

**ANÁLISIS DE QTLs EN UNA POBLACIÓN DE MAPEO DE GIRASOL  
(*Helianthus annuus* L.) DERIVADA DE UN CRUZAMIENTO  
INTER-SUBESPECÍFICO**

**CARLOS JAVIER LOZANO CAVAZOS**

**T E S I S**

**Presentada como Requisito Parcial  
para Obtener el Grado de:**

**DOCTOR EN CIENCIAS  
EN FITOMEJORAMIENTO**



**Universidad Autónoma Agraria  
“Antonio Narro”**

**PROGRAMA DE GRADUADOS**

**Buenavista, Saltillo, Coahuila, México.**

**ABRIL DE 2008.**

**UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO**  
**SUBDIRECCIÓN DE POSTGRADO**

**ANÁLISIS DE QTLs EN UNA POBLACIÓN DE MAPEO DE GIRASOL**  
**(*Helianthus annuus* L.) DERIVADA DE UN CRUZAMIENTO INTER-**  
**SUBESPECÍFICO**

**T E S I S**

**Por**

**CARLOS JAVIER LOZANO CAVAZOS**

Elaborada bajo la supervisión del Comité Particular de Asesoría y aprobada como  
requisito parcial para obtener el grado de:

**DOCTOR EN CIENCIAS EN FITOMEJORAMIENTO**

**Comité Particular**

Asesor principal:

\_\_\_\_\_  
**Dr. M. Humberto Reyes Valdés**

Asesor:

\_\_\_\_\_  
**Dr. Juan Manuel Martínez Reyna**

Asesor:

\_\_\_\_\_  
**Dr. Octavio Martínez de la Vega**

Asesor:

\_\_\_\_\_  
**Dr. Alberto Flores Olivas**

Asesor:

\_\_\_\_\_  
**Dr. Rosalinda Mendoza Villarreal**

\_\_\_\_\_  
**Dr. Jerónimo Landeros Flores**  
Director de Postgrado

Buenavista, Saltillo, Coahuila, México. Abril de 2008.

## AGRADECIMIENTOS

**A mi Universidad Autónoma Agraria Antonio Narro**, por haber sido un recinto fundamental en mi preparación profesional.

**Al Dr. M. Humberto Reyes Valdés**, por permitirme formar parte de su equipo de trabajo de investigación, por su asesoría, apoyo, confianza y amistad.

**Al Dr. Juan Manuel Martínez Reyna**, por formar parte de mi Comité de Asesoría, así como sus valiosas sugerencias en la realización del presente documento.

**A la Dra. Rosalinda Mendoza Villarreal**, por su valiosa colaboración y gran apoyo incondicional en la realización de trabajos de laboratorio relacionados a la presente investigación, así como en la revisión de tesis.

**Al Dr. Octavio Martínez de la Vega y al Dr. Alberto Flores Olivas**, por formar parte de mi Comité de Asesoría.

**A Miguel Rodríguez** por su valiosa ayuda en la realización del trabajo de campo y su gran amistad.

**A la M.C. Alma Patricia García Villanueva**, por su valiosa colaboración en el desarrollo de la presente investigación.

**A los Maestros del programa en Fitomejoramiento**, que me brindaron sus conocimientos, los cuales contribuyeron en gran medida en mi formación profesional.

**A todos mis compañeros y amigos** del Programa de Postgrado en Fitomejoramiento por su gran apoyo y amistad a lo largo de mi estancia en el mismo.

## DEDICATORIA

A **DIOS** por todas sus bendiciones y amor infinito para superar este reto profesional.

A mi esposa **Aurora Gabriela Dávila Valdés**, por su valioso amor, apoyo incondicional y comprensión en todo momento de mi vida.

A mis hijos **Gabriela Estefanía y Carlos Alexis Lozano Dávila**, por ser en todo momento una luz de amor y fortaleza.

A mis padres **Sr. Alejandro Javier Lozano del Río y Elsa Cavazos Dávila**, por su gran amor, consejos, sacrificios y apoyo en cada meta que me he trazado en la vida.

A mi hermano **Eloy Alejandro Lozano Cavazos**, por sus consejos, amor y apoyo en todo momento y proyecto de mi vida.

## COMPENDIO

ANÁLISIS DE QTLs EN UNA POBLACIÓN DE MAPEO DE GIRASOL (*Helianthus annuus* L.) DERIVADA DE UN CRUZAMIENTO INTER-SUBESPECÍFICO

Por

CARLOS JAVIER LOZANO CAVAZOS

DOCTORADO EN CIENCIAS EN FITOMEJORAMIENTO

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO  
BUENAVISTA, SALTILLO, COAHUILA, MÉXICO. ABRIL DE 2008.

Dr. M. Humberto Reyes Valdés---Asesor---

**Palabras clave:** Girasol cultivado, girasol silvestre, mapa de ligamiento, cruza inter-subespecífica, QTL.

El mejoramiento del girasol, la segunda especie con mayor producción de aceite a nivel mundial, depende en gran medida de la introducción de diversidad genética de material silvestre. El descubrimiento de restos de girasol domesticado de aproximadamente 4,000 años de antigüedad en San Andrés, Tabasco, implica un origen de domesticación más antiguo y posiblemente independiente en México con relación al este de Norteamérica. El objetivo del presente trabajo fue detectar posiciones de loci de caracteres cuantitativos (QTL's) y estimar sus efectos genéticos, especialmente aquellos relacionados con la domesticación del girasol. Se trabajó con ocho caracteres: altura de planta, número y diámetro de capítulos, número y peso de achenios, días a floración,

días a madurez fisiológica, y contenido de aceite de aquenios, con el enfoque de mapeo por intervalos en una población de 149 líneas F<sub>2:3</sub>, derivada de un cruzamiento de girasol cultivado “HA89” (*Helianthus annuus* L. var. *macrocarpus* (DC.) Cockerell) y silvestre “Ac-8-2” (*Helianthus annuus* L. ssp. *texanus* Heiser). Para la evaluación de la población F<sub>2:3</sub> en condiciones de campo, incluyendo los dos progenitores, se utilizó el diseño experimental de bloques incompletos al azar con dos repeticiones. Para llevar a cabo los análisis de QTL’s se utilizó el software MapQTL 4.0, tomando como base un mapa de ligamiento genético de baja densidad basado en marcadores AFLP, el cual cubre 581.452 cM en 17 grupos de ligamiento, y de acuerdo a los diferentes mapas genéticos publicados, este representa aproximadamente el 43.73% del promedio del genoma de *Helianthus annuus* L. Se consideró un nivel de significancia estadística de amplitud genómica de 0.05 en la detección de QTL’s. Para establecer los valores críticos de los estadísticos de prueba se realizaron pruebas de permutación, solamente para aquellos grupos de ligamiento con un LOD score >1.5. Para determinar alguna posible asociación carácter-marcador se realizó un ANOVA por locus individual no ligado para cada carácter evaluado en el software de análisis de datos R (versión 2.6.0). Se identificó un QTL significativo con amplitud genómica de 0.017 para peso de aquenios. Asimismo, se identificaron cinco QTL’s sugestivos (aquellos mostrando una significancia de amplitud de grupo de ligamiento < 0.05, pero con valor de amplitud genómica > 0.05) a través del mapeo por intervalos para altura de planta, número y diámetro de capítulos, número y peso de aquenios. Los QTL’s detectados (incluyendo los sugestivos) explicaron de 7.1 al 11.9% de la varianza fenotípica para cinco caracteres evaluados. El análisis de marcadores individuales identificó cuatro loci no ligados que afectan al carácter número de aquenios y días a madurez fisiológica en forma altamente significativa ( $P < 0.001$ ). El

contenido de aceite de akenios en 83 líneas de girasol fluctuó de 20.16 a 43.10% a través del método Blight y Dyer. Este estudio contribuirá al conocimiento de la genética de la domesticación del girasol, así como en la generación de marcadores para selección con fines de mejoramiento genético, dirigido a la introgresión de caracteres silvestres a variedades de girasol elite.

## ABSTRACT

QTL ANALYSIS IN A SUNFLOWER MAPPING POPULATION (*Helianthus annuus*  
L.) DERIVED FROM AN INTER-SUBSPECIFIC CROSS

By

CARLOS JAVIER LOZANO CAVAZOS

DOCTOR OF SCIENCE IN PLANT BREEDING

UNIVERSIDAD AUTÓNOMA AGRARIA ANTONIO NARRO  
BUENAVISTA, SALTILLO, COAHUILA, MÉXICO. APRIL, 2008

Ph.D. M. Humberto Reyes Valdés--Adviser---

**Key words:** Cultivated sunflower, wild sunflower, linkage map, inter-subspecific cross,  
QTL.

The improvement of cultivated sunflower, the second largest oilseed crop in the world, depends largely on the introduction of genetic diversity from wild species. The discovery of 4,000 years old domesticated sunflower remains from San Andrés, Tabasco, implies an earlier and possibly independent origin of domestication in México than eastern North America. The purpose of the present study was to detect quantitative trait loci (QTL's) positions and its genetic effects, especially those related with sunflower domestication. Eight traits were evaluated and analyzed by interval mapping: plant height, branching, head diameter, number of achenes, achene weight, days to flowering, days to physiological maturity, and seed oil content in a population of 149

lines  $F_{2:3}$ , derived from the cross of a cultivated sunflower line “HA89” (*Helianthus annuus* L. var. *macrocarpus* (DC.) Cockerell), and a wild sunflower accession “Ac-8-2” (*Helianthus annuus* L. ssp. *texanus* Heiser). The  $F_{2:3}$  population and its progenitors were grown in field conditions, by using a random incomplete block design with two replications. QTL analyses were carried out in MapQTL 4.0, in a low-density genetic linkage map based on AFLP markers, which covers 581.452 cM on 17 linkage groups, and it represents approximately 43.37 % of the average *Helianthus annuus* L. genome, according to the genetic linkage maps published up to date. A genome-wide statistical significance of 0.05 was used to detect QTL’s. To determine the empirical significance threshold values, permutation tests were performed, only in linkage groups showing a LOD score  $>1.5$ . Single marker-analyses using ANOVA per unlinked individual locus for each trait evaluated was carried out to determine any possible association marker-trait in R (versión 2.6.0). One QTL with genome-wide significance of 0.017 for achene weight was identified. Likewise, interval mapping identified five putative QTL’s (those showing a linkage group-wide statistical significance  $< 0.05$ , but with a genome-wide statistical significance  $> 0.05$ ) for plant height, branching, head diameter, number of achenes, and achene weight. All of QTL’s detected explained 7.1 to 11.9% of the phenotypic variance for five traits evaluated. Single-marker analyses identified highly significantly ( $P < 0.001$ ) four unlinked loci for number of achenes and days to physiological maturity. The seed oil content in 83 sunflower lines ranged from 20.16 to 43.10 % by the Blight and Dyer method. This work will contribute to the knowledge of the genetics of sunflower domestication, and it will generate markers for selection with breeding purposes, aimed to wild trait introgression in elite sunflower cultivars.

## ÍNDICE DE CONTENIDO

	Página
ÍNDICE DE CUADROS -----	<i>xiii</i>
ÍNDICE DE FIGURAS -----	<i>xiv</i>
I. INTRODUCCIÓN -----	1
1.1. Objetivos -----	4
II. REVISIÓN DE LITERATURA -----	5
2.1. Origen del girasol cultivado -----	5
2.2. Taxonomía -----	6
2.3. Distribución geográfica-----	7
2.4. Especiación del girasol -----	7
2.5. Poblaciones de mapeo -----	9
2.6. Selección del modelo genético -----	11
2.7. Mapeo de QTL's-----	12
2.7.1. Métodos estadísticos para el mapeo de QTL's -----	13
2.7.1.1. Análisis de marcador individual -----	13
2.7.1.2. Mapeo por intervalos -----	15
2.8. LOD score -----	19
2.9. Significancia estadística (Prueba de hipótesis) -----	20
2.9.1. Comportamiento del estadístico de prueba bajo la hipótesis nula -----	22

2.9.2. Pruebas de permutación-----	23
2.9.3. Valores de $P$ ajustados -----	25
2.9.4. Intervalos de confianza para QTL's -----	27
2.10. QTL's múltiples -----	28
2.10.1. Métodos de mapeo de QTL's múltiples -----	29
2.10.1.1 Regresión múltiple -----	29
2.10.1.2. Propiedades del análisis de regresión múltiple -----	29
2.11. Mapeo múltiple por intervalos -----	31
III. MATERIALES Y MÉTODOS -----	34
3.1. Antecedentes de investigación -----	34
3.1.1. Formación de la población de mapeo -----	34
3.1.2. Genotipificación -----	34
3.1.3. Construcción del mapa genético -----	35
3.1.4. Análisis de ligamiento -----	35
3.2. Localización experimental -----	37
3.3. Material genético -----	37
3.4. Evaluación fenotípica replicada -----	38
3.5. Caracteres evaluados -----	40
3.6. Diseño experimental y análisis estadístico -----	46
3.7. Análisis genético -----	47
IV. RESULTADOS Y DISCUSIÓN -----	48
V. CONCLUSIONES -----	58
VI. RESUMEN -----	59

VII. LITERATURA CITADA -----	63
VIII. APÉNDICE -----	74

## ÍNDICE DE CUADROS

Cuadro		Página
4.1	Cuadrados medios para seis caracteres relacionados a la domesticación del girasol en una población segregante $F_2:F_3$ -----	50
4.2	Resultados del análisis de QTL's a través de mapeo por intervalos para caracteres relacionados a la domesticación del girasol en una población segregante $F_2:F_3$ . -----	54
4.3	Loci marcadores no ligados con significancia al 0.001 de probabilidad, para caracteres relacionados a la domesticación del girasol en una población segregante $F_2:F_3$ a través del método de marcador individual (ANOVA) -----	56

## ÍNDICE DE FIGURAS

<b>Figura</b>		<b>Página</b>
<b>3.1</b>	Líneas F <sub>2:3</sub> (fenotipos replicados) evaluadas en el Campo Agrícola Experimental de la Universidad Autónoma Agraria Antonio Narro en Navidad, N. L.-----	39
<b>3.2</b>	Extracción de lípidos a través del método propuesto por Bligh y Dyer. -----	43
<b>4.1</b>	Perfiles de un QTL con significancia de amplitud genómica para peso de aquenios en el grupo de ligamiento 11, denominado PA-G11. (A) <i>Lod score</i> , (B) efectos aditivos, (C) varianza fenotípica explicada por el QTL (%), y (D) media estimada de la distribución del carácter peso de aquenios (g) -----	52

## I. INTRODUCCIÓN

El este de Norteamérica es una de al menos seis regiones del mundo donde se piensa que la agricultura se originó de manera independiente. La principal evidencia de esta hipótesis proviene de cambios morfológicos en el registro arqueobotánico específicamente del cultivo de girasol. Recientemente, el girasol se ha considerado como el único cultivo domesticado en el este de Norteamérica. Sin embargo, el descubrimiento de hace aproximadamente 4000 años de restos de girasol domesticado de San Andrés, Tabasco, implica un origen más antiguo y posiblemente independiente de domesticación con relación al este de Norteamérica, lo cual ha estimulado nuevas investigaciones sobre el origen geográfico del girasol domesticado (Lentz *et al.* 2001; Pope *et al.* 2001).

Una de las herramientas más poderosas en el aprovechamiento de los recursos genéticos silvestres son los mapas genéticos basados en marcadores de DNA. El mapeo genético es una herramienta muy importante en el entendimiento de la organización genómica y en la relación entre genes y fenotipo. Un mapa genético es una representación lineal o circular del orden de marcadores y distancias entre estos en un cromosoma(s) de un organismo. Uno de los objetivos en la construcción de un mapa genético es detectar y localizar los elementos genéticos responsables de la variación en un fenotipo de interés. El diseño del experimento de mapeo involucra el seleccionar el

tipo de cruce, las líneas progenitoras involucradas, un método para medir el fenotipo, y una estrategia para la clasificación de genotipos.

Tradicionalmente, las estrategias para clasificar genotipos y fenotipos han sido evaluadas en términos de su poder para detectar un efecto genético. Esto depende del tamaño del efecto genético y la información en el experimento (Sen *et al.* 2005). El enfoque estándar para loci genéticos (QTL's) que contribuyen a la variación en un carácter cuantitativo hace uso de la suposición que la variación ambiental residual sigue una distribución normal (Lander y Botstein, 1989).

La identificación de LOCI DE CARACTERES CUANTITATIVOS (QTL's) es de suma importancia para proporcionar un medio de respuesta con respecto al control genético de los caracteres cuantitativos. La base para detectar QTL's es asociar la variación del fenotipo con los genotipos marcadores en una población segregante a través de métodos estadísticos. Muchos factores, tales como el tamaño del genoma, densidad del mapa genético, informatividad de marcadores, y la proporción de datos perdidos, pueden afectar la distribución del estadístico de prueba.

Los métodos de mapeo de QTL's pueden ser clasificados en diferentes grupos de acuerdo al número de marcadores, modelos genéticos y enfoques analíticos aplicados. Con el objetivo de detectar y/o localizar QTL's, se pueden considerar dos métodos estadísticos comunes. El primer enfoque se basa en el análisis de varianza (ANOVA), o regresión lineal simple, llevando a cabo pruebas estadísticas basadas únicamente en la información de los marcadores de DNA. No se requiere ningún mapa genético para el

análisis de marcador individual, y los cálculos se basan en medias fenotípicas y varianzas dentro de cada una de las clases genotípicas. El segundo, y más utilizado, se basa en los marcadores de DNA, en la estimación de la fracción de recombinación entre estos (o la distancia genética estimada), y en el cálculo basado en máxima verosimilitud o en un modelo de regresión incluyendo marcadores múltiples (dos o más) como variables independientes. El mapeo de loci de caracteres cuantitativos (QTL's) es un análisis de la relación entre el genotipo en varias localizaciones genómicas y el fenotipo para un conjunto de caracteres cuantitativos en términos de número, posiciones genómicas, efectos, interacción y pleiotropía de QTL's, considerando la interacción ambiental.

La detección de un QTL depende de: (1) el tipo de cruce, (2) el tamaño del efecto del QTL, (3) el número de individuos obtenidos, (4) la densidad de marcadores, y (5) y la rigidez del valor crítico seleccionado (*i.e.*, nivel de significancia). Los principios en los cuales se basa el mapeo de QTL's son: (1) disectar la herencia compleja de los caracteres cuantitativos en factores mendelianos, (2) localizar en el genoma dichos factores, y (3) estimar los efectos de los QTL's. El mapeo de QTL's tiene como objetivos fundamentales: (1) contar con información para clonación basada en mapas y, (2) contar con marcadores para selección en mejoramiento genético.

El girasol cultivado (*Helianthus annuus* L. var. *macrocarpus* (DC.) Cockerell) especie diploide anual ( $2n = 2x = 34$ ), es una de las principales fuentes para extracción de aceite de alta calidad en el mundo. Se han realizado análisis de QTL's para contenido de aceite en girasol (León *et al.* 2003; Mokrani *et al.* 2002). El contenido de aceite en la

semilla es un carácter determinado predominantemente por el genotipo materno, como lo han demostrado cruzamientos recíprocos utilizando líneas de bajo y alto contenido de aceite en aquenios de girasol (Pawlowski, 1964; Thompson *et al.* 1979; Haro-Ramírez *et al.* 2007).

