplicidad se establecen las familias que pueden ser obtenidas en la segunda generación de un par de líneas homocigotas para los distintos alelos de los genes de interés y para los marcadores (Lynch y Walsh 1998). Las líneas homocigotas son utilizadas como progenitores P_1 y P_2 , el cruce de estas líneas origina una población F_1 la cual presenta un genotipo heterocigoto para los genes de interés y los marcadores. Esta última población puede ser trabajada de dos formas ya sea que se autofecunde y origine una población F_2 o bien que se cruce con uno de los progenitores iniciales obteniéndose así una población de retrocruza. Un ejemplo de este tipo de población es la retrocruza 1, que proviene del cruzamiento de la F_1 con uno de sus progenitores. Estas dos últimas poblaciones F_2 o retrocruza son la progenie a evaluar.

Una vez construido el mapa se puede hacer el análisis de QTLs, el cual está basado en la asociación entre la segregación de caracteres cuantitativos y marcadores.

Fracción de Recombinación.

La distancia genética es una función lineal del número promedio de entrecruzamientos que ocurre entre dos loci. Como no es técnicamente factible contabilizar la cantidad de quiasmas entre los loci individuales, se toma como

camino alternativo el estimar primero la recombinación entre loci, que es el resultado del proceso de entrecruzamiento y posteriormente transformar a distancia mapa.

Para intervalos cortos entre los loci la frecuencia de recombinación es aproximadamente igual a la distancia genética, sin embargo esta igualdad se pierde cuando el intervalo es grande. Mientras la distancia crece linealmente y sin límite conforme el número de intercambios, la frecuencia de recombinación tiene un límite superior de 50 por ciento. La frecuencia de recombinación no es aditiva. Bajo la suposición de independencia entre intervalos, la frecuencia de recombinación entre dos loci extremos es menor a la suma de los dos intervalos componentes. La diferencia es el doble del producto de la frecuencia de recombinación de dichos intervalos bajo no interferencia. Lo anterior se expresa por medio de la fórmula de Trow (1913), para tres loci consecutivos A, B y C.

$$r_{AC} = r_{AB} + r_{BC} - 2r_{AB}r_{BC} \quad (1)$$

donde r_{AC} es frecuencia de recombinación entre los loci A y C; r_{AB} es la frecuencia de recombinación entre los loci A y B, r_{BC} es la frecuencia de recombinación entre B y C.

Lo anterior se basa en el principio de que la formación de un quiasma es independiente en los dos intervalos y se fundamenta en la ausencia de interferencia, este último proceso notado y nombrado así por Muller (1916). La

interferencia, se presenta cuando la ocurrencia de los entrecruzamientos no tiene una independencia estadística, pues la presencia de uno inhibe la ocurrencia de otro en una región vecina; es decir, la probabilidad de que dos entrecruzamientos se presenten en un intervalo es menor al producto de las probabilidades de la ocurrencia de cada uno (Owen, 1950)

Funciones Mapa

Las funciones mapa pretenden transformar el valor de la frecuencia de recombinación a una distancia mapa, expresada en unidades denominadas por Haldane (1919) como Morgans o centiMorgan (cM), en reconocimiento a T. H. Morgan. La distancia mapa es aditiva, a diferencia de la fracción de recombinación entre los intervalos de los loci. El primero en establecer una función que realiza la conversión de la recombinación a distancia mapa fue Haldane (1919), bajo la suposición de no interferencia entre los entrecruzamientos:

$$m = -\frac{1}{2} \text{Ln}(1 - 2r), \quad (2)$$

donde r es la frecuencia de recombinación y m es la distancia mapa en cM

Debe tenerse en cuenta que no existe una relación exacta entre la distancia genética y la distancia física entre loci. La distancia genética esta

basada en el entrecruzamiento, el cual tiene una tasa variable, dependiendo del organismo, cromosoma o región cromosómica. La frecuencia de recombinación está bajo control genético, por medio de genes modificadores donde algunos presentan una influencia a través de todo el genoma mientras otros sólo controlan una región específica del cromosoma (Brooks, 1988).

Métodos estadísticos de Análisis de QTLs.

Para el análisis e interpretación de las investigaciones en QTLs se tienen diferentes métodos estadísticos y diseños experimentales. Las metodologías estadísticas más empleadas son basadas en máxima verosimilitud y mínimos cuadrados.

Máxima Verosimilitud

Es una de las herramientas estadísticas más poderosas, al emplearse tanto para estimación como para prueba de hipótesis. Para un vector de observaciones independientes Y_1, Y_2, \dots, Y_n la función de verosimilitud de un parámetro θ se define como:

$$L(\theta) = \prod_{j=1}^n f_{Y_j}(Y_j; \theta) \quad (3)$$

Para el caso en el cual el parámetro conocido es r tenemos:

$$L(r) = \prod_{j=1}^n f_{Y_j}(Y_j; r) \quad (4)$$

donde r es la frecuencia de recombinación parámetro a estimar.

En el mapeo de QTLs el método de máxima verosimilitud consiste en maximizar la función de verosimilitud con respecto al parámetro desconocido r . Esto es el estimador de verosimilitud \hat{r} de la frecuencia de recombinación, que maximiza la función de verosimilitud. Para facilitar los cálculos se trabaja con el logaritmo de la función, esto se justifica ya que el logaritmo natural es una función monótonica, Maximizar el logaritmo de la verosimilitud equivale a maximizar la verosimilitud misma (Lynch y Walsh, 1997):

$$l = Ln[L(r)] \quad (5)$$

donde l es el logaritmo del máximo de la función de verosimilitud.

Para realizar pruebas de hipótesis se emplea la prueba generalizada de razón de verosimilitudes. (GLTR) la cual esta dada por:

$$GLTR = \frac{\max lr(y)}{\max l(y)}, \quad (6)$$

donde GLTR es la razón de verosimilitud; $lr(y)$ es el modelo reducido de verosimilitud donde el valor de los parámetros se conoce; y $l(y)$ es el modelo completo de verosimilitud en el cual p parámetros son estimados.

Al obtener el logaritmo tanto del modelo completo como del reducido se tiene el logaritmo de la razón de verosimilitud; en la cual se examina si el QTL esta ligado al marcador o marcadores. La función

$$LR = -2Ln \left[\frac{\max lr(y)}{\max l(y)} \right], \quad (7)$$

proporciona un estadístico de prueba para la detección de QTLs. El estadístico es usualmente desplegado gráficamente a través del uso de los perfiles de verosimilitud. La razón de verosimilitud es graficada como una función de la posición mapa del supuesto QTL dada una fracción de recombinación. Esta aproximación a la posición de un QTL fue introducida por Lander y Botstein (1989), quienes proponen graficar los LOD (logaritmo de los momios):

$$LOD(r) = \text{Log}_{10} \left[\frac{\max l(y)}{\max lr(y)} \right] \quad (8)$$

El estimado de máxima verosimilitud de la posición mapa es donde se produce un máximo significativo del LOD. En una investigación empleando verosimilitud si el LOD excede un valor crítico se tiene la evidencia de un posible QTL, ubicando su posición estimada en el pico de la gráfica. En caso contrario si el pico de la gráfica no excede el valor crítico, entonces no existe evidencia de un QTL en esa vecindad.

Mapeo por Intervalos.

Una de las primeras metodologías para la investigación de QTLs fue el mapeo por intervalos (Lander y Botstein, 1989), auxiliándose de la máxima verosimilitud. Dicho método basa su análisis en la construcción de una gráfica o curva de verosimilitud para cada intervalo comprendido entre dos loci marcadores, a lo largo de todo el cromosoma, teniendo por lo tanto $n-1$ intervalos en los cuales se realizan las pruebas de verosimilitud (n es el número de marcadores). Al unir todos los intervalos se genera una sola gráfica para el cromosoma bajo estudio.

Dicha técnica tiene las ventajas que presenta la estimación de máxima verosimilitud. La desventaja es que para aplicarlo en la generalidad de las veces es necesario utilizar métodos numéricos; por lo tanto su demanda de tiempo computacional es grande. Aunado a lo anterior se tiene la desventaja de que a medida que aumentan los parámetros por estimar, el modelo aumenta su complejidad grandemente.

Mínimos Cuadrados

Dadas las dificultades del empleo de máxima verosimilitud Haley y Knott (1992) y Martínez y Curnow (1992) estudiaron como alternativa el empleo de mínimos cuadrados para el análisis de QTLs. Mostraron como la regresión de

mínimos cuadrados proporciona estimaciones muy similares a las obtenidas con máxima verosimilitud, pero con las ventajas adicionales de ser de fácil manejo e interpretación, además de no acrecentar en gran medida su complejidad al aumentar los parámetros por estimar.

El método de regresión determina el tipo de relación que existe entre una variable dependiente Y y k variables independientes X_k , la cual puede ser lineal o no lineal. El modelo tiene las siguientes suposiciones que lo fundamentan:

- 1). Los valores de la variable independiente son fijos
- 2). Todos los posibles valores de Y se distribuyen normalmente, con variancias iguales y son estadísticamente independientes (Daniel, 1995).

Si $k = 1$ entonces se está trabajando con una regresión simple y las anteriores suposiciones pueden expresarse por medio de la ecuación:

$$Y = \alpha + \beta X + \varepsilon \quad (9)$$

donde Y es la variable dependiente ; X es la variable independiente; α es el intercepto; β es el coeficiente de regresión.

Estos parámetros se estiman buscando el mejor ajuste para la distribución de los valores de las variables X y Y . Los mínimos cuadrados en regresión lineal nos dan una solución del mejor ajuste, por medio de la minimización del valor promedio de las desviaciones cuadráticas entre el valor

observado y el estimado por el modelo. El teorema de Gauss-Markoff asegura que son los mejores estimadores lineales e insesgados. Además presentan la propiedad adicional de maximizar la cantidad de varianza en \mathbf{y} que puede ser explicada por un modelo lineal.

El coeficiente de determinación (R^2) estima la proporción de la varianza en Y que es explicada por el modelo, al asumir que $E(X|Y)$ es lineal. La varianza de la variable de respuesta tiene dos componentes: cantidad de varianza explicada por el modelo (varianza de la regresión) y la varianza residual (no explicada por el modelo de regresión). El valor de R^2 proporciona una estimación puntual de ρ^2 que es el coeficiente de determinación de la población.

Para el caso de la retrocruza (Martínez y Curnow, 1992), la regresión se lleva al cabo a lo largo de toda la longitud entre el intervalo flanqueado por dos marcadores:

$$Y_{ij} = \alpha + \beta_1 P_j + \varepsilon_{ij} \quad (10)$$

donde $P_j(\mathbf{X})$ es la probabilidad de que un alelo hipotético del QTL Q ó q , este presente en un individuo de un particular grupo de marcadores; P es función de la posición del QTL en el intervalo; α es el valor para el homocigoto recesivo; β_i es el valor del efecto aditivo y ε_{ij} es el término residual.

Haley y Knott en (1992) relacionan la gráfica de R^2 , obtenida de la regresión, con la gráfica de máxima verosimilitud. En un análisis de máxima verosimilitud, una prueba para p parámetros se obtiene al maximizar $\frac{L_1}{L_0}$ donde L_1 es la verosimilitud en el modelo completo y L_0 verosimilitud el modelo reducido. La función $2\text{Log}_e\left(\frac{L_1}{L_0}\right)$ proporciona una prueba estadística de la razón de verosimilitud, que se distribuye aproximadamente como una χ^2 con p grados de libertad. La regresión de mínimos cuadrados es máxima verosimilitud cuando los errores son independientes y se distribuyen normalmente (Draper y Smith 1966). Por lo tanto la función de razón de verosimilitud puede escribirse en términos de suma de cuadrados residuales:

$$LR = n\text{Log}_e\left(\frac{SCT}{SCE}\right) \quad (11)$$

donde SCT es la suma de cuadrados del total, SCE es al suma de cuadrados del error y n es el número de observaciones.

Asumiendo normalidad e independencia en los errores Haley y Knott (1992), usaron como modelo la progenie F_2 , tomando como indicador de la posible posición de un QTL el valor más grande de R^2 que se obtiene de todas las regresiones realizadas a lo largo del cromosoma. No obstante que la suposición de normalidad de los errores dentro de los genotipos marcadores no

se cumple, Hayle y Knott (1992), Martínez y Curnow (1992) y Rebae et al. (1995) mostraron que las estimaciones obtenidas con mínimos cuadrados son muy cercanos a los logrados por el empleo de máxima verosimilitud.

Martínez y Curnow (1992) demostraron, estudiando los valores esperados de los estimadores en el caso de una retrocruza, que las estimaciones de posición y efecto de QTLs obtenidas por medio de regresión de mínimos cuadrados son muy similares a las conseguidas por la razón de verosimilitud, cuando se emplea la técnica de mapeo por intervalos en un cromosoma. Ellos mostraron como la gráfica del valor de SCE (suma de cuadrados residual) contra la posición en el cromosoma, se comporta en una forma inversamente proporcional a la gráfica de LODs. Con lo anterior concluyen que un claro mínimo del valor de SCE, podría evidenciar la presencia de un QTL. Ellos argumentan que la diferencia entre los dos métodos surge solamente por que los errores no tienen una distribución normal, sin embargo las estimaciones obtenidas son bastante aproximadas a las proporcionadas por máxima verosimilitud.

Xu (1998) realizó una comparación de las estimaciones obtenidas entre la regresión de mínimos cuadrados y máxima verosimilitud y corroboró que no existen diferencias significativas entre los resultados obtenidos por ambos métodos. Menciona que el método de regresión es particularmente útil cuando se emplean pruebas de permutación (Churchill y Doerge 1994) y bootstrap

(Visscher et al., 1996) para la determinación de valor crítico y la construcción de intervalos de confianza respectivamente. Ambos métodos involucran múltiples análisis de datos, que para el caso de máxima verosimilitud constituyen una limitante si el tamaño del conjunto de datos y del genoma tienen un gran tamaño.

