



UNIVERSIDAD DE BUENOS AIRES

Facultad de Ciencias Exactas y Naturales

Departamento de Ecología, Genética y Evolución

**DESARROLLO DE UN MAPA GENÉTICO Y FÍSICO INTEGRADO
PARA GIRASOL CULTIVADO Y SU APLICACIÓN
EN LA CARACTERIZACIÓN DE REGIONES GENÓMICAS INVOLUCRADAS
EN LA RESISTENCIA A ENFERMEDADES**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires
en el área Ciencias Biológicas

Lic. Paola M. Talia

Directora de Tesis: Dra. Ruth Heinz
Directora Asistente: Dra. Norma Paniego
Consejero de Estudios: Dr. H. Esteban Hopp

Instituto de Biotecnología CICV y A INTA Castelar

Buenos Aires, 2008

Desarrollo de un mapa genético y físico integrado para girasol cultivado y su aplicación en la caracterización de regiones genómicas involucradas en la resistencia a enfermedades

Resumen

En el presente trabajo se describe la construcción de un mapa genético saturado de referencia para girasol (*Helianthus annuus*) y la localización física de grupos de ligamiento que incluyen genes y/o regiones genómicas de interés agronómico.

Se obtuvo un mapa genético a partir de una población de líneas recombinantes endocriadas (RILs) derivadas del cruzamiento de RHA266 x PAC2 que incluye 547 marcadores (231 SSR, 9 EST-SSR, 3 InDels y 304 AFLPs ordenados en 17 grupos de ligamiento (GL) con un valor LOD de 4. La longitud total estimada del mapa genético es de 1.942,3 cM con una densidad media de 3,6 cM por *locus*. La incorporación de SSRs permitió relacionar el mapa con otros mapas públicos unificando la nomenclatura para la denominación de los GLs.

Se caracterizó el complemento cromosómico de *H annuus* mediante el mapeo físico de regiones repetitivas propias de la especie utilizando la técnica BAC-FISH. Se localizaron físicamente marcadores SSR ubicados en los GL 10 y 16 usando la técnica BAC-FISH y marcadores anclados en los GL 7 y 17 mediante C-PRINS. Los GL 10, 16 y 17 fueron seleccionados en base a su asociación con QTLs de resistencia a *Sclerotinia sclerotiorum* y el GL 7 por su asociación con QTLs de tolerancia a estrés hídrico.

La integración de estrategias basadas en mapas genéticos saturados, mapas físicos y genómica comparativa permite expandir los alcances de la investigación hacia el análisis de QTLs, el clonado posicional de genes, y el mapeo físico de regiones de interés agronómico.

Palabras clave: *Helianthus annuus*, SSR, EST-SSR, mapa genético, mapa físico, cariotipo, BAC-FISH, PRINS.

Development of an integrated genetic and physical map for cultivated sunflower and its application in the characterization of genomic regions involved in disease resistance

Abstract

This work presents the construction of a reference saturated-genetic-linkage-map for cultivated sunflower (*Helianthus annuus*) and the physical localization of linkage groups that include genes or genomic regions of agricultural interest.

The reference genetic map was obtained using a recombinant inbred line (RIL) mapping population generated from the cross between (RHA266 x PAC2). It consisted of 547 markers (231 SSR, 9 EST-SSR, 3 InDels and 304 AFLPs) distributed in 17 linkage groups (LG), with a LOD score of 4. The map spanned 1942.3 centimorgans (cM) and had a mean density of 3.6 cM per *locus*. The incorporation of SSR allowed the complete cross reference with the public maps, unifying the nomenclature for the LGs.

Sunflower chromosome complement was characterized using the BAC-FISH technique. Physical mapping of SSR of the LGs 10 and 16 was accomplished using BAC-FISH procedure and SSR of LGs 7 and 17 were localized by the C-PRINS technique. LGs 10, 16 and 17 were selected based on their association with QTLs for resistance to *Sclerotinia sclerotiorum* and LG 7 based on its association with QTLs for tolerance to water stress.

The integration of saturated genetic maps, physical map, and comparative genomics approaches will allow expanding the scope of the present research to the analysis of QTLs, map-based gene cloning, and the physical mapping of agricultural important regions.

Key words: *Helianthus annuus*, SSR, EST-SSR, genetic map, physical map, karyotype, BAC-FISH, PRINS.

AGRADECIMIENTOS

El siguiente es un apartado difícil de escribir, ya que seguramente es la única sección que la mayoría de la gente lee. Todos esperan ver cómo el autor distribuye los agradecimientos, qué le debe a cada persona, en lo personal o en lo profesional, y qué le dice a cada uno. Conformar a todos los que me apoyaron durante esta tesis es un tarea más que ardua, y en ese sentido me siento en la obligación de escribir lo que siento hoy, que no es exactamente igual a lo que sentí durante el transcurso de todos estos años.

A mis directoras, Dra. Ruth Heinz y Dra. Norma Paniego, por acompañarme y brindarme todo su apoyo y confianza, por estar dispuestas a ayudarme siempre que fue necesario y compartir sinceramente cada logro.

Al Dr. Esteban Hopp, por su estímulo constante y por su apoyo incondicional.

A Vero Nishi, por su valiosa colaboración durante los primeros años de esta tesis, por ser tan buena y generosa.

Al Dr. Eduardo Greizerstein, por brindarme el apoyo necesario en las técnicas de citogenética molecular, porque sin él no hubiera podido hacer nada, por su paciencia y amabilidad.

A la Dra. Lidia Poggio, por abrirme las puertas de su laboratorio, por su aliento y asesoramiento en la interpretación de los resultados de las técnicas de citogenética molecular.

Al Dr. Gerardo Cervigni por su valiosa colaboración en el manejo de los programas de mapeo genético y al Dr. Javier Calcagno, por su asesoramiento en el análisis estadístico de los datos.

A la Dra. Margarita Cano y el Dr. Mario Poli por su colaboración en la técnica de genotipificación mediante métodos radioactivos.

A Dra. Elisa Carrillo, directora del Instituto de Biotecnología, por permitirme desarrollar este trabajo con total libertad y a Lela por escucharme y contenerme los días posteriores al accidente, que fueron tan duros para mí.

Al INTA, por ofrecerme sus instalaciones y darme la oportunidad de desarrollar mi tesis en ella.

Agradezco profundamente a las entidades que financiaron este proyecto por su aporte a la realización de esta tesis, a la Agencia Nacional de Promoción Científica y Técnica, a través del PAV2003-137, PID 267 y PID 024 Bid1201 y al INTA mediante el PN 1952 y el PE AEBIO 1433.

A Luis Fernández, por pasarme la mayor parte de la literatura necesaria para desarrollar esta tesis y por su buena predisposición en el soporte informático.

A los chicos del invernáculo: Agustín, Juan Ignacio, Matías, Juan, Valeria Peralta, José y Javier por gran ayuda en el mantenimiento y la multiplicación de la población de mapeo y a la gente del secuenciador, Apu, Vero y Pablo por ser tan estrictos en su trabajo y por su entusiasmo y disposición a ayudarme.

A Paula Fernández, por estar siempre atenta a lo que necesito tanto en lo laboral como en lo personal, por pasarme información sobre cursos y congresos.

A Daniela Tosto, por tratarme como si fuéramos hermanas, por escucharme y orientarme. Por robarme tantos *tips* y permitir que use los suyos también ¡¡¡Me encanta tenerte cerca!!!

A Carolina Díaz Quijano por su participación en el marco del proyecto en colaboración sobre mapeo físico de regiones de interés agronómicas.

A Corina Fusari por los 3 InDels.

A Mariana del Vas, por ser tan participativa, franca y responsable, por contenerme y aconsejarme.

A Carolina Martínez, por ser la ministra de los hermosos eventos que se hacen en el Instituto, deseo que nunca pierdas esa chispa de alegría que nos hace pasar momentos tan amenos.

A Marisa LB, por ser tan franca y sensible durante los meses que compartimos espacio en el locutorio, por ofrecerme y prestarme dinero cuando lo necesité.