El presente trabajo de investigación contribuirá al conocimiento de la genética de la domesticación del girasol, generando marcadores para llevar a cabo la selección asistida por marcadores (MAS), a través del análisis de QTL's. De acuerdo a lo anterior se plantearon los siguientes objetivos.

### **1.1. Objetivos**

1. Formación de líneas endogámicas recombinantes de girasol (RIL's).
2. Detección, localización y estimación de efectos de QTL's relacionados con la domesticación del girasol (*Helianthus annuus* L. var. *macrocarpus* (DC.) Cockerell).
3. Determinar el contenido de aceite en aquenios de líneas de girasol para determinar la posible presencia de QTL's que afecten esta característica.

## II. REVISIÓN DE LITERATURA

### 2.1. Origen del girasol cultivado

Antes de la colonización humana en Norte América, *Helianthus annuus* estaba restringido en forma silvestre, a lo que hoy es el sureste de los Estados Unidos (Heiser, 1978). En lo que respecta al girasol cultivado, se han encontrado achenios en investigaciones arqueológicas realizadas en la parte central de Estados Unidos (Seiler y Rieseberg, 1997). Por otro lado, existen nuevas evidencias que sugieren que el girasol fue domesticado primeramente en México, aunque se postula que puede ser posible un origen separado en los Estados Unidos (Piperno 2001; Burke *et al.* 2002).

Esta idea del origen de domesticación múltiple entra en conflicto con los resultados del análisis de aloenzimas y DNA de cloroplastos, que indica un virtual monomorfismo en las líneas cultivadas, lo cual sugiere un origen simple del girasol domesticado a partir de *Helianthus annuus* silvestre (Rieseberg y Seiler, 1990). Sin embargo, Harter *et al.* (2004), argumenta que las variedades actuales de girasol tienen un origen único en Estados Unidos.

## 2.2. Taxonomía

De acuerdo con el Departamento de Agricultura de los Estados Unidos de América (USDA) 2008, el girasol se clasifica taxonómicamente de la siguiente manera:

Reino: Plantae

División: Magnoliophyta

Clase: Magnoliopsida

Subclase: Asterideae

Orden: Asterales

Familia: Asteraceae

Tribu: Heliantheae

Género: *Helianthus*

Especie: *H. annuus*.

Los girasoles silvestres y cultivados exhiben un gran número de diferencias fenotípicas. Los girasoles silvestres presentan múltiples ramificaciones, hojas pequeñas, capítulos de 3.0 a 5.0 cm de diámetro, y una longitud de aquenios de 5.0 a 5.2 mm. En contraste, las plantas de girasol cultivado no presentan ramificaciones (monocéfalas), producen hojas grandes, capítulos de 9.0 a 11.0 cm de diámetro, y una longitud de aquenios de 9.5 a 10.5 mm. (Burke *et al.* 2002).

### 2.3. Distribución geográfica

El girasol cultivado (*Helianthus annuus* L. var. *macrocarpus* (DC.) Cockerell) se encuentra ampliamente adaptado a diversas áreas geográficas, desde subtropicales hasta subárticas (Jan *et al.* 1998; Lawson *et al.* 1998).

Las especies silvestres de girasol están adaptadas a un amplio rango de hábitat y poseen una considerable variación para la mayoría de las características agronómicas, calidad de semilla, resistencia a insectos, enfermedades y tolerancia a condiciones ambientales desfavorables (Seiler, 1992). El girasol silvestre *H. annuus* L. se encuentra ampliamente distribuido en Norteamérica y presenta la mayor variación morfológica y de hábitat (Arias y Rieseberg, 1995).

Schilling y Heiser (1981) señalaron que las especies silvestres de girasol *H. arizonensis*, *H. ciliaris* y *H. laciniatus* se presentan en el norte de México; *H. californicus* en Baja California Norte; *H. annuus* en el norte de México y *H. niveus* en el norte de México y Baja California.

### 2.4. Especiación del girasol

La especiación es el origen del aislamiento reproductivo y divergencia entre poblaciones, de acuerdo al concepto biológico de especie por Mayr (West-Eberhard, 2005). Los evolucionistas creen que la especiación inicia usualmente con la subdivisión de especies dentro de poblaciones aisladas por una barrera geográfica

(alopatría). La especiación simpátrica difiere de manera significativa de la especiación geográfica en que ésta es generada por interacciones ecológicas (Darwin, 1858). Se han formulado hipótesis con relación a la arquitectura genética de la adaptación y especiación (Orr y Coyne, 1992; Orr, 1998 a, b; Noor *et al.* 2001; Navarro y Barton, 2003).

*Helianthus annuus* y *Helianthus petiolaris* comparten una amplia distribución simpátrica a lo largo del centro y oeste de los Estados Unidos. Estas son fácilmente distinguibles por diversos caracteres morfológicos y cromosómicos y por relaciones de ligamiento divergente causado por un mínimo de 20 rompimientos y fusiones cromosómicas (Burke *et al.* 2004). En general, *H. annuus* está restringido a suelos arcillosos pesados y *H. petiolaris* a suelos arenosos. Sin embargo, estos dos habitat se encuentran a lo largo del centro y oeste de Norteamérica, lo que ha producido innumerables “mosaicos” de zonas híbridas. Además, *H. annuus* y *H. petiolaris* han originado al menos tres especies diploides estabilizadas que son fuertemente divergentes de sus progenitores con respecto a sus habitat (dunas areniscas, desiertos y habitat de pantanos salinos) (Rosenthal *et al.* 2002). *Helianthus paradoxus*, pudo haber tenido su origen en la hibridación entre *H. annuus* y *H. petiolaris* (Rieseberg *et al.* 1990; Rieseberg, 1991).

Rieseberg *et al.* (2003), describieron cómo la hibridación entre dos especies de girasol generó genotipos, los cuales se adaptaron a condiciones muy diferentes a la de sus progenitores. Esto dio lugar a tres especies híbridas diploides de girasol, las cuales están aisladas ecológicamente tanto uno de otro como de sus progenitores. Lo

anterior, proporciona la evidencia de que la hibridación interespecífica puede ser adaptativa.

## 2.5. Poblaciones de mapeo

Para llevar a cabo un mapeo de QTL's, usualmente se construye una población de mapeo experimental, la cual se deriva de un cruzamiento entre líneas endogámicas homocigóticas. Se utilizan comúnmente cuatro tipos de poblaciones para la construcción del mapa genético y para el mapeo de QTL's: Población  $F_2$  de intercruza, población  $F_2$  de retrocruza (BC), dobles haploides (DH), y líneas endogámicas recombinantes (RIL's).

Población  $F_2$  de intercruza: es una población de mapeo muy popular. Tales poblaciones conservan todas las combinaciones posibles de alelos parentales (Lander *et al.* 1987). Sin embargo, cada individuo  $F_2$  posee un genotipo diferente, por lo que ningún diseño experimental puede ser empleado para controlar de manera efectiva la influencia ambiental. Para resolver este problema, se puede utilizar la evaluación de progenies heterogéneas derivadas de individuos segregantes a través de la autofecundación (tal como la  $F_3$ ); sin embargo, se afecta la precisión de mapeo debido a la heterogeneidad genética (Paterson *et al.* 1990; Paterson, 1997).

Población de  $F_2$  retrocruza: se deriva del cruzamiento del híbrido  $F_1$  a uno de los dos progenitores, y posee ventajas y desventajas similares a las poblaciones  $F_2$  de intercruza. Esta población contiene menos información genética que la  $F_2$  de intercruza,

ya que los efectos aditivos y algunos tipos de efectos epistáticos no pueden ser diferenciados de los efectos de dominancia.

Población de dobles haploides (DH): es un tipo de población utilizada comúnmente en especies que pueden estar sujetas al cultivo de anteras (usualmente de plantas  $F_1$ ), seguido por duplicación cromosómica. Los individuos (DH) son completamente homocigóticos, y pueden ser autopolinizados para producir una progenie numerosa, la cual es genéticamente idéntica. Esto permite evaluaciones de fenotipos replicados. De este modo, una población (DH) puede llamarse también una población permanente. Este método presenta algunas desventajas: (1) no es posible estimar los efectos de dominancia y tipos de epistasis relacionados a estos, (2) pueden variar las proporciones en genotipos de polen que a su vez generan plantas (DH), por lo tanto, puede ser causado una distorsión en la segregación y falso ligamiento entre algunos loci marcadores.

Población de líneas endogámicas recombinantes (RIL's): son poblaciones homocigóticas "permanentes". Estas pueden ser generadas por medios tradicionales a través de autofecundación o apareamiento de estructuras familiares por muchas generaciones, iniciando desde  $F_2$  por el método de descendencia de una semilla hasta que casi todos los loci segregantes lleguen a ser homocigóticos. La principal desventaja de esta población es similar a la de la población de dobles haploides: datos perdidos en efectos de dominancia y epistasis. El desarrollo de la población (RIL) lleva un tiempo considerable y no es posible que la totalidad de los individuos sean homocigotos en todos los loci segregantes a través de generaciones limitadas de autofecundaciones ó

apareamientos de estructuras familiares, lo cual disminuye la eficiencia del mapeo de QTL's.

## 2.6. Selección del modelo genético

Un modelo genético cuantitativo relaciona el valor genotípico de un individuo con los alelos de los loci que contribuyen a la variación en una población en términos de efectos aditivos, de dominancia y epistáticos. Esta partición de efectos genéticos se relaciona a la partición de la varianza genética.

Se han propuesto numerosos modelos para describir esta relación; algunos se basan en la partición ortogonal (el efecto de un QTL es definido consecuentemente en una población de referencia en equilibrio para uno, dos, o más loci) de la varianza genética en una población en equilibrio (Zeng *et al.* 2005). La ortogonalidad asegura que los efectos aditivos, de dominancia y epistáticos puedan ser estimados de manera independiente para uno o más loci en la población de referencia donde el modelo es definido e interpretado.

Muchas publicaciones en genética cuantitativa (Falconer y Mackay, 1996) utilizan el siguiente modelo para interpretar los efectos genéticos entre genotipos  $AA$ ,  $Aa$ , y  $aa$  en un locus:

$$G_2 = \mu + a, \quad G_1 = \mu + d, \quad G_0 = \mu - a$$

En este modelo,  $a$  es el efecto aditivo definido como la mitad de la diferencia entre los dos valores genotípicos homocigotos,  $d$  es el efecto de dominancia definido como la diferencia entre el valor genotípico heterocigoto y la media del valor genotípico homocigoto, y  $\mu$  es una constante. De esta manera, los efectos genéticos son definidos solamente como una función de los valores genotípicos.

El objetivo del modelo es el de proporcionar una forma para resumir e interpretar las diferencias entre los valores genotípicos y también la variación genética observada en una población bajo estudio. La selección del modelo es el componente clave del análisis, siendo la base para la estimación de los parámetros genéticos y la interpretación de los datos. Tradicionalmente, el modelo genético aditivo se basa en el principio de mínimos cuadrados con efectos genéticos definidos como desviaciones de la media poblacional (Cockerham, 1954; Kempthorne, 1957). En tal modelo, los efectos aditivos, de dominancia y epistáticos están en función de un arreglo de valores genotípicos. Es de suma importancia tomar en cuenta información biológica y experimental, como la heredabilidad, cobertura de los marcadores y tamaño de muestra, para establecer criterios en cuanto a selección de modelos para el análisis de mapeo de QTL's.

## **2.7. Mapeo de QTL's**

El principal objetivo del mapeo de QTL's es localizar regiones cromosómicas que afectan de manera significativa la variación de caracteres cuantitativos en una población. Esta localización es importante para la identificación de genes responsables en el entendimiento de la base genética de la variación de los caracteres cuantitativos

(Zeng, 2005). Uno de los aspectos importantes en el mapeo de QTL's es considerar el tamaño de muestra. El número de individuos proporciona información para la estimación de medias fenotípicas y varianzas.

El mapeo de QTL's se lleva a cabo comúnmente utilizando poblaciones biparentales. Por ejemplo, para caracteres de plantas de baja heredabilidad, se sugiere utilizar los valores fenotípicos promedio de la progenie  $F_3$  derivada de plantas autofecundadas  $F_2$  en lugar del fenotipo  $F_2$  por sí mismo. Toda la progenie  $F_3$  derivada de la misma planta  $F_2$  pertenece a la misma familia  $F_{2:3}$ . Si el tamaño de cada familia  $F_{2:3}$  (el número de progenie  $F_3$ ) es suficientemente grande, el valor promedio de la familia representará el valor genotípico de la planta  $F_2$ , y por tanto, puede ser incrementado el poder en el mapeo de QTL's de manera significativa (Zhang y Xu, 2004). Los procedimientos de mapeo de QTL's tales como el mapeo por intervalos (IM) (Lander y Botstein, 1989) y el mapeo compuesto por intervalos (CIM) (Jansen, 1993; Zeng, 1993, 1994) involucran pruebas de hipótesis nula.

## **2.7.1 Métodos estadísticos para el mapeo de QTL's**

### **2.7.1.1. Análisis de marcador individual**

El análisis de varianza (ANOVA), prueba de t y regresión lineal simple en loci marcadores, se consideran como métodos simples para el mapeo de QTL's. El primero de estos fue el primer método para la detección de asociaciones entre marcadores genéticos y caracteres cuantitativos (Elston y Steward, 1973; Hill, 1975; Mcmillan y Robertson, 1974; Soller *et al.* 1976). Los tres enfoques mencionados no proporcionan

una estimación del parámetro de la población para la localización de un QTL, o la frecuencia de recombinación entre el marcador y el QTL (Doerge, 2002). Lo anterior se debe a que la localización y el efecto del QTL no pueden ser estimados por separado.

Los análisis de marcador individual son aún utilizados como un medio para identificar marcadores que están segregando con un carácter. La mayor parte de estas aplicaciones se enfocan en la detección de marcadores individuales, en lugar de regiones genómicas, y representan un medio eficiente y rápido para examinar poblaciones grandes para caracteres específicos, tales como resistencia a enfermedades (Timmerman *et al.* 1996; Varshney *et al.* 2000).

El método de regresión lineal simple se considera el más sencillo en cuanto a la asociación de marcadores con la variación de un carácter cuantitativo (Soller *et al.* 1976). Evidentemente, este simple enfoque presenta algunas desventajas, tales como: **(1)** el método no revela si los marcadores están asociados con uno o más QTL's, **(2)** no estima las posiciones probables de los QTL's, **(3)** los efectos de los QTL's son probablemente subestimados ya que estos son confundidos con las frecuencias de recombinación y a causa de lo anterior, **(4)** el método no es muy poderoso y se requiere de muchos individuos para realizar la prueba (Lander y Botstein, 1989). No obstante, este enfoque es aún considerado como un método muy útil para detectar las asociaciones entre marcadores de DNA y caracteres cuantitativos.

El enfoque de máxima verosimilitud fue introducido para mejorar el poder para estimar los efectos de QTL's, por lo cual, se han propuesto diversos procedimientos de

máxima verosimilitud basados en marcadores individuales (Weller *et al.* 1988; Luo y Kearsey, 1991; Darvasi y Weller, 1992). El método de máxima verosimilitud en marcadores individuales no puede determinar la posición del QTL con respecto al marcador.

La ventaja principal del análisis de varianza en un loci marcador es su simplicidad. Además, no se requiere un mapa genético para los marcadores, y el método puede ser fácilmente aplicable para loci múltiples. Otra ventaja sería la inclusión de cofactores (*e.g.*, tratamientos o efectos ambientales). El análisis de varianza para el mapeo de QTL's tiene tres desventajas principales: **(1)** no se obtienen estimaciones por separado en cuanto a localizaciones y efectos de QTL's (la localización del QTL se indica observando solamente, cuales marcadores proporcionan las mayores diferencias entre las medias de grupo del genotipo). El efecto aparente del QTL en un marcador será mas pequeño que el efecto real del QTL como resultado de la recombinación entre el marcador y el QTL; **(2)** se descartan individuos cuyos genotipos no aparecen en el marcador (datos perdidos) y, **(3)** puede haber una disminución en el poder de detección de QTL's cuando los marcadores se encuentran ampliamente espaciados.

#### **2.7.1.2. Mapeo por intervalos**

Este método fue propuesto primeramente por Lander y Botstein (1989), en el cual se examinan los intervalos entre pares adyacentes de marcadores a lo largo de un cromosoma y se determina el perfil de verosimilitud de una posición de un QTL en el genoma, calculando la razón de verosimilitud. El principio de máxima verosimilitud se

basa en: (1) que puede calcularse para un conjunto de parámetros (particularmente el efecto y posición del QTL) proporcionando los datos observados en fenotipos y genotipos marcadores, (2) las estimaciones de los parámetros corresponden a mayores verosimilitudes, y (3) el estadístico de prueba es la razón de verosimilitud (LR) (evidencia de un QTL), comparando la máxima verosimilitud con la verosimilitud de un modelo reducido, donde el modelo reducido se refiere a la hipótesis nula (no efecto del QTL). La diferencia entre máxima verosimilitud y regresión es que este último asume normalidad dentro de un grupo de marcadores. Si el perfil de verosimilitud supera un valor crítico predefinido en una región, se indica la presencia de un QTL en los límites del perfil de verosimilitud máximo con la amplitud del límite definido por el llamado intervalo de soporte (Lander y Botstein, 1989). Estos autores, generaron una prueba de razón de verosimilitud basada en la hipótesis:

$$H_0 : b^* = 0 \quad \text{y} \quad H_1 : b^* \neq 0$$

Donde  $b^*$  es el efecto de un QTL sugestivo expresado como una diferencia en efectos entre el homocigoto y heterocigoto, asumiendo que el QTL sugestivo fue localizado en el punto de consideración.

Haley y Knott (1992); y Martínez y Curnow (1992), introdujeron un procedimiento de mapeo de QTL's basado en el método de regresión de mínimos cuadrados. El principio básico es estimar las probabilidades de genotipos QTL's en los intervalos definidos por marcadores, y de estas probabilidades calcular los coeficientes

de regresión para ser utilizados en la estimación de la localización y efecto de los QTL's.

Esta metodología resultó ser una buena aproximación a los métodos basados en máxima verosimilitud y redujo en gran magnitud la demanda computacional del mapeo de QTL's, simplificando de igual manera, las modificaciones al modelo básico y puede ser ejecutado en paquetes estadísticos estándar. Uno de las principales desventajas de este enfoque es: **(1)** ligamiento entre QTL's, **(2)** este método trata con los efectos de QTL's adicionales como la variación de muestreo, la cual puede causar una estimación sesgada de los QTL's disminuyendo el poder de la prueba.

Haley *et al.* (1994), ampliaron el método para análisis de cruzamientos entre líneas exogámicas. Kao (2000) investigó de manera analítica y numérica a través de simulación, las diferencias entre el mapeo de QTL's basado en máxima verosimilitud y regresión lineal. Su estudio indicó que los métodos basados en máxima verosimilitud pueden ser más exactos, precisos y poderosos a costa de un incremento en demanda computacional.