Prueba de Permutación

Un problema que se presenta en todos los métodos de detección de QTLs es el de determinar el valor crítico del estadístico de prueba. Al emplear cualquier metodología estadística, verosimilitud o regresión, se presentan dos problemas. Primero el determinar la distribución de la prueba estadística y segundo la gran cantidad de pruebas correlacionadas que se presentan. Se necesitan múltiples pruebas de hipótesis para examinar un cromosoma y determinar la posible localización de un QTL (Haley 1994; Jansen 1993; Jansen y Stram 1994; Zeg 1993,1994). Por lo anterior se propuso un procedimiento, basado en la prueba de permutación inicialmente propuesta por Fisher (1935). Dicho procedimiento determina una distribución empírica de los datos y define un valor crítico y se denomina como prueba de permutación (Churcill y Doerge 1994).

La prueba de permutación propuesta por Fisher (1935) se basa en permutaciones aleatorias repetidas de los valores de la variable dependiente y la generación de una muestra de la prueba estadística de una distribución nula

apropiada. El procedimiento es estadísticamente válido y puede emplearse en mínimos cuadrados y en máxima verosimilitud, para cualquier distribución de la característica cuantitativa. Para estimar el valor crítico de un estadístico por medio de la prueba de permutación los individuos son enumerados de 1 a n ; posteriormente los datos son desordenados al efectuarse permutaciones aleatorias. En seguida se asigna el i -ésimo valor de la característica al individuo que ahora tiene el i -ésimo número. Los datos reordenados son analizados para estimar la ubicación y el efecto del QTL; todo el procedimiento es repetido N veces y en cada una de la repeticiones se obtiene el valor máximo de F . Al final de todas las pruebas efectuadas se cuenta con N valores máximos los que son ordenados en forma descendente. El valor de F , correspondiente al cuantil del nivel y de significancia de amplitud cromosómica (α_c) deseado representa el valor crítico. Por su parte el nivel de significancia de amplitud genómica se aproxima como sigue:

$$\alpha_g = \alpha_c \varphi \quad (12)$$

donde φ es el número de grupos de ligamiento en el genoma analizado; α_c es la probabilidad del error Tipo I de amplitud cromosómica y α_g es la probabilidad del error Tipo I de amplitud genómica.

El valor crítico utilizado para detectar la presencia de un QTL es el equivalente a un error Tipo I menor o igual que α_g . Este valor crítico es por lo tanto mayor que muchos de los valores por comparar lo que en ocasiones conduce a una pérdida en el poder de detección de los efectos del QTL

(Churchill y Doerge, 1994). Con base a su experiencia Churchill y Doerge (1994) recomiendan realizar cuando menos 1000 permutaciones para un $\alpha = 0.05$ y 10,000 si $\alpha = 0.01$.

La descripción anterior de la prueba de permutación involucra estimación de un valor crítico para cada cromosoma. Otro método es la búsqueda directa de un valor crítico para α_g a través de permutaciones con maximización a todo lo largo del genoma.

Sesgo en las Estimaciones

Existe un problema asociado con la metodología en las investigaciones de QTLs. Es el sesgo en la estimación del número y efecto de estos. Dicho error es muy difícil de eliminar y no se había hecho claramente evidente hasta que Beavis (1994) reportó un comportamiento no adecuado en las inferencias de los estudios de QTLs, a partir de inconsistencias entre las investigaciones por él realizadas con resultados previamente reportados en análisis de las mismas características cuantitativas en maíz.

Los resultados reportados de trabajos en el análisis de QTLs típicamente presentan uno o varios loci con un gran efecto genético y varios loci adicionales explicando una pequeña cantidad de la variación genotípica. Esto se encuentra determinado en gran medida por el nivel de significancia, al sólo reportar

aquellas estimaciones que resultan ser significativas. Por lo tanto a medida que se emplea un nivel de significancia más severo se incrementa la posibilidad de tener un mayor sesgo (George et al., 1995, Burns, 1997). Otro de los factores que se ha visto que influyen en el sesgo de las estimaciones es que sólo se logran detectar unos pocos genes (no más de 10) que explican la característica del total que realmente influye (Visscher y Haley, 1998). Lo anterior se fundamenta en simulaciones diseñadas de tal forma que la variabilidad fenotípica es explicada por mucho más de 10 QTLs y en las que sólo se han logrado detectar a lo sumo 10 de estos QTLs (Beavis, 1994).

Beavis (1994) después de hacer varios estudios y eliminar causas probables que provoquen el sesgo como es el caso de: número de marcadores, grado de endogamia en la progenie, interacción entre QTLs y medio ambiente, y fuentes parentales empleadas para generar la población, concluye que la causa más probable de la presencia de sesgo en las estimaciones así como la pérdida de congruencia entre las investigaciones es un factor de muestreo.

Lo anterior no debe tomarse de una forma drástica que nos lleve a concluir que los QTLs que hasta ahora se han localizado constituyan una mala detección en cuanto a ubicación del gen, pero si es conveniente tener ciertas reservas con los efectos estimados. No obstante debe recordarse que no existe una evidencia del número de loci que afectan a la mayoría de las características (Beavis, 1994) lo cual hace que se manifieste una inconsistencia, en número y efecto, en la detección de QTLs que afectan a una

característica. Sin embargo se ha mostrado una convergencia en la detección al menos en aquellos loci con efectos mayores.

Factores que influyen en la estimación

La mayoría de las técnicas biométricas diseñadas para estimar el número de factores genéticos responsables de la variación de una característica continua requieren una estructura y tamaño de población que en pocas ocasiones se satisface (Zeng, 1992). Una de las limitantes más importantes en la implementación y desarrollo de una investigación es el recurso económico con que se cuenta, el cual por lo general es muy restringido. Lo anterior obliga a los investigadores a trabajar bajo las condiciones mínimas y no óptimas, o bien con pocas posibilidades de variación en la investigación.

Número de Marcadores

La densidad requerida de marcadores empleados se encuentra en función del objetivo de la investigación. Si lo que se persigue es detectar las regiones del genoma que expliquen una proporción significativa de la variación, son necesarios pocos marcadores. La densidad de marcadores debe ser mayor cuando el objetivo es estimar localización y efecto de genes que determinan las características cuantitativas o bien cuando se pretende separar grupos de QTLs ligados (Darvasis y Soller, 1994).

Dependiendo de la cantidad de varianza genética y su naturaleza (un QTL o varios) se busca tener un óptimo espaciamiento entre los marcadores. Para un cromosoma de 100 cM de longitud y QTLs distribuidos uniformemente se recomienda una separación entre marcadores entre 20-30 cM (Visscher y Haley, 1998). En lo que respecta al número de marcadores se sugiere una densidad que va de 2 a 5 marcadores por cromosoma (Visscher y Haley, 1998). Sin embargo debe tenerse cuidado cuando se emplean sólo dos marcadores ya que no puede inferirse la presencia de más de un QTL con el método de mínimos cuadrados (Whittaker et al., 1996), mientras la inferencia con máxima verosimilitud es comprometida por la alta correlación entre las pruebas (Visscher y Haley, 1998).

Cuando no se colocan marcadores en los extremos de un cromosoma se corre el riesgo de que el QTL no sea flanqueado por los dos marcadores, lo que provoca que se lleguen a tener malas inferencias sobre la ubicación y efectos del gen (Martínez y Curnow, 1992).

Heredabilidad

Es uno de los parámetros más útiles en genética cuantitativa. Se define como la proporción de la varianza fenotípica total que es atribuida a los efectos de los genes (Kersey, 1996). La heredabilidad en sentido amplio se obtiene como:

$$H^2 = \frac{V_G^2}{V_F^2} \quad (13)$$

donde V_G es la varianza genética; V_F es la varianza fenotípica.

En la mayoría de las aplicaciones genéticas cuantitativas se le da mayor importancia al componente aditivo genético conocido como heredabilidad en sentido estrecho (h) y se expresa por la razón:

$$h^2 = \frac{V_A^2}{V_F^2}, \quad (14)$$

donde V_A es la varianza aditiva y V_F es la varianza fenotípica

El parámetro (h^2) describe la variación concomitante entre los progenitores y lo progenie para una determinada característica.

La heredabilidad de un carácter es en gran medida determinante en la detección de QTLs. Entre menor sea su valor hay menos probabilidades de que sean detectados correctamente el número y efecto de los genes que explican el carácter.

Tamaño de Muestra

El tamaño de muestra es una de los factores que más influyen en los resultados obtenidos en una investigación de caracteres cuantitativos. Se ha

encontrado que un tamaño pequeño de muestra provoca baja precisión en la estimación del número de QTLs además de que se tiende a sobreestimar los efectos de estos genes.

El tamaño deseable de la muestra depende, entre otras cosas, del número de marcadores empleados simultáneamente para la detección del QTL. La relación entre estas dos condiciones del experimento, distribución de marcadores y tamaño de muestra, es directa. A medida que se emplean más marcadores simultáneamente, la muestra debe ser de un mayor tamaño (se incrementa en forma geométrica) para poder tener presentes todos los genotipos posibles.

Beavis (1994) concluyó que las modificaciones en las técnicas de análisis en las investigaciones de QTLs podrían tener un menor impacto en los resultados obtenidos, comparados con el efecto que se tendría en los resultados cuando las modificaciones se realizan en las características relacionadas con el diseño del experimento.

Simulación

La simulación es una técnica empleada para realizar un gran número de evaluaciones en investigaciones en un tiempo relativamente corto y donde se modifican diferentes factores del experimento.

La simulación ha sido aplicada en casi todos los campos de la actividad humana. Es una de las herramientas más útiles en la investigación, pero probablemente una de las más difíciles en su aplicación apropiada y en la elaboración de conclusiones a partir de los resultados (Phillips et al, 1976). La simulación consiste en modelos que representan sistemas reales. Considerando a un sistema como el conjunto de características con una interacción bien definida entre ellas y a partir de las cuales se selecciona la mínima cantidad de ellas que permitan predecir el comportamiento futuro del sistema. Por lo tanto un modelo se construye en base a aquellas características representativas del sistema real, de tal forma que su desempeño sea lo más apegado a la realidad y con la menor complejidad en su manejo.

Adkins y Pooch (1977) citan las siguientes ventajas en el empleo de la simulación:

1. Permite el control de experimentos al poder cambiar la información inicial y así evaluar el comportamiento del sistema bajo diferentes condiciones.
2. No provoca disturbios en el sistema real, al no alterar dicho sistema cuando este es evaluado bajo diferentes condiciones.
3. Proporciona ahorro en el tiempo empleado en un estudio, al realizar en un tiempo mucho menor aquellas actividades que en situaciones reales requieren de grandes períodos de tiempo.

4. Es una de las herramientas más efectivas para la educación y capacitación.

Uno de los problemas más comúnmente encontrados al modelar un sistema real es la poca frecuencia con que se presentan los comportamientos predecibles al manifestarse en una gran mayoría cambios aleatorios o de variación probabilística. Esto hace necesario el uso de modelos o estructuras matemáticas para la adecuada descripción de los sistemas, utilizando frecuentemente modelos de probabilidad así como métodos para inferencias estadísticas y generadores de números aleatorios. Debido a lo anterior la mayoría de los modelos son elaborados en programas de computación. Es importante la selección del lenguaje en el cual el programa será escrito, buscando elegir aquel donde la ejecución de las rutinas sea factible y además el consumo de tiempo sea el menor posible.

Si el experimento y la obtención de la muestra estadística son realizados en una computadora a través de la generación de números aleatorios, el procedimiento es denominado como muestreo empírico o Monte Carlo (Kleijnen,1974). El procedimiento fue llamado así por el principado de Monaco “la capital del juego del azar”, al considerar a una ruleta como un generador de números aleatorios. El nombre y el desarrollo sistemático de los métodos de Monte Carlo datan aproximadamente de 1944 y su uso real como herramienta de investigación, se sitúa en los trabajos dirigidos a la elaboración de la bomba atómica durante la Segunda Guerra Mundial, al implementarse con este

método una simulación directa de problemas probabilísticos de hidrodinámica concernientes a la difusión de neutrones aleatorios en el material de fusión.

Bajo el nombre de “Método Monte Carlo” o “Simulación Monte Carlo” se agrupan una serie de procedimientos numéricos que analizan la distribución de variables aleatorias utilizando la simulación de números aleatorios. Es una de las herramientas más útiles en diversas ramas de la ciencia, principalmente en problemas donde el tiempo de evaluación, para una solución del problema es muy grande.

Mathematica®

Mathematica® es un sistema general de software para técnicas computacionales desarrollado por Wolfram (1988); es empleado por científicos, ingenieros y analistas. Su aplicación abarca áreas de la ciencia, tecnología y negocios. La primera versión apareció en 1988. Es un software general de computación y lenguaje dirigido hacia las matemáticas y otras aplicaciones y puede ser empleado de las siguientes formas:

- Calculadora numérica y simbólica.
- Visualización de sistemas de funciones y datos.

- Lenguaje de programación para la elaboración de programas de diferentes dimensiones.
- Software de plataforma para correr paquetes construidos para aplicaciones específicas.
- Creación de documentos interactivos donde se mezcle texto, gráficos animados y sonidos.
- Modelar y analizar datos de medioambiente.
- Un lenguaje de control para programas y procesos externos.

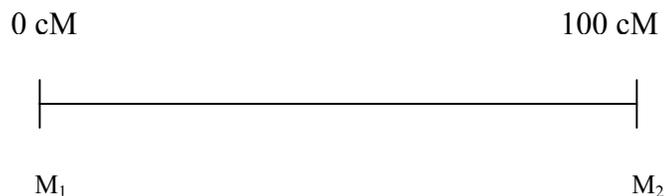
El programa Mathematica® permite diferentes estilos de programación como son: los de estructura en bloque, condicionales, iteraciones y recursivos, así como programación funcional y orientada a objetos.