A Betiana Parodi por ser una fuente inagotable de consultas sobre ortografía y gramática tanto en español como en inglés.

A las chicas del laboratorio 110 de la Facu, Gi, Geo, Aye, Maia, Maria y Alba, por ayudarme en todos los detalles y hacerme sentir tan cómoda...y también a los chicos! Sergio, Diego, Juan, Diego, Pablo y Arturo por tantas veces que los llame para que me ayuden a “enfocar” en el microscopio.

A Ana Distéfano por su constancia y rigurosidad en el trabajo.

A Flor del Viso, Ceci Vazquez, Ari y Fer Carrari por su constante interés en saber como van mis resultados.

A Angel Cataldi, por ser una persona que dice y actúa en consecuencia, cosa tan poco frecuente por estos días.

A Guille y Paula Ruybal, por su entusiasmo y disposición para ayudarme con los programas de diseño.

A Isabel, Fabián, Vilma, Majo y Roxana por estar siempre dispuestos a ayudar y festejar todos mis chistes.

A las chicas de tenis, que hicieron que mis semanas en el INTA fueran mucho más animosas. Juntas logramos un motivo más para ir contentas a laburar.

A todas las personas del Instituto que estuvieron dispuestas a todo para que esto sea realidad y que compartieron conmigo estos lindos años.

A mis padres, a los que extraño y necesito más que nunca. A mi papá, que si bien no pudo ir a la facultad daba clases todos los días, era un apasionado de la literatura y la política. De vos aprendí las ganas de aprender, debatir, el amor por la lectura y los libros. A mi mamá, que fue la persona más “hermosa” que existió y la más emprendedora. Siempre te admiré, gracias por todo lo que me diste mientras estuviste conmigo y todo lo que me dejaste para que pueda hacer mi camino.

A mi hermana Karina, por enseñarme a ver la vida desde otro ángulo. Por haberme “vestido” en sus viajes a Italia, por ser tan buena tía, por hacerse cargo del emprendimiento familiar y haberlo hecho bien, te felicito!

A mi hermana Ale, por querer regresar a vivir a este país para estar cerca nuestro. Quiero que sepas que para mí sería la mayor felicidad.

A Susy, por ser “la mejor suegra” que tuve y ¡la mejor que hay!, gracias por estar siempre presente y a la vez no invadirnos y respetarnos. Saber que podemos contar con vos nos tranquiliza.

A Maneco, por su comprensión, ayuda y colaboración durante cada etapa de la tesis, por estar dispuesto a apoyarme a pesar de todos los inconvenientes, por quererme como soy.

A Fermín, por ser la estrella que ilumina mi camino, por esperarme hasta tarde y estar enojado durante más de 10 minutos, y luego darme los abrazos más lindos del mundo. Por llenar mi vida de sonrisas. Por darme la posibilidad de sentir más amor del que jamás hubiera imaginado podría sentir; mucho, mucho amor, del amor más grande, más puro y más pleno.

A Maneco, a Fermín y a los que vendrán...
A mi mamá y a mi papá, que ya no me acompañan,
pero me dejaron todo su apoyo.
A mis hermanas Ale y Karina, a mi sobrino Eze y a mi suegra Susy.

Además, dedico este trabajo de tesis a todos
esos jóvenes que transitan vidas duras llenas de piedras,
para alentarlos a continuar con sus proyectos ya que con esfuerzo, paciencia,
dedicación y perseverancia todo puede ser posible.

ÍNDICE GENERAL

ABREVIATURAS Y SIGLAS	16
1. INTRODUCCIÓN	19
1.1 Generalidades del girasol.....	19
1.1.1 Descripción taxonómica.....	19
1.1.2 Importancia económica.....	19
1.1.3 El girasol cultivado (<i>Helianthus annuus L.</i>).....	20
1.1.4 Girasol en Argentina.....	21
1.1.5 Estado de avance del conocimiento sobre la genómica de girasol.....	22
1.2 Herramientas moleculares para el análisis genómico.....	23
1.2.1 Marcadores moleculares.....	23
1.2.1.1 Microsatélites.....	25
1.2.1.2 Marcadores funcionales.....	26
1.2.2 Métodos de genotipificación de los marcadores moleculares.....	34
1.2.2.1 Errores en el proceso de genotipificación, causas y consecuencias.....	35
1.2.3 Mapeo Genético.....	37
1.2.3.1 Construcción de mapas de ligamiento.....	42
1.2.3.1.1 Mapas genéticos de girasol.....	45
1.2.4 Mapeo físico.....	47
1.2.4.1 Características de los genomas de las plantas.....	48
1.2.4.2 Los mapas físicos.....	49
1.2.4.3 La técnica de FISH.....	50
1.2.4.4 La técnica de GISH.....	55
1.2.4.5 La técnica de PRINS.....	55
1.2.4.6 Avances en citogenética molecular en girasol.....	56
1.3 Hipótesis de trabajo.....	58
1.4 Objetivos.....	59
1.4.1 Objetivo general.....	59
1.4.2 Objetivos específicos.....	59

2. MATERIALES Y MÉTODOS	60
2.1 Material vegetal.....	60
2.2 Extracción de ADN.....	61
2.2.1. Extracción por CTAB con el agregado de agentes para extraer polisacáridos.....	61
2.3 Cuantificación del ADN de la población RIL.....	62
2.4 Selección de los marcadores microsatélites y análisis de polimorfismo en los parentales.....	63
2.5 Utilización de programas bioinformáticos aplicados al descubrimiento de marcadores moleculares	63
2.5.1 Estrategia para explotar la base de datos de EST de girasol para detectar marcadores microsatélites	63
2.5.2 Estrategias para detectar polimorfismos de tipo SNP/InDels en los unigenes de girasol	65
2.6 Condiciones de amplificación de fragmentos mediante la reacción en cadena de la polimerasa (PCR)	65
2.6.1 Condiciones de amplificación de los marcadores moleculares.....	65
2.6.2 Reacciones de amplificación para uso de sondas específicas.....	67
2.7 Electroforesis y visualización de ADN en geles de agarosa.....	67
2.8 Purificación de fragmentos de ADN de geles de agarosa.....	68
2.9 Electroforesis y visualización de los fragmentos en geles de poliacrilamida.....	68
2.9.1 Métodos manuales.....	68
2.9.2 Métodos automáticos.....	70
2.9.2.1 Marcación directa.....	71
2.9.2.2 Marcación indirecta.....	72
2.10 Análisis de microsatélites usando el Software GeneMapper® 4.0 (Applied Biosystem, EUA)	74
2.11 Caracterización de los marcadores funcionales.....	78
2.11.1 Análisis de los marcadores EST-SSR.....	78
2.11.2 Análisis de los SNPs por medio de la técnica de CAPS.....	78
2.11.3 Identificación de SNPs e InDels mediante la técnica de resecuenciación de genes candidatos a partir de colecciones de ESTs.....	79
2.12 Análisis estadístico de los datos obtenidos.....	79
2.12.1 Análisis de segregación de los alelos en la población estudiada.....	79
2.12.2 Utilización de herramientas bioinformáticas para el desarrollo	