El mapeo por intervalos (IM) tiene diversas ventajas con relación al análisis de varianza en que este: **(1)** proporciona una curva, la cuál indica la evidencia de un QTL, **(2)** toma en cuenta la deducción de QTL's a posiciones entre marcadores, **(3)** proporciona mejores estimaciones en cuanto a efectos de QTL's, y **(4)** la ejecución de este enfoque toma en cuenta datos de genotipos marcadores incompletos. Sin embargo, este enfoque presenta algunas desventajas: **(1)** este método no es una prueba por

intervalos como tal (una prueba, la cual pudiera distinguir si existe o no un QTL dentro de un intervalo definido, y que deba ser independiente de los efectos de QTL's que se encuentren fuera de una región definida). Aún cuando no exista un QTL dentro de un intervalo, el perfil de verosimilitud en un intervalo, puede exceder aún el valor crítico de manera significativa si existe un QTL en alguna región cercana en el cromosoma, (2) si existe más de un QTL en un cromosoma, el estadístico de prueba en la posición probada será afectado por todos estos QTL's, y las posiciones y efectos estimados de los QTL's identificados por este método pueden ser probablemente sesgados (Haley y Knott, 1992; Martínez y Curnow, 1992), y (3) no es eficiente utilizar únicamente dos marcadores a la vez para llevar a cabo la prueba, mientras que la información de otros marcadores no es utilizada. Estos problemas también se aplican a muchos otros métodos de mapeo de QTL's comparables (Knapp *et al.* 1990).

Xu (1995), señaló que aunque el método de Hayley y Knott (1992), no demanda mucho tiempo computacional y produce resultados similares con respecto a los obtenidos por IM, la estimación de la varianza residual es sesgada (tiende a dar valores más bajos para la prueba de razón de verosimilitud), y puede ser afectado el poder de detección de QTL's. Para cada posición en el genoma es determinada la verosimilitud para la presencia de un QTL segregante (verosimilitud bajo la hipótesis alternativa  $H_1$ ). De igual manera se calculan los efectos genéticos del QTL y la varianza residual.

El enfoque del mapeo por intervalos considera un QTL a la vez, por lo tanto, IM puede sesgar la identificación y estimación de QTL's cuando se localizan múltiples QTL's en el mismo grupo de ligamiento (Lander y Botstein, 1989; Haley y Knott, 1992;

Zeng, 1994). Existen algunos problemas con éste método, los cuales son: **(1)** el estadístico de prueba en un intervalo puede ser afectado por QTL's localizados en otras regiones del cromosoma, por lo que las posiciones estimadas y efectos de QTL's identificados por éste método pueden presentar un sesgo y, **(2)** no es eficiente utilizar solamente dos marcadores a la vez para realizar la prueba, mientras que no es utilizada la información de otros marcadores.

En el mapeo por intervalos, la distribución del fenotipo es modelada como una mezcla de dos o más componentes, los cuales corresponden a los dos o más genotipos diferentes en el QTL sugestivo (Lander y Botstein, 1989). En casos donde la distribución fenotípica se desvía de la distribución normal, puede resultar en falsos positivos, los cuales pueden ser obtenidos en regiones de baja información genotípica (ejemplo., marcadores ampliamente espaciados, bajo nivel de polimorfismo, o muchos datos perdidos de marcadores). Rebaï *et al.* (1994, 1995), destacaron que para el mapeo por intervalos, la posición del QTL es un parámetro que se presenta únicamente bajo la hipótesis alternativa.

## **2.8. LOD score**

La evidencia de un QTL en una posición cromosómica en particular puede evidenciarse como un mapa de verosimilitud, la prueba estadística de razón de verosimilitud es graficada con respecto a la posición mapa del QTL. Lander y Botstein (1989) introdujeron por primera vez el concepto de mapas de verosimilitud (utilizar el LOD score como un estadístico de prueba).

El LOD score (logaritmo de los momios, es el logaritmo base 10 de la verosimilitud de que dos loci estén ligados sobre la verosimilitud de que estén segregando de manera independiente) muestra la evidencia de la presencia de un QTL en una localización determinada. Cuando un LOD score excede el valor crítico predeterminado en un grupo de ligamiento, un QTL es detectado.

En el mapeo por intervalos, el efecto del QTL es descrito por un modelo normal mixto. La evidencia de un QTL sugestivo en una posición en el genoma es medido por la razón de verosimilitud del modelo mixto comparado a una distribución normal (LOD score) (Feenstra y Skovgaard, 2004). Este enfoque puede causar ocasionalmente LOD score aparentes o falsos en regiones de baja información genotípica (*e.g.*, marcadores ampliamente espaciados), especialmente si la distribución fenotípica se desvía de manera marcada de la distribución normal.

## **2.9. Significancia estadística (Prueba de hipótesis)**

Sin tomar en cuenta el método utilizado para estimar y localizar QTL's individuales o múltiples, una vez que son calculados los estadísticos de prueba se determina la verosimilitud del evento. La base estadística de estas comparaciones depende de los supuestos del modelo, el más común de los cuales requiere de los valores del carácter cuantitativo para ser distribuidos de manera normal. Sin embargo, esta distribución no es normal en realidad, y necesita ser considerado como una mezcla de distribuciones normales (Doerge *et al.* 1997).

Un enfoque para obtener la distribución del estadístico de prueba es el de utilizar un método de simulación para producir los datos (Lander y Botstein, 1989; Jansen, 1993; Rebai *et al.* 1994). Los estadísticos de prueba en conjunto muestran el comportamiento de la prueba y, por lo tanto, representan la distribución estadística del estadístico de prueba en particular. A partir de esta distribución, se elige el nivel de significancia estadístico o valor crítico, sobre el cual los resultados son considerados significativos de manera estadística (válidos). Los métodos de remuestreo no paramétricos han proporcionado una alternativa útil para valores críticos basados en simulación (Lander y Kruglyak, 1995).

La prueba de permutaciones (Fisher, 1935; Churchill y Doerge, 1994) y el método bootstrap (Efron y Tibshirani, 1993) han sido aplicados como un medio de aleatorización de los datos fenotípicos (carácter) con el objetivo de evaluar cualquier estadístico de prueba bajo la hipótesis nula.

En ambos métodos de remuestreo, las asignaciones de los genotipos marcadores permanecen en su estado original, por lo tanto, no ocurre ningún cambio en el mapa genético. Una implicación a esto último es que se conserva toda la información genotípica y de la población (*e.g.* distorsión de la segregación, datos perdidos y fracciones de recombinación) (Doerge, 2002). Situaciones en las cuales los efectos biológicos y estadísticos son minimizados (*e.g.* distorsión en la segregación, variación ambiental, tamaño pequeño de muestra y datos incompletos), los valores críticos basados en métodos de remuestreo o en términos teórico son generalmente los mismos (Van Ooijen, 1999; Doerge y Rebai, 1996).

### **2.9.1. Comportamiento del estadístico de prueba bajo la hipótesis nula.**

Por lo general, se puede tolerar un margen de error en la detección de los QTL's, e.g. 5% de los casos. Este problema puede ser explicado a través de un ajuste de prueba múltiple, tales como Bonferroni, Scheffe ó Tukey, el cual corrige el nivel de significancia de acuerdo a la cantidad de pruebas estadísticas independientes que sean realizadas (Doerge, 2002). Lander y Botstein (1989) analizaron el tema del valor crítico del estadístico de prueba (utilizando el LOD score) para el mapeo por intervalos. Sin embargo, es diferente el valor crítico del estadístico de prueba para el mapeo compuesto por intervalos.

Para lograr un nivel de significancia total  $\alpha$  para la prueba con  $M$  intervalos, puede utilizarse un nivel simbólico de significancia  $\alpha/M$  para la prueba en cada intervalo. Sin embargo, no es clara la distribución de la máxima razón de verosimilitud sobre un intervalo bajo la hipótesis nula. Esta distribución depende en gran medida del tamaño de muestra, el número de marcadores ajustados en el modelo y el tamaño genético de cada intervalo.

Muchos factores como el tamaño de muestra, tamaño del genoma, la densidad del mapa genético, informatividad de los marcadores y la proporción de datos perdidos, pueden afectar la distribución del estadístico de prueba bajo la hipótesis nula (Lander y Botstein, 1989; Zeng, 1994; Churchill y Doerge, 1994; Kao, 1995).

Se le denomina valor de  $P$  a la probabilidad de obtener un LOD score mayor con respecto al observado. LOD scores mayores resultan en valores pequeños de  $P$ ; los cuales a su vez, indican que la hipótesis nula es falsa (existencia de un QTL). Un LOD score de 3 indica que la verosimilitud de obtener los datos observados, es 1,000 veces mas probable que si no existiera ningún QTL dado que existe un QTL en una posición específica.

Se han realizado diversos esfuerzos en cuanto a la estimación del valor crítico, el cual corresponde a un valor de  $P$  del 5%. La distribución nula del LOD score máximo, depende de: **(1)** el tipo de cruce, **(2)** el tamaño del genoma, **(3)** el número y espaciado entre marcadores genéticos, **(4)** la cantidad y patrón en cuanto a la información de genotipos perdidos, y **(5)** la distribución real de los fenotipos.

### **2.9.2. Pruebas de permutación**

Para evitar cálculos asimétricos aproximados de manera asintótica, se puede utilizar la prueba de permutaciones (Churchill y Doerge, 1994). La idea es replicar el análisis original una gran cantidad de veces en grupos de datos generados a través de un remuestreo aleatorio de los datos originales del carácter mientras permanecen intactos los datos de marcadores.

La evaluación de la significancia estadística de amplitud genómica es un problema importante y difícil en el análisis de ligamiento multipunto. La permutación es ampliamente utilizada para determinar la significancia de amplitud genómica (Zou *et al.*

2004). Churchill y Doerge (1994), sugirieron originalmente al menos 1,000 permutaciones para estimar los valores críticos, lo cual corresponde a un valor de  $P$  ajustado de 0.05; sin embargo, señalan que se pueden requerir hasta 10,000 permutaciones para obtener estimaciones estables de un valor crítico, correspondiente a un valor de  $P$  ajustado de 0.01. La prueba de permutaciones tiene la ventaja que no hace suposiciones en la distribución del fenotipo. Sin embargo, los valores críticos dependen de los datos observados, de modo que estos deben ser analizados por el método de Monte Carlo para cada estudio.

El algoritmo para generar un valor crítico basado en permutaciones es el siguiente: **(a)** En el experimento los individuos son designados con números únicos (1,...,  $n$ ), **(b)** los datos fenotípicos son mezclados, tomando una permutación al azar de los índices 1,...,  $n$ , e igualando el  $i$ -ésimo valor del carácter fenotípico a el individuo con el índice dado por el  $i$ -ésimo elemento de los índices permutados. Este vector permutado de caracteres es igualado con la información original de los genotipos (no permutado) para todos los individuos, **(c)** se realiza un análisis del genoma para efectos de QTL's en el conjunto de datos permutados resultantes, registrando el mayor estadístico de prueba (LOD score), **(d)** los pasos **a** y **b** se repiten un total de  $N$  veces ( $N$  es frecuentemente 1000), produciendo  $N$  estadísticos de prueba máximos, uno de cada análisis genómico, **(e)** los estadísticos de prueba  $N$  máximos son ordenados en orden ascendente (menor a mayor) y, **(f)** el percentil  $100(1-\alpha)$  de los valores  $N$  ordenados, es el valor crítico estimado del experimento para controlar el error tipo I a nivel  $\alpha$ . Con  $\alpha=0.05$  y  $N=1000$ , el valor 950 de los estadísticos de prueba máximos ordenados, será el valor

crítico estimado. Se puede observar que aún cuando  $N=1000$ , es suficiente para obtener un valor crítico para  $\alpha=0.05$  se recomienda el conjunto de datos mezclados en el orden de  $N=10,000$  con  $\alpha=0.01$ , para obtener estimaciones estables (Churchill y Doerge, 1994).

### **2.9.3. Valores de $P$ ajustados**

Los valores de  $P$  ajustados proporcionan significancia de amplitud genómica de LOD score individuales, y estos se obtienen a través de una modificación del algoritmo para generar valores críticos basados en permutaciones. El LOD score actúa como un perfil de verosimilitud, revelando en cada punto de análisis el soporte relativo por la presencia de un QTL en esa posición (Lystig, 2003).

Dos de los enfoques más comunes para estimar los valores críticos son: (1) el método analítico (Lander y Botstein, 1989; Lander y Kruglyak, 1995) y, (2) el método empírico basado en pruebas de permutación (Churchill y Doerge, 1994; Doerge y Churchill, 1996; Nettleton y Doerge, 2000). Los valores críticos se calculan para controlar la tasa de falsos positivos a nivel  $\alpha$  para el análisis completo del genoma. Un valor crítico no proporciona un nivel preciso de significancia de amplitud genómica para los valores LOD score individuales. Es importante conocer la significancia de amplitud genómica o valores de  $P$  ajustados de cada uno de los puntos de análisis representados en la curva LOD score (Westfall y Young, 1993).

La interpretación de un valor de  $P$  ajustado para un LOD score en particular dentro de un análisis genómico, se refiere a la probabilidad de observar el máximo LOD score de un análisis genómico que fue al menos tan grande como el LOD score en cuestión, dado que la hipótesis nula es verdadera (Lystig, 2003).

Lo anterior difiere de un valor de  $P$  estándar no ajustado, el cual considera solamente la significancia marginal del LOD score para un punto de análisis en particular, sin explicar otros LOD score obtenidos para el análisis completo del genoma. Además, tal algoritmo modificado también proporciona un cálculo aproximado Monte Carlo a la distribución nula del LOD score máximo proveniente del análisis completo del genoma. Este algoritmo es apropiado cuando la hipótesis nula plantea que no hay presencia de algún QTL en una posición en particular del genoma, de manera que la información fenotípica y genotípica es mutuamente independiente.

Para calcular los valores de  $P$  ajustados, la única parte del algoritmo descrito por (Lystig, 2003), que debe ser modificado es el inciso **(f)**: en lugar de extraer un valor simple de los valores  $N$  ordenados para formar un valor crítico, se calcula simplemente (para cada LOD score estadístico observado “ $x$ ”, del análisis original del genoma) la proporción de los estadísticos de prueba  $N$  máximos que son mayores o iguales a “ $x$ ”. Se puede facilitar la determinación del número de estadísticos máximos mayores que un LOD score estadístico observado, ya que los estadísticos de prueba  $N$  máximos fueron ordenados secuencialmente en el inciso **(e)**.

Para obtener valores de  $P$  ajustados es necesario incorporar una modificación al algoritmo descrito por (Lystig, 2003), como sigue: para cada LOD score “ $x$ ” del análisis genómico de los datos originales no permutados, se calcula el número de estadísticos de prueba máximos que son mayores o iguales a “ $x$ ” y dividir por  $N$ .

Los valores ajustados de  $P$  calculados a través de la modificación del algoritmo, pueden utilizar la distribución total de los estadísticos de prueba máximos a partir de un conjunto  $N$  de análisis genómico, mientras un valor crítico basado en permutaciones utiliza solamente el percentil  $100(1-\alpha)$  de esos estadísticos. Es importante notar que además de un valor crítico  $\alpha=0.05$ , puede calcularse cualquier otro cuantil de interés a partir del conjunto de LOD score máximos  $N$ . Los cuantiles de la distribución nula aproximada del LOD score máximo pueden ser utilizados para determinar valores críticos múltiples, los cuales en tal caso corresponden a valores ajustados de  $P$  (Lystig, 2003). En resumen, los valores ajustados de  $P$  proporcionan niveles precisos de significancia para todos los puntos de análisis en el genoma. Estos valores representan un suplemento valioso para las curvas LOD score, así como un recurso para modelos genéticos de caracteres cuantitativos.

#### **2.9.4. Intervalos de confianza para QTL's**

Un intervalo de confianza puede ser utilizado para identificar una región cromosómica en la cual se ejecuta una búsqueda para localizar de manera exacta un QTL. Entre los diversos métodos que existen para construir un intervalo de confianza

para la localización de QTL's, uno de los más importantes es el método bootstrap descrito por (Visscher *et al.* 1996), el cual es un procedimiento de remuestreo y es utilizado para determinar un intervalo de confianza empírico para la localización del QTL, asumiendo que existe el efecto del QTL.

## 2.10. QTL's múltiples

Los enfoques estadísticos para localizar loci de caracteres cuantitativos múltiples son más poderosos que los enfoques que consideran QTL's individuales ya que estos pueden diferenciar de manera potencial entre QTL's ligados y/o interacción entre estos.

Cuando los alelos de dos o más QTL's interactúan (epistasia), se altera en gran medida el carácter cuantitativo de una manera en la cual es difícil predecir. Uno de los casos más extremos y más simples es la pérdida completa en la expresión del carácter en la presencia de una combinación en particular de alelos en QTL's múltiples.

Uno de los retos fundamentales en la búsqueda de QTL's múltiples es (1) considerar cada posición en el genoma de manera simultánea, para la localización de un QTL potencial que pudiera actuar de manera independiente, (2) estar ligado o interactuar epistáticamente con otro QTL (Doerge, 2002).

La aplicación de algoritmos genéticos (*i.e.* procedimientos de optimización numéricos basados en principios evolutivos, tales como mutación, delección y selección) (Carlborg *et al.* 2000; Foster, 2001) permite un muestreo de los modelos de QTL's a

través de números desiguales de QTL's, y puede ser utilizado en conjunción con cualquier metodología de mapeo de QTL's que es implementada para una búsqueda multidimensional genómica.

Existen diversas ventajas utilizando métodos que modelan QTL's múltiples de manera simultánea: **(1)** se puede controlar la presencia de un QTL, reducir la variación residual y obtener mayor poder para detectar QTL's adicionales, **(2)** se pueden separar QTL's ligados, y **(3)** se puede identificar interacciones entre QTL's (epistasis) (Shrimpton y Robertson, 1988; Frankel y Schork, 1996; Mackay, 1996).

### **2.10.1. Métodos de mapeo de QTL's múltiples**

#### **2.10.1.1 Regresión múltiple**

Este método comparte muchas de las desventajas del análisis de varianza en loci marcadores, y lo más importante, este requiere de datos completos de genotipos marcadores. Este método tiene la ventaja potencial de controlar la variabilidad fenotípica debido a QTL's múltiples; sin embargo, posee la desventaja de que el éxito del control depende de colocar fortuitamente los marcadores junto a un QTL verdadero (Zeng, 1993).

#### **2.10.1.2. Propiedades del análisis de regresión múltiple**

**(a)** En el análisis de regresión múltiple, asumiendo aditividad de los efectos de QTL's entre loci (i.e., ignorando epistasis), el coeficiente de regresión parcial esperado

del carácter en un marcador depende solamente de aquellos QTL's localizados en el intervalo definido por los dos marcadores flanqueantes, y no es afectado por los efectos de QTL's localizados en otros intervalos. Esta propiedad indica esencialmente que puede ser construida una prueba condicional (intervalo), basándose en el coeficiente de regresión parcial, y esta a la vez probaría el efecto del ligamiento de solamente aquellos QTL's que están localizados dentro del intervalo definido (Zeng, 1993).

(b) Dependiendo de los marcadores no ligados en el análisis de regresión múltiple, se reducirá la varianza de muestreo del estadístico de prueba, controlando alguna variación genética residual y por tanto se incrementará el poder del mapeo de QTL's. Esto significa que incluso marcadores no ligados contienen información útil, la cual puede ser utilizada para incrementar el poder estadístico de la prueba y la eficiencia del mapeo genético (Zeng, 1993).