III. MATERIALES Y METODOS.

En la presente investigación se trabajó con tres factores experimentales: tamaño de muestra, número de marcadores y heredabilidad. El efecto de las diferentes condiciones se evaluó en una población de mapeo formada por una retrocruza 1.

Se utilizaron tres tamaños de muestra: 100, 200 y 300 individuos. En lo que respecta a la heredabilidad, se emplearon dos valores 10 por ciento y 5 por ciento. En cada caso, se estableció un grupo de ligamiento de 100 cM con 2, 4 y 6 loci marcadores distribuidos uniformemente.

Siempre se colocó un marcador en cada extremo del cromosoma, esto con la finalidad de evitar que el QTL se ubique fuera de un intervalo entre dos marcadores. La distribución de los marcadores se muestra en la Figura 3.1.



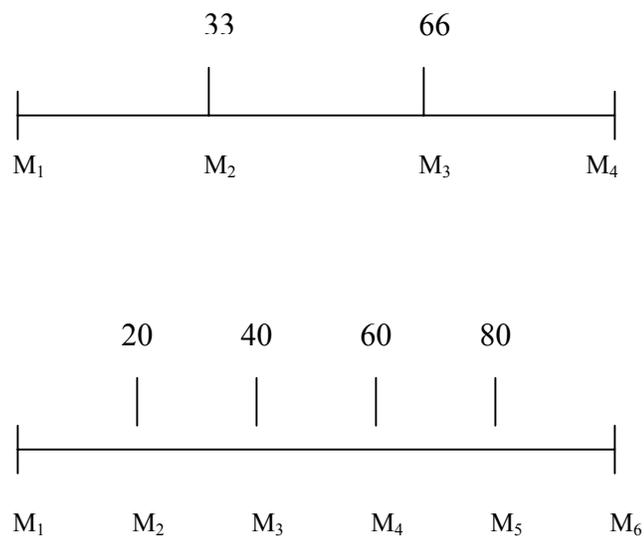


Figura 3.1 Cromosoma de 100 cM y distribución de cada uno de los marcadores (M_i), conforme tres diferentes densidades.

En la figura 3.2 se representan el pedigrí de la población segregante, así como un QTL hipotético segregando con un locus marcador. Los genotipos del QTL hipotético se ubican en una escala

$$P_1 \ m_1q_m_2 \setminus m_1q_m_2 \times P_2 \ M_1QM_2 \setminus M_1QM_2$$

$$F_1 \ M_1QM_2 \setminus m_1q_m_2$$

$$F_1 \times P_1 = BC_{1.1}$$

$$\begin{array}{ccc} \text{-----} & \mathbf{a} & \text{-----} \\ \mathbf{qq} & & \mathbf{Qq} \\ \mathbf{2b} & & \mathbf{B + b} \end{array}$$

Figura 3.2 Pedigrí y valores genotípicos en una retrocruza 1. Los valores son 2b para el genotipo qq y B+b para el genotipo Qq.

Al combinarse los factores del experimento se tuvieron 18 diferentes condiciones experimentales que se evaluaron en 100 repeticiones cada una.

En este trabajo se utilizaron simulaciones de Monte Carlo para obtener valores de variables aleatorias con distribución discreta (Bernoulli y Uniforme) así como variables con distribución continua (Normal).

Las simulaciones se realizaron con el programa Mathematica y con la utilización de los comandos de simulaciones genéticas del paquete MathGenome (Reyes-Valdés 2003). Con estas simulaciones se generó el genotipo y valor fenotípico de cada uno de los individuos que constituyen la muestra.

Cada simulación se inicia se inicia con la construcción *in silico* del cromosoma. Como primer paso se establece la posición del QTL, por medio de la generación de números aleatorios de una distribución uniforme discreta en el intervalo de 0 a 100, teniéndose como única restricción que el QTL no coincida exactamente con la posición de un locus marcador. A partir de la información de la posición y número de marcadores así como la ubicación del QTL, se procede a la formación de los haplotipos que aportaran los progenitores a cada individuo de la población segregante. Se simula el proceso biológico en el que se realiza la formación de los gametos con la generación de números aleatorios de una distribución Bernoulli con una probabilidad de 1/100 (1 por ciento) de que exista recombinación a cada intervalo de 1cM. Este proceso se realiza para el

progenitor heterocigoto. Después de efectuado el proceso anterior se cuenta con la información genética de dos gametos por progenitor y la elección del gameto que aporta cada progenitor se realiza por la generación de números aleatorios de una distribución Bernoulli con una probabilidad de $\frac{1}{2}$. Al tenerse el gameto que aporta cada progenitor se procede a conformar al individuo que forma parte de la generación segregante. Dicho individuo es evaluado para determinar que genotipo presenta en el QTL, ya sea Qq o qq, con la finalidad de otorgar un valor fenotípico. El valor fenotípico se asigna a partir de la generación de números pseudoaleatorios de una distribución Normal, donde los valores de los parámetros μ y σ se definieron de acuerdo a las condiciones iniciales de heredabilidad y efecto aditivo del QTL.

Para el caso de la media (μ) el valor se dio buscando sólo que cumpliera con la condición de que la diferencia entre los genotipos, homocigoto (qq) y heterocigoto (Qq), sea igual a el efecto aditivo (a) que en la presente investigación se fijó en uno. Lo anterior se representa esquemáticamente en la Figura 3.2.

Para la obtención del otro parámetro, la desviación estándar residual (σ_E), su valor correspondiente se calculó a partir de la ecuación de heredabilidad:

$$h^2 = \frac{\sigma_A^2}{\sigma_A^2 + \sigma_E^2} \quad (15)$$

donde σ_A^2 es la varianza genética aditiva y σ_E^2 es la varianza residual.

La varianza aditiva es determinada a partir de la formula

$$\sigma_A^2 = E(G^2) - \mu_G^2 \quad (16)$$

donde G es el valor genotípico aditivo y μ_G es la media del valor genotípico estimado

En el caso de una retrocruza donde sólo se tiene efecto aditivo tenemos que:

$$E[G] = m + \frac{1}{2}a$$

$$E[G^2] = m^2 + ma + \frac{1}{2}a^2$$

donde m es el valor promedio de los progenitores

Sustituyendo en 16 se tiene:

$$\sigma_A^2 = \left[m^2 + ma + \frac{1}{2}a^2 \right] - \left[m + \frac{1}{2}a \right]^2$$

Por lo tanto

$$\sigma_A^2 = \frac{1}{4}a^2 \quad (17)$$

Entonces $\sigma_A^2 = \frac{1}{4}$ para $a = 1$

Despejando σ_E^2 de la ecuación 15 y sustituyendo los valores de heredabilidad y σ_A^2 se calcula el valor de la varianza residual, para después obtener su raíz cuadrada y así tener el valor del parámetro σ de la distribución normal.

En base a lo anterior los parámetros que se emplearon en la distribución Normal para generar los valores fenotípicos se muestran en el Cuadro 3.1.

Cuadro 3.1 Valores de los parámetros (μ y σ) de una distribución Normal, para el caso de dos diferentes heredabilidades y los genotipos del QTL obtenidos en una retrocruza.

Heredabilidad	Genotipo del QTL	Valores de los parámetros de V.A~N(μ , σ)
10%	Qq	N(10 , 1.5)
10%	qq	N(9 , 1.5)
5%	Qq	N(10 , 2.179)
5%	qq	N(9 , 2.179)

El procedimiento antes citado se realizó en las 100 repeticiones de cada una de las 18 condiciones experimentales investigadas.

Al analizar a los individuos de la muestra para aplicar el método de mapeo por intervalos por medio de regresión, la información con la que se cuenta obtenida a partir de la anterior simulación es valor fenotípico y el genotipo marcador presente. Por lo anterior es importante remarcar lo siguiente:

de los dos progenitores que intervienen en la retrocruza (F_1 y P_1), el progenitor que determina la variación genotípica en la progenie es el heterocigoto (F_1). A partir de éste se definen los cuatro posibles haplotipos que contienen el alelo de interés del QTL, los cuales son el producto de la presencia o ausencia de recombinación entre los marcadores y el QTL: M_1QM_2 ; M_1Qm_2 ; m_1QM_2 ; m_1Qm_2 . Mediante el siguiente diagrama, de dos cromosomas del progenitor F_1 (Figura 3.3), se puede visualizar como se originaron los cuatro genotipos en un cierto intervalo.

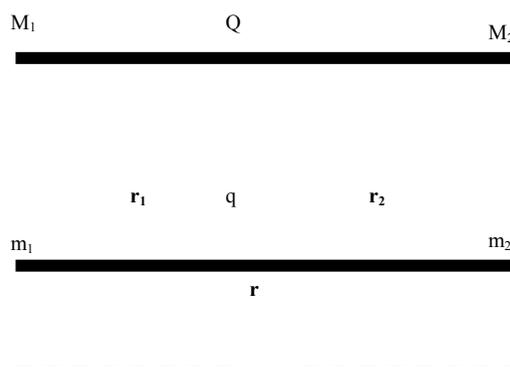


Figura 3.3 Haplotipos presentes en el progenitor F_1 , así como los puntos donde es posible se realice la recombinación, simple en r_1 (entre el marcador 1 y el QTL) o r_2 (marcador 2 y el QTL), mientras r representa la recombinación entre los marcadores 1 y 2. Se originan 4 posibles haplotipos con el alelo Q.

Con lo anterior podemos empezar a definir como se aplicó el método de regresión en la estimación del efecto y posición del QTL. En primer lugar se define que el efecto del alelo Q es B y el efecto del alelo q es b. Si Y_{ij} representa la observación en la i -ésima planta ($i = 1, 2, \dots, n$) en el j -ésimo grupo de genotipo marcador ($j = 1, 2, 3, 4$) se tiene que el valor esperado de Y_{ij} se obtiene con base en lo siguiente: (Martínez y Curnow, 1992)

$$E[qq] = 2b$$

$$E[Qq] = B + b$$

por lo tanto:

$$\begin{aligned} E[Y_{ij}] &= 2b(1 - \rho_j) + (B + b)\rho_j \\ &= 2b - 2b\rho_j + (B + b)\rho_j \\ &= 2b + (B + b - 2b)\rho_j \\ &= 2b + (B - b)\rho_j \end{aligned} \tag{18}$$

donde ρ_j es la probabilidad de que un determinado haplotipo de la F_1 contenga el alelo Q en un cierto sitio.

Si 18 se sustituye en el modelo de regresión lineal se tiene:

$$Y_{ij} = \beta_0 + \beta_1\rho_j + \varepsilon_{ij} \tag{19}$$

donde β_0 igual a $2b$ y es el intercepto de la recta de regresión con la ordenada, β_1 igual a $(B-b)$ que es la pendiente de la recta de regresión y tiene el valor de a que es el efecto aditivo; ε_{ij} es el error residual.

La ecuación de regresión empleada en el presente trabajo fue equivalente a:

$$Y_{ij} = \beta_0 + a\rho_j + \varepsilon_{ij}, \tag{20}$$

donde a es el efecto aditivo

Con la determinación de la anterior ecuación y bajo el supuesto de que los errores se distribuyen normalmente con media cero y varianzas iguales, se aplicó la regresión en el mapeo por intervalos, evaluando el cromosoma totalmente cada centimorgan.

Para la obtención del coeficiente de β_1 , las probabilidades condicionales (ρ_j) del genotipo hipotético Qq dados diferentes genotipos marcadores se obtuvieron empleando las ecuaciones de Reyes-Valdés (2000), quien asumió que la coincidencia toma valor de $c = 1$ es decir; que no existe interferencia. Las expresiones para dichas probabilidades condicionales son:

$$\rho_1 = P[M_1QM_2 | M_1M_2] = 1 - \frac{r_1r_2}{1-r} \quad (21)$$

$$\rho_2 = P[M_1Qm_2 | M_1m_2] = \frac{r_2 - r_1r_2}{r} \quad (22)$$

$$\rho_3 = P[m_1QM_2 | m_1M_2] = \frac{r_1 - r_1r_2}{r} \quad (23)$$

$$\rho_4 = P[m_1Qm_2 | m_1m_2] = \frac{r_1r_2}{1-r} \quad (24)$$

Las variables r_1 , r_2 y r están definidas en la Figura 3.3. Estas ecuaciones son equivalentes a las empleadas por Martínez y Curnow (1992), sólo que estos autores representan a r_2 en función de r_1 y r .

Como se mencionó anteriormente se asume que no existe interferencia de quiasma para la presencia de un entrecruzamiento entre una posición del cromosoma y otra. Es por esto que es posible emplear la inversa de la función mapa de Haldane (1919) y con esta hacer el cambio de distancia en cM a frecuencia de recombinación, para fines de cálculo de heredabilidades condicionales. La función de Haldane (1919) es:

$$H(d) = \frac{1}{2}(1 - e^{-2d}) \quad (25)$$

donde d es la distancia mapa en cM

Después de aplicar la regresión a lo largo del cromosoma se determina como posición estimada del QTL el punto donde se encuentre el máximo valor de F ajustada. En esta región se estima el efecto y los valores de R^2 , R^2 ajustada a través del ANOVA para regresión. El mapeo por intervalos aplicando la regresión es realizó por medio del programa SATORI (Reyes Valdés, 2003), en cada una de las 18 condiciones experimentales con 100 repeticiones cada una.