del mapa genético	80
2.12.3 Distribución de los marcadores moleculares en el mapa genético.....	81
2.13 Preparación de sondas de copia única y de ADN repetitivo para hibridación <i>in situ</i>	82
2.13.1 Preparación de sondas de ADN repetitivo para ser utilizadas en hibridación <i>in situ</i>	83
2.13.1.1 Preparación de sondas con similitud a genes ribosomales	83
2.13.1.2 Preparación de sondas con similitud a retrotransposones.....	84
2.13.1.3. Utilización de secuencias clonadas en Cromosomas Artificiales de Bacterias (BACs) conteniendo ADN repetitivo como sondas para la caracterización del complemento cromosómico	88
2.13.2 Preparación de sondas de copia única.....	90
2.13.2.1 Preparación de sondas a partir de RGAs.....	90
2.13.2.2 Preparación de sondas correspondientes a marcadores SSR de localización única en el genoma de girasol	92
2.13.2.2.1 Diseño de sondas tipo overgo (<i>overlapping oligonucleotide probes</i>)	92
2.13.2.2.2 Marcación de sondas.....	93
2.13.2.2.3 Hibridación, lavados y exposición de membranas.....	94
2.13.2.2.4 Identificación de clones positivos y confirmación por PCR.....	95
2.13.2.2.5 BAC- <i>Southern blot</i>	95
2.14 Preparados citológicos, hibridación mediante la técnica FISH y amplificación de regiones específicas por PRINS y/o C-PRINS	96
2.14.1 Obtención de preparados citológicos.....	96
2.14.1.1 Obtención y fijación del material vegetal.....	96
2.14.1.2 Técnica de aplastado (<i>squash</i>).....	97
2.14.2 Hibridación <i>in situ</i> (FISH).....	97
2.14.2.1 Acondicionamiento de los preparados para FISH.....	97
2.14.2.2 Marcado de sondas para FISH.....	98
2.14.2.3 Hibridación.....	98
2.14.2.4 Lavados post-hibridación.....	99
2.14.2.5 Detección visual de los sitios de hibridación.....	100
2.14.2.6 Obtención y análisis de imágenes.....	100
2.14.3. <i>Primed in situ labelling</i> (PRINS)	100
2.14.4 <i>Cycling-Primed in situ labelling</i> (C-PRINS).....	101

2.14.4.1 Preparación de la mezcla de extensión.....	101
2.14.4.2 Método indirecto de detección de la reacción.....	102
3. RESULTADOS.....	103
3.1 Población de mapeo: Caracterización y evaluación genotípica.....	103
3.2 Desarrollo de marcadores funcionales.....	104
3.2.1 Búsqueda y caracterización de SSR en bases de datos de unigenes y análisis en cuatro genotipos	104
3.2.2 Análisis de marcadores funcionales mediante CAPS.....	109
3.3 Análisis de marcadores utilizando diferentes estrategias de genotipificación.....	112
3.4 Análisis de marcadores moleculares.....	114
3.4.1 Análisis de marcadores microsatélites genómicos.....	114
3.4.2 Análisis de marcadores EST-SSR.....	122
3.5 Construcción del mapa genético de ligamiento.....	123
3.5.1 Distribución de los marcadores moleculares.....	132
3.6 Contribución mediante diversas sondas en la caracterización del complemento cromosómico de <i>Helianthus annuus</i> utilizando la técnica FISH	133
3.6.1 Sondas derivadas de la región ribosomal 26S.....	133
3.6.2 Sonda derivada de una secuencia altamente repetitiva de girasol.....	134
3.6.3 Utilización de un clon de BAC con alto contenido de ADN repetitivo como sonda en experimentos de FISH	134
3.6.4 Contribución mediante diversas sondas repetitivas a la caracterización del complemento cromosómico de <i>Helianthus annuus</i> utilizando la técnica de PRINS	139
3.7 Localización de secuencias de copia única mediante BAC-FISH, PRINS y C-PRINS	141
3.7.1 Utilización de clones de BAC conteniendo ADN microsatélites como sonda en experimentos de hibridación <i>in situ</i>	141
3.7.2 Localización de secuencias de bajo número de copias o de copia única mediante PRINS y C-PRINS (<i>Cycling-Primed in situ labelling</i>).....	147
3.7.2.1 Localización de secuencias microsatélites.....	147
3.7.2.2 Localización de secuencias de RGAs (Análogos de genes de resistencia)	150
4. DISCUSIÓN.....	151

4.1 Población de mapeo.....	151
4.2 Desarrollo de marcadores moleculares funcionales en girasol.....	152
4.3 Evaluación de diferentes estrategias de Genotipificación.....	155
4.4 Análisis de polimorfismo de marcadores neutros y funcionales.....	157
4.5 Mapa genético de ligamiento.....	159
4.6 Caracterización del complemento cromosómico de <i>H. annuus</i>	162
4.7 Localización de secuencias repetitivas, de bajo número de copias y de copia única mediante las técnicas de BAC-FISH, PRINS o C-PINS.....	164

5. CONCLUSIONES.....	169
-----------------------------	------------

6. REFERENCIAS BIBLIOGRÁFICAS.....	173
---	------------

7. ÍNDICE DE TABLAS

Tabla 2.1 Origen de las líneas parentales de girasol utilizadas en el mapeo genético de marcadores moleculares.....	60
Tabla 2.2 La tabla detalla la concentración final de los componentes usados para la extracción de ácidos nucleicos.....	62
Tabla 2.3. Posibles combinaciones de los di-y trinucleótidos.....	64
Tabla 2.4 Regiones amplificadas con similitud a retrotransposones. La tabla detalla el nombre, tamaño y la secuencia de los oligonucleótidos seleccionados.....	88
Tabla 2.5 Secuencia de los oligonucleótidos con similitud a genes de resistencia seleccionados para su utilización como sondas en hibridación <i>in situ</i>	92
Tabla 3.1 Frecuencia de SSR en la base de datos de EST de girasol.....	105
Tabla 3.2 Clasificación según Gene Ontology (GO) de los EST de girasol conteniendo microsatélites genotipificados en la población de mapeo.....	107
Tabla 3.3 Resultados obtenidos de la evaluación de 127 EST- SSR en cuatro líneas de girasol RHA266, PAC2, HA89 y RHA801.....	108
Tabla 3.4 Ejemplos de regiones candidatas seleccionadas para el mapeo de EST.....	110
Tabla 3.5 Resultado del análisis de 594 marcadores SSR en los parentales RHA266 y PAC2.....	114
Tabla 3.6 La tabla muestra los datos de la EC de seis marcadores que	

conforman un panel	119
Tabla 3.7 Distribución y densidad de los marcadores en cada grupo de ligamiento.....	128
Tabla 3.8 Nombre, tipo y motivo de la repetición del marcador microsatélite incluido en el clon de BAC.....	142
Tabla 3.9 La tabla muestra tamaño, tipo de repetición, asociación a QTLs y el grupo de ligamiento al cual están anclados los microsatélites seleccionados para poner a punto las técnicas de PRINS y C-PRINS.....	147
Tabla anexa Información completa de los marcadores SSR, EST-SSR e EST-InDeL incluidos en la matriz de datos.....	193

8. ÍNDICE DE FIGURAS

Figura 1.1 Generación de líneas recombinantes endocriadas (RILs).....	40
Figura 1.2 Método para diseñar overgos.....	54
Figura 2.1 Fragmentos del estándar de peso molecular.....	71
Figura 2.2 Ejemplificación del método de marcado indirecto.....	73
Figura 2.3 Parámetros de análisis del proyecto utilizando el programa GenMapper [®] 4.0 (Applied Biosystem, EUA).....	75
Figura 2.4 Alelos indicados con los bins y el rango en pb del marcador seleccionado	76
Figura 2.5 Visualización simultánea de dos genotipos en el Genotype Plot.....	77
Figura 2.6 Secuencia completa del clon EF235. En rojo se indica la secuencia amplificada utilizada como sonda.....	83
Figura 2.7 Secuencia completa Ha785.....	84
Figura 2.8 Secuencia completa del retrotransposón de girasol perteneciente a la familia <i>Gypsy-like</i>	87
Figura 2.9 Secuencia completa del retrotransposón de girasol perteneciente a la familia <i>copla-like</i>	88
Figura 2.10 Distribución de los arreglos por duplicado del total de los clones que componen la genoteca.....	89
Figura 3.1 Multiplicación de las 124 líneas recombinantes endocriadas en las instalaciones del Instituto de Biotecnología CICVyA INTA Castelar.....	103
Figura 3.2 Distribución de los diferentes tipos de repetición de EST-SSR presentes en unigenes de girasol.....	105
Figura 3.3 Distribución de la frecuencia de los motivos de los EST-SSR presentes en unigenes de girasol.....	106