(c) Dependiendo de los marcadores ligados en el análisis de regresión múltiple se reducirá la oportunidad de interferencia de posibles QTL's ligados múltiples en la prueba de hipótesis y en la estimación de parámetros, pero con un posible incremento en la varianza de muestreo. Un intervalo de prueba puede traer consigo una pérdida en el poder estadístico de la prueba ya que esta es una prueba condicional (Zeng, 1993).

(d) Generalmente, dos coeficientes de regresión parcial de la muestra para el valor del carácter en dos marcadores en un análisis de regresión múltiple no están correlacionados a menos que los dos marcadores generalmente sean marcadores

adyacentes. Esto se relaciona a la correlación entre dos estadísticos de prueba en dos intervalos para un intervalo de prueba.

Se ha mostrado que para un intervalo de prueba, un estadístico de prueba en un intervalo generalmente no está correlacionado de manera asintótica al estadístico de prueba en otro intervalo a menos que dos intervalos sean adyacentes entre sí (Zeng, 1993). Incluso cuando los dos intervalos sean adyacentes, la correlación entre dos estadísticos de prueba en dos intervalos es usualmente muy pequeña. Esta propiedad está relacionada con la determinación de un valor crítico apropiado de un estadístico de prueba bajo una hipótesis nula para una prueba global que cubre un genoma completo (Zeng, 1993).

### **2.11. Mapeo múltiple por intervalos**

Este enfoque proporciona un medio detallado para estimar los parámetros genéticos por medio de la diferencia entre poblaciones progenitoras y entre poblaciones segregantes. Esto proporciona una estimación de efectos aditivos, de dominancia y epistáticos de QTL's y la partición de la varianza explicada por el QTL (Zeng *et al.* 1999).

El método puede también proporcionar una estimación eficiente de los valores genotípicos para individuos basado en datos de marcadores, los cuales pueden ser utilizados para la selección asistida por marcadores. El modelo MIM se basa en el modelo de Cockerham (1954), para interpretar los parámetros genéticos y en el método

de máxima verosimilitud para estimar los parámetros genéticos. Con éste método, se puede mejorar la precisión y poder del mapeo de QTL.

De igual manera, la epistasis entre QTL's, los valores genotípicos de individuos y las heredabilidades de caracteres cuantitativos pueden ser estimados y analizados más fácilmente. La idea de este método es ajustar los efectos de QTL's sugestivos múltiples con los efectos epistáticos asociados de manera directa en un modelo para facilitar la búsqueda, prueba y estimación de posiciones, efectos e interacciones de QTL's múltiples y sus contribuciones a la varianza genética.

MIM nos permite: **(1)** deducir la ubicación de QTL's a posiciones entre marcadores, **(2)** considera datos genotípicos perdidos y, **(3)** permite interacciones entre QTL's. Existen otros enfoques, incluyendo el método Bayesiano (Hoeschele y VanRaden, 1993; Satagopan *et al.* 1996; Uimari y Hoeschele, 1997; Sillanpää y Arjas, 1999), y el uso del algoritmo genético (Carlborg *et al.* 2000). MIM consiste de cuatro componentes:

- a.** Un procedimiento de evaluación diseñado para analizar la verosimilitud de los datos dando un modelo genético (número, posiciones, y términos epistáticos de QTL's).
- b.** Una estrategia de búsqueda optimizada para seleccionar el mejor modelo genético.
- c.** Un procedimiento de estimación para todos los parámetros genéticos de los caracteres cuantitativos (número, posiciones, efectos y epistasis de QTL's; varianzas genéticas y covarianzas explicadas por los efectos de los QTL's) dado el modelo genético seleccionado.

**d.** Un procedimiento de predicción para estimar o predecir los valores genotípicos de individuos y su descendencia basado en el modelo genético seleccionado y en los valores de los parámetros genéticos estimados para la selección asistida por marcadores.

### III. MATERIALES Y MÉTODOS

#### 3.1. Antecedentes de investigación

##### 3.1.1. Formación de la población de mapeo

La construcción de la población de mapeo “Correcaminos F<sub>2:3</sub>” se conformó por 149 líneas derivadas de F<sub>2</sub> en F<sub>3</sub> provenientes de un cruzamiento entre la línea cultivada HA89 (*Helianthus annuus* L. var. *macrocarpus* (DC.) Cockerell) y una planta silvestre identificada como Ac-8-2 (*H. annuus* L. ssp. *texanus* Heiser), procedente de una población colectada en Saltillo, Coahuila. La población de mapeo se desarrolló en la Universidad Autónoma Agraria “Antonio Narro” en el periodo de 1996 al 2002. Las plantas F<sub>1</sub> obtenidas se cruzaron fraternalmente y las plantas F<sub>2</sub> fueron autofecundadas para formar las 149 líneas. El criterio de selección de estas líneas fue que contaran con una producción de 50 aquenios como mínimo (Rodríguez de la Paz, 2005).

##### 3.1.2. Genotipificación

La técnica molecular se llevó a cabo a través del sistema AFLP (polimorfismo en longitud de fragmentos amplificados) propuesto por Vos *et al.* (1995), y se empleó el protocolo “IRDye Fluorecent AFLP<sup>®</sup> Kit for Large Plant Genome Analysis LI-COR

(Biosciences)” en el Laboratorio de Genética del CINVESTAV, IPN (Centro de investigación y de Estudios Avanzados) en Irapuato, Gto, México (Castillo, 2005).

### **3.1.3. Construcción del mapa genético**

Para la construcción del mapa de ligamiento, se empleó el software JoinMap® versión 3.0 (Van y Voorrips, 2001), desarrollado para poblaciones experimentales en especies diploides, el cual maneja poblaciones con padres no homocigotos. Se realizó el registro de polimorfismos, codificando con el número 1 la presencia, y 0 la ausencia de la banda. Se realizaron pruebas de  $X^2$  cuadrada para determinar la distorsión en la segregación para cada locus con el uso del sistema software Matemática 5.0 (Wolfram, 2003). Se formaron grupos de ligamiento con un LOD de 3.0 y una frecuencia de recombinación máxima de 0.45 (Castillo, 2005). La conversión de frecuencias de recombinación a distancia mapa se realizó a través de la función de mapeo de Kosambi (1944).

### **3.1.4. Análisis de ligamiento**

De las 11 combinaciones de AFLP con iniciadores selectivos en 145 líneas, se obtuvieron 295 bandas polimórficas. De todas las bandas polimórficas encontradas, 194 procedieron del progenitor HA89 y 101 del progenitor Ac-8-2, con un promedio de 26.8 polimorfismos por combinación, con un peso molecular entre 64 y 506 bases (Castillo, 2005). Dicho promedio de bandas polimórficas es superior a los 21.5 con un peso molecular entre 200 y 1000 bases reportado por Langar *et al.* (2003) en girasol.

En cuanto a la tasa de distorsión en la segregación en las bandas polimórficas obtenidas a partir de las 11 combinaciones de iniciadores selectivos utilizados, se eliminaron los 87 polimorfismos con mayor distorsión, bajo el criterio de un valor de  $P$  para la  $X^2$  cuadrada menor o igual a 0.002, lo cual representó un 29.4% de polimorfismos eliminados. Lo anterior resultó en un total de 208 bandas polimórficas para realizar el análisis de ligamiento. De estas, 144 bandas (69.23%) segregaron 3:1 y 64 bandas (30.76%) lo hicieron 1:1. Por otra parte, se obtuvo una media de frecuencia de bandas de 0.65, por lo que se descartaron todas aquellas líneas con una frecuencia  $<0.45$ , lo cual resultó en la eliminación de 12 líneas. Por lo anterior, se consideraron solamente 133 líneas para el análisis de mapeo (Castillo, 2005).

En el **Cuadro A.1** se presentan los 208 loci utilizados para obtener el mapa de ligamiento genético, de los cuales el 19.23% (40 loci) se agruparon en 17 grupos de ligamiento (**Figura A.1**) con un LOD score mínimo de 3.0 y una frecuencia de recombinación máxima de 0.45, lo cuál resultó en un mapa de baja densidad.

El poder estadístico del experimento no fue suficiente para detectar ligamiento entre un número considerable de loci. Sin embargo, los marcadores no ligados también son útiles para detectar QTL's. El número de loci mapeados por cada grupo de ligamiento varió de 2 a 6 con tamaños de grupo de 21 a 65 cM. De esta manera se obtuvo una distancia promedio entre loci de 14.65 cM. Los loci mapeados tuvieron un origen de banda del 20 % para Ac-8-2 y 80% para HA89 (Castillo, 2005). Esto sugiere que una porción substancial del genoma del progenitor macho podría estar sub-representado en el mapa genético. En este contexto, Langar *et al.* (2003) encontraron

una sub-representación del genoma del progenitor masculino (LR4) de girasol cultivado, debido probablemente a la alta heterocigosidad del progenitor silvestre; por lo que muchas bandas procedentes de loci heterocigotos con distorsión en la segregación en esta investigación se eliminaron del análisis estadístico.

El mapa obtenido cubre 581.452 cM, lo cual representa aproximadamente el 43.37% del promedio del genoma de *Helianthus annuus* L., de acuerdo a los diferentes mapas de ligamiento genético de girasol cultivado publicados (Berry *et al.* 1995; Gentzbittel *et al.* 1995; Jan *et al.* 1998; Seung-Chul y Rieseberg, 1999; Tang *et al.* 2002; Langar *et al.* 2003; Yu *et al.* 2003).

### **3.2. Localización experimental**

El presente trabajo se desarrolló en la Universidad Autónoma Agraria “Antonio Narro” cuyas coordenadas geográficas son 25° 22’ de latitud y 101° 00’ de longitud W, con altitud de 1754 m. para la producción de plántula. Por otra parte, las labores de transplante se realizaron en el Campo Agrícola Experimental de la Universidad Autónoma Agraria Antonio Narro en Navidad, N. L., localizado a 27° 04’ N y 100° 56’ O, con altitud de 1895 m.

### **3.3. Material genético**

Se utilizó la población de mapeo “Correcaminos F<sub>2:3</sub>”, conformada por 149 líneas derivadas de F<sub>2</sub> en F<sub>3</sub>, incluyendo los dos progenitores (línea cultivada HA89 y un

material silvestre proveniente de una colecta realizada en Saltillo, Coahuila, identificado como Ac-8-2).

### **3.4. Evaluación fenotípica replicada**

Se realizó la siembra de las líneas  $F_2:F_3$  para la producción de plántulas en primavera de 2004 en vasos de polipropileno utilizando ocho semillas por cada línea, en condiciones de vivero. El transplante en campo se realizó el día 12 de mayo de 2004. La distancia entre plantas, surcos y grupos fue de 0.70, 0.80, y 1.60 m, respectivamente (**Figura 3.1**).

Se utilizó la dosis de fertilización 60-40-00, para lo cual se aplicó urea y superfosfato triple como fuente de nitrógeno y fósforo, respectivamente. La frecuencia de riegos varió dependiendo de las condiciones de humedad del suelo a lo largo del ciclo del cultivo.

Dada la limitada cantidad de semilla, se establecieron cuatro plantas por línea por parcela en cada una de las dos repeticiones, de las cuales se consideraron solamente dos plantas para la evaluación de todos los caracteres, ya que presentaron competencia completa a lo largo del ciclo del cultivo.



**Figura 3.1** Líneas F<sub>2:3</sub> (fenotipos replicados) evaluadas en el Campo Agrícola Experimental de la Universidad Autónoma Agraria Antonio Narro en Navidad, N. L.

Cada parcela estuvo bordeada por plantas de la variedad “Primavera” a través de una siembra directa a una distancia de 0.70 m de cada planta orillera por parcela. Lo anterior se realizó dos semanas después del transplante de las líneas  $F_2:F_3$ . Se estableció un lote de autofecundación para 29 líneas adicionales, con cuatro plantas por parcela y una repetición, con la finalidad de formar líneas endogámicas recombinantes de girasol (RIL's), para lo cual se realizaron tres a cuatro autofecundaciones en botones florales de cada planta correspondiente a una línea adicional, utilizando sólo un aplicador de franela por planta para evitar cualquier contaminación por polen.

### **3.5. Caracteres evaluados**

En la etapa de floración se cubrieron botones florales con bolsas de papel, previa identificación de cada línea  $F_2:F_3$ , en las dos plantas consideradas para la evaluación de todos los caracteres. Se realizaron tres a cuatro autofecundaciones en botones florales para cada planta evaluada por línea, utilizando sólo un aplicador de franela por planta para evitar cualquier contaminación por polen.

Los caracteres evaluados en condiciones de campo fueron: días a floración, días a madurez fisiológica, altura de planta, número y diámetro de capítulos. Para la evaluación de este último, se llevó a cabo un muestreo aleatorio de cinco capítulos por planta, los cuales fueron autofecundados para la producción de achenios. El diámetro de capítulos se basó en dos mediciones perpendiculares, considerando solamente el promedio de estas mismas en cada repetición para los análisis estadísticos. En las 133 líneas  $F_{2:3}$  (fenotipos replicados), así como los progenitores no se observaron diferencias en cuanto

a días a floración y días a madurez fisiológica, por lo cual, se consideró tomar un solo dato cuantitativo para cada línea, por lo que no se realizaron los análisis de varianza respectivos a estos caracteres. El peso, número y contenido de aceite de aquenios, se evaluaron en condiciones de laboratorio, tomando como referencia los capítulos muestreados para la evaluación del diámetro de capítulos.

En el **Cuadro A.2** se muestran las medias solamente para seis caracteres evaluados (altura de planta, número de capítulos, diámetro de capítulos, número de aquenios, peso de aquenios, y contenido de aceite de aquenios).

La altura de planta se evaluó cuando se presentó el 50% de apertura floral para cada línea, realizando mediciones desde la base hasta el punto de crecimiento de cada planta utilizando una cinta métrica. Los capítulos se contaron cuando las plantas presentaron madurez fisiológica. Desde la siembra hasta que el 50% de los capítulos de cada planta por línea evaluada alcanzó la etapa R5.5 (50% de apertura de flores liguladas) se determinaron los días a floración, mientras que los días a madurez fisiológica se tomaron cuando las brácteas y una gran proporción del capítulo se tornaron de color amarillo y café (etapa R9) (Schneiter y Miller, 1981). Los aquenios fueron cosechados, contabilizados y pesados de cada capítulo muestreado para evaluar el número y peso de aquenios.

Para determinar el contenido de aceite de aquenios se utilizó el método para extracción de lípidos propuesto por Bligh y Dyer (1959), el cual requiere muy poca

cantidad de muestra (0.5 g), permite una excelente estimación de porcentaje de lípidos y es rápido (Figura 3.2). El protocolo de este método es el siguiente:

- a) Pesar 0.5 g de muestra previamente molida + 40 ml de la mezcla de solventes (70 ml de cloroformo, 25 ml de metanol y 5 ml de agua destilada).
- b) Transferir la mezcla de solventes + muestra a un embudo de separación.
- c) Añadir agua destilada (1 ml) mezclar y dejar formar la bicapa.
- d) Colectar la capa de cloroformo dentro de un matraz bola fondo plano a peso constante.

#### ***Primera Extracción***

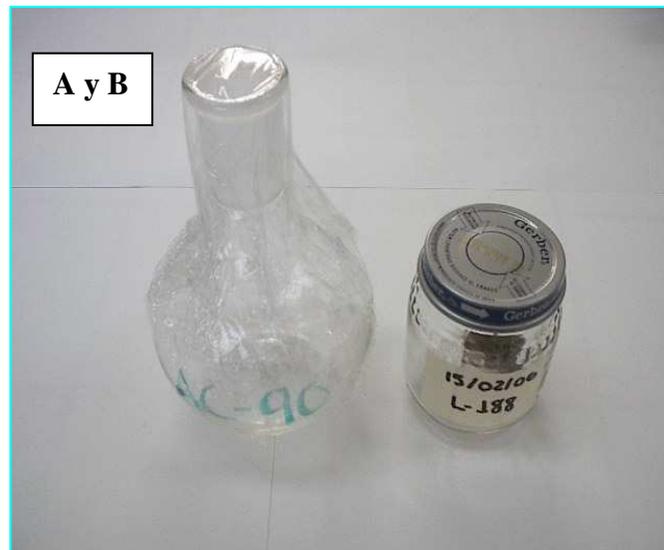
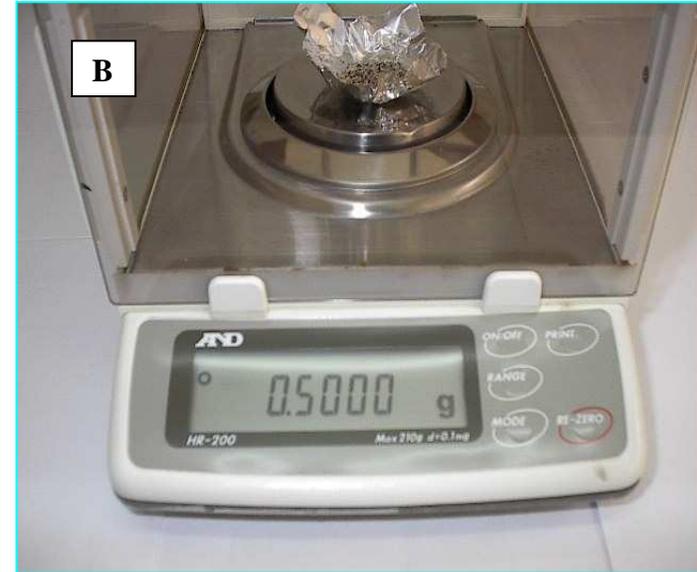
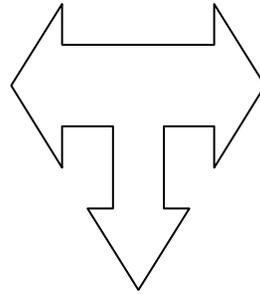
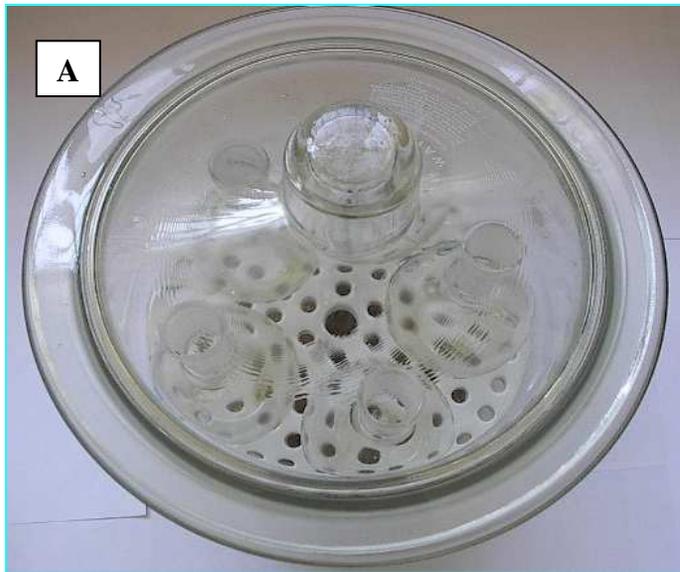
- e) Añadir 5 ml de Metanol y 40 ml de Cloroformo, mezclar y colectar la capa de cloroformo dentro de un matraz bola fondo plano a peso constante.