SATORI es un programa para el análisis de QTLs en pedigrees de tres generaciones. Los padres iniciales pueden ser endogámicos o no. Se permite el

empleo de alélos múltiples para los marcadores, pudiendo ser de tipo codominante. Son necesarios los genotipos marcadores de los individuos que conforman a los abuelos, padres y progenie a evaluar. El valor fenotípico sólo es necesario en la tercera generación. La introducción de los datos se realiza en tres archivos de tipo texto: el primero contiene la información de genotipos marcadores y valor fenotípico de la progenie; el segundo presenta los datos de los abuelos y los padres; y un último archivo tiene la información referente a la posición mapa de los marcadores para un grupo de ligamiento, siempre iniciando la lista del mapa con el primer marcador en la posición cero cM y colocando el último marcado al final de la longitud del cromosoma.

SATORI es un programa que se ejecuta en Mathematica® de Wolfram Research (1988) y tiene las siguientes opciones de salida:

- Generación de Tablas de valores del estadístico de F para datos analizados con regresión múltiple o simple. Proporciona como salida una lista de los valores de F en cada cM y la posición donde se encuentra el valor máximo.
- Realiza análisis de regresión simple o múltiple. Usualmente el coeficiente de regresión es el valor estimado del efecto aditivo en regresión simple.
- Proporciona un reporte completo ya sea de regresión múltiple o simple para la posición donde se encontró el valor mayor de F.

- Permite la reparametrización de las variables de los progenitores, así como la estimación de efectos aditivos.
- Elabora una gráfica de la lista de F para regresión simple o múltiple, graficando la posición del cromosoma contra el valor de F correspondiente.
- Construye una gráfica para los LODs aproximados.
- Trabaja con bootstrap para la elaboración de un intervalo de confianza para la posición mapa, basándose en un remuestreo donde es aplicada la regresión. La distribución empírica obtenida después de aplicado el remuestreo es desplegada gráficamente.
- Permite la realización de prueba de permutación. Este método numérico se basa en la permutación aleatoria de los fenotipos sobre los genotipos marcadores de la progenie para después aplicar la regresión. El procedimiento es llevado a cabo N veces para finalmente obtener un valor de F que constituye el umbral crítico, para probar la hipótesis de la presencia de uno o más QTLs.

Para la realización de las permutaciones se fijan los valores de los genotipos marcadores, mientras los datos fenotípicos son reordenados al permutarse. Con este nuevo arreglo se vuelve a realizar el mapeo por medio de la regresión. Se toma el valor de F ajustado más alto, obtenido en la investigación de todo el cromosoma, repitiéndose la prueba de permutación en

un cierto número de veces. Los valores de F ajustada de cada permutación se ordenan en forma descendente tomando como umbral crítico el valor de F que se encuentre en el cuantil correspondiente a el nivel cromosómico de 0.005 (en la presente investigación), lo cual corresponde a una significancia de 0.05 en amplitud genómica para los grupos de ligamiento. El procedimiento de permutación se realizó para 3 repeticiones de las 100 en cada una de las diferentes condiciones experimentales. Las 3 cantidades de F obtenidas fueron promediadas para obtener así una sola cantidad crítica contra la cual comparar los resultados de las 100 repeticiones de cada una de las distintas combinaciones experimentales. Sólo los valores de F ajustada que resultaron ser mayores al umbral crítico, así obtenido, fueron considerados como significativos. Con la anterior comparación se define el poder de detección de QTL conforme varían los factores experimentales: tamaño de muestra, número de marcadores y heredabilidad, al determinar el porcentaje de estimaciones que resultaron ser significativas.

Además de obtener el poder de detección, se estimó el sesgo en el efecto, restando el efecto aditivo previamente establecido (en este caso de uno) al estimado por medio de la regresión. También fue estimado el sesgo condicional a partir de aquellos resultados que fueron significativos, donde al promedio del efecto de las estimaciones significativas se le restó el efecto aditivo real del QTL. Adicionalmente para los resultados del total de la muestra y en los que fueron significativos se obtuvo el promedio de R^2 en ambos casos.

Con el promedio de los valores significativos se estimó el sesgo condicional de R^2 al restarle a este valor la heredabilidad correspondiente al QTL.

Se realizaron análisis de varianza para ver la influencia de los tres diferentes factores estudiados sobre el valor de F ajustada y sesgo condicional. Para el caso del ANOVA entre los tres factores experimentales y el valor de F ajustada FA, el modelo empleado fue el siguiente:

$$FA_{ijk} = \mu + \tau_i + \Delta_j + H_k + (\tau\Delta)_{ij} + (\tau H)_{ik} + (\Delta H)_{jk} + (\tau\Delta H)_{ijk} + \varepsilon_{ijk} \quad (26)$$

donde FA es el valor de F ajustada; τ_i representa el tamaño de muestra; Δ_j representa la densidad de marcadores; H_k es la heredabilidad y E_{ijk} es el error residual

En el caso del ANOVA de los factores experimentales y el sesgo condicional el modelo empleado fue:

$$sesgo_{ijk} = \mu + \tau_i + \Delta_j + H_k + (\tau\Delta)_{ij} + (\tau H)_{ik} + (\Delta H)_{jk} + \varepsilon_{ijk} \quad (27)$$

donde τ_i representa el tamaño de muestra; Δ_j representa la densidad de marcadores; H_k es la heredabilidad y E_{ijk} es el error residual

También se construyeron diferentes gráficas para visualizar más claramente la influencia de los tres factores del experimento en la definición de la localización y efecto estimado del QTL.

IV. RESULTADOS Y DISCUSIÓN

En el Cuadro 4.2 se presenta el concentrado de los resultados obtenidos en cada una de las diferentes combinaciones de densidad de marcadores, tamaño de muestra y heredabilidad. En la columna DEP se tiene la raíz cuadrada de la desviación cuadrática media entre la localización real y la estimada del QTL. Es notoria la reducción constante de la magnitud de la desviación, a medida que en las distintas combinaciones de los factores experimentales se trabaja con valores cada vez mayores. En la columna EF se tiene el promedio del efecto estimado y se observa como los valores tienden a estar cercanos al valor real del efecto del QTL que es uno. Sin embargo no se puede establecer una tendencia definida en el comportamiento de los valores de estas estimaciones al no apreciarse una influencia bien establecida con respecto al cambio de magnitud de los tres factores experimentales. Lo que sí se puede señalar es la aparente tendencia que siguen la mayoría de los promedios de la estimación del efecto, a presentar un ligero sesgo hacia la derecha y esto se establece con base en los valores de la media del efecto de todas las combinaciones posibles para las distintas magnitudes de los tres factores experimentales, así como su desviación estándar ($\bar{x} = 1.036742$, $s_{\bar{x}} = .0080454$). Al observar la columna de EEE (error estándar del efecto

Cuadro 4.1 Concentrado de los resultados para las tres diferentes densidades de marcadores (2, 4 y 6) y tres tamaños de muestra (100, 200 y 300) en ambos valores de heredabilidad (0.05 y 0.1). **Hered** = heredabilidad; **NM** = número de loci marcadores; **TM** = tamaño de muestra; **EF** = efecto estimado; **DEP** = raíz cuadrada de la desviación cuadrática media entre el valor real de la posición y el estimado; **EEE** = error estándar del efecto estimado; **EFC** = estimación condicional del efecto; **SESGO-C** = sesgo condicional; **R2** = valor de R^2 en todas las repeticiones; **R2C** = valor de R^2 condicional; **SCR2** = sesgo condicional de R^2 ; **R2A** = valor de R^2 ajustada; **FA** = promedio de F ajustada; **PD** = poder de detección con $p=0.005$ a nivel cromosómico.

Hered	NM	TM	DEP	EF	EEE	SESGO	EFC	SESGO-C	R2	R2-C	SCR2	R2A	FA	PD (%)
0.05	2	100	35.96248	1.040912	0.80975	-0.040912	2.41485	1.41485	0.04064	0.119797	0.069797	0.030846	4.276759	8
	2	200	28.61049	1.018401	0.55877	-0.018401	1.59670	0.59670	0.02890	0.067959	0.017959	0.023996	5.980698	17
	2	300	22.39888	1.071162	0.50914	-0.071162	1.53546	0.53546	0.02940	0.054409	0.004409	0.026139	9.183721	36
	4	100	33.37095	1.026482	0.69029	-0.026482	1.85794	0.85794	0.06066	0.131855	0.081855	0.051070	6.483926	19
	4	200	17.90587	1.062531	0.40469	-0.062531	1.40948	0.40948	0.05165	0.079727	0.029727	0.046859	10.902699	40
	4	300	15.04792	1.046184	0.31687	-0.046184	1.19573	0.19573	0.04772	0.058989	0.008989	0.044529	15.057945	69
	6	100	28.34255	1.107441	0.56641	-0.107441	1.76943	0.76943	0.06777	0.133199	0.083199	0.058254	7.310719	17
	6	200	19.22134	1.065657	0.35705	-0.065657	1.35037	0.35037	0.05458	0.079678	0.029678	0.049808	11.567359	45
	6	300	11.88949	1.020457	0.26795	-0.020457	1.15588	0.15588	0.04775	0.058401	0.008401	0.046039	15.691283	70
0.1	2	100	28.59545	1.058952	0.69060	-0.058952	1.67180	0.67180	0.06867	0.140102	0.040102	0.059163	7.481595	25
	2	200	21.15680	0.962850	0.39050	0.037150	1.22101	0.22101	0.05099	0.083271	-0.016729	0.046195	10.932480	44
	2	300	18.72859	1.058621	0.37724	-0.058621	1.18214	0.18214	0.05657	0.070835	-0.029165	0.053406	18.330383	71
	4	100	20.39068	0.974001	0.39808	0.025999	1.31845	0.31845	0.08815	0.143462	0.043462	0.078841	9.712853	36
	4	200	15.29673	1.016712	0.30604	-0.016712	1.11773	0.11773	0.08548	0.097399	-0.002601	0.080865	18.767192	79
	4	300	9.90606	1.015960	0.19610	-0.015960	1.02605	0.02605	0.08398	0.085594	-0.014406	0.080906	27.589857	97
	6	100	18.48892	1.042352	0.35348	-0.042352	1.34396	0.34396	0.10587	0.158085	0.058085	0.096743	11.963673	45
	6	200	9.88888	1.042523	0.28169	-0.042523	1.11831	0.11831	0.09972	0.110495	0.010495	0.095178	22.333966	85
	6	300	6.91375	1.030168	0.19157	-0.030168	1.03511	0.03511	0.09298	0.093607	-0.006393	0.08994	30.860716	99

estimado) para cada combinación de factores, se observa una disminución a medida que los factores experimentales trabajan con niveles mayores, es decir la presencia de una relación negativa entre EEE y los valores de los factores experimentales.

con el total de repeticiones. En cuanto a los promedios del efecto condicional estimado, es decir aquel estimado únicamente cuando se detectó el QTL (columna EFC) se distingue como se acerca cada vez más al valor real del efecto a medida que se incrementa el nivel de cada uno de los factores. Se puede señalar sobre la base de los resultados anteriores, como el tamaño de muestra representa un componente con una influencia notoria en el efecto condicional, pues comparando los resultados de trabajar utilizando un número no muy grande de marcadores con un tamaño de muestra grande, 4 marcadores-300 individuos, se logran resultados que son similares y en ocasiones mejores que cuando se emplea una densidad mayor de marcadores pero con una muestra de menor tamaño, 6 marcadores-200 individuos. En la columna R2 se tienen los datos del valor promedio de R^2 para las 100 repeticiones, mientras en la columna R2C se presenta el promedio de R^2 sólo para los resultados significativos. Se observa como el valor de R^2 tiende a estar alrededor de la magnitud de la heredabilidad. En el caso de 4 y 6 marcadores la aproximación mejora a medida que el tamaño de muestra incrementa, tanto en el caso de trabajar con el total de los experimentos (R2) así como al considerar sólo las significativas (R2C). Sin embargo los promedios de R2C presentan valores más alejados de la magnitud de la heredabilidad en comparación con

los promedios reportados en la columna R2 es decir, Esto último apoya lo mencionado en cuanto a que la sobreestimación del efecto del QTL se ve favorecida cuando se emplean únicamente las estimaciones significativas. Como se sabe el valor de R2 representa aquella varianza que es explicada por el modelo y que en esta investigación corresponde a la varianza debida sólo a causas genéticas; por lo tanto si este valor se esta incrementando el efecto que se estima también es mayor que el real.

Dentro de las densidades de 4 y 6 marcadores es importante hacer notar como el utilizar una muestra de 100 individuos no genera buenos resultados, ya que se alejan mucho del valor de la heredabilidad real. Para el caso de trabajar con sólo dos marcadores, en virtud de los resultados obtenidos, resulta poco recomendable el empleo de esta densidad, ya que se estiman magnitudes bastante alejadas de la heredabilidad real. Lo anterior se reafirma al analizar la columna SCR2 donde están contenidos los valores del sesgo condicional para R^2 .

En la columna FA se tiene el valor promedio de la F ajustada máxima obtenida en las regresiones llevadas acabo en cada experimento. Se distingue como la magnitud de FA incrementa conforme los factores experimentales trabajan con valores mayores. El conocer el comportamiento de FA es importante pues como se sabe la FP (F crítica) que se obtiene después de la prueba de permutaciones es de valor grande cuando se usan valores pequeños

de p ; por lo tanto sí las FA en las estimaciones presenta una magnitud grande se tendrá un mayor poder de detección, lo que se muestra en la columna PD.

A continuación se reporta el análisis del comportamiento del valor de FA máximo para la regresión sobre los valores de los tres factores experimentales, para cada población. En el Cuadro 4.2 se reportan los resultados del análisis de varianza para FA, con los tres factores estudiados: tamaño de muestra, densidad de marcadores y heredabilidad. Al observar el cuadro se aprecia como cada uno de los factores en forma individual así como la interacción de las combinaciones de dos de ellos a la vez presentan una alta significancia; esto nos lleva a inferir la gran trascendencia de los tres componentes experimentales en el valor de FA. Por su parte la interacción de los tres factores no resultó significativa.