Figura 3.4 Patrón de amplificación de 7 marcadores EST-SSR sobre distintos genotipos de girasol.....	109
Figura 3.5a Vista de la pantalla de resultados del programa SNP2CAPS para el alineamiento de las secuencias correspondientes al fragmento EF421. 3.5b Gel de poliacrilamida 6% desnaturalizante, teñido con SYBR [®] Gold (Molecular Probes, EUA) mostrando el polimorfismo entre los parentales de la población de mapeo RHA266 x PAC2 de los marcadores F205 y F234.....	112
Figura 3.6 Identificación de polimorfismos de SSRs genómicos en los genotipos RHA266 y PAC2.....	115
Figura 3.7 Fotografía de un gel de poliacrilamida 6% teñido con nitrato de plata mostrando la segregación de los alelos del marcador microsatélite ORS510 en los parentales (RHA266 y PAC2) y en 28 líneas recombinantes de la progenie.....	116
Figura 3.8 Fotografía de un gel de poliacrilamida 6% mostrando la segregación del marcador microsatélites HA3204 utilizando métodos radioactivos (γ -P ³²) en los parentales de la población de mapeo (RHA266 y PAC2) y en 46 RILs.....	116
Figura 3.9 Fotografía de un gel de agarosa 1,5% mostrando el producto de amplificación de los marcadores microsatélites. 3.9a HA2605 y HA503 utilizando los fluoróforos HEX y 6-FAM. 3.9b HA3886, ORS432 y HA3319 utilizando el fluoróforo NED.....	117
Figura 3.10a y b Electroferogramas del marcador microsatélite ORS297.....	118
Figura 3.11 Co-electroforesis de seis marcadores microsatélites que conforman un panel....	120
Figura 3.12 Visualización simultánea de tres genotipos en el Genotype Plot.....	121
Figura 3.13 Fotografía de un gel de agarosa 1,5% mostrando el producto de amplificación de los marcadores TC18194 (6-FAM) y TC22997 (HEX).....	122
Figura 3.14 Visualización simultánea correspondiente a la amplificación de tres líneas para el marcador TC11140.....	123
Figura 3.15 Mapa de ligamiento de SSRs, EST-SSR, EST-InDels y AFLPs sobre una población de líneas recombinantes del cruzamiento de RHA266 x PAC2.....	129
Figura 3.16 Metafases mitóticas hibridadas con sonda EF235 correspondiente a la región ribosomal 26S.....	134
Figura 3.17 Metafase mitótica: hibridación utilizando el clon BAC772.....	136
Figura 3.18 Metafases mitóticas de la línea HA89 de girasol.....	137
Figuras 3.19a y b Cariograma de girasol caracterizado a partir de la célula de la figura 3.16. 3.19c Idiograma correspondiente al cariograma.....	138
Figura 3.20 BAC- <i>Southern blot</i> del clon repetitivo BAC772 utilizando como sondas	

secuencias repetitivas de girasol.....	139
Figura 3.21 Metafase mitótica de hibridación mediante PRINS utilizando el oligonucleótido Retro2.....	140
Figura 3.22 Autorradiografía de un filtro de BAC de alta densidad (Clemson, CUGI, HA_HBa), hibridado con un grupo de diez sondas overgo.....	142
Figura 3.23 Prometafase mitótica: se observan señales de localización dispersa en ensayo de BAC-FISH utilizando como sonda el clon de BAC3.....	143
Figura 3.24 BAC- <i>Southern blot</i> de los 6 clones BAC seleccionados.....	144
Figura 3.25 Metafase mitótica: las flechas indican señales de hibridación en cromosomas homólogos en un ensayo de BAC-FISH utilizando como sonda el clon de BAC3 conteniendo al microsatélite HA4222 (GL 16).....	146
Figura 3.26 Metafase mitótica: las flechas indican dos señales teloméricas en cromosomas homólogos en un ensayo de BAC-FISH utilizando como sonda el clon de BAC6 conteniendo al microsatélite HA2600 (GL 10).....	146
Figura 3.27 Metafase mitótica: las flechas indican la localización del microsatélite ORS1247 usando C-PRINS.....	148
Figura 3.28 a: Metafase mitótica b: núcleo interfásico las flechas señalan la localización del microsatélite HA3103 usando C-PRINS.....	149

ABREVIATURAS Y SIGLAS

ADN	ácido desoxirribonucleico
ADNc	ácido desoxirribonucleico copia
AN	número de acceso en el Genbank
ARN	ácido ribonucleico
AFLP	Polimorfismos de longitud de fragmentos amplificados
BACs	Cromosomas Artificiales de Bacterias
BIBACs	Cromosomas Artificiales de Bacterias Binarios
BSA	Albúmina de suero bovino
CAPS	<i>Cleaved Amplified Polymorphic Sequence</i>
CC	<i>coiled coil</i>
cm	centímetros
CMS	esterilidad citoplasmática masculina
cM	centiMorgan
CTAB	Cetil-trimetil-bromuro de amonio
DAPI	4'6- diamidino-2-phenylindole
DHPLC	Cromatografía Líquida de alta Performance Desnaturalizante
dH₂O	agua destilada
dNTP	desoxirribonucleótido trifosfato
DO	Densidad Óptica
Dr	Doctor
DS	Desvío standard
EC	electroforesis capilar
EDTA	Etilen-diamino-tetra-acetato
EEA	Estación Experimental Agropecuaria
EST	<i>expressed sequence tags</i>
EST-SSR	microsatélites incluidos en las secuencias de ESTs
EtBr	bromuro de etidio
EU	unidades enzimáticas
EUA	Estados Unidos de América

Fig	figura
FISH	Hibridación <i>in situ</i> fluorescente
FITC	isotiocianato de fluoresceína
g	gramo
GL	grupo de ligamiento
h	hora
HAGI	<i>Helinathus annuus gene index</i>
InDels	<i>insertions-deletions</i>
INRA	<i>Institut National de la Recherche Agronomique (France)</i> .
INTA	Instituto Nacional de Tecnología Agropecuaria
kpb	kilo pares de bases
LD	desequilibrio de ligamiento
Lic.	Licenciado
LRR	repeticiones ricas en leucina o <i>leucine-rich repeats</i>
LZ	<i>leucine zipper</i>
mA	miliamper
ml	mililitro
mg	miligramo
min	minuto
M	molar
M	Morgan
mM	milimolar
NBS	sitio de unión a nucleótido o <i>nucleotide binding site</i>
N°	número
N-P-K	Nitógeno-Fósforo-Potasio
ng	nanogramo
nt	nucleótido
overgos	<i>overlapping oligos</i>
pb	pares de bases
PCR	Reacción en cadena de la polimerasa
PHC	Podredumbre húmeda del capítulo
pmol	pico mol

ppm	partes por millón
PVP	Polivinilpolipirrolidona
QTL	<i>loci</i> de características cuantitativas
RAPD	ADN Polimórfico amplificado al Azar
RFLP	Polimorfismos en la Longitud de los Fragmentos de Restricción
RFU	rango de fluorescencia
RGA	Análogos de Genes de Resistencia
RU	Reino Unido
SCAR	<i>Sequence-Characterized Amplified Region</i>
SDS	Dodecil-sulfato-sódico
SNP	Polimorfismo de nucleótido simple (<i>Single Nucleotide Polymorphisms</i>).
SSH-EST	hibridación sustractiva de secuencias
SSR	<i>Simple Sequence Repeats</i> . Secuencias Simples Repetidas o microsatélites
STS	<i>Sequence Tagged Site</i>
rpm	revoluciones por minuto
seg	segundo
TAE	<i>buffer</i> de electroforesis TRIS-acetato/EDTA
TBE	<i>buffer</i> de electroforesis TRIS-borato/EDTA
TIR	<i>Toll receptor interleukin</i>
USDA	<i>United States Department of Agriculture</i>
μl	microlitro
UV	ultravioleta
V	volt

1. INTRODUCCIÓN

1.1. Generalidades del girasol

1.1.1 Descripción taxonómica

El girasol cultivado (*Helianthus annuus* L.) es una especie diploide, perteneciente a la familia Asteraceae (*Compositae*), subfamilia *Asteriodae*, tribu *Helianthae*, subtribu *Helianthinae*, género *Helianthus* (Seiler & Rieseberg, 1997).