#### ***Segunda Extracción***

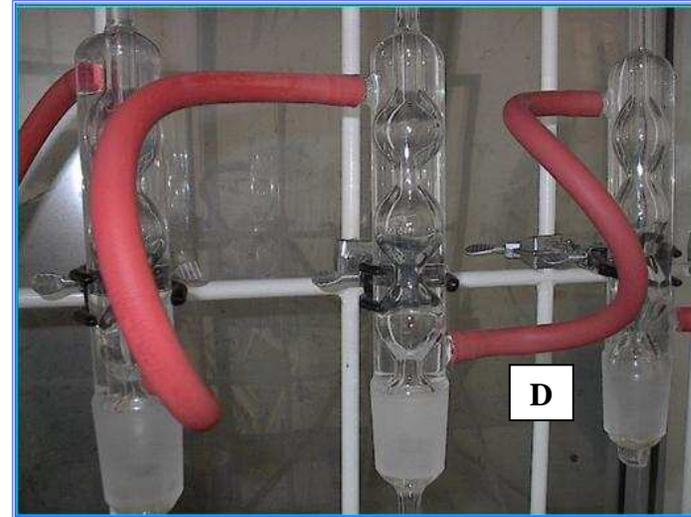
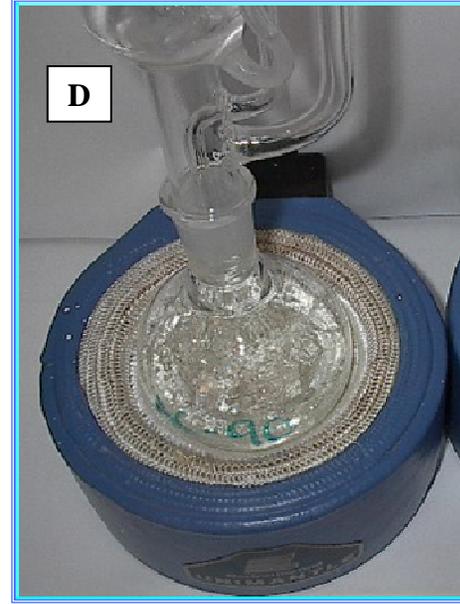
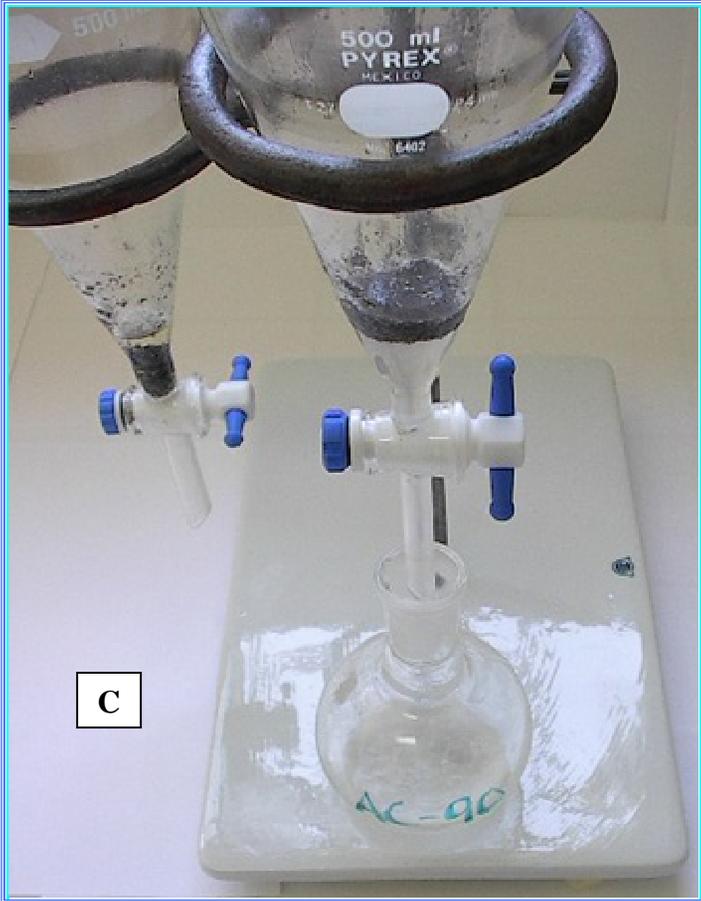
- f) Añadir 5 ml de Metanol y 20 ml de Cloroformo, mezclar y colectar la capa de cloroformo dentro de un matraz bola fondo plano a peso constante.

#### ***Tercera Extracción***

- g) Extracción de solventes (Soxhlet) y cálculo del % de lípidos.



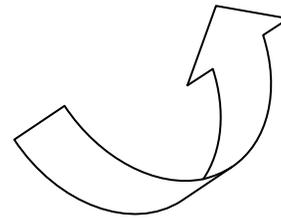
**Figura 3.2** Extracción de lípidos a través del método propuesto por Bligh y Dyer. **A)** Matraz a peso constante sin aceite, **B)** peso de muestra (g).



**Figura 3.2** Continuación..... C) Extracción de aceite en embudos de separación, D) Extracción de solvente (Soxhlet).



$$\% \text{ Aceite} = \frac{\text{Peso Cte (Matraz + Aceite)} - \text{Peso Cte (Matraz)}}{\text{Gramos de Muestra Seca}} \times 100$$



**Figura 3.2** Continuación..... **E)** Matrazes con aceite en estufa, **F)** Matrazes con aceite a peso constante.

### 3.6. Diseño experimental y análisis estadístico

El diseño experimental utilizado para la evaluación fenotípica fue bloques incompletos al azar con dos repeticiones, utilizando cuatro grupos (bloques) de 38 parcelas por repetición. Cada parcela consistió de cuatro plantas. Los factores incluyeron repeticiones, grupos y familias y se aplicó el siguiente modelo lineal:

$$Y_{ijk} = \mu + G_i + R_j + F_{ki} + (GR)_{ij} + E_{ijk}$$

Donde:

$\mu$  = media general.

$i = 1, 2, \dots, g$  (grupos)

$j = 1, 2$  (repeticiones)

$k = 1, 2, \dots, f$  (familias)

$G_i$  = efecto del  $i$ -ésimo grupo.

$R_j$  = efecto de la  $j$ -ésima repetición.

$F_{ki}$  = efecto de la  $k$ -ésima familia anidada en el  $i$ -ésimo grupo.

$(GR)_{ij}$  = efecto de la interacción del  $i$ -ésimo grupo por la  $j$ -ésima repetición.

$E_{ijk}$  = efecto del error experimental.

$Y_{ijk}$  = observación de la  $k$ -ésima familia anidada en el  $i$ -ésimo grupo, de la  $j$ -ésima repetición.

Se realizaron análisis de varianza en el software de análisis de datos R Versión 2.6.0 (R Development Core Team, 2006), para altura de planta, número y diámetro de capítulos, número y peso de aquenios y contenido de aceite de aquenios para determinar diferencias estadísticas entre las 133 líneas consideradas para el análisis de QTL's, incluyendo asimismo los dos progenitores utilizados en los cruzamientos inter-subespecíficos.

### **3.7. Análisis genético**

Se realizó un ANOVA por locus individual no ligado (análisis de marcadores individuales) en el software para análisis de datos R Versión 2.6.0 (R Development Core Team, 2006), para determinar alguna posible asociación carácter-marcador para cada carácter evaluado.

En cuanto a los análisis de QTL's se tomó como base el mapa genético basado en marcadores AFLP. Se consideró un nivel de significancia estadística de amplitud genómica de 0.05 en la detección de QTL's. Se utilizó el enfoque analítico de mapeo por intervalos (Lander y Botstein, 1989) para la identificación de QTL's y estimación de sus efectos fenotípicos, con el uso del paquete MapQTL® Versión 4.0 (Van Ooijen *et al.* 2002).

Se realizaron pruebas de permutación (Churchill y Doerge, 1994; Doerge y Churchill, 1996) para establecer los valores críticos de los estadísticos de prueba, solamente para aquellos grupos de ligamiento con un LOD score mayor que 1.5.

## IV. RESULTADOS Y DISCUSIÓN

### Líneas endogámicas recombinantes (RIL's)

De las plantas correspondientes a las 29 líneas adicionales (líneas derivadas de F<sub>3</sub> en F<sub>4</sub>) para obtener las líneas endogámicas recombinantes, se obtuvo un promedio de 93 aquenios por línea, considerando solamente el 75.86% de las líneas con respecto al total. El resto de las líneas adicionales (24.14%), presentó un 100% de mortalidad de plantas.

### Análisis estadístico

#### Evaluación fenotípica replicada

Se observaron diferencias significativas en cuanto a repeticiones solamente para el carácter contenido de aceite de aquenios con un valor de  $P$  igual a  $2.2 \times 10^{-16}$  (**Cuadro 4.1**). Se observaron diferencias significativas entre grupos para altura de planta, número de capítulos y contenido de aceite de aquenios con un valor de  $P$  igual a 0.0004464,  $2.237 \times 10^{-06}$  y  $5.896 \times 10^{-07}$ , respectivamente. En cuanto a diámetro de capítulos, número y peso de aquenios, los resultados obtenidos del análisis no indican significancia entre grupos (**Cuadro 4.1**). En cuanto a la interacción de repeticiones por grupos se observó significancia altura de planta y número de aquenios con un valor de  $P$

igual a 0.0443265 y 0.08072. Sin embargo, para los caracteres número y diámetro de capítulos, así como peso de aquenios y contenido de aceite de aquenios no hubo significancia en la interacción de repeticiones por grupos (**Cuadro 4.1**). Se detectaron diferencias significativas entre familias dentro de grupos para los caracteres altura de planta, número de capítulos, mostrando un valor de  $P$  igual a  $2.2 \times 10^{-16}$  para los dos caracteres. Asimismo, para el diámetro de capítulos, número y peso de aquenios, y contenido de aceite de aquenios se detectaron diferencias significativas con un valor de  $P$  igual a 0.0001479, 0.02641,  $2.0 \times 10^{-16}$ , y  $2.72 \times 10^{-15}$ , respectivamente entre familias dentro de grupos (**Cuadro 4.1**).

De acuerdo a las diferencias significativas encontradas entre familias dentro de grupos para seis caracteres evaluados, se infiere la existencia de variación genética en la población para los caracteres evaluados. Por último, el contenido de aceite de aquenios en 83 líneas de girasol fluctuó de 20.16 a 43.10%. En este contexto, Burke *et al.* (2005), reportaron un contenido de aceite de 40% en girasol cultivado y 26 % en girasol silvestre.

### **Análisis genético**

Se detectó un QTL con significancia de amplitud genómica de 0.017 para peso de aquenios denominado (PA-G11) (**Figura 4.1**). Este QTL asociado al locus marcador M208P1E-ACC/M-CAC (posición del marcador 0 cM en grupo de ligamiento), presentó un efecto aditivo estimado de 0.206332 y explicó el 10.2% de la varianza fenotípica (**Figuras 4.1B y 4.1C**).

**Cuadro 4.1** Cuadrados medios para seis caracteres relacionados a la domesticación del girasol en una población segregante F<sub>2</sub>:F<sub>3</sub>.

<i>F.V.</i>	<i>Caracteres evaluados</i> <sup>†</sup>					
	<i>AP</i>	<i>NC</i>	<i>DC</i>	<i>NA</i>	<i>PA</i>	<i>CA</i>
Repeticiones	0.0519 NS	178 NS	0.160 NS	87 NS	0.129 NS	0.08397 ***
Grupos	0.1214 ***	3695 ***	1.660 NS	312 NS	0.311 NS	0.00928 ***
Rep x Grupos	0.0525 **	589 NS	1.720 NS	1627 *	0.267 NS	0.00132 NS
Familias/grupos	0.0895 ***	2860 ***	2.470 ***	1003 **	1.214 ***	0.00465 ***
Error Residual	0.0190	341	1.138	708	0.166	0.00072

<sup>†</sup> AP = Altura de planta (m); NC = número de capítulos; DC = diámetro de capítulos (cm); NA = número de aquenios; PA = peso de aquenios (g), CA = contenido de aceite de aquenios (%).

\* Significativo al 0.05 de probabilidad.

\*\* Significativo al 0.01 de probabilidad.

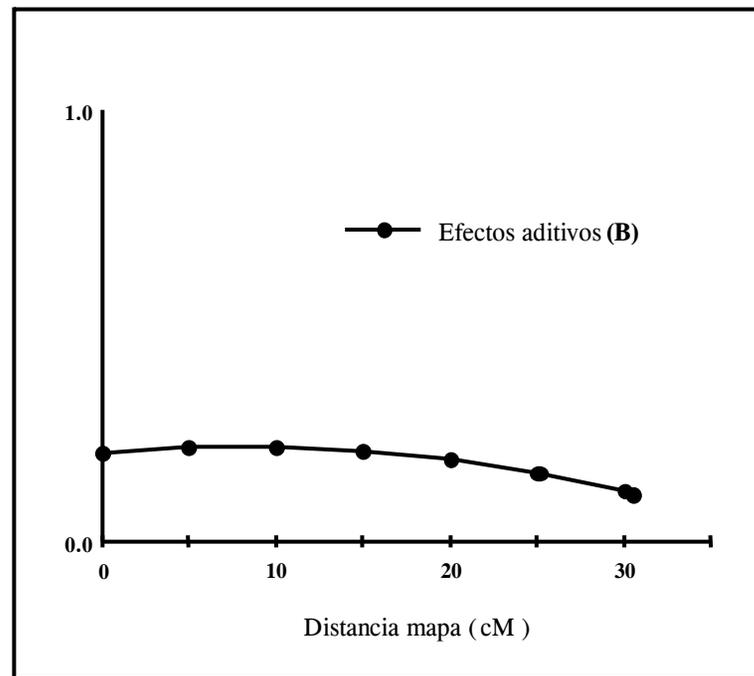
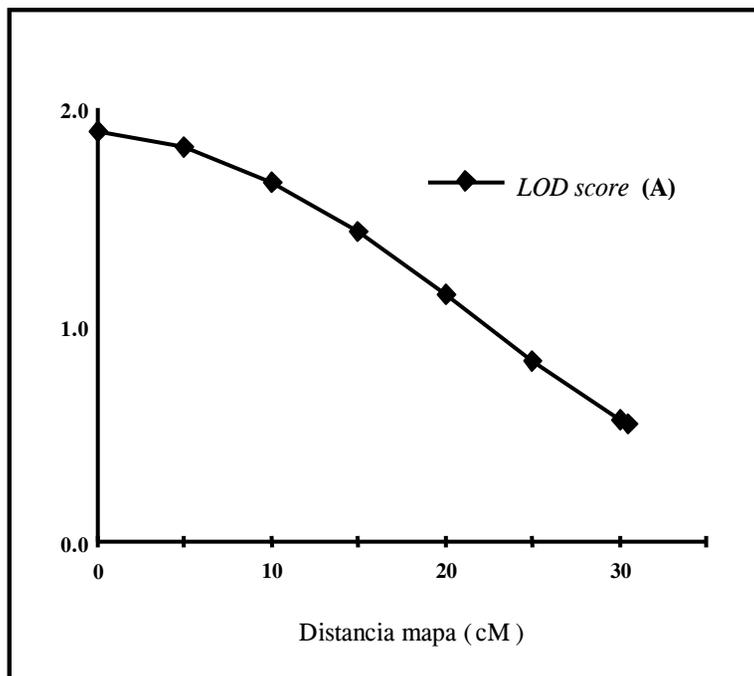
\*\*\* Significativo al <0.001 de probabilidad.

NS = No significativo.

Asimismo, se detectó un QTL sugestivo para peso de aquenios (PA-G3) ligado al marcador M263P1E-ACC/M-CAT (posición del marcador 0 cM en grupo de ligamiento), el cual explicó un porcentaje mayor de la varianza fenotípica con respecto a PA-G11 (**Cuadro 4.2; Figura 4.1C**).

Estos dos QTL's explicaron de manera simultánea el 22.1% de la variación fenotípica observada para este carácter. Se detectaron otros cuatro QTL's sugestivos distribuidos en cuatro de los 17 grupos de ligamiento, afectando los caracteres altura de planta (AP-G2), número de capítulos (NC-G4), diámetro de capítulos (DC-G12) y número de aquenios (NA-G17), los cuales explicaron de 7.1 al 10.3% de la varianza fenotípica observada (**Cuadro 4.2**).

Dos QTL's sugestivos para los caracteres número y diámetro de capítulos (NC-G4 y DC-G12, respectivamente), presentaron efectos aditivos negativos con respecto al progenitor cultivado. El efecto negativo observado en el carácter diámetro de capítulos pudo deberse a un fenómeno biológico o a error de muestreo. Los valores bajos en cuanto a LOD score ( $\leq 1.14$ ) detectados para el contenido de aceite se debieron probablemente al tamaño de muestra, ya que de los 133 individuos considerados en el mapa genético, solamente el 62.4% de estos contaron con datos cuantitativos para este carácter. Se determinó al efecto aditivo de cada QTL identificado, considerando el efecto como positivo cuando la media estimada del carácter cuantitativo asociado a los alelos (HA89) fue mayor con respecto a la media de los alelos silvestres (AP-G2, NA-G17, PA-G3) y (PG-G11) (**Cuadro 4.2; Figuras 4.1B y 4.1D**).



**Figura 4.1** Perfiles de un QTL con significancia de amplitud genómica para peso de achenios en el grupo de ligamiento 11, denominado PA-G11. **(A)** *Lod score*, **(B)** efectos aditivos, **(C)** varianza fenotípica explicada por el QTL (%), y **(D)** media estimada de la distribución del carácter peso de achenios (g).