Cuadro 4.2 Análisis de Varianza del valor de F ajustada (FA) con respecto a la influencia de los tres factores experimentales. **TM** = tamaño de muestra, **NM** = número de marcadores y **HERED** = heredabilidad

Fuentes de Variación	de G.l	Suma de Cuadrados	de Cuadrado Medio	F	$p(F)$
TM	2	39683.5	19841.77	304.0601	0.0000
NM	2	16545.9	8272.94	126.7765	0.0000
HERED	1	29037.7	29037.64	444.9801	0.0000
TM NM	4	2279.7	569.93	8.7337	0.0000
TM HERED	2	5748.8	2874.4	44.048	0.0000
NM HERED	2	1662.2	831.11	12.7361	0.0000
TM NM HERED	4	377.1	94.27	1.4446	0.2168
Residual	1782	116286.3	65.26		

Como complemento de lo anterior, en la Figura 4.1 se visualiza como actúa cada factor experimental en el valor promedio de FA. Se observa como FA presenta una relación positiva con tamaño de muestra, número de marcadores y heredabilidad. Si se examina lo correspondiente al tamaño de muestra (Figura 4.1A) se advierte como el cambio de un nivel a otro se presenta un incremento proporcional, por lo tanto su influencia es constante y determinante en el valor promedio de FA en el espacio de tamaños de muestra estudiados. En lo que respecta a la influencia de la heredabilidad en la magnitud promedio de FA (Figura 4.1B) se establece como este componente es definitivo en la investigación, ya que en la gráfica se registra un cambio drástico en dicho valor promedio al trabajar con una y otra heredabilidad y esto se refleja directamente en el poder de detección de QTLs como ya se había citado anteriormente. Para el caso de densidad de marcadores (Figura 4.1C) se tiene que en el cambio de 2 a 4 marcadores el aumento en la magnitud promedio de FA es más notorio en comparación con lo que se registra al pasar de 4 a 6 marcadores. Se espera que en un momento dado el aumento en el número de marcadores ya no presentara una gran relevancia en lo que respecta a la diferencia que se puede obtener en el valor de FA. Con base en lo antes mencionado se establece que a medida que se empleen magnitudes mayores en cada uno de los componentes experimentales, principalmente tamaño de muestra y heredabilidad, favorece a que una mayor cantidad de resultados sean significativos y por lo tanto empleados para las estimaciones del QTLs.

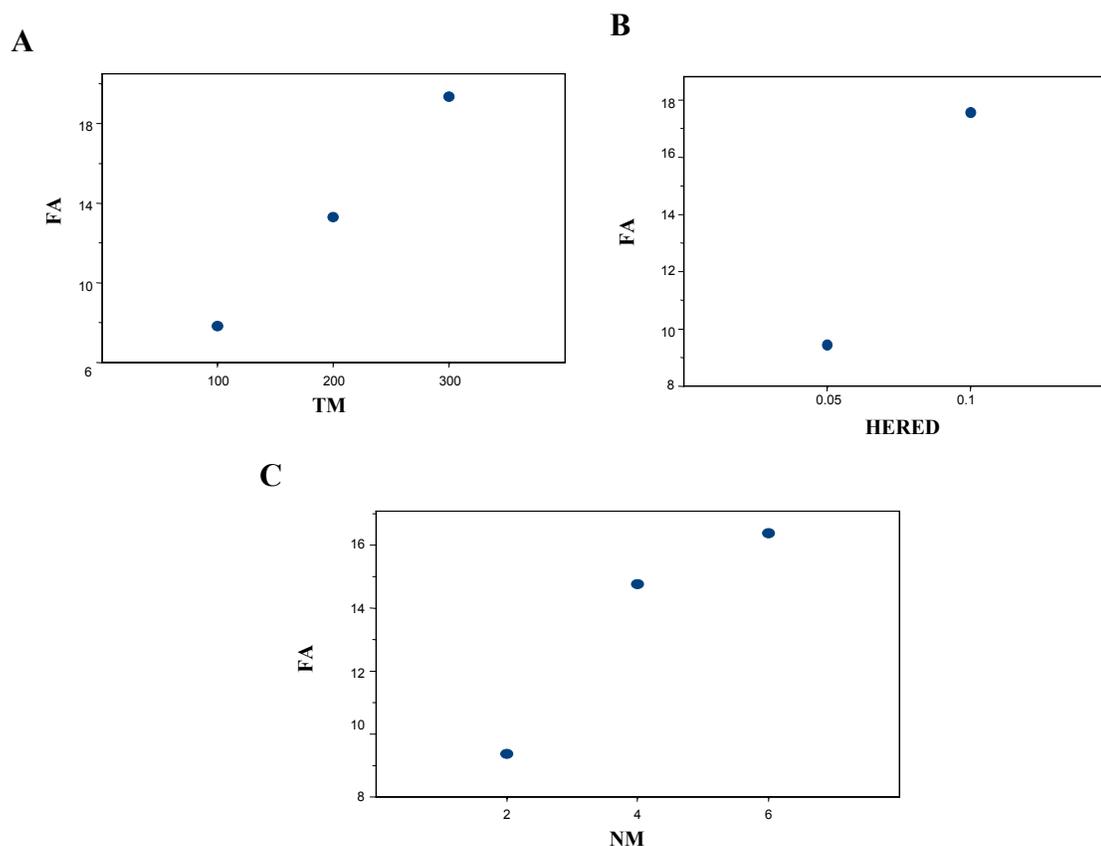


Figura 4.1 Comportamiento del valor promedio de FA con respecto a los tres factores experimentales estudiados. **A.** Tamaño de muestra (**TM**). **B.** Heredabilidad (**HERED**). **C.** Número de Marcadores (**NM**).

En las Figuras 4.2 y 4.3 se combinan los diferentes tamaños de muestra y números de marcadores para cada heredabilidad. En principio es importante hacer notar que para el QTL con una heredabilidad de 0.1 la dispersión entre las diferentes estimaciones del efecto fue menor en comparación con la dispersión observada para una heredabilidad de 0.05. La dispersión disminuye conforme aumenta el número de marcadores al igual que cuando se incrementa el tamaño de muestra, el cual manifiesta una gran influencia. Entre el valor de

FA y la estimación del efecto se ve como se establece una relación cuasi-lineal. Se observa además como en las combinaciones de 4 marcadores-300 individuos y 6 marcadores-200 individuos se presenta un comportamiento muy similar. Es importante hacer notar que en los casos de 4 o 6 marcadores la magnitud del efecto estimado continúa ascendiendo conforme lo hace FA pero en una menor proporción y sus estimaciones del efecto tienen un mayor acercamiento al valor real del efecto que es uno. Por lo tanto, a medida que el tamaño de muestra, la densidad de marcadores y la heredabilidad del QTL aumenta, las estimaciones del efecto de los resultados significativos son menos sesgados. Esto se infiere al observar como la pendiente del diagrama de dispersión también disminuye a medida que el número de marcadores y el tamaño de muestra se incrementan en magnitud.

Sobre la base de lo antes establecido se puede inferir la gran influencia que tiene el valor de FA en la presencia de sesgo en la estimación del efecto, principalmente por que sólo se trabaja con resultados significativos, provenientes de una comparación entre FA y la FP. Como ya se mencionó, las estimaciones que presentan un valor de FA grande son a la vez las que tienden a una estimación del efecto mayor, reduciéndose esta tendencia a medida que los factores se trabajan con magnitudes cada vez más grandes (Figuras 4.2 y 4.3). Esto concuerda con lo mencionado por Georges *et al.* (1995) quien señala que a medida que se utilizan niveles significativos más severos, mayores serán los sesgos en el efecto estimado del QTL debido al valor tan grande que alcanza FP. Por lo tanto una forma de inducir la disminución del sesgo de la

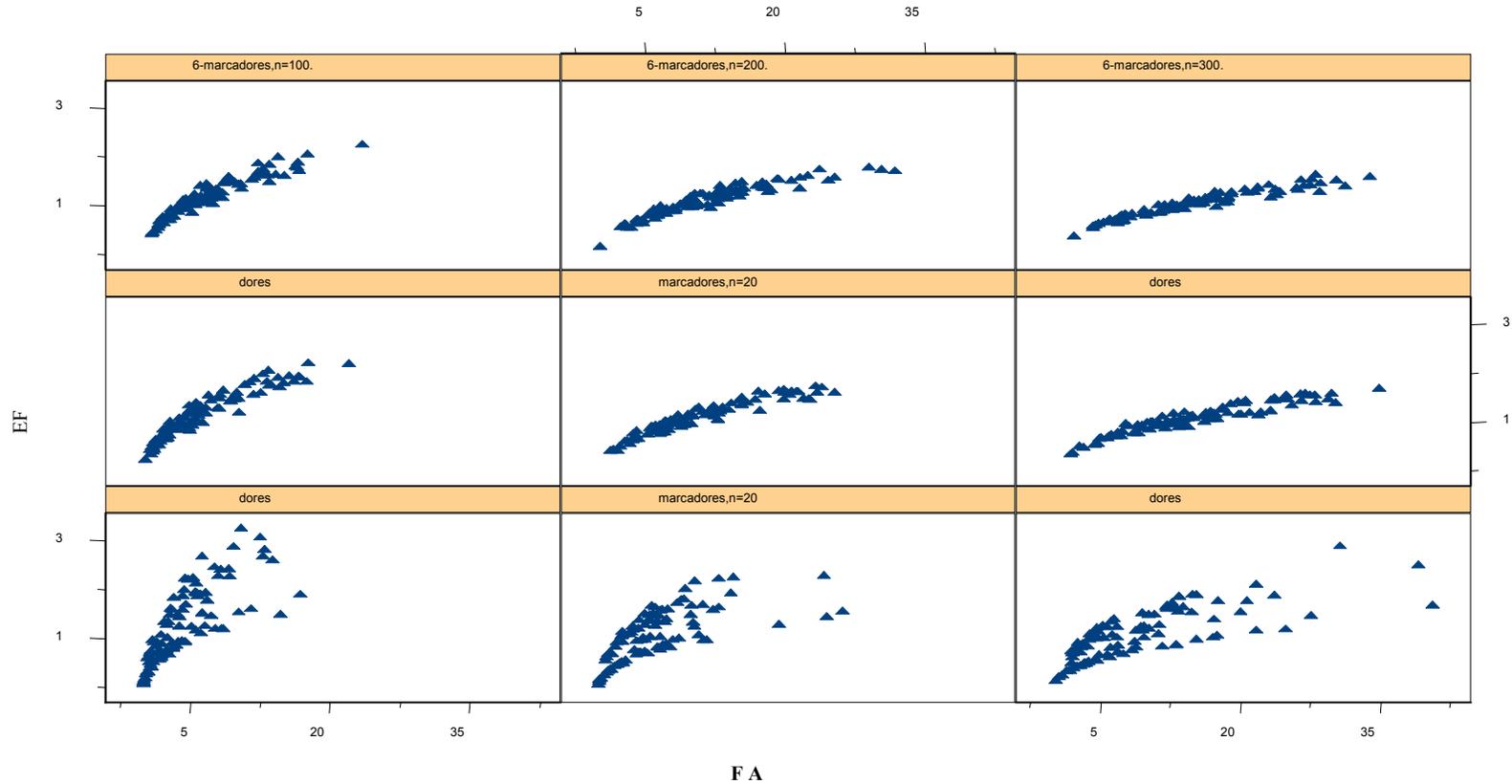


Figura 4.2 Efecto estimado (EF) del QTL en todas las repeticiones, de acuerdo con el valor promedio de la F ajustada (FA) para cada una de las combinaciones resultantes de los tres diferentes tamaños de muestra y tres densidades de marcadores, para una heredabilidad de 5%. Se emplearon los valores absolutos de las estimaciones del efecto.