La familia *Compositae* es una de las familias más grandes y diversas de todas las plantas con flor, conteniendo un décimo de todas las especies de angiospermas. Se caracteriza por tener una flor compuesta (inflorescencia), de la cual deriva su nombre. La familia probablemente se originó en América del Norte hace 30 a 100 millones de años (Raven & Axelrod, 1974) y sufrió una rápida y extensiva diversificación produciendo un gran espectro de taxones.

La mayoría de las especies de la familia *Compositae* analizadas hasta la fecha tienen genomas más grandes que 1,1 picogramos de ADN /1C (Kesseli & Michelmore, 1997) con excepción de algunas especies como *Leontodon taraxacoi* cuyo genoma es de 0.3 pg (<http://compgenomics.ucdavis.edu/cwp/draft.php>).

1.1.2 Importancia económica

La familia *Compositae* contiene más de cuarenta especies económicamente importantes para la vida del hombre, entre ellas: lechuga (*Lactuca sativa* L.; alimentación), girasol (*Helianthus annuus* L.; alimentación y ornamental), topinambur (*Helianthus tuberosus* L., alimentación, forrajero, prebiótico), cártamo (*Carthamus tinctorius* L., alimentación, uso industrial), camomila (medicinal), crisantemo, dalia, margarita y *Zinnia* (ornamentales), entre otras.

Particularmente del girasol se obtiene un aceite comestible de alta calidad con bajo contenido de ácidos grasos saturados y alto contenido de ácidos grasos mono y di insaturados. Existen tres calidades de aceite de girasol, dependiendo de su composición, algunas de las cuales no necesitan hidrogenación, característica que los hace ideales

para ser usados en frituras (aceite medio oleico y alto oleico) y para elaborar comida para bebé, atendiendo a su estabilidad durante los procesos de elaboración.

El aceite de girasol es considerado de gran valor debido a su composición de ácidos grasos insaturados (oleico y linoleico) y su bajo contenido en ácido linolénico. Es usado además, en jabones y detergentes en Europa del Este, como “transportador” en agroquímicos, particularmente pesticidas hidrofóbicos, surfactantes, adhesivos, plásticos, suavizantes y lubricantes. Además el aceite de girasol es usado en masajes, aromaterapia y tiene un gran potencial como insumo en la producción de biodisel.

1.1.3 El girasol cultivado (*Helianthus annuus* L.)

El nombre científico del girasol es *Helianthus annuus* L., *Helianthus* deriva de dos palabras griegas *helios* que significa sol y *anthos* que significa flor. Es una especie diploide ($2x = 2n = 34$), postulándose un número básico $x = 17$. Se propone que $x = 17$ sería un número básico derivado, siendo *Vigueira* ($x = 8$) y otro género desconocido ($x = 9$) sus probables progenitores (Blake, 1918). Posee un genoma haploide medio de aproximadamente 3.500 Mpb (Baack et al., 2005). El centro de diversidad de dicha especie es América del Norte (Heiser, 1976; Heiser & Smith, 1955). El género *Helianthus* está compuesto por treinta y seis especies perennes y doce especies anuales, estas especies son muy variables en cuanto a morfología y adaptaciones ecológicas, algunas especies silvestres de *Helianthus* han sido importantes fuentes de recursos genéticos para la identificación de genes de resistencia a enfermedades, tolerancia a sequía, esterilidad citoplasmática masculina (CMS) entre otras características agronómicas importantes (Jan et al., 1998).

El girasol es naturalmente una especie alógama y auto-incompatible, su domesticación ha permitido la autocompatibilidad y un aumento en el tamaño y en el vigor de la semilla. El girasol cultivado puede ser cruzado sin dificultad, tiene una alta producción de semillas (aproximadamente 1.000 semillas/planta) y un tiempo generacional corto de 70 a 100 días. Esto hace que el girasol cultivado tenga varias características que facilitan su análisis genético, a su vez el desarrollo de germoplasma autopolinizable ha favorecido el avance de los programas de mejoramiento incluyendo el desarrollo de híbridos.

El desarrollo del girasol cultivado por cruzamientos interespecíficos (*H. annuus* x *H. petiolaris*) con incorporación del citoplasma androesteril (CMS) de *H. petiolaris* y los genes restauradores de la fertilidad de *H. annuus* abrieron el camino de la industria de la semilla híbrida de girasol (Fick & Zimmer, 1976; Kinman, 1970; Leclercq, 1969). La identificación de nuevos genes restaurados de especies silvestres como *Helianthus maximiliani* L, recientemente reportados, abre nuevas posibilidades de aplicación al mejoramiento de esta especie (Fen & Jann, 2008).

1.1.4 Girasol en Argentina

El cultivo de girasol es el segundo en importancia dentro del mercado Argentino de producción de semillas oleaginosas. Argentina es el tercer productor mundial de girasol, luego de Rusia y Ucrania y el primer exportador mundial de aceite de girasol.

La producción de girasol ha venido recuperando protagonismo en los últimos años a un ritmo de 6,2% anual acumulativo, hasta la campaña 2005/06. Durante la campaña 2006/07 se observó un incremento del 8%, pero la producción fue afectada seriamente por el clima. La producción de girasol de la campaña 2007/2008 fue de 4,51 millones de toneladas, casi un millón de toneladas más que en la anterior. En las 2,64 millones de hectáreas cosechadas (97,8% de las sembradas), el rinde promedio nacional se ubicó en 17,1 quintales por hectárea (Asociación Argentina de Girasol, <http://www.asagir.org.ar/>).

La importancia del cultivo para la Argentina cobra mayor relevancia si se considera la producción a nivel internacional. De acuerdo a los datos del Departamento de Agricultura de Estados Unidos (USDA), para la campaña 2007/2008 (hemisferio norte) la producción mundial de girasol cayó de 29,8 millones a 27 millones de toneladas (9,3%). Los tres grandes oferentes del hemisferio norte (Rusia, Unión Europea y Ucrania) produjeron en conjunto 3,9 millones de toneladas menos.

Por su parte, las producciones de aceite y harina cayeron 7,7 y 8,7%, respectivamente. Sin embargo, el organismo estadounidense estima un crecimiento del 16,5% en la producción para la campaña 2008/2009, con una recuperación importante en la oferta de los tres principales competidores y un aumento de 6% en la producción argentina (www.ellitoral.com/index.php/diarios/2008/06/28/laregion/REG-15.html).

Dada la relevancia del girasol y el desarrollo agroindustrial derivado del mismo para la economía del país, es fundamental que la Argentina promueva este cultivo definiendo una política agrícola robusta e independiente y desarrollando un importante paquete tecnológico que favorezca la competitividad a nivel mundial. En este contexto considerando los problemas tecnológicos y la escasez de herramientas disponibles para los planes de mejoramiento asistido al momento del inicio de la presente tesis, se consideró la construcción de un mapa genético de ligamiento integrando los mapas existentes, incluyendo los marcadores disponibles para la especie así como el desarrollo de nuevos marcadores y su correlación a un mapa físico.