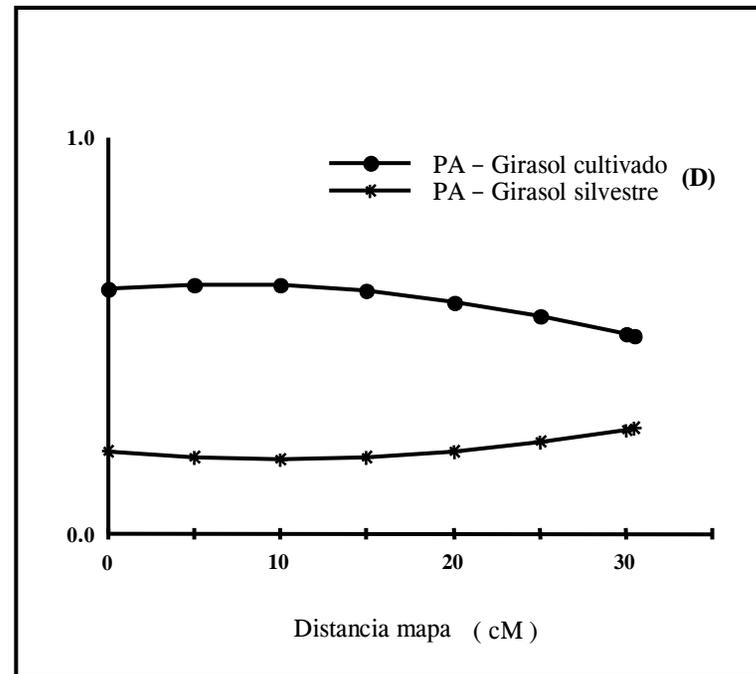
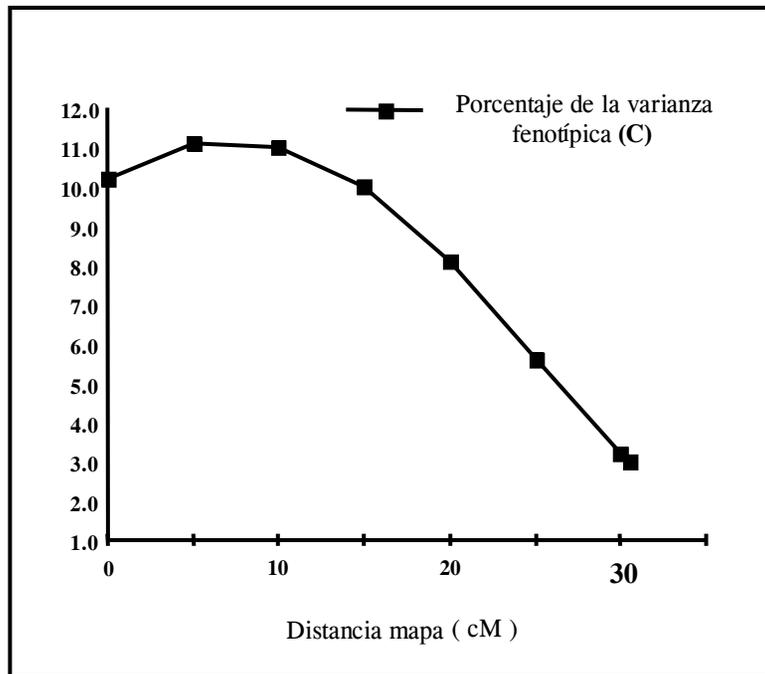


Figura 4.1 Continuación.....

**Cuadro 4.2** Resultados del análisis de QTL's a través de mapeo por intervalos para caracteres relacionados a la domesticación del girasol en una población segregante F<sub>2</sub>:F<sub>3</sub>.

<b>Mapeo por Intervalos</b>								
<i>Carácter</i> <sup>†</sup>	<i>QTL</i> <sup>††</sup>	<i>Posición</i> ( <i>cM</i> ) <sup>†††</sup>	<i>Locus Marcador</i> <i>Flanqueante</i>	<i>Lod score</i>	<i>PVF</i> <sup>¶</sup>	<i>AG</i> <sup>φ</sup>	<i>AC</i> <sup>φφ</sup>	<i>Efectos</i> <i>Aditivos</i>
AP	AP-G2	5.0	M243P1E-AGC/M-CAT	1.94	10.3	0.204	0.012	0.0953084
NC	NC-G4	0.0	M434P1E-ACG/M-CAA	2.24	9.0	0.085	0.005	-8.21146
DC	DC-G12	0.0	M323P1E-AGG/M-CAC	1.62	7.1	0.170	0.010	-0.452724
NA	NA-G17	22.5	M307P1E-AGC/M-CAT	1.53	8.5	0.187	0.011	8.92463
PA	PA-G3	0.0	M263P1E-ACC/M-CAT	1.92	11.9	0.085	0.005	0.238776
PA	PA-G11	0.0	M208P1E-ACC/M-CAC	1.89	10.2	0.017	0.001	0.206332

<sup>†</sup> AP = Altura de planta (m); NC = número de capítulos; DC = diámetro de capítulos (cm); NA = número de aquenios; PA = peso de aquenios (g).

<sup>††</sup> La descripción de cada QTL se basa de acuerdo a la abreviación de cada carácter y al grupo de ligamiento en el mapa de ligamiento genético.

<sup>†††</sup> Posición del QTL con respecto al primer locus marcador flanqueante.

<sup>¶</sup> Porcentaje de la varianza fenotípica explicada por el QTL.

<sup>φ</sup> Significancia estadística de amplitud genómica.

<sup>φφ</sup> Significancia estadística de amplitud cromosómica.

Por el contrario, se consideró como efecto negativo (QTL's con efectos aditivos en dirección opuesta) cuando la media estimada del carácter cuantitativo asociado al genotipo silvestre fue mayor con respecto al genotipo cultivado (NC-G4, DC-G12), por lo que la direccionalidad de los efectos alélicos dependen en gran parte del fondo genético.

El análisis de marcadores individuales (asociación locus marcador-QTL) identificó 96 loci significativos de los 168 loci polimórficos no ligados al mapa genético, asociados a los ocho caracteres evaluados de los cuales uno se encuentra potencialmente asociado al carácter número de aquenios (M120P1E-AGC/M-CAA) y tres al carácter días a madurez fisiológica (M290P1E-ACC/M-CTT, M486P2E-ACC/M-CTT, M177P1E-ACC/M-CTT) a una significancia igual a  $P < 0.001$  (**Cuadro 4.3**). Lo anterior, representa el 46.15% de los 208 bandas polimórficas consideradas para el análisis de ligamiento.

Por otra parte, se registraron asociaciones QTL-marcador de manera significativa al 0.001 de probabilidad a través del análisis de marcadores individuales en 28 loci marcadores para siete caracteres evaluados, de los cuales tres: M217P2E-ACC/M-CTT, M281P1E-ACA/M-CAG y M372P2E-ACC/M-CTT, se asociaron a dos caracteres diferentes (DMF-CA, NA-PA y AP-DMF, respectivamente) (**Cuadro 4.3**). Por otra parte, tomando como base las 28 bandas polimórficas, 21 de estas provinieron del progenitor HA89 y siete del progenitor silvestre (Ac-8-2) representando el 75.0% y 25.0%, respectivamente (**Cuadro 4.3**).

**Cuadro 4.3** Loci marcadores no ligados con significancia al 0.001 de probabilidad, para caracteres relacionados a la domesticación del girasol en una población segregante  $F_2:F_3$  a través del método de marcador individual (ANOVA).

<b>Marcador Individual (ANOVA)</b>			
<i>Carácter</i> <sup>†</sup>	<i>Locus marcador</i>	<i>Carácter</i> <sup>†</sup>	<i>Locus marcador</i>
AP	M84P1E-ACG/M-CAA	DMF	M469P1E-ACC/M-CTT
AP	M372P2E-ACC/M-CTT	DMF	M342P1E-ACC/M-CTT
DC	M312P2E-ACA/M-CAG	DMF	M243P1E-ACC/M-CTT
DC	M160P1E-AAG/M-CAA	DMF	M194P1E-ACC/M-CTT
DC	M253P1E-ACG/M-CAA	DMF	M189P1E-ACC/M-CTT
DC	M221P1E-AGC/M-CTT	DMF	M155P1E-ACC/M-CTT
NA	M281P1E-ACA/M-CAG	DMF	M151P1E-ACC/M-CTT
PA	M281P1E-AGC/M-CAT	DMF	M134P1E-ACC/M-CTT
PA	M281P1E-ACA/M-CAG	DMF	M89P1E-ACC/M-CTT
DF	M180P1E-ACG/M-CAA	DMF	M372P2E-ACC/M-CTT
DMF	M416P1E-ACC/M-CTT	DMF	M217P2E-ACC/M-CTT
DMF	M166P1E-ACC/M-CTT	CA	M209P1E-AGC/M-CAA
DMF	M324P2E-ACC/M-CTT	CA	M217P2E-ACC/M-CTT
DMF	M305P2E-ACC/M-CTT	CA	M155P2E-AGC/M-CTT
DMF	M364P1E-ACG/M-CAA	CA	M92P2E-ACC/M-CAC
DMF	M162P1E-AGG/M-CAC		

<sup>†</sup> AP = Altura de planta (m); DC = diámetro de capítulos (cm); NA = número de achenios; PA = peso de achenios (g); DF = días a floración; DMF = días a madurez fisiológica; CA = contenido de aceite de achenios (%).

Se detectaron dos loci marcadores (bandas polimórficas) con igual número de pares de bases y una significancia al 0.001 de probabilidad a través del análisis de marcadores individuales para peso de aquenios con diferente combinación de bases restrictivas del oligonucleótido EcoRI y MseI (**Cuadro 4.3**).

## V. CONCLUSIONES

Se obtuvo un promedio de 93 aquenios por línea adicional (líneas derivadas de F<sub>3</sub> en F<sub>4</sub>) en el 75.86% de estas con respecto al total en la formación de líneas endogámicas recombinantes (RIL's).

La evaluación fenotípica demuestra gran variabilidad genética entre la población.

Se encontraron 96 loci significativos asociados a ocho caracteres específicamente al carácter número de aquenios y días a madurez fisiológica a una significancia igual a  $P < 0.001$ .

El mapeo por intervalos identificó un QTL con significancia de amplitud genómica de 0.017 para peso de aquenios en el grupo de ligamiento 11 y cinco QTL's a nivel sugestivo.

El contenido de aceite de aquenios en 83 líneas de girasol fluctuó de 20.16 a 43.10% a través del método Blight y Dyer.

## VI. RESUMEN

La identificación de loci de caracteres cuantitativos (QTL's) es de suma importancia para proporcionar un medio de respuesta con respecto al control genético de los caracteres cuantitativos.

El objetivo fundamental del presente trabajo fue detectar posiciones de QTL's y estimar sus efectos genéticos, especialmente aquellos relacionados con la domesticación del girasol. Se trabajó con ocho caracteres: altura de planta, número y diámetro de capítulos, número y peso de aquenios, días a floración, días a madurez fisiológica, y contenido de aceite de aquenios. Para este último se utilizó el método para extracción de lípidos propuesto por Bligh y Dyer (1959).

El presente trabajo se desarrolló en la Universidad Autónoma Agraria "Antonio Narro" cuyas coordenadas geográficas son 25° 22' de latitud y 101° 00' de longitud W, con altitud de 1754 m. para la producción de plántula. Por otra parte, las labores de transplante (evaluación fenotípica replicada) se realizaron en el Campo Agrícola Experimental de la Universidad Autónoma Agraria Antonio Narro en Navidad, N. L., localizado a 27° 04' N y 100° 56' O, con altitud de 1895 m.

Se utilizó la población de mapeo “Correcaminos F<sub>2:3</sub>”, conformada por 149 líneas derivadas de F<sub>2</sub> en F<sub>3</sub>, incluyendo los dos progenitores (línea cultivada HA89 y un material silvestre proveniente de una colecta realizada en Saltillo, Coahuila, identificado como Ac-8-2). El diseño experimental utilizado para la evaluación fenotípica fue bloques incompletos al azar con dos repeticiones, utilizando cuatro grupos (bloques) de 38 parcelas por repetición. Cada parcela consistió de cuatro plantas. Los factores incluyeron repeticiones, grupos y familias.

Se estableció un lote de autofecundación para 29 líneas adicionales, con cuatro plantas por parcela y una repetición, con la finalidad de formar líneas endogámicas recombinantes de girasol (RIL's), de las cuales se obtuvo un promedio de 93 aquenios por línea.

Para llevar a cabo los análisis de QTL's se utilizó el software MapQTL 4.0, tomando como base un mapa de ligamiento genético de baja densidad basado en marcadores AFLP. Se consideró un nivel de significancia estadística de amplitud genómica de 0.05 en la detección de QTL's. Para establecer los valores críticos de los estadísticos de prueba se realizaron pruebas de permutación, solamente para aquellos grupos de ligamiento con un LOD score >1.5. Para determinar alguna posible asociación carácter-marcador se realizó un ANOVA por locus individual no ligado para cada carácter evaluado en el software de análisis de datos R (versión 2.6.0).

El análisis de marcadores individuales (asociación locus marcador-QTL) identificó 96 loci no ligados (representando el 46.15% de los 208 bandas polimórficas

consideradas para el análisis de ligamiento) distribuidos en los ocho caracteres evaluados, de los cuales uno se encuentra potencialmente asociado ( $P < 0.001$ ) al carácter número de achenios (M120P1E-AGC/M-CAA) y tres al carácter días a madurez fisiológica (M290P1E-ACC/M-CTT, M486P2E-ACC/M-CTT, M177P1E-ACC/M-CTT).

Se registraron asociaciones QTL-marcador de manera significativa al 0.001 de probabilidad a través del análisis de marcadores individuales en 28 loci marcadores para siete caracteres evaluados, de los cuales tres se asociaron a dos caracteres diferentes. Tomando como base las 28 bandas polimórficas, 21 de estas provinieron del progenitor HA89 y siete del progenitor silvestre (Ac-8-2) representando el 75.0% y 25.0%, respectivamente.

Se detectó un QTL con significancia de amplitud genómica de 0.017 para peso de achenios. Se detectaron cinco QTL's sugestivos distribuidos en cuatro de los 17 grupos de ligamiento, afectando los caracteres altura de planta, número de capítulos, diámetro de capítulos, número de achenios y peso de achenios. Los QTL's detectados (incluyendo sugestivos) explicaron de 7.1 al 11.9% de la varianza fenotípica para cinco caracteres evaluados.

El contenido de aceite de achenios en 83 líneas de girasol fluctuó de 20.16 a 43.10%. Este estudio contribuirá al conocimiento de la genética de la domesticación del girasol, así como en la generación de marcadores para selección con fines de

mejoramiento genético, dirigido a la introgresión de caracteres silvestres a variedades de girasol elite.

## VII. LITERATURA CITADA

- Arias, M. D., and L. H. Rieseberg. 1995. Genetic relationships among domesticated and wild sunflowers (*Helianthus annuus*, Asteraceae). *Economic Botanic* 49:239-248.
- Berry, S. T., A. J. Leon, C. C. Hanfrey, P. Challis, A. Burkholz, S. R. Barnes, G. K. Rufener, M. Lee, and P. D. S. Caligari. 1995. Molecular marker analysis of *Helianthus annuus* L.2. Construction of a RFLP linkage map for cultivated sunflower. *Theoretical and Applied Genetics* 91:195-199.
- Bligh, E. G., and W. J. Dyer. 1959. A rapid method of total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* 37(8):911-917.
- Burke, J. M., S. Tang, S. J. Knapp, and L. H. Rieseberg. 2002. Genetic analysis of sunflower domestication. *Genetics* 161:1257-1267.
- Burke, J. M., Z. Lai, M. Salmaso, T. Nakazato, S. X. Tang, A. Heesacker, S. J. Knapp, and L. H. Rieseberg. 2004. Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* 167:449-457.
- Burke, J. M., S. J. Knapp, and L. H. Rieseberg. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics* 171:1933-1940.
- Carlborg, O., L. Andersson, and B. Kinghorn. 2000. The use of a genetic algorithm for simultaneous mapping of multiple interacting quantitative trait loci. *Genetics* 155:2003-2010.

- Castillo, R. F. 2005. Construcción de un mapa genético de girasol (*Helianthus annuus* L.) basado en marcadores AFLP's. Tesis Maestría en Ciencias. UAAAN. Buenavista, Saltillo, Coahuila, México. 46 p.
- Churchill, G. A., and R. W. Doerge. 1994. Empirical threshold values for quantitative trait mapping. *Genetics* 138:963-971.
- Cockerham, C. C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* 39:859-882.
- Darvasi, A., and J. I. Weller. 1992. On the use of the moments method of estimation to obtain approximate maximum likelihood estimation of linkage between a genetic marker and a quantitative locus. *Heredity* 68:43-46.
- Darwin, C. 1858. On the Origin of Species by Means of Natural Selection (Murray, London).
- Doerge, R. W., and G. A. Churchill. 1996. Permutation tests for multiple loci affecting a quantitative character. *Genetics* 142:285-294.
- Doerge, R. W., and A. Rebai. 1996. A significance thresholds for QTL mapping tests. *Heredity* 76:459-464.
- Doerge, R. W. 2002. Mapping and analysis of quantitative trait loci in experimental populations. *Nat. Rev.* 3:43-52.
- Doerge, R. W., Z-B Zeng, and B. S. Weir. 1997. Statistical issues in the search for genes affecting quantitative traits in experimental populations. *Stat. Sci.* 12(3):195-219.
- Efron, B., and R. J. Tibshirani. 1993. An Introduction to the Bootstrap (Chapman and Hall, New York).
- Elston, R. C., and J. Steward. 1973. The analysis of quantitative traits for simple genetic models from parental,  $F_1$  and backcross data. *Genetics* 73:695-711.

- Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*, Ed. 4. Longman, Harlow, UK.
- Feenstra, B., and Ib M. Skovgaard. 2004. A quantitative trait locus mixture model that avoids spurious LOD score peaks. *Genetics* 167:959-965.
- Fisher, R. 1935. *The Design of Experiments* 3rd edn (Oliver and Boyd, London).
- Foster, J. A. 2001. Evolutionary computation. *Nature Rev. Genet.* 2:428-436.
- Frankel, W. N., and N. J. Schork. 1996. Who's afraid of epistasis? *Nat. Genet.* 14:371-373.
- Gentzbittel, L., F. Fear, Y-X Zhang, A. Bervillé, and P. Nicolas. 1995. Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.) *Theor. Appl. Genet.* 90:1079-1086.
- Haley, C. S., and S. A. Knott. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity.* 69:315-324.
- Haley, C. S., S. A. Knott, and J. M. Elsen. 1994. Mapping quantitative trait loci in crosses between outbred lines using least squares. *Genetics* 136:1195-1207.
- Haro-Ramírez, P. A., M. C. J. García, y M. H. Reyes-Valdés. 2007. Determinación maternal del contenido de aceite en semillas de girasol. *Rev. Fitotec. Mex.* 30(1):39-42.
- Harter, A. V., K. A. Kardnert, D. Falush, D. L. Lentz, R. A. Bye, and L. H. Rieseberg. 2004. Origin of extant domesticated sunflowers in eastern North America. *Nature* 430:201-204.
- Heiser, C. B. 1978. *Taxonomy of Helianthus and origin of the domesticated sunflower.* Sunflowers Science and Technology ASA, CSSA and ASSA, Madison WI. pp: 31-53.

- Hill, A. P. 1975. Quantitative linkage: A statistical procedure for its detection and estimation. *Ann. Hum. Genet. Lond.* 38:439-559.
- Hoeschele, I., and P. M. VanRaden. 1993. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. *Theor. Appl. Genet.* 85: 953-960.
- Jan, C. C., B. A. Vick, J. F. Miller, A. L. Kahler, and E. T. Butler. 1998. Construction of an RFLP linkage map for cultivated sunflower. *Theor. Appl. Genet.* 96:15-22.
- Jansen, R. C. 1993. Interval mapping of multiple quantitative trait loci. *Genetics* 135: 205-211.
- Kao, C-H. 1995. Statistical methods for locating positions and analyzing epistasis of multiple quantitative trait genes using molecular marker information. Dissertation of Department of Statistics at North Carolina State University.
- Kao, C-H. 2000. On the differences between maximum likelihood and regression interval mapping in the analysis of quantitative trait loci. *Genetics* 156:855-865.
- Kempthorne, O. 1957. An Introduction to Genetic Statistics. The Iowa State University Press, Ames, Iowa.
- Kosambi, D. D. 1944. The estimation of map distances from recombination values. *Annals of Eugenics* 12:172-175.
- Knapp, S. J., J. W. C. Bridges, and D. Brikes. 1990. Mapping quantitative trait loci using molecular marker linkage maps. *Theor. Appl. Genet.* 79:583-592.
- Lander, E. S., P. Green, J. Abrahamson, A. Barlow, M. J. Daly, S. E. Lincoln, and L. Newburg. 1987. MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* 1:174-181.
- Lander, E. S., and D. Botstein. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* 121:185-199; erratum 136, 705 (1994).