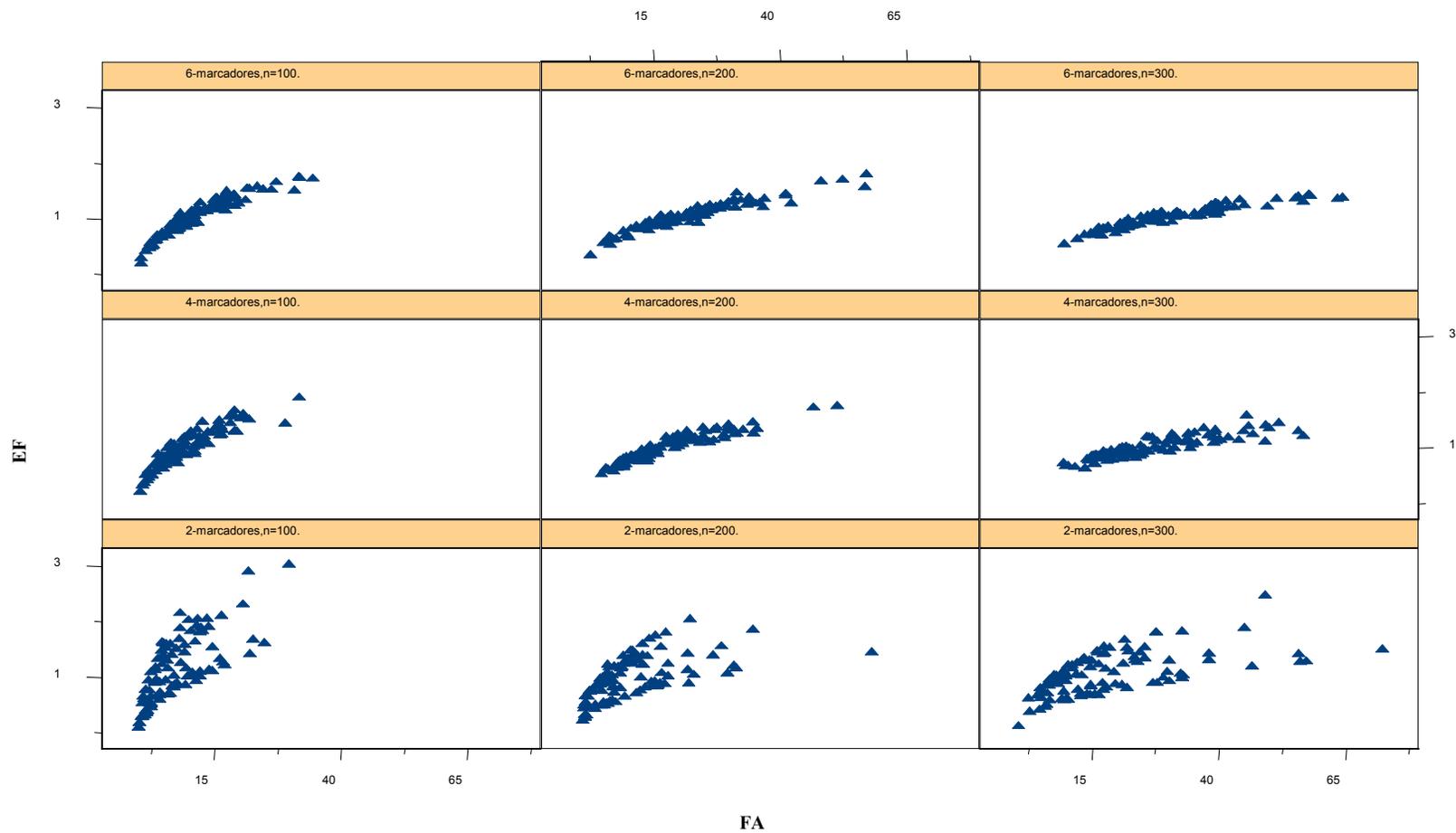


Figura 4.3 Efecto estimado (EF) del QTL en todas las repeticiones, de acuerdo con el valor promedio de la F ajustada (FA) para cada una de las combinaciones resultantes de los tres diferentes tamaños de muestra y tres densidades de marcadores, para una heredabilidad de 10%. Se emplearon los valores absolutos de las estimaciones del efecto.

Aquí se examinan tres aspectos: dispersión (Figura 4.4), precisión (Figura 4.5) y sesgo en la estimación (Figura 4.6) en el comportamiento de los resultados. El análisis de los tres aspectos antes citados se realiza en lo que respecta a la influencia que manifiesta cada uno de los tres factores experimentales en la estimación del efecto del QTL investigado. En lo que respecta a la dispersión en las estimaciones, la Figura 4.4 muestra el comportamiento de las estimaciones en las 100 repeticiones, para cada una de las 18 diferentes combinaciones posibles de los distintos niveles de los tres factores experimentales trabajados. Se observa como la heredabilidad constituye un componente determinante en el grado de dispersión presente en todas las estimaciones del efecto. La manera de visualizarlo es comparando el comportamiento de los resultados para un mismo número de marcadores y tamaño de muestra pero diferente heredabilidad. Con base en esta gráfica se hace evidente la relación inversa que se presenta entre la dispersión de las estimaciones del efecto y los tres componentes experimentales. El empleo de sólo dos marcadores en la estimación del efecto no es muy apropiado a menos que los otros dos factores experimentales tengan magnitudes grandes. Esto último que puede apreciarse en el Cuadro 4.3 donde se ve como la varianza en la estimación con el empleo de dos marcadores se ve mejorada a medida que se aumenta el tamaño de muestra o la heredabilidad y más aun si son ambos componentes los que se incrementan.

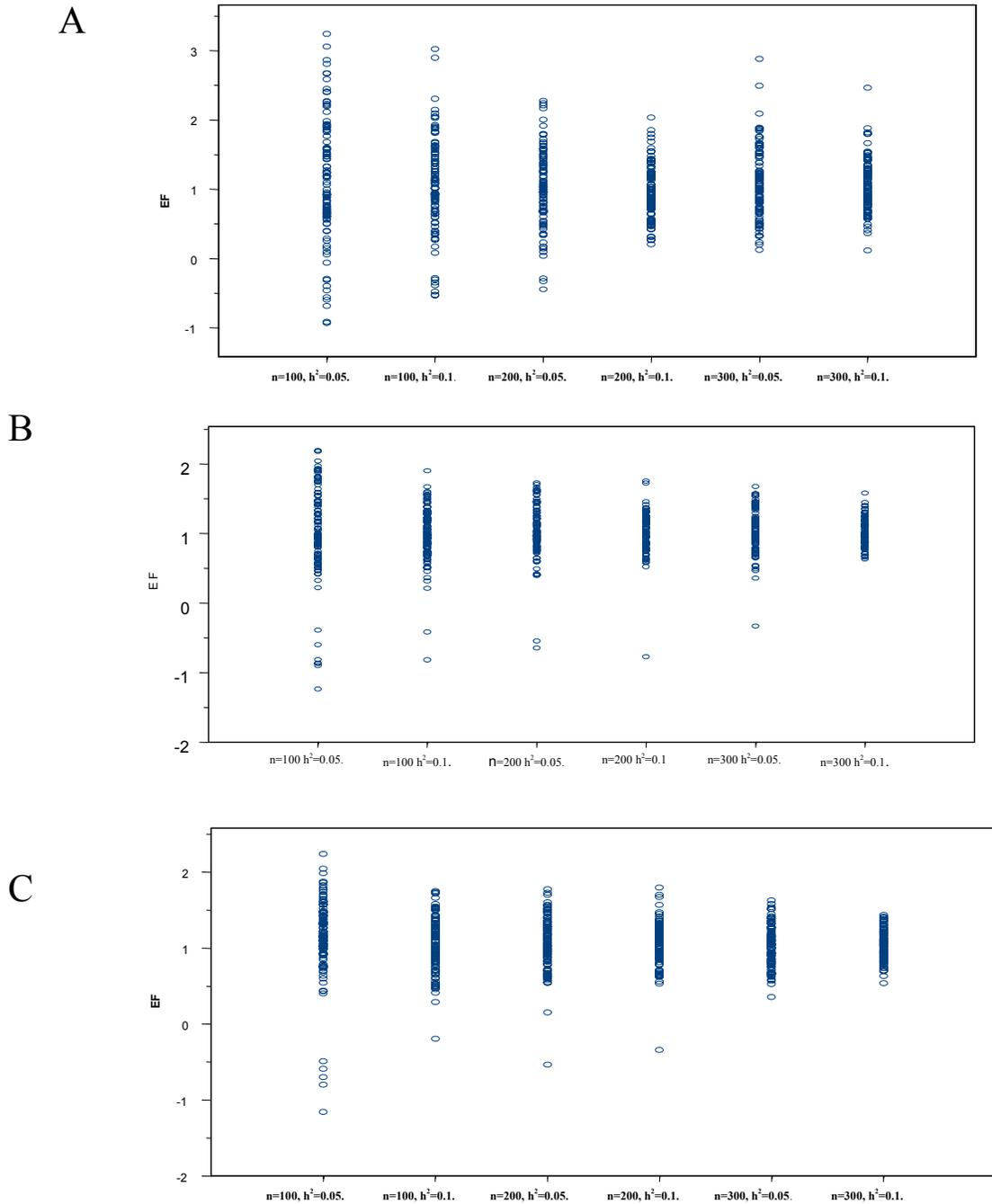


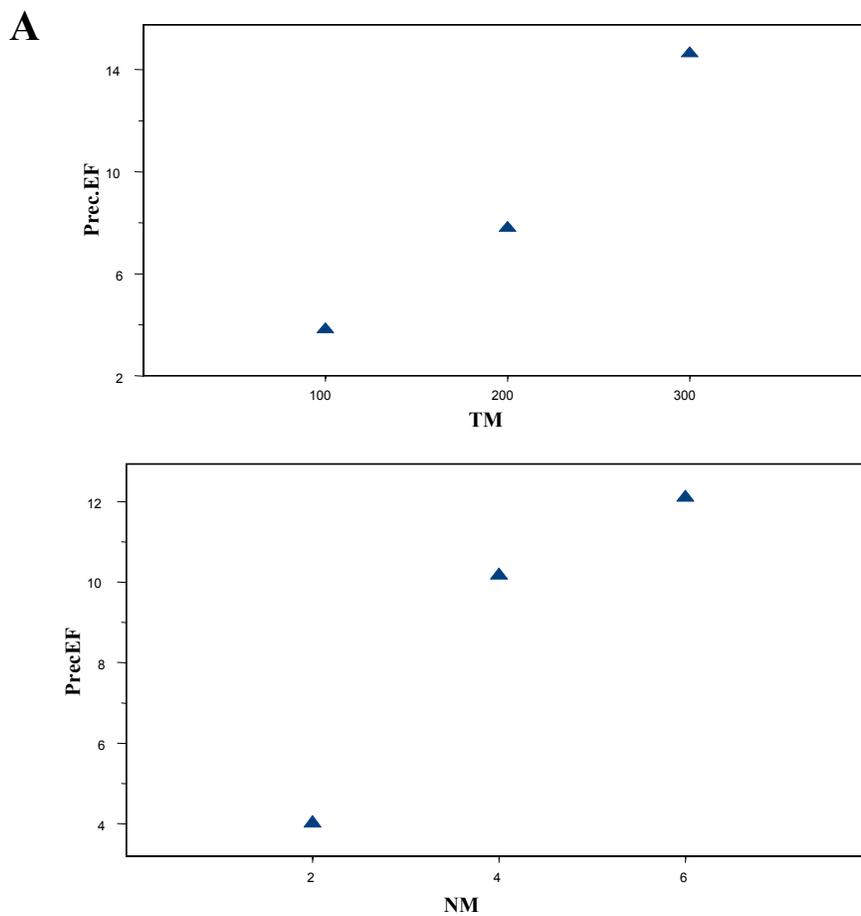
Figura 4.4. Comportamiento medio en cuanto a dispersión de las estimaciones del efecto con 100 repeticiones. **A:** Dos marcadores. **B:** Cuatro marcadores. **C:** Seis marcadores. Se mantuvo fijo el número de marcadores mientras los otros dos factores se combinaron de acuerdo a sus distintas magnitudes.

Cuadro 4.3 Varianza para el efecto estimado en los diferentes condiciones.

NM	Hered.	TM	Varianza en el efecto estimado	
2	0.05	100	0.655694	
		200	0.312229	
		300	0.259224	
	0.1	100	0.476922	
		200	0.152487	
		300	0.142311	
	4	0.05	100	0.476501
			200	0.163776
			300	0.100405
0.1		100	0.158466	
		200	0.093662	
		300	0.038456	
6		0.05	100	0.320823
			200	0.127483
			300	0.071795
	0.1	100	0.124948	
		200	0.079351	
		300	0.036701	

En cuanto a la precisión en la estimación del efecto, evaluada como la inversa de la varianza de esta estimación, en la Figura 5 se observa la influencia de cada uno de los tres factores experimentales. Sobre la base de esta gráfica se puede inferir como la relación entre los tres factores y la precisión es positiva, ya que a medida que el componente trabaja con una magnitud más grande la precisión en la estimación del efecto es mayor. Si se

analiza como afecta el tamaño de muestra (Figura 4.5A) es notorio como el cambio de una magnitud a otra en forma ascendente siempre reporta un beneficio en el rango de tamaño de muestra estudiado. Para el caso de número de marcadores (Figura 4.5B), por la relación positiva presente también se manifiesta una tendencia favorecedora continua para la precisión en la estimación. Sin embargo, el mayor beneficio se tuvo cuando se cambió de 2 a 4 marcadores, mientras en el cambio de 4 a 6 marcadores la respuesta fue menor.



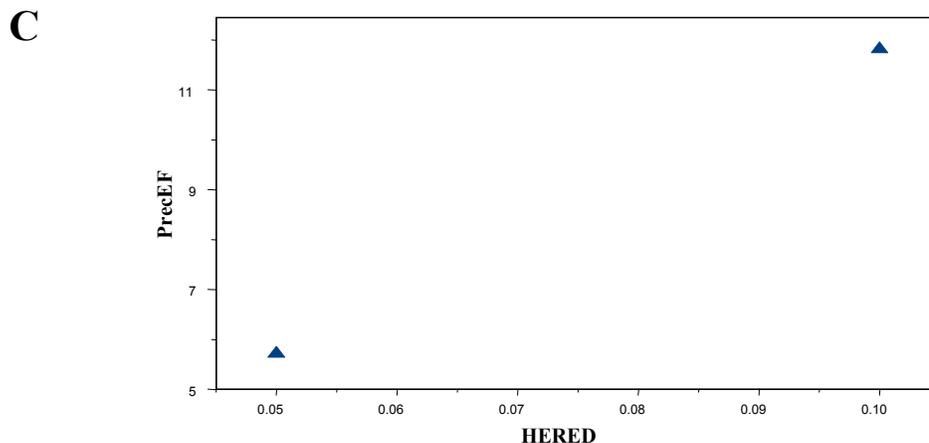


Figura 4.5 Influencia de los factores estudiados sobre la precisión en la estimación del efecto del QTL. **A.** Influencia del tamaño de muestra. **B.** Influencia del número de marcadores. **C.** Influencia de la heredabilidad.

En lo que respecta a la influencia de la heredabilidad (Figura 4.5C) claramente se ve como este factor experimental es determinante en la precisión de las estimaciones del efecto, al igual que lo fue en el aspecto de dispersión. Como no todos los cambios en los niveles de los factores experimentales contribuyen de igual forma, para favorecer la precisión, esto nos da la opción de señalar, dentro de cada uno de los factores experimentales, cuales cambios de magnitudes y combinaciones de ellos resultan ser más benéficas para la precisión.