1.1.5 Estado de avance del conocimiento sobre la genómica de girasol

Pese a la importancia económica del girasol a escala mundial, el grado de avance de los conocimientos públicos sobre el genoma de girasol es limitado cuando se lo compara con cultivos como el arroz, el maíz, la soja, el tomate, el trigo, la papa y la cebada, entre otros. Sin embargo, en los últimos diez años, se ha producido un avance interesante en relación a los estudios genómicos y post-genómicos aplicados al girasol, lo cual ha incrementado significativamente la disponibilidad de herramientas moleculares, tales como microsatélites o SSR (*Simple Sequence Repeats*) (Paniego et al., 2002; Tang et al., 2003, Tang et al., 2002), EST (*Expressed Sequence Tags*) (Fernández et al., 2003; Gentzittel et al., 1999; <http://cgpdb.ucdavis.edu/>), mapas genéticos de distinto nivel de saturación (Al-Chaarani et al., 2004; Kiani et al., 2007a; Lai et al., 2005; Tang et al., 2003, Tang et al., 2002, Yu et al., 2003), colecciones genómicas de clones de BACs (*Bacterial Artificial Chromosome*) (Bouzidi et al., 2006; Feng et al., 2006; Gentzittel et al., 2002; Özdemir et al., 2004) y microarreglos de ADN (Alignan et al., 2006; Hewezi et al., 2006; Fernández et al., 2008; Lee & Mitchell-Olds, 2006). Estos desarrollos, junto con el advenimiento de técnicas de citogenética molecular, han impactado positivamente sobre el conocimiento de esta especie a nivel genómico. Sin embargo, al momento de inicio del presente trabajo, tanto el desarrollo de un mapa genético de ligamiento de referencia que integrara los mapas construidos independientemente, por distintos grupos a nivel internacional, así como el desarrollo de mapas físicos, permitiendo correlacionar grupos de ligamiento con regiones cromosómicas específicas a través de la utilización de las colecciones genómicas en

BACs y del desarrollo de técnicas de citogenética molecular no habían sido ampliamente desarrollados. El desarrollo de un mapa consenso que unifique los mapas publicados de manera independiente por distintos autores constituye una herramienta fundamental para estudios de genómica aplicada. Un mapa de referencia saturado permite orientar la búsqueda de regiones candidatas para la localización de polimorfismos asociados a caracteres agronómicos de interés, los cuales permiten abordar estudios de asociación genotipo-fenotipo y conducir a la localización génica de QTLs que gobiernan un determinado carácter, con su consiguiente aplicación en programas de mejoramiento a través de selección asistida por marcadores (MAS).

1.2 Herramientas moleculares para el análisis genómico

En los últimos veinte años se ha desarrollado una serie de técnicas moleculares para abordar el análisis genómico, lo cual permite explorar la información genética que los organismos poseen. El análisis genómico para la identificación de los genes implicados en un proceso biológico es esencial para entender la biología de una célula o de un organismo en su conjunto. Este tipo de análisis, que ha supuesto una auténtica revolución en el abordaje de los problemas biológicos, conlleva un cambio radical en las aproximaciones experimentales y requiere el desarrollo de nuevas tecnologías de alta eficiencia para el análisis global de los genomas, así como de herramientas bioinformáticas que permitan el procesamiento de los datos, la generación de información y la extracción del conocimiento. Entre estas herramientas, particularmente los marcadores moleculares, funcionan como señaladores de diferentes regiones del genoma, permitiendo identificar un polimorfismo directamente en el ADN del individuo, y asociarlo a un determinado fenotipo resultante de caracteres gobernados por bases genéticas con distinto grado de complejidad.

1.2.1 Marcadores moleculares

El término “marcador” deriva de la utilización de marcas genéticas que pueden estar asociadas a un gen de interés. Sin embargo, los marcadores moleculares no necesariamente representan un gen sino que funcionan como etiquetas para señalar una posición determinada sobre el genoma, dado que ocupan posiciones genómicas

específicas dentro de los cromosomas llamadas *locus*. Se denomina marcador a todo *locus* que sea susceptible de ser mapeado. Un marcador permite identificar la presencia de por ejemplo, genes de resistencia a cierta enfermedad sin necesidad de evaluar la respuesta a un patógeno, o bien de un carácter que sólo se exprese en fase adulta o de difícil o costosa identificación.

Los marcadores moleculares permiten detectar variaciones (polimorfismos) en la secuencia de ADN entre dos individuos. Estas variaciones pueden modificar o no su fenotipo y se originan a partir de diferentes clases de mutaciones en el ADN, tales como sustituciones, rearrreglos (inserciones y deleciones) o errores en la replicación de las repeticiones en tándem del ADN.

Cuando un marcador polimórfico puede discriminar entre un genotipo homocigoto y uno heterocigoto es llamado marcador codominante, en cambio, se llama marcador dominante a aquel que es incapaz de diferenciar un genotipo heterocigoto. Esto es debido a que el polimorfismo en un marcador codominante es causado por una diferencia en el tamaño de los fragmentos o bandas visualizadas en un gel, mientras que un marcador dominante sólo aparece como presencia o ausencia de la banda. Las diferentes formas de un marcador, visualizadas como bandas en un gel, son consideradas alelos de ese marcador, por lo cual, mientras un marcador codominante puede tener múltiples alelos, en uno dominante sólo existen dos alelos posibles.

Las características deseables de un buen marcador son: la detección fácil y rápida de todas las clases genotípicas posibles, poseer alta reproducibilidad de detección, bajo requerimiento de material biológico para su identificación, distribución homogénea a lo largo del genoma (en el caso de utilizarlos para la construcción de mapas) y un alto nivel de polimorfismos (Cubero, 2003).

En la actualidad, los marcadores más utilizados son los marcadores moleculares basados en PCR. Se los denomina marcadores de PCR porque se basan en la reacción en cadena de la polimerasa o PCR. Estos marcadores son los más ampliamente estudiados debido principalmente a su abundancia.

Los diferentes tipos de marcadores moleculares pueden ser clasificados en tres categorías: a) marcadores de primera generación, que incluyen a los Polimorfismos de Longitud de Fragmentos de Restricción o RFLPs (*Restriction Fragment Length*

Polymorphisms), Polimorfismos Amplificados al Azar o RAPDs (*Random Amplified Polymorphic DNA*) y sus modificaciones; b) marcadores de segunda generación, que incluyen microsatélites o SSR (*Simple Sequence Repeats*), Polimorfismos de Longitud de Fragmento Amplificados o AFLPs (*Amplified Fragment Length Polymorphic*) y sus modificaciones y c) marcadores de tercera generación, que incluyen a los ESTs (*Expressed Sequence Tags*) y los Polimorfismo de Nucleótido Simple o SNP (*Single Nucleotide Polymorphisms*). Esta categoría fue propuesta por Gupta et al. (2001).

Los marcadores moleculares son usados en programas de mejoramiento vegetal (desarrollo de mapas genéticos, identificación de QTLs y estudios de mapeo comparativo), en programas de selección asistida por marcadores (selección indirecta, introgresión asistida por marcadores, desarrollo de líneas puras y mejora de caracteres cuantitativos). Además, son usados en la conservación y uso de recursos genéticos, en la identificación de variedades y en el desarrollo de plantas transgénicas (Nuez & Carrillo, 2002).

Para llevar adelante estudios de genómica aplicada y conducir programas de mejoramiento moderno es fundamental la disponibilidad de distintos tipos de marcadores moleculares para la especie en estudio, como así también conocer su localización en el genoma mediante su incorporación en mapas genéticos.

1.2.1.1 Microsatélites

Los microsatélites o SSR (*Simple Sequence Repeats*) son repeticiones en tándem de 1 a 5 nucleótidos distribuidas a lo largo del genoma. Poseen una región repetitiva corta, generalmente menor a 100 pb, y en su mayoría pueden ser amplificados por PCR estándar. Los microsatélites ganan y pierden unidades de repetición por deslizamiento de la ADN polimerasa. Este mecanismo mutacional es complejo y específico de secuencias repetidas en tándem (Ellegren, 2000; Schlötterer, 2000). Los SSRs, son altamente polimórficos, abundantes, codominantes y se encuentran distribuidos en las regiones eucromáticas del genoma, por lo que no tienen efecto de presión de selección y, por este motivo se consideran marcadores neutros. En la mayoría de los casos son cromosomas específicos, multi-alélicos, altamente polimórficos y poseen alta capacidad de procesamiento durante la genotipificación.

Estas características favorecen su utilización en la construcción de mapas de ligamiento saturados y en la selección asistida por marcadores, en los estudios de paternidad, en los análisis forenses y en los estudios de genética de poblaciones (Goldstein & Schlötterer, 1999; Schlötterer, 2004).