- Lander, E., and L. Kruglyak. 1995. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat. Genet.* 11:241-247.
- Langar, K., M. Lorieux, E. Desmarais, Y. Griveau, Gentzbittel and Bervillé. 2003. Combined mapping of DALP and AFLP markers in cultivated sunflower using F<sub>9</sub> recombinant inbred lines. *Theor. Appl. Genet.* 106:1068-1074.
- Lawson, W. R., K. C. Goulter, R. J. Henry, G. A. Kong, and J. K. Kochman. 1998. Marker-assisted selection for two rust resistance genes in sunflower. *Mol. Breed.* 4:227-234.
- León, A. J., F. H. Andrade, and M. Lee. 2003. Genetic analysis of seed-oil concentrations across generations and environments in sunflower (*Helianthus annuus* L.). *Crop Science* 43:135-140.
- Lentz, D. L., M. E. D. Pohl, K. O. Pope, and A. R. Wyatt. 2001. Prehistoric sunflower (*Helianthus annuus* L.) domestication in Mexico. *Econ. Bot.* 55:370-376.
- Luo, Z. W., and M. J. Kearsey. 1991. Maximum likelihood estimation of linkage between a marker gene and a quantitative locus. II. Application to backcross and doubled haploid populations. *Heredity* 66:117-124.
- Lystig, T. C. 2003. Adjusted *P* values for genome-wide scans. *Genetics* 164:1683-1687.
- Mackay, T. F. C. 1996. The nature of quantitative genetic variation revisited: Lessons from *Drosophila* bristles. *BioEssays* 18:113-121.
- Martínez, O., and R. N. Curnow. 1992. Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theor. Appl. Genet.* 85:480-488.
- Mcmillan, I., and A. Robertson. 1974. The power of methods for the detection of major genes affecting quantitative traits. *Heredity* 32:349-356.
- Mokrani, L., L. Gentzbittel, F. Azanza, L. Fitamant, G. Al-Chaarani, and A. Sarrafi. 2002. Mapping and analysis of quantitative trait loci for grain oil content and

- agronomic traits using AFLP and SSR in sunflower (*Helianthus annuus* L.). *Theor Appl Genet* 106:149-156.
- Navarro, A., and N. H. Barton. 2003. Accumulating postzygotic isolation genes in parapatry: a new twist on chromosomal speciation. *Evolution* 57:447-459.
- Nettleton, D., and R. W. Doerge. 2000. Accounting for variability in the use of permutation testing to detect quantitative trait loci. *Biometrics*. 56:52-58.
- Noor, M. A. F., K. L. Grams, L. A. Bertucci, and J. Reiland. 2001. Chromosomal inversions and the persistence of species. *Proc. Natl. Acad. Sci. USA* 98:12084-12088.
- Orr, H. A. 1998a. The population genetics of adaptation: the distribution of factors fixed during adaptive evolution. *Evolution* 52:935-949.
- Orr, H. A. 1998b. Testing natural selection vs. genetic drift in phenotypic evolution using quantitative trait locus data. *Genetics* 149:2099-2104.
- Orr, H. A., and J. A. Coyne. 1992. The genetics of adaptation revisited. *Am. Nat.* 140:725-742.
- Paterson, A. H. 1997. Comparative mapping of plant phenotypes. *Plant Breed Rev.* 14:13-37.
- Paterson, A. H., J. W. Deverna, B. Lanini, and S. D. Tanksley. 1990. Fine mapping of quantitative trait loci using selected overlapping recombinant chromosomes in an interspecies cross of tomato. *Genetics* 124:735-742.
- Pawlowski, S. H. 1964. Seed genotype and oil percentage relationship between seeds of a sunflower. *Can J Genet Cytol* 6:293-297.
- Piperno, D. R. 2001. ARCHAEOLOGY: On Maize and the Sunflower. *Science* 292: 2260-2261.

- Pope, K O., M. E. D. Pohl, J. G. Jones, D. L. Lentz, C. von Nagy, F. J. Vega, and I. R. Quitmyer. 2001. Origin and environmental setting of ancient agriculture in the lowlands of Mesoamerica. *Science* 292:1370-1373.
- R Development Core Team. 2006. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. <http://www.R-project.org>.
- Rebaï, A., B. Goffinet, and B. Mangin. 1994. Approximate thresholds of interval mapping tests for QTL detection. *Genetics* 138:235-240.
- Rebaï, A., B. Goffinet, and B. Mangin. 1995. Comparing power of different methods for QTL detection. *Biometrics* 51:87-99.
- Rieseberg, L. H., and G. J. Seiler. 1990. Molecular evidence and the origin and development of the domesticated sunflower (*Helianthus annuus*, Asteraceae). *Econ. Bot.* 44:79-91.
- Rieseberg, L. H., R. Carter, and S. Zona. 1990. Molecular tests of the hypothesized hybrid origin of two diploid *Helianthus* species (Asteraceae). *Evolution* 44: 1498-1511.
- Rieseberg, L. H. 1991. Homoploid reticulate evolution in *Helianthus*: evidence from ribosomal genes. *Am. J. Bot.* 78:1218-237.
- Rieseberg, L. H., O. Raymond, D. M. Rosenthal, Z. Lai, K. D. Livingstone, T. Nakazato, J. L. Durphy, A. E. Schwarzbach, L. A. Donovan, and C. Lexer. 2003. Major ecological transitions in wild sunflowers facilitated by hybridization. *Science* 301:1211-1216.
- Rodriguez de la Paz, J. 2005. Formación de una población de mapeo genético en girasol (*Helianthus annuus* L.) y evaluación comparativa de apareamiento cromosómico. Tesis Doctor en Ciencias. UAAAN. Buenavista, Saltillo, Coahuila, México. 73 p.

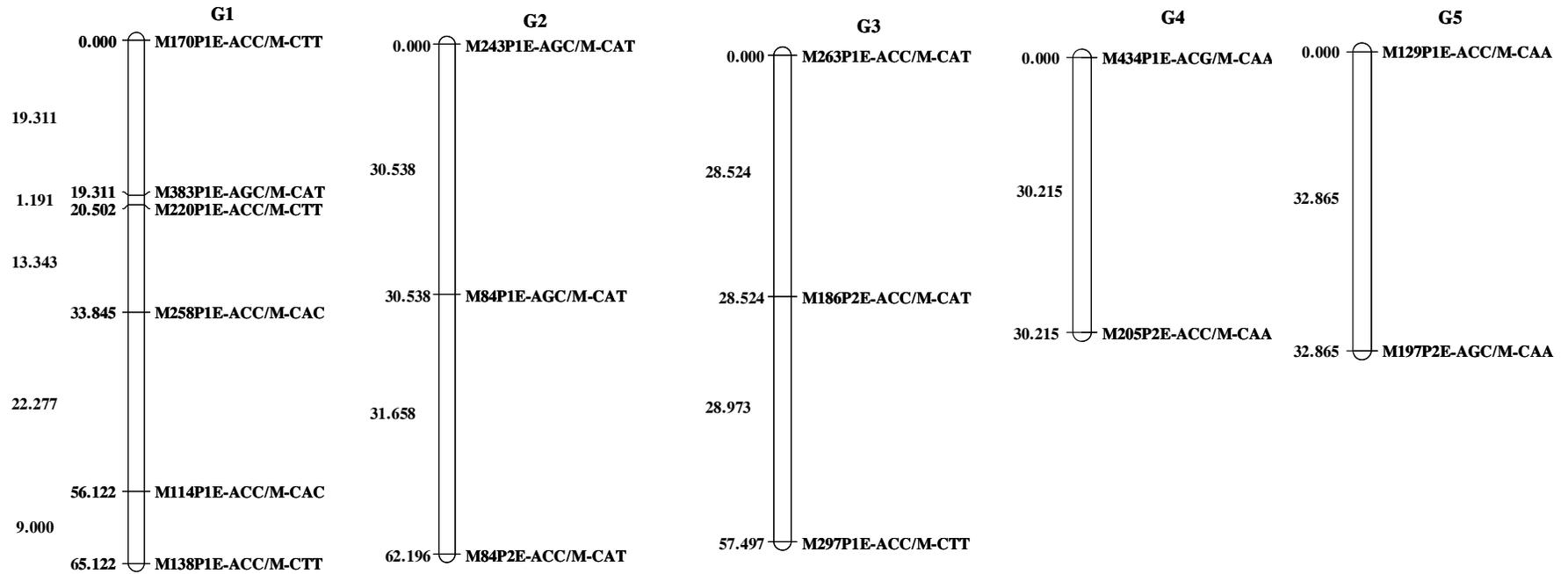
- Rosenthal, D., A. E. Schwarzbach, L. A. Donovan, O. Raymond, and L. H. Rieseberg. 2002. Phenotypic differentiation between three ancient hybrid taxa and their parental species. *Int. J. Plant Sci.* 163:387-98.
- Satagopan, J. M., B. S. Yandell, M. A. Newton, and T. C. Osborn. 1996. A Bayesian approach to detect quantitative trait loci using Markov Chain Monte Carlo. *Genetics* 144: 805-816.
- Schneiter, A. A., and J. F. Miller. 1981. Description of sunflower growth stages. *Crop Science* 21:901-903.
- Seiler, G. J. 1992. Utilization of wild sunflower species for the improvement of cultivated sunflower. *Field Crop Research* 30:195-230.
- Seiler, G. J., and L. H. Rieseberg. 1997. Systematics, origin and germoplasm resources of the wild and domesticated sunflower. ASA, CSSA, SSSA. Sunflower Technology and Production, Agronomy Monograph no. 35. Madison, WI. USA. pp:21-65.
- Sen, S., J. M. Satagopan, and G. A. Churchill. 2005. Quantitative trait locus study design from an information perspective. *Genetics* 170:447-464.
- Seung-Chul, L. and H. Rieseberg. 1999. Genetic architecture of species differences in annual sunflowers: Implications for adaptive trait introgression. *Genetics* 153:965-977.
- Schilling, E. E., and C. B. Heiser. 1981. Intrageneric classification of *Helianthus* (Compositae). *Taxon* 30:393-403.
- Shrimpton, A. E., and A. Robertson. 1988. The isolation of polygenic factors controlling bristle score in *Drosophila melanogaster*. I. Allocation of third chromosome sternopleural bristle effects to chromosome sections. *Genetics* 118:437-443.
- Sillanpää, M. J., and E. Arjas. 1999. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics*. 151:1605-1619.

- Soller, M., T. Brody, and A. Genizi. 1976. In the power of experimental designs for the detection of linkage between marker loci and quantitative loci in crosses between inbred lines. *Theor. Appl. Genet.* 47:35-39.
- Tang, S., J. K. Yu, M. B. Slabaugh, D. K. Shintani, and S. J. Knapp. 2002. Simple sequence repeat map of the sunflower genome. *Theor. Appl. Genet.* 105:1124-1136.
- Thompson, T. E., G. N. Fick, and J. R. Cedeno. 1979. Maternal control of seed oil percentage in sunflower. *Crop Science* 19:617-619.
- Timmerman-Vaughan, G. A., J. A. McCallum, T. J. Frew, N. F. Weeden, and A. C. Russell. 1996. Linkage mapping of quantitative trait loci controlling seed weight in pea (*Pisum sativum* L.). *Theor. Appl. Genet.* 93:431-439.
- Uimari, P., and I. Hoeschele. 1997. Mapping-linked quantitative trait loci using Bayesian analysis and Markov Chain Monte Carlo algorithms. *Genetics* 146:735-743.
- United States Department of Agriculture (USDA). 2008. <http://plants.usda.gov>. Rev. 04 de marzo de 2008.
- Van Ooijen, J. W. 1999. LOD significance thresholds for QTL analysis in experimental populations of diploid species. *Heredity* 83:613-624.
- Van Ooijen, J. W., and R. E. Voorrips. 2001. JoinMap® 3.0: Software for the calculation of genetic linkage maps. Plant Research International, Wageningen, the Netherlands.
- Van Ooijen, J. W., M. P. Boer, R. C. Jansen, and C. Maliepaard. 2002. MapQTL® 4.0: Software for the calculation of QTL positions on genetic maps. Plant Research International, Wageningen, The Netherlands.
- Varshney, M., R. Prasad, N. Kumar, H. S. Harjit-Singh, and P. K. Gupta. 2000. Identification of eight chromosomes and a microsatellite marker on 1AS

- associated with QTL for grain weight in bread wheat. *Theor. Appl. Genet.* 8:1290-1294.
- Visscher, P. M., R. Thompson, and C. S. Haley. 1996. Confidence intervals in QTL mapping by bootstrapping. *Genetics.* 143:1013-1020.
- Vos, P., R. Hogers, M. Bleeker, M. Reijans, T. Van de Lee, M. Hornes, A. Frijters, J. Pot, J. Peleman, M. Kuiper, and M. Zabeau. 1995. AFLP: A new technique for DNA fingerprinting. *Nucleic Acids Res.* 23:4407-414.
- Weller, J. I., M. Soller, and T. Brody. 1988. Linkage analysis of quantitative traits in an interspecific cross of tomato (*Lycopersicon esculentum* × *Lycopersicon pimpinellifolium*) by means of genetic markers. *Genetics* 118:329-339.
- West-Eberhard, M. J. 2005. Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci. USA* 102:6543-6549.
- Westfall, P. H., and S. S. Young. 1993. Resampling-Based Multiple Testing: Examples and Methods for P-Value Adjustment. John Wiley and Sons, New York.
- Wolfram, S. 2003. Software Mathematica version 5.0. Wolfram Research, Inc.
- Xu, S. 1995. A comment on the simple regression method for interval mapping. *Genetics* 141:1657-1659.
- Yu, J. K., S. Tang, M. B. Slabaugh, A. Heesacker, G. Cole, M. Herring, J. Soper, F. Han, W. C. Chu, D. M. Webb, L. Thompson, K. J. Edwards, S. Berry, A. J. Leon, G. Grondona, C. Olungu, N. Maes, and S. J. Knapp. 2003. Towards a saturated molecular genetic linkage map for cultivated sunflower. *Crop Science* 43:367-387.
- Zeng, Z-B. 1993. Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proc. Natl. Acad. Sci. USA.* 90:10972-10976.
- Zeng, Z-B. 1994. Precision mapping of quantitative trait loci. *Genetics* 136:1457-1468.

- Zeng, Z-B. 2005. QTL mapping and the genetic basis of adaptation: recent developments. *Genetica* 123:25-37.
- Zeng, Z-B., C-H. Kao, and C. J. Basten. 1999. Estimating the genetic architecture of quantitative traits. *Genet. Res.* 74:279-289.
- Zeng, Z-B., T. Wang, and W. Zou. 2005. Modeling quantitative trait loci and interpretation of models. *Genetics* 169:1711-1725.
- Zhang, Y-M., and S. Xu. 2004. Mapping quantitative trait loci in F<sub>2</sub> incorporating phenotypes of F<sub>3</sub> progeny. *Genetics* 166:1981-1993.
- Zou, F., J. P. Fine, J. Hu, and D. Y. Lin. 2004. An efficient resampling method for assessing genome-wide statistical significance in mapping quantitative trait loci. *Genetics* 168:2307-2316.

## **VIII. APÉNDICE**



**Figura A.1** Mapa de ligamiento genético de girasol basado en una población de 133 líneas  $F_{2:3}$ , derivada de un cruzamiento entre girasol cultivado (*Helianthus annuus* L. var. *macrocarpus* (DC.) Cockerell, HA89) y silvestre (*Helianthus annuus* L. ssp. *texanus*, Heiser) con marcadores AFLP, con 40 loci distribuidos en 17 grupos de ligamiento.

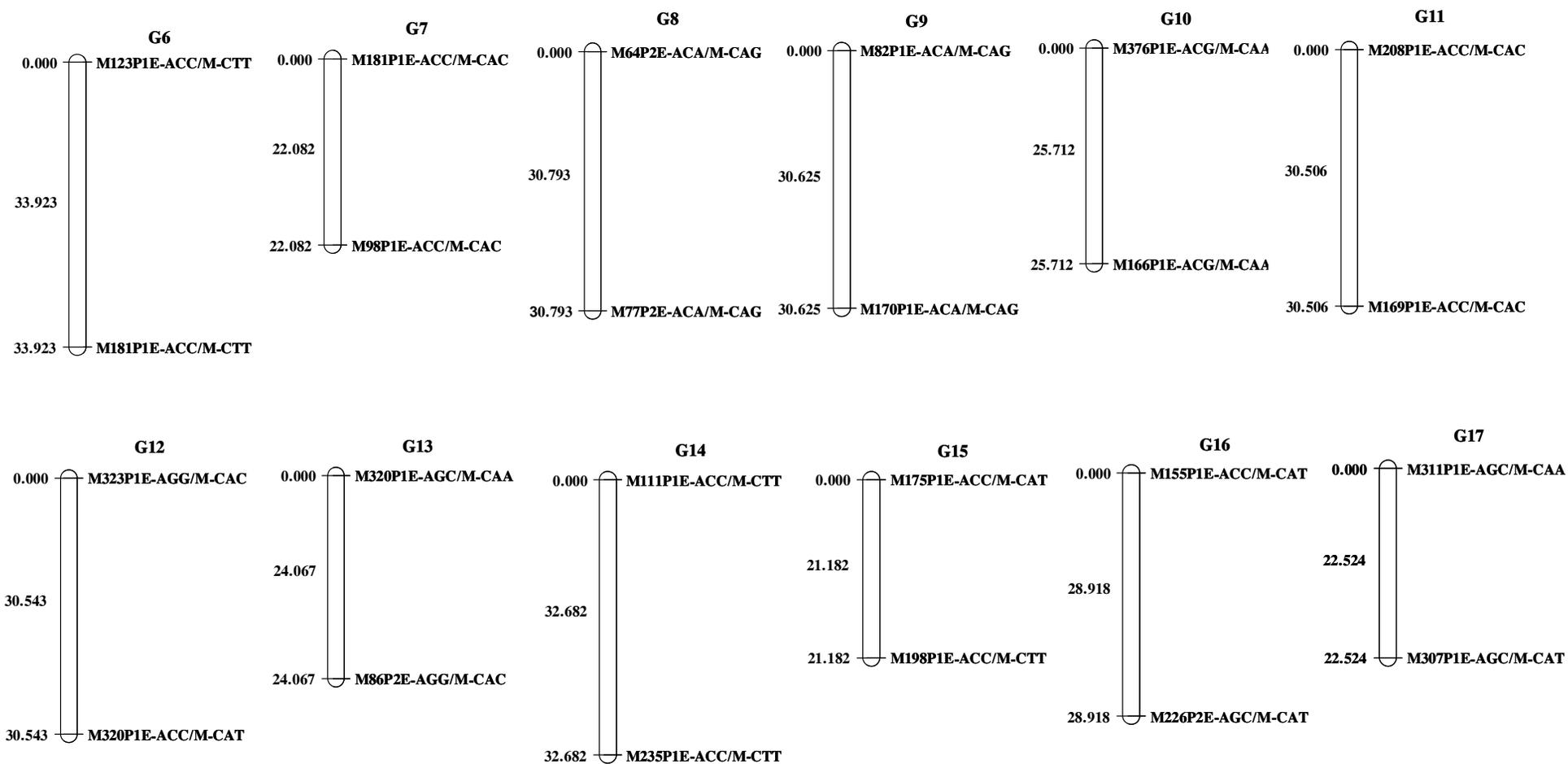


Figura A.1 Continuación.....