Para el caso del sesgo condicional es decir, cuando el efecto se estima únicamente en los experimentos donde se detectó el QTL, los resultados se presentan en la Figura 4.6 En ella se muestra como actuó cada componente experimental en el sesgo condicional, de acuerdo con los niveles evaluados de

cada factor. Se aprecia como este sesgo y cada uno de los distintos factores presentan una relación negativa, esto por que al aumentar el valor con que trabaja cada factor se tiene una repercusión negativa en el sesgo condicional al reducirse su magnitud. Puede notarse que la influencia del tamaño de muestra (Figura 4.6A) sobre la dimensión en el sesgo condicional fue más amplia cuando se utilizan 200 individuos en lugar de 100. Mientras tanto se tuvo una disminución menor en el sesgo, en comparación con el anterior, cuando la muestra fue de 300 individuos en vez de 200. Para el caso de la influencia del número de marcadores en el sesgo condicional (Figura 4.6B) es notorio como se presentó un mayor efecto favorecedor en la reducción del sesgo cuando se cambió de 2 a 4 marcadores, en comparación con el beneficio logrado por utilizar 6 en lugar de 4 marcadores. El factor heredabilidad (Figura 4.6C) vuelve a manifestar una gran trascendencia en los valores que del sesgo condicional de la estimación, manifestando así su gran importancia en al investigación de QTLs. Como ya se mencionó, el sesgo se encuentra muy relacionado con el comportamiento del efecto estimado en función del valor de FA (Figuras 4.2 y 4.3); a medida que el valor promedio de FA es grande, (lo cual coincide con condiciones experimentales favorables) la pendiente de la relación entre FA y el efecto estimado disminuye, y por ende el sesgo tiende a disminuir.

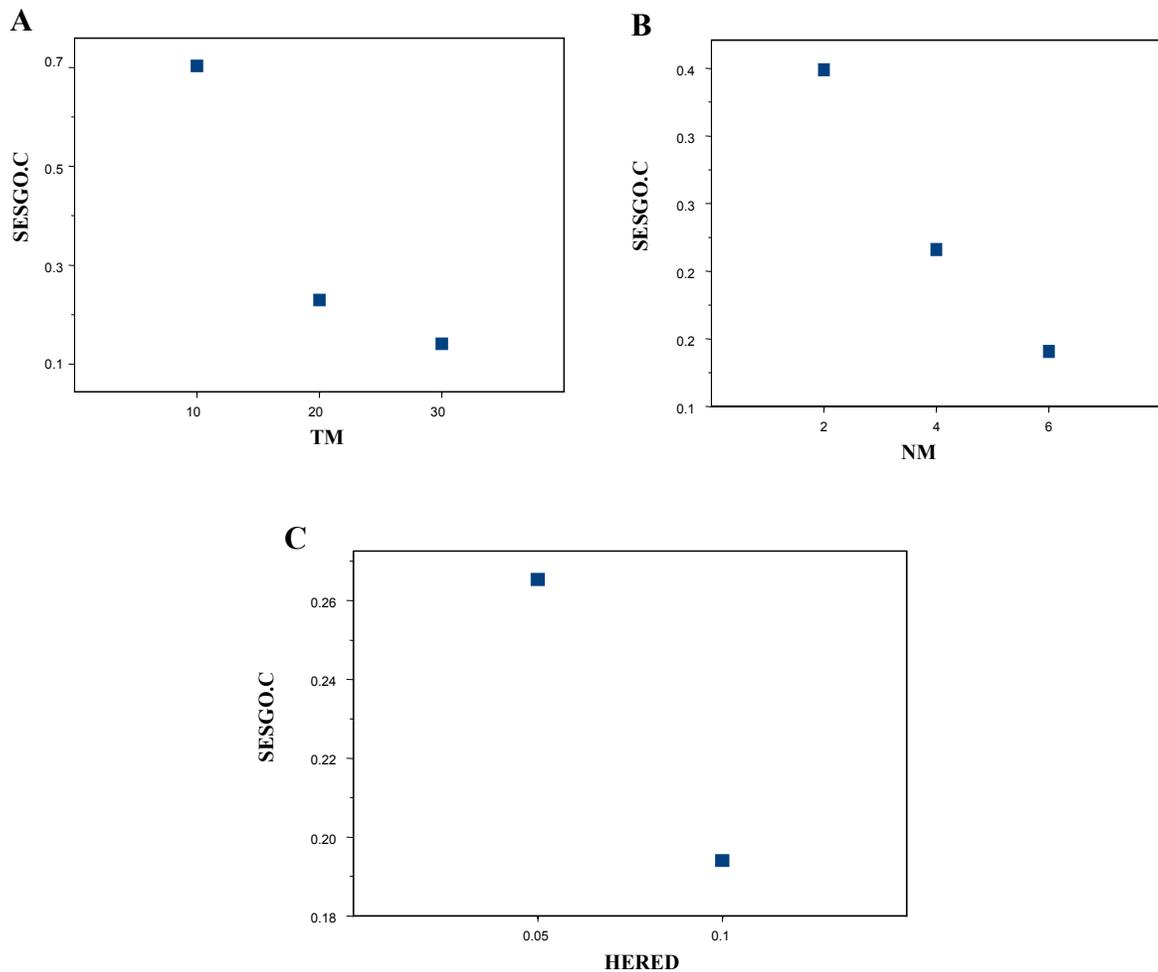


Figura 4.6 Influencia de cada factor sobre el valor promedio del sesgo condicional. **A.** Efecto de los diferentes tamaños de muestra en el sesgo condicional. **B.** Efecto de las diferentes densidades de marcadores en el sesgo condicional. **C.** Efecto de cada uno de los valores de la heredabilidad en el sesgo condicional.

En el Cuadro 4.4 Se presentan los resultados del análisis de regresión para probar la influencia de cada factor experimental en forma individual así como cuando interactúan dos de ellos a la vez, en el sesgo condicional al estimar el efecto del QTL. Primero es importante señalar como cada factor en

forma individual presenta una alta significancia en la presencia de sesgo condicional en las estimaciones, corroborando así lo observado en la Figura 4.6. En lo que respecta a la interacción entre dos factores puede verse como todas ellas presentan una alta significancia.

Cuadro 4.4. Análisis de regresión para el sesgo condicional en cada combinación de los tres factores experimentales. Por sólo emplear las estimaciones del efecto significativas, la cantidad de datos para cada una de las diferentes combinaciones experimentales es variable. τ_i Efecto del tamaño de muestra. Δ_j Efecto del densidad de marcadores. H_k Efecto de la heredabilidad.

Efecto	Estimación	Error Estandar	Valor de	
			t	p
Intercepto	2.3028	0.1825	12.6166	0.0000
τ	-0.0046	0.0006	-7.62	0.0000
Δ	-0.1812	0.031	-5.8548	0.0000
H	-0.5873	1.7897	-6.4744	0.0000
$\tau\Delta$	0.0002	0.0001	2.618	0.0090
τH	0.0123	0.0054	2.2833	0.0226
ΔH	0.8735	0.2613	3.343	0.0009

Al generalizar el comportamiento de las estimaciones del efecto con respecto a dispersión de las estimaciones, así como precisión y sesgo condicional de estas estimaciones se infiere lo siguiente: en lo que respecta al tamaño de muestra es evidente como este factor tiene una gran relevancia en los resultados en cada uno de los tres aspectos analizados en la estimación del

efecto. Esto concuerda con lo ya mencionado por Beavis (1994) donde sugiere que las estimaciones del valor del efecto genético son afectadas por el número de progenie.

Al presentarse un efecto favorable en los cambios en sentido ascendente de un tamaño de muestra a otro, mejoran las estimaciones. Por esto se justifica el trabajar con muestras grandes en las investigaciones de QTLs, siempre que esto sea posible. En lo que respecta al número de marcadores se señala como no es muy recomendable el uso de marcadores espaciados a 100 cM a menos que los otros dos factores experimentales trabajen con valores grandes. Con esto se pretende el compensar el efecto no favorable de emplear una densidad de marcadores baja. Para el caso de 4 y 6 marcadores se observa en la Figura 4.6 y Cuadro 4.1 (columna SESGO-C), como estas densidades reportan una gran reducción en el sesgo condicional promedio. También se presenta una disminución en la dispersión de las estimaciones (Figura 4.4), dispersión que disminuye e incluso llega a ser muy semejante entre las densidades de 4 y 6 marcadores cuando el tamaño de muestra se encuentre entre 200 o 300 individuos. En cuanto a la precisión en la estimación, la ganancia lograda de 4 a 6 marcadores es menos radical en comparación con la que se tiene por el cambio de 2 a 4 marcadores (Figura 4.5). Por lo antes expuesto, se establece que una buena densidad de marcadores, para un cromosoma de 100 cM, se encuentra entre un rango de 4 a 6 marcadores equidistantes. Lo anterior coincide en gran medida con lo recomendado por Visscher y Haley (1998).

En lo que respecta al factor heredabilidad se hace patente como resulta definitivo en cada uno de los tres aspectos analizados en la estimación del efecto del QTL. Y esto se distingue claramente en las tres gráficas donde se evalúan los resultados reportados según la heredabilidad trabajada, observándose una amplia diferencia en los resultados promedio de cada una (Figura 4.4, Figura 4.5 y Figura 4.6).

En la Figura 4.7 se describe el comportamiento de la precisión en la estimación de la localización del QTL. Se observa como la precisión en la estimación de la posición del QTL se ve afectada por cada factor experimental y dentro de ellos por los diferentes niveles que este componente puede asumir. Examinando en forma individual el componente tamaño de muestra (Figura 4.7A) se hace notorio la mejora que se alcanza en precisión cuando se trabaja con una muestra de 300 individuos en vez de 200, mientras que el beneficio en la precisión al cambiar de un tamaño de 100 a uno de 200 fue menor comparado con el anterior. Esta discrepancia pudo obedecer a error de muestreo en las simulaciones. La densidad de marcadores provocó una mejora en la precisión de la estimación de la localización constante al cambiar de una densidad a otra. Sin embargo no se espera que esta tendencia siga igual a medida que se sigue aumentando el número de marcadores. El factor heredabilidad se manifiesta como determinante para la calidad de resultados logrados en la investigación. Con base en lo mostrado en esta figura, se puede señalar que la precisión en la estimación de la posición y cada factor experimental presentan una relación positiva, pues se tiene una mejora

constante en la precisión a medida que los diferentes componentes asumen valores mayores en la investigación.

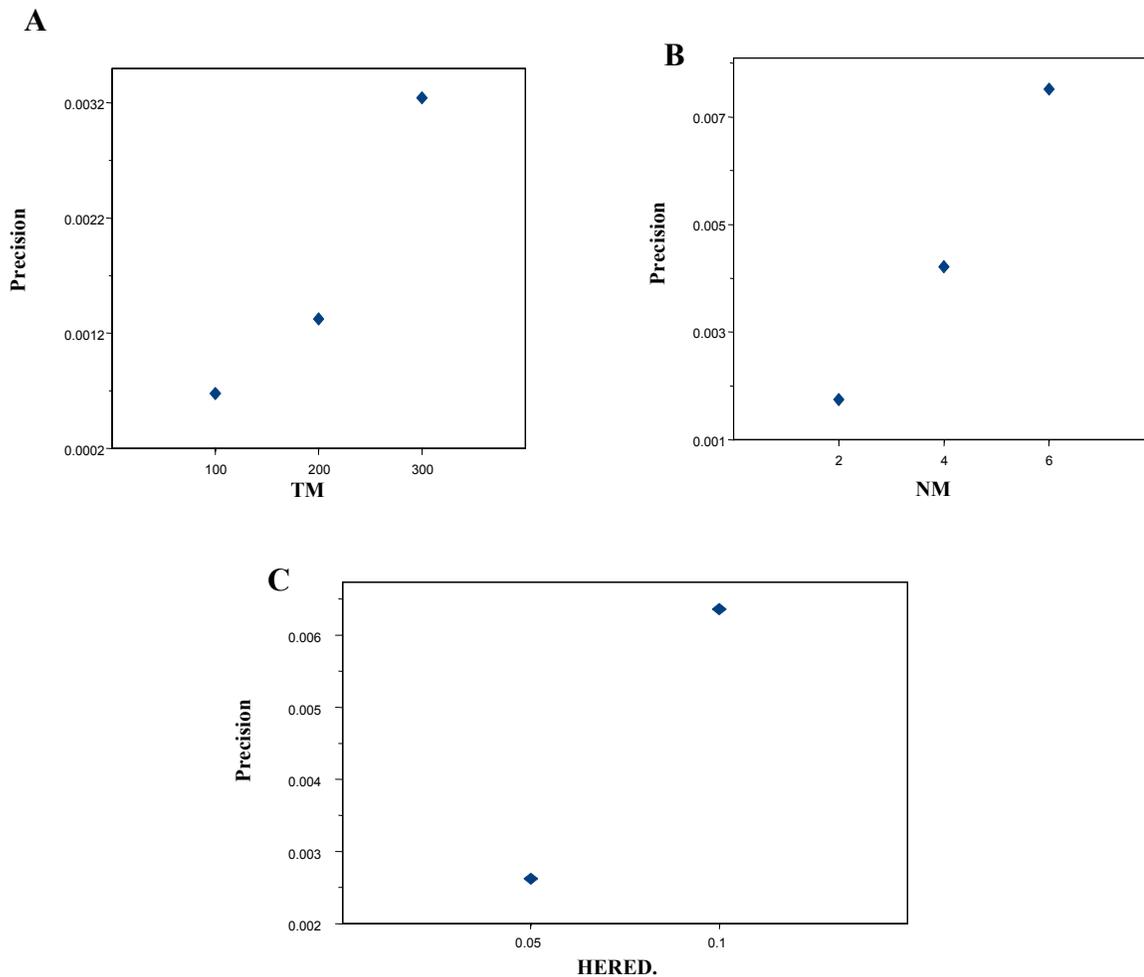


Figura 4.7 Comportamiento de la precisión en la estimación de la ubicación del QTL, de acuerdo a los diferentes factores experimentales investigados. **A.** Influencia del tamaño de muestra. **B.** Influencia del número de marcadores. **C.** Influencia del valor de la heredabilidad.