Las desventajas más destacables están relacionadas con el proceso de desarrollo el cual es costoso, laborioso y demandante en tiempo, y a la inestabilidad en el tiempo de algunos *loci*. Además, problemas técnicos tales como artefactos de la PCR (*stutter bands*) complican la automatización del análisis. Finalmente, a pesar de la gran cantidad de microsatélites presentes en la mayoría de los genomas eucariotas, se ha reportado que la densidad de *loci* informativos podría ser baja para algunas aplicaciones de mapeo genético (Dieringer & Schlötterer, 2003; Tóth et al., 2000).

En cuanto al girasol, distintas iniciativas internacionales dieron lugar al desarrollo de este tipo de marcadores. La Argentina a través de INTA generó la colección denominada HAx; Europa, a través del grupo Cartisol, generó la colección denominada CRSx y Estados Unidos, a través de la Universidad de Oregon, la colección denominada ORSx. En la actualidad se dispone de 2.040 marcadores microsatélites para girasol, (Dehemer & Friedt, 1998; Gedil et al., 1999; Paniego et al., 2002; Tang et al., 2002; Yu et al., 2003), algunos de los cuales han sido integrados a distintos mapas genéticos (Al-Chaarani et al., 2003; 2004; Tang et al., 2002; Yu et al., 2003).

1.2.1.2 Marcadores funcionales

En los últimos años se ha desarrollado una extensa variedad de marcadores moleculares funcionales, considerados así porque corresponden a regiones del genoma que se expresan (Gupta et al., 2002; Phillips & Vasil, 2001). Esto fue posible, en primer lugar, debido a la generación de un gran número de secuencias que se expresan (ESTs) para especies vegetales en bases de datos públicas. Las secuencias ESTs son generadas tomando clones al azar de una colección de ADNc y secuenciando las mismas en una lectura con una extensión promedio de 300-500 bases. Hasta el momento, se han desarrollado colecciones de EST para más de ciento cuarenta especies de plantas económicamente importantes, incluyendo al girasol (<http://www.ncbi.nih.gov>; <http://www.tigr.org>).

Aproximadamente el 1% de todas las secuencias de ESTs depositadas en el NCBI pertenecen a la familia de las Compuestas, lo que las convierte en una de las familias de plantas mejor representadas. Sin embargo, los proyectos de *Arabidopsis*, arroz, maíz, tomate y soja aún reúnen más del 50% del total de las entradas.

En la base de datos GenBank existen, hasta diciembre de 2007, 283.749 secuencias de EST para el género *Helianthus*, depositadas mayoritariamente por la iniciativa Proyecto Genómico de las Compuestas o CGP *Compositae Genome Project* (<http://cgpdb.ucdavis.edu>). De estas secuencias, 93.278 corresponden a la especie *H. annuus* y representan 36.743 secuencias únicas (liberadas el 17 de junio de 2006) (Índice Genes de girasol de TIGR, HAGI, <http://www.dfci.harvard.edu/>). Otras contribuciones al banco de ESTs de girasol la han aportado iniciativas menores conducidas por grupos de investigación franceses y argentinos, pertenecientes a Genoplante y ENSAT (Francia) y al INTA (Argentina). El grupo del INTA llevó a cabo la caracterización de una colección de ESTs órgano-específico aislado de colecciones de ADN copia usando la tecnología de hibridación sustractiva de secuencias (SSH-ESTs) (Fernández et al., 2003). Esta colección sustentó la búsqueda e identificación de genes candidatos asociados a respuestas a estreses abióticos mediante la estrategia de análisis concertado de expresión génica a través de la tecnología de microarreglos de ADN (Fernández et al., 2008), como así también la identificación de marcadores funcionales descripta en el desarrollo de esta tesis.

El análisis las secuencias de EST permite la identificación de mutaciones puntuales del tipo SNP, INDELS y motivos de repetición del tipo SSRs, denominados EST-SSR.

Los marcadores EST-SSR han sido desarrollados a partir de una gran variedad de plantas, tales como *Arabidopsis*, maíz, soja, arroz, trigo y más recientemente girasol (Cardley et al., 2000; Eujayl et al. 2002; Gao et al., 2003; Hackauf & Wehling 2002; Heesack et al., 2008; Kantety et al., 2002; Kumpatla & Mukhopadhyay, 2005; Morgante et al., 2002; Nicot et al., 2004; Pashley et al., 2006; Thiel et al., 2003; Tóth et al., 2000; Varshney et al., 2002). Ellos han sido usados frecuentemente en mapeo genético, mapeo comparativo y análisis de diversidad genética y funcional (Varshney et al., 2005b). Si bien en varias especies se ha reportado que estos marcadores son menos polimórficos que los SSR genómicos (Becker & Heun., 1996; Cho et al., 2000; Eujayl et al., 2000; Lee et al., 2004), tienen varias ventajas, entre las cuales podemos mencionar: su

desarrollo es relativamente económico gracias a la disponibilidad de secuencias de ESTs en base de datos públicas y la frecuencia estimada de SSR es mayor en las regiones génicas, fundamentalmente intergénicas que en las regiones heterocromáticas de ADN (Morgante et al., 2002). Los marcadores EST-SSR son marcadores funcionales y tienen mayor probabilidad de poder ser transferidos entre especies, ya que derivan de regiones del genoma relativamente más conservadas (Varshney et al., 2005a).

La frecuencia de SSR descrita de manera general en las secuencias de ESTs es de 2 a 5% (Kantety et al., 2002; Varshney et al., 2005b) sin embargo, se observan inconsistencias en las estimaciones de frecuencias descritas en diferentes estudios, incluso para una misma especie, lo cual se relaciona con una multiplicidad de factores, incluyendo las herramientas y criterios utilizados para la identificación de repeticiones en tándem y el tamaño del conjunto de datos analizados. Los criterios de búsqueda utilizados, tales como repeticiones perfectas e imperfectas, número mínimo de repeticiones, longitud de los espacios para las repeticiones compuestas, etc., a menudo llevan a diferencias significativas en el número de SSRs identificados en una especie dada y entre diferentes estudios usando el mismo conjunto de datos.

La estructura y la longitud de los microsatélites son los principales factores que determinan su variabilidad (Brinkmann et al., 1998; McMurray, 1995). En general, los *loci* con más repeticiones tienden a ser más polimórficos (Cho et al., 2000; Goldstein & Clark 1995; Innan et al., 1997; La Rota et al., 2005; Schug et al., 1998; Temnykh et al., 2001; Weber 1990). En algunas especies se ha demostrado que las repeticiones de dinucleótidos mutan más frecuentemente que las de tri y tetra nucleótidos (Schug et al., 1998). Sin embargo, en humanos, fueron observados porcentajes similares de mutación entre di-, tri- y tetranucleótidos (Chakraborty et al., 1997). En este contexto es frecuente que las búsquedas de EST-SSR se limiten a motivos de repetición que se expandan por más de dieciocho bases teniendo en cuenta que en muchas especies el nivel de polimorfismo disminuye con SSR más cortos que esa longitud (Cho et al., 2000).

En general, los diferentes motivos de repetición se presentan en frecuencias similares en los ESTs de diferentes especies. Entre los motivos dinucleótidos, el motivo GA/CT ha sido reportado como el más abundante, y el GC/CG como el menos abundante en trigo, cebada, maíz, sorgo y arroz (Kantety et al., 2002, Temnykh et al., 2000). El motivo trinucleótido más abundante en cebada, maíz, sorgo y arroz fue

GGC/CCG, y en trigo fue AAC/GTT. Morgante et al. (2002) y Vershney et al. (2005) demostraron que el motivo dinucleótido más frecuente en ESTs de plantas fue AG.

Kumpatla & Mukhopadhyay, (2005) analizaron EST-SSR en veinte dicotiledóneas incluyendo el girasol. En dicho estudio, el criterio para la identificación de SSR fue poco astringente, e incluyeron motivos de mononucleótidos y dinucleótidos de más de cinco repeticiones. Los mononucleótidos fueron informados como los motivos más frecuentes, seguidos por los dinucleótidos y los trinucleótidos. El motivo dinucleótido más frecuente fue el AG/CT en todas las especies investigadas, el motivo trinucleótido más frecuente fue el AAG/CTT en dieciséis de las veinte especies estudiadas (no incluye al girasol). Estas características representan una tendencia general en los EST-SSR de dicotiledóneas, mientras que el motivo CCG/CGG es el más frecuente en monocotiledóneas (Temnykh et al., 2001). Según lo reportado por estos autores en girasol el más frecuente fue ACC/GGT.