**Cuadro A.1** Polimorfismos utilizados para obtener el mapa de ligamiento.

Marcador	Locus	Marcador	Locus	Marcador	Locus	Marcador	Locus
1	M431P1E-AAG/M-CAA	14	M116P1E-AGG/M-CAC	27 <sup>¶</sup>	M155P1E-ACC/M-CAT	40	M198P1E-ACA/M-CAG
2	M357P1E-AAG/M-CAA	15	M110P1E-AGG/M-CAC	28	M149P1E-ACC/M-CAT	41	M186P1E-ACA/M-CAG
3	M317P1E-AAG/M-CAA	16	M201P1E-ACG/M-CAA	29 <sup>¶</sup>	M383P1E-AGC/M-CAT	42	M181P1E-ACA/M-CAG
4	M437P1E-AGC/M-CAA	17	M84P1E-ACG/M-CAA	30	M353P1E-AGC/M-CAT	43	M130P1E-ACA/M-CAG
5	M328P1E-AGC/M-CAA	18	M416P1E-ACC/M-CTT	31	M281P1E-AGC/M-CAT	44 <sup>¶</sup>	M82P1E-ACA/M-CAG
6 <sup>¶</sup>	M320P1E-AGC/M-CAA	19	M290P1E-ACC/M-CTT	32	M262P1E-AGC/M-CAT	45	M371P1E-AAG/M-CAA
7 <sup>¶</sup>	M311P1E-AGC/M-CAA	20 <sup>¶</sup>	M220P1E-ACC/M-CTT	33 <sup>¶</sup>	M243P1E-AGC/M-CAT	46	M255P1E-AGC/M-CAA
8	M75P1E-AGC/M-CAA	21	M166P1E-ACC/M-CTT	34	M193P1E-AGC/M-CAT	47	M215P2E-ACC/M-CAC
9 <sup>¶</sup>	M208P1E-ACC/M-CAC	22 <sup>¶</sup>	M111P1E-ACC/M-CTT	35	M410P1E-ACA/M-CAG	48	M118P2E-ACC/M-CAA
10	M193P1E-ACC/M-CAC	23	M288P1E-AGC/M-CTT	36	M281P1E-ACA/M-CAG	49	M115P2E-ACC/M-CAA
11	M85P1E-ACC/M-CAC	24	M183P1E-AGC/M-CTT	37	M261P1E-ACA/M-CAG	50	M304P2E-ACG/M-CAA
12 <sup>¶</sup>	M323P1E-AGG/M-CAC	25	M137P1E-AGC/M-CTT	38	M235P1E-ACA/M-CAG	51	M486P2E-ACC/M-CTT
13	M129P1E-AGG/M-CAC	26 <sup>¶</sup>	M175P1E-ACC/M-CAT	39	M224P1E-ACA/M-CAG	52	M324P2E-ACC/M-CTT

<sup>¶</sup> Loci marcadores ligados al mapa genético.

**Cuadro A.1** Continuación.....

Marcador	Locus	Marcador	Locus	Marcador	Locus	Marcador	Locus
53	M305P2E-ACC/M-CTT	66 <sup>¶</sup>	M181P1E-ACC/M-CAC	79	M252P1E-AAG/M-CAA	92	M215P1E-AGC/M-CAA
54	M435P2E-AGC/M-CTT	67 <sup>¶</sup>	M169P1E-ACC/M-CAC	80	M232P1E-AAG/M-CAA	93	M209P1E-AGC/M-CAA
55	M212P2E-AGC/M-CTT	68	M364P1E-ACG/M-CAA	81	M223P1E-AAG/M-CAA	94	M120P1E-AGC/M-CAA
56	M147P2E-AGC/M-CTT	69	M343P1E-ACG/M-CAA	82	M217P1E-AAG/M-CAA	95	M94P1E-AGC/M-CAA
57	M63P2E-AGC/M-CTT	70 <sup>¶</sup>	M235P1E-ACC/M-CTT	83	M199P1E-AAG/M-CAA	96	M328P1E-ACC/M-CAC
58	M301P2E-ACC/M-CAT	71 <sup>¶</sup>	M123P1E-ACC/M-CTT	84	M179P1E-AAG/M-CAA	97 <sup>¶</sup>	M258P1E-ACC/M-CAC
59	M210P2E-AGC/M-CAT	72	M341P1E-AGC/M-CTT	85	M160P1E-AAG/M-CAA	98	M254P1E-ACC/M-CAC
60	M89P2E-AGC/M-CAT	73	M307P1E-AGC/M-CTT	86	M154P1E-AAG/M-CAA	99	M202P1E-ACC/M-CAC
61	M312P2E-ACA/M-CAG	74	M298P1E-AGC/M-CTT	87	M109P1E-AAG/M-CAA	100 <sup>¶</sup>	M114P1E-ACC/M-CAC
62	M305P2E-ACA/M-CAG	75	M68P1E-AGC/M-CTT	88	M298P1E-AGC/M-CAA	101	M475P1E-AGG/M-CAC
63 <sup>¶</sup>	M64P2E-ACA/M-CAG	76	M165P1E-AGC/M-CAT	89	M293P1E-AGC/M-CAA	102	M469P1E-AGG/M-CAC
64	M284P2E-AGC/M-CAA	77	M385P1E-ACA/M-CAG	90	M276P1E-AGC/M-CAA	103	M181P1E-AGG/M-CAC
65	M318P1E-ACC/M-CAC	78	M299P1E-AAG/M-CAA	91	M232P1E-AGC/M-CAA	104	M174P1E-AGG/M-CAC

<sup>¶</sup> Loci marcadores ligados al mapa genético.

**Cuadro A.1** Continuación.....

Marcador	Locus	Marcador	Locus	Marcador	Locus	Marcador	Locus
105	M162P1E-AGG/M-CAC	118	M180P1E-ACG/M-CAA	131	M151P1E-ACC/M-CTT	144	M213P1E-ACC/M-CAT
106	M136P1E-AGG/M-CAC	119 <sup>¶</sup>	M166P1E-ACG/M-CAA	132 <sup>¶</sup>	M138P1E-ACC/M-CTT	145	M205P1E-ACC/M-CAT
107	M274P1E-ACC/M-CAA	120	M132P1E-ACG/M-CAA	133	M134P1E-ACC/M-CTT	146	M182P1E-ACC/M-CAT
108	M213P1E-ACC/M-CAA	121	M469P1E-ACC/M-CTT	134	M89P1E-ACC/M-CTT	147 <sup>¶</sup>	M307P1E-AGC/M-CAT
109 <sup>¶</sup>	M129P1E-ACC/M-CAA	122	M342P1E-ACC/M-CTT	135	M221P1E-AGC/M-CTT	148	M297P1E-AGC/M-CAT
110	M122P1E-ACC/M-CAA	123	M243P1E-ACC/M-CTT	136	M199P1E-AGC/M-CTT	149	M154P1E-AGC/M-CAT
111	M107P1E-ACC/M-CAA	124 <sup>¶</sup>	M198P1E-ACC/M-CTT	137	M169P1E-AGC/M-CTT	150 <sup>¶</sup>	M84P1E-AGC/M-CAT
112	M464P1E-ACG/M-CAA	125	M194P1E-ACC/M-CTT	138	M133P1E-AGC/M-CTT	151	M64P1E-AGC/M-CAT
113	M439P1E-ACG/M-CAA	126	M189P1E-ACC/M-CTT	139	M359P1E-ACC/M-CAT	152	M322P1E-ACA/M-CAG
114 <sup>¶</sup>	M434P1E-ACG/M-CAA	127 <sup>¶</sup>	M181P1E-ACC/M-CTT	140	M341P1E-ACC/M-CAT	153	M285P1E-ACA/M-CAG
115 <sup>¶</sup>	M376P1E-ACG/M-CAA	128	M177P1E-ACC/M-CTT	141 <sup>¶</sup>	M320P1E-ACC/M-CAT	154	M268P1E-ACA/M-CAG
116	M311P1E-ACG/M-CAA	129 <sup>¶</sup>	M170P1E-ACC/M-CTT	142 <sup>¶</sup>	M263P1E-ACC/M-CAT	155	M244P1E-ACA/M-CAG
117	M253P1E-ACG/M-CAA	130	M155P1E-ACC/M-CTT	143	M220P1E-ACC/M-CAT	156	M192P1E-ACA/M-CAG

<sup>¶</sup> Loci marcadores ligados al mapa genético.

**Cuadro A.1** Continuación.....

Marcador	Locus	Marcador	Locus	Marcador	Locus	Marcador	Locus
157	M174P1E-ACA/M-CAG	170	M268P2E-ACC/M-CAC	183	M155P2E-AGC/M-CTT	196	M163P2E-ACA/M-CAG
158 <sup>¶</sup>	M170P1E-ACA/M-CAG	171 <sup>¶</sup>	M86P2E-AGG/M-CAC	184	M280P2E-ACC/M-CAT	197	M259P2E-AGC/M-CAA
159	M160P1E-ACA/M-CAG	172	M334P2E-ACC/M-CAA	185	M239P2E-ACC/M-CAT	198	M92P2E-ACC/M-CAC
160	M153P1E-ACA/M-CAG	173	M316P2E-ACC/M-CAA	186 <sup>¶</sup>	M186P2E-ACC/M-CAT	199	M75P2E-AGG/M-CAC
161	M224P1E-AGC/M-CAA	174	M282P2E-ACC/M-CAA	187	M164P2E-ACC/M-CAT	200	M325P2E-ACC/M-CAA
162	M173P1E-AGC/M-CAA	175	M251P2E-ACC/M-CAA	188	M122P2E-ACC/M-CAT	201	M96P2E-AGC/M-CTT
163	M115P1E-AGC/M-CAA	176 <sup>¶</sup>	M205P2E-ACC/M-CAA	189 <sup>¶</sup>	M84P2E-ACC/M-CAT	202	M294P2E-AGC/M-CAT
164 <sup>¶</sup>	M98P1E-ACC/M-CAC	177	M85P2E-ACC/M-CAA	190	M442P2E-AGC/M-CAT	203	M372P2E-ACC/M-CAC
165 <sup>¶</sup>	M297P1E-ACC/M-CTT	178	M380P2E-ACG/M-CAA	191	M363P2E-AGC/M-CAT	204	M340P2E-ACC/M-CAA
166	M197P2E-AAG/M-CAA	179	M372P2E-ACC/M-CTT	192	M311P2E-AGC/M-CAT	205	M242P2E-ACC/M-CAA
167	M172P2E-AAG/M-CAA	180	M217P2E-ACC/M-CTT	193 <sup>¶</sup>	M226P2E-AGC/M-CAT	206	M424P2E-AGC/M-CAT
168	M100P2E-AAG/M-CAA	181	M319P2E-AGC/M-CTT	194	M301P2E-ACA/M-CAG	207	M366P2E-AGC/M-CAT
169 <sup>¶</sup>	M197P2E-AGC/M-CAA	182	M193P2E-AGC/M-CTT	195	M227P2E-ACA/M-CAG	208 <sup>¶</sup>	M77P2E-ACA/M-CAG

<sup>¶</sup> Loci marcadores ligados al mapa genético.

**Cuadro A.2** Medias de seis caracteres evaluados de las 133 líneas F<sub>2:3</sub>, incluyendo los dos progenitores (girasol cultivado y silvestre). **AP** = Altura de planta (m); **NC** = Número de capítulos; **DC** = Diámetro de capítulos (cm); **NA** = Número de aquenios; **PA** = Peso de aquenios (g); **CA** = Contenido de aceite de aquenios (%).

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
2	1.14	27	3.27	27	0.47	34.92
7	1.15	12	3.55	8	0.18	33.12
9	0.99	40	4.59	74	2.37	31.97
11	1.14	87	1.33	5	0.07	30.94
13	1.23	NA	NA	NA	NA	NA
22	1.05	58	3.65	8	0.10	35.31
25	1.16	35	1.12	1	0.01	NA
27	1.46	10	7.91	49	1.66	32.54
29	1.00	19	2.47	3	0.11	NA
32	0.70	12	2.60	11	0.25	28.68
33	0.94	31	1.57	22	0.28	40.44
42	1.27	19	1.75	3	0.06	NA
43	1.17	42	3.48	32	0.72	37.3
54	1.07	9	4.50	13	0.31	30.56
59	1.10	22	2.00	6	0.09	NA

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
60	1.22	51	NA	NA	NA	NA
68	0.84	40	2.95	7	0.12	NA
69	1.14	46	1.25	1	0.01	NA
71	1.09	5	3.12	20	0.24	22.85
74	0.77	6	4.08	43	0.53	NA
79	1.47	40	4.52	72	0.45	30.93
83	1.23	19	3.17	2	0.05	NA
84	1.43	34	6.05	49	0.30	NA
87	1.09	26	2.52	6	0.04	NA
88	1.06	31	3.16	31	0.37	31.8
99	0.74	18	3.60	43	0.68	36.76
102	1.37	90	2.93	24	0.36	38.44
107	1.43	23	6.13	12	0.22	33.31
109	1.42	51	4.42	25	0.49	32.79
110	1.23	11	2.75	11	0.20	NA

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
111	0.98	12	1.12	7	0.15	NA
112	1.18	21	1.75	7	0.11	NA
114	0.85	14	2.62	7	0.13	33.07
117	0.92	10	3.54	17	0.43	30.16
118	0.83	11	2.53	15	0.17	33.8
121	NA	1	1.12	5	0.10	NA
123	1.22	81	2.52	8	0.12	33.99
125	1.04	31	3.01	12	0.24	31.17
127	1.18	32	3.50	4	0.06	NA
130	1.55	13	4.63	84	2.49	26.24
137	1.28	12	1.42	1	0.01	NA
138	0.91	22	1.35	3	0.05	NA
141	1.17	20	1.95	11	0.21	NA
145	1.23	18	4.37	27	0.49	29.88
147	0.92	18	2.38	2	0.03	NA

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
148	1.12	18	1.10	2	0.04	NA
169	1.51	48	4.25	16	0.36	29.51
175	1.66	75	1.62	4	0.06	NA
178	0.75	12	1.42	5	0.09	37.3
180	0.97	13	4.74	32	0.88	37.55
187	1.04	37	2.85	23	0.30	35.34
188	1.33	57	3.54	30	0.53	38.51
190	0.95	41	1.25	1	0.01	NA
192	1.13	20	3.37	59	1.25	37.8
194	1.00	20	3.59	32	0.84	30.49
196	0.75	30	1.73	23	0.46	35.2
197	1.27	62	2.94	7	0.10	32.25
198	0.93	28	4.86	27	0.47	31.27
202	0.99	9	NA	NA	NA	NA
203	0.72	17	1.30	1	0.26	38.44

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
204	0.99	22	2.86	34	0.58	38.29
209	1.26	51	3.25	9	0.17	36.23
210	1.07	9	2.34	2	0.03	NA
211	1.20	65	NA	1	0.01	NA
216	0.89	6	1.97	5	0.09	NA
217	1.18	47	1.62	1	0.01	39.11
218	1.08	38	2.65	5	0.11	37.48
219	0.97	14	3.65	13	0.21	32.79
226	1.24	8	7.33	17	0.53	32.53
227	1.35	29	3.73	16	0.41	34.8
228	0.98	37	3.13	17	0.34	30.56
229	0.74	22	3.05	19	0.17	38.88
230	0.90	15	1.93	25	0.56	25.98
234	1.00	33	1.19	3	0.04	NA
235	0.85	23	1.44	6	0.11	32.96

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
237	0.80	19	3.43	23	0.35	NA
239	1.23	17	2.00	3	0.09	NA
246	1.16	5	2.53	12	0.47	32.07
247	1.08	59	3.22	4	0.06	NA
250	0.82	28	2.48	30	0.55	32.74
253	1.21	15	4.17	11	0.26	22.22
255	1.03	23	3.53	35	1.44	34.65
256	0.88	24	2.84	24	0.36	34.17
258	0.96	28	3.61	45	0.76	33.89
259	1.08	42	3.69	23	0.40	35.15
260	1.04	8	3.22	47	0.93	37.47
263	1.34	25	4.31	35	0.72	34.74
264	0.97	57	2.99	6	0.08	37.53
280	0.93	22	3.28	66	1.14	33.38
281	0.88	10	1.37	1	0.01	NA

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
284	1.34	8	NA	NA	NA	NA
286	0.67	19	2.93	24	0.34	27.65
287	1.17	56	1.62	2	0.03	NA
289	1.14	50	3.83	8	0.23	39.9
291	1.72	44	3.23	17	0.22	20.39
293	0.89	23	3.62	66	1.36	32.64
295	0.98	6	NA	0	NA	NA
299	1.06	28	2.68	17	0.32	36.19
302	0.72	11	3.55	1	0.02	NA
303	1.30	21	NA	0	NA	NA
309	0.98	10	1.84	38	0.73	43.1
313	1.23	42	1.62	1	0.01	NA
314	1.09	55	1.61	1	0.01	NA
316	1.12	NA	3.50	130	2.93	35.64
317	1.07	18	4.23	7	0.21	34.29

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
320	1.07	22	1.63	9	0.17	33.01
321	1.18	31	2.87	11	0.20	37.69
323	0.92	16	1.56	6	0.15	NA
326	0.70	21	2.79	17	0.31	30.58
327	1.06	14	5.20	16	0.37	21.99
328	1.15	30	2.69	8	0.10	29.65
329	0.83	23	3.89	12	0.31	32.4
330	0.99	22	1.46	2	0.04	NA
333	1.21	12	5.36	11	0.26	35.4
338	0.85	17	2.59	7	0.13	NA
340	0.88	51	3.23	6	0.11	25.13
342	0.66	10	2.94	4	0.08	NA
344	1.20	38	1.54	14	0.20	21.58
351	0.93	59	3.57	6	0.11	NA
353	1.10	12	3.17	15	0.27	34.32

**Cuadro A.2** Continuación.....

<b>Línea</b>	<b>AP</b>	<b>NC</b>	<b>DC</b>	<b>NA</b>	<b>PA</b>	<b>CA</b>
354	0.90	73	3.59	19	0.22	28.88
355	0.91	22	1.57	1	0.01	NA
360	1.06	28	NA	NA	NA	NA
361	1.27	20	4.39	51	1.21	35.05
362	0.80	33	3.12	21	0.36	24.28
363	1.07	21	2.95	6	0.11	31.79
365	0.72	23	3.17	21	0.36	20.16
367	1.17	27	4.54	21	0.46	NA
369	1.11	13	1.76	31	0.71	34.59
370	0.79	17	3.23	2	0.02	NA
372	1.17	61	2.83	56	1.10	28.35
373	1.17	97	1.80	19	0.26	28.37
374	1.63	22	3.65	7	0.13	NA
Ac-8-2	1.21	405	1.34	2	0.01	NA
HA89	0.81	NA	5.06	93	4.34	44.33