A diferencia de las estimaciones del efecto, en las de ubicación del QTL- al menos para las condiciones experimentales estudiadas en el presente trabajo- no se aprecia un punto donde la ganancia en los resultados para cada uno de los factores tienda a estabilizarse. Una extensión del espacio de niveles

de los factores aquí estudiados permitiría el fijar en una cantidad óptima en la cual un cambio en el factor no reporta muchos beneficios.

Es importante señalar que la tendencia de la desviación cuadrática media entre la posición real y a la estimada, es hacia alcanzar el valor de cero a medida que el valor de FA es mayor (Cuadro 4.1). Esto repercute en beneficio de las estimaciones del efecto pues, como ya se mencionó al final, sólo se toman en cuenta los resultados que son significativos y estos son precisamente los que presentan los valores mayores de FA siendo a la vez las estimaciones que reportan una menor desviación de la estimación.

Con base en los resultados presentados y lo señalado a partir de ellos se infiere la gran influencia que los tres factores experimentales: heredabilidad, tamaño de muestra y número de marcadores, tienen en el análisis de QTLs, en cuanto a: detección, precisión en la localización, así como precisión y sesgo condicional del efecto estimado del QTL. La heredabilidad presenta el gran inconveniente de no ser un factor cuya magnitud pueda ser establecida arbitrariamente por el investigador. Por lo tanto su efecto debe ser disminuido por medio de condiciones experimentales óptimas para la evaluación fenotípica.

El tamaño de muestra también resulta ser determinante para la obtención de buenos resultados en todas las estimaciones realizadas, con la ventaja de que la magnitud de este factor si es definida por el investigador. Sin embargo, no siempre es posible trabajar con el tamaño de muestra adecuado

ya sea por causas económicas o de manejo. De ahí la importancia de investigar si con el manejo de otro factor experimental es posible trabajar con un tamaño de muestra no muy grande y a la vez alcanzar resultados satisfactorios. Una alternativa que se emplea es el trabajar con una muestra pequeña para detectar regiones del genoma donde es probable la presencia del QTL de interés. Después de tener detectada la probable región, adicionar marcadores que sean específicos para esta región del genoma, así como individuos. Finalmente en esta segunda muestra se realizan las estimaciones para localización y efecto del QTL.

La densidad de marcadores aun y cuando es mencionada en tercer término no implica que tenga una menor importancia, por otro lado es un factor que presenta flexibilidad para ser manejado por el investigador.

V. CONCLUSIONES

Existe un ligero sesgo positivo en la estimación del valor absoluto del efecto de un QTL en experimentos de retrocruzas por medio del método de mínimos cuadrados. Este sesgo tiende a crecer cuando la estimación se realiza únicamente en los casos de detección significativa del QTL. Esto se debe a una correlación entre el estadístico de prueba (F) y el valor del estimado. De aquí que entre más estricto sea el nivel de significancia impuesto, el sesgo tenderá a ser mayor.

Dicho sesgo es más acusado cuando se utilizan tamaños de muestras pequeños y bajas densidades de marcadores. Por lo tanto, es de esperarse que en la literatura científica de experimentos de mapeo por intervalos, exista una tendencia a sobreestimar los efectos de los QTLs.

VI. RESUMEN

El estudio de QTLs en una población segregante por medio de marcadores moleculares se encuentra limitado por las características del experimento y la magnitud de los efectos individuales de cada gen. Los resultados reportados de trabajos en el análisis de QTLs típicamente presentan uno o más loci con un gran efecto genético y varios loci adicionales explicando una pequeña cantidad de la variación genotípica. Al detectarse sólo los loci con efectos mayores se propicia la sobreestimación de los efectos de QTLs. Este error es muy difícil de eliminar y no se había hecho evidente hasta que Beavis (1994) lo reportó.

El objetivo del presente estudio fue evaluar el sesgo en la estimación de los efectos de QTLs, bajo diferentes condiciones experimentales en mapeo de regresión por intervalos. Se trabajó con tres factores experimentales: tamaño de muestra (100, 200 y 300), heredabilidad (10 y 5 por ciento) y densidad de marcadores (2, 4 y 6). Se estableció un grupo de ligamiento de 100 cM en el que se distribuyeron uniformemente cada una de las tres densidades de marcadores. Con la combinación de los distintos niveles de cada uno de los tres factores

experimentales se obtuvieron 18 diferentes condiciones por estudiar, sometiéndose cada una de ellas a 100 repeticiones. Con la utilización de los comandos de simulación genética del paquete MathGenome (Reyes-Valdés, 2003) se obtuvo la población segregante para evaluar. El análisis de los datos generados se realizó aplicando el método de mapeo por intervalos con regresión. Esto con la aplicación del programa SATORI (Reyes-Valdés, 2003). Además se realizaron pruebas de permutación con 3 de las 100 repeticiones de cada una de las condiciones experimentales investigadas, con la finalidad de obtener un umbral crítico contra el cual comparar el estadístico de prueba obtenido en la regresión.

Con el análisis de los resultados se puede establecer que existe un ligero sesgo positivo en la estimación del valor absoluto del efecto de un QTL en experimentos de retrocruzas por medio del método de mínimos cuadrados. Este sesgo tiende a crecer cuando la estimación se realiza únicamente en los casos de detección significativa del QTL. Esto se debe a una correlación entre el estadístico de prueba (F) y el valor del estimado. De aquí que entre más estricto sea el nivel de significancia impuesto, el sesgo tenderá a ser mayor. Dicho sesgo es más acusado cuando se utilizan tamaños de muestras pequeños y bajas densidades de marcadores.

VI. LITERATURA CITADA

- Adkins, Gerald, and Pooch, Udow. 1977 "Computer Simulation: A Tutorial", *Computer*, 10, 4, April.
- Beattie, CW. 1994 Livestock genome maps. *Trends Genet.* 10: 334-338
- Beavis W.D. 1994 The power and deceit of QTL experiments ; lessons from comparative QTL studies. *Jurnal Title: Report of annual Corn and Sorghum. Research Conference Pages: 250-266 ARIEL*
- Beckmann, J.S. and Soller, M. 1983 Restriction fragment length polymorphisms in genetic improvement: methodologies, mapping and costs. *Theor. Appl. Genet.* 67: 35-43.
- Brooks, L.D. 1988 The evolution of recombination rates. *In R.E. Michod and B.R. Levin (eds), The evolution of sex*, p.p. 87-105. Sinauer Assoc., Sunderland, MA.
- Burns, M. 1997 Quantitative trait loci Mapping in Arabidopsis; *Theory and Practice*. PhD, University of Birmingham.
- Churchill G.A. y Doerge R.W. 1994 Empirical threshold values for quantitative trait mapping. *Genetics* 138: 963-971.
- Crow, J.F. 1993a Francis Galton: counts and measure, measure and count. *Genetics* 135: 1-4.
- Daniel, Wayne W. 1995 Bioestadística, base para el análisis de las ciencias de la salud. Editorial UTEHA.
- Darvasi, A. and M. Soller. 1994b Optimum spacing of genetic markers for determining linkage between marker loci and quantitative trait loci. *Theor. Appl. Genet.* 89: 351-357.
- Fisher, R.A. 1918 The correlation between relatives on the supposition of Mendelian inheritance. *Trans. Roy. Soc. Edinburgh.* 52: 399-433.
- Fisher, R.A. 1935 *The designing experiments*. 8th Ed. Hafner, N.Y.

- Galton, F. 1865 Hereditary talent and character. *Macmillan's Magazine*, 12, 157-166; 318-327.
- Galton, F. 1889 Natural Inheritance. Macmillan, London.
- Gelderman, H. 1975 Investigation on inheritance of quantitative characters in animals by gene markers. I. Methods. *Theor. Appl. Genet.* 46: 300-319
- Georges M., Nielsen D., Mackinnon M., Mishira A., Okimoto R., Zao X., Womack J. y Hoeschle I. 1995 Mapping quantitative trait loci controlling milk production in dairy cattle by exploiting progeny testing. *Genetics* 39: 907-920.
- Haldane, J.B.S. 1919 The combination of linkage values, and the calculation of distance between the loci linked factors. *J. Genetics* 8: 299-309.
- Haley C.S. y Knott S.A. 1992 A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* 69: 315-324.
- Haley, C.S., S.A. Knott and J.M. Elsen 1994 Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136: 1195-1205.
- Jansen, R.C. 1993a Maximum likelihood in a generalized linear finite mixture model by using the EM algorithm. *Biometrics* 49: 227-231.
- Jansen, R.C. and P. Stam. 1994 High resolution of quantitative trait into multiple loci via interval mapping. *Genetics* 136: 1447-1455.
- Kearsey M.J. and Pooni H.S. 1996 The genetical analysis of quantitative traits. CHAPMAN & HALL.
- Kersey M.J. and Farquhar A.G. 1998 QTL analysis in plants; where are we now? short review. *Heredity* 80: 137-142
- Kleinjnen, J.P.C. 1974 *Statistical Thechniques in Simulation*. Part1, Dekker.
- Knott S.A., Neale D., Sewell M. y Haley C. 1997 Multiple markers mapping of quantitative trait loci in an outbred pedigree of loblolly pine. *Theor. Appl. Genet* 94: 810-820.
- Lander, E.S. and Botstein, D. 1989 Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121: 185-199.

- Lewis P.A. y Orav E. 1989 Simulation methodology for statisticians, operations analysts and engineers Vol I Wadsworth & Brooks/Cole Statistics Probability Series.
- Lynch M. y Walsh B. 1998 Geneticas and analysis of quantitative traits. Sinauer Associates, Inc. Publishers.
- Martínez O. y Curnow R. N. 1992 Estimating the locations and sizes of effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* 85: 480-488.
- Muller, H.J. 1916 *Amer. Nat.* 50: 193-434
- Owen A.R.G. 1950 The theory of genetical recombination. *Adv. Genet.* 3: 117-157.
- Paterson, A.H., Lander E. Hewitt, J.D., Peterson S. Lincoln, S.E., Tanksley, S.D. 1988 Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* 355: 721-726
- Phillips, D.T., Ravindran, A., and Solberg, J.J. 1976 Operations Research Principles and Practice, New York: John Wiley and Sons.
- Pooch U.W. y Wall J.A. 1993 Discrete event simulation a practical approach. CRC Press, Inc.
- Rebaï, A. B. Goffinet, and B. Mangin 1995. Comparing power of different methods of QTL detection. *Biometrics.* 51: 87-99.
- Reyes-Valdés M.H. 2000 A model of marker-based selection in gene introgression breeding programs. *Crops Sci.* 40: 91-98
- Reyes-Valdés, M. H., 2003. Guide to use MathGenome. <<http://www.uaaan.mx/~mhreyes/mtgnm.htm>>
- Reyes-Valdés, M.H. 2003. SATORI: a Mathematica application for QTL análisis in three-generation outbred pedigrees. Universidad Autónoma Agraria Antonio Narro <<http://www.uaaan.mx/~mhreyes/sat.htm>>
- Sax, K. 1923 The association of size differences with seed-coat pattern and pigmentation in *Phaseolus vulgaris*. *Genetics.* 8: 552-560.
- Stingler, S.M. 1986 *The history of statistics*. Harvard Univ. Press, Cambridge, M.A.

- Trow, A.H. 1923. Forms of reduplication: primary and secondary. *J.Genetics* 2: 313-324.
- Utz,H.F., and A.E. Melchinger 1994. Comparison of different approaches to interval mapping of quantitative trait loci. In J.W. van Ooijen and J. Jansen (eds), *Biometrics in plant breeding: applications of molecular markers*, p.p. 195-204. CPRO-DLO, Netherlands.
- Visscher, P.M., Thompson, R.N., and Haley, C.S. 1996 Confidence intervals in QTL mapping by bootstrapping. *Genetics*. 143: 1013-1020.
- Visscher P.M. y Haley C.S. 1998 Power of chromosomal test to detect genetic variation using genetic markers. *Heredity* 81: 317-326
- Weissenbach, J.A. 1993 A second generation linkage map of the human genome based on highly informative microsatellite loci. *Gene* 135: 275-278.
- Whittaker, J.C., Thompson, R., and Visscher, P.M. 1996 On the mapping of QTL by regression of phenotype on marker type. *Heredity* 77: 23-32
- Wilks, S.S. 1938 The large sample distribution of the linkelihood ratio for testing composite hypotheses. *Ann. Math. Stat.* 9: 60-62
- Williams, J.G.K., A.R. Kubelik, K.J. Livak, J.A. Rafalski, and S.V. Tingey 1990 DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucl. Acids Res.* 18: 6531-6535.
- Wolfram, Stephen. 1988 Mathematica® A System for Doing Mathematics by Computer. Second Edition Addison-Wesley Publishing Company.
- Xu S. 1998 Further investigation on the regression method of mapping quantitative trait loci. *Heredity* 80: 364-373.
- Zebeau M, Vos P. 1993 Selective restriction fragment amplification: a general meted for DNA fingerprinting. EPO Patent No. 0534858A1
- Zeng, Z-B 1992 Correcting the bias of Wright's estimates of the number of genes affecting a quantitative character: a further improved method. *Genetics*. 131: 987-1001.
- Zeng, Z-B 1993 Theoretical basis for separation of multiple linked gene effects in mapping of quantitative trait loci. *Proc. Natl. Acad. Sci. USA* 90: 10972-10976.
- Zeng, Z-B 1994 Precision mappig of quantitative trait loci. *Genetics* 136: 1457-1468.