Pashley et al. (2006) desarrollaron marcadores EST-SSR, los cuales fueron utilizados para el análisis genético de especies poco frecuentes de girasol silvestre. Estos autores seleccionaron di-, tri y tetranucleótidos con una repetición mínima de 5, 4 y 3 subunidades, respectivamente. El tipo de repetición más abundante fue el de trinucleótidos, seguido por el de tetranucleótidos y luego el de dinucleótidos, siendo ATG/TGA el motivo trinucleótido más abundante.

Recientemente, Heesacker et al. (2008) identificaron SSR e InDels a partir de la base de datos de ESTs de girasol y realizaron estudios de transferibilidad entre distintas especies de Compuestas (*Helianthus*, *Carthamus* y *Lactuca*). Para tal fin fueron utilizadas dieciséis accesiones tanto de germoplasma silvestre como de líneas elite, algunas de las cuales son parentales de poblaciones de mapeo. En el citado trabajo se analizó una colección de 17.904 genes únicos obtenidos a partir del ensamblado de 89.225 ESTs provenientes de *H. annuus*, *H. argophyllus* y *H. paradoxus* en búsqueda de motivos de SSR. El criterio de búsqueda utilizado se basó en un número mínimo de repeticiones (n) considerando los valores de $n \geq 4$ y $n \geq 5$. El número total de SSRs encontrado fue de 16.643 y 4.755 SSR, respectivamente; en ambos casos, el motivo más frecuente en la colección fue el de dinucleótidos, seguido por el de trinucleótidos y por último el de tetranucleótidos. Se diseñaron oligonucleótidos para una selección de 484 EST-SSRs, de los cuales 427 produjeron perfiles de alta calidad para el genotipado en

H. annuus. Para estos marcadores se analizó la capacidad de transferencia a nivel de especie y de género mediante la genotipificación en distintas accesiones de *Helianthus* silvestres, *Carthamus* y *Lactuca*, observándose por un lado, que la transferencia es alta dentro del género *Helianthus* ($\geq 82\%$) lo que facilita el mapeo comparativo y el mapeo a partir de poblaciones derivadas de cruzamientos interespecíficos y, por otro, que es bajo el potencial de transferencia a nivel de otros miembros de la familia *Asteraceae* ($\leq 16\%$). Finalmente, de los EST-SSR desarrollados en este estudio, el 55,1% fueron polimórficos en poblaciones derivadas de cruzamientos entre líneas elite, mientras que el 83,3% fueron polimórficos en poblaciones derivadas de cruzamientos entre líneas elite y germoplasma silvestre, concluyendo que aproximadamente un 10% de los *loci* que se transcriben puede ser mapeados genéticamente usando EST-SSRs.

Los marcadores SNP (Single Nucleotide Polymorphisms) representan el polimorfismo más abundante en la mayoría de los organismos y tiene su origen en un cambio de base o pequeñas inserciones o deleciones en una posición particular de una secuencia dada.

En general, son marcadores bialélicos, lo que representa una desventaja respecto a otros marcadores que son multialélicos como los SSR, pero tienen como ventaja la abundancia, la estabilidad y la posibilidad de ser analizados por técnicas de genotipificación de alto desempeño a un costo relativamente bajo.

La información sobre la frecuencia, naturaleza y distribución de los SNPs en plantas superiores está en progresivo aumento y, en consecuencia, este tipo de marcador representa un importante sistema para mejoradores y genetistas.

El desarrollo de los marcadores SNP consiste usualmente de tres partes, de manera cronológica: la resecuenciación de una región específica del genoma a partir de un pequeño grupo de individuos, el alineamiento múltiple de las secuencias obtenidas, la validación en un conjunto mayor (por ejemplo, remover los falsos positivos debido a errores de secuenciación o debido a la presencia de secuencias homólogas/parálogas) y la genotipificación en una población de mayor tamaño (Chagné et al., 2007).

Los ESTs representan una valiosa fuente para el descubrimiento de SNP, y han sido usados en remolacha (Schneider et al., 2001), maíz (Batley et al., 2003; Ching et al., 2002), arroz (Nasu et al., 2002), soja (Zhu et al., 2003), caña de azúcar (Grivet et al.,

2003), girasol (Fusari et al., 2008; Heesacker et al., 2008; Kolkman et al., 2004; 2007; Lai et al., 2005; Liu & Burke, 2006), entre otras especies de importancia agronómica.

La frecuencia de los SNPs varía en diferentes regiones dentro de cada genoma, siendo la distribución de los sitios polimórficos generalmente no al azar. Las regiones genómicas con bajo porcentaje de recombinación generalmente presentan reducido nivel de polimorfismo (Rafalski & Morgante, 2004). Los SNPs pueden presentarse en regiones codificantes y no codificantes sin embargo, en la mayoría de los organismos estudiados se ha observado mayor presencia de SNPs en regiones no codificantes del genoma (Edwards et al., 2007).

Los SNP son utilizados en estudios de diversidad nucleotídica, de filogenia, de evolución, de mapeo genético, de mapeo fino, de clonado posicional, de mapeo por asociación a nivel poblacional, y de selección asistida a partir de la identificación de variantes alélicas causativas (QTNs: *Quantitative Trait Nucleotides*), de las cuales deriva una estrategia de selección asistida directa basada en genes candidatos (Fusari et al., 2008; Koebner, 2004; Kwok, 2001; Rafalski, 2002).

En girasol se han descrito 605 polimorfismos de tipo SNPs caracterizados *in silico* a partir del análisis de secuencias de ESTs provenientes de los genotipos RHA280 y RHA801, validados y mapeados mediante el análisis de formación de heteroduplex (Lai et al., 2005). Recientemente, Heesacker et al. (2008) genotipificaron 39 InDels identificados a partir de la base de datos de ESTs de girasol y realizaron estudios de transferibilidad entre distintas especies de Compuestas (*Helianthus*, *Carthamus* y *Lactuca*).

La identificación de SNPs mediante la técnica de resecuenciación de genes se ha reportado en varios trabajos. En primer lugar, Kolkman et al. (2004), caracterizaron alelos de la enzima que confiere resistencia a imidazolinas y sulfonilureas en girasol, y desarrollaron un sistema de marcadores moleculares para identificar nuevas variantes alélicas y para conducir trabajos de selección asistida de genes de resistencia a herbicidas en girasol. Recientemente, y con el propósito de analizar la diversidad funcional presente en poblaciones de girasol silvestre y cultivado, Liu & Burke (2006) y Kolkman et al. (2007) caracterizaron en total 90 *loci*, detectando 1.522 SNP. Finalmente, Fusari et al. (2008) llevaron a cabo la caracterización de SNPs amplificando

las regiones codificantes y no codificantes correspondientes a 28 genes candidatos relacionados a respuestas a estrés sobre 19 accesiones de girasol cultivado. En este último trabajo se analizaron aproximadamente 14 kb utilizando programas públicos y/o disponibles en el laboratorio para definir los polimorfismos existentes entre genotipos, reconociéndose 69 SNPs informativos y 43 InDels. En promedio, esto representa 1 SNP cada 69 pb. Los valores de diversidad nucleotídica y LD (Desequilibrio de Ligamiento) estimados sobre el conjunto de líneas estudiadas fueron comparables a los publicados por Liu & Burke (2006) y Kolkman et al. (2007), a pesar de que en estos trabajos se aplicaron diferentes estrategias de muestreo y se analizaron diferentes genotipos. Estos trabajos describen que el desequilibrio de ligamiento medido en regiones acotadas del genoma decae hasta valores de 0,3 a las 5.500 pb, y concluyen que a partir de estos valores es factible conducir estudios de mapeo poblacional de caracteres agronómicos basado en estudios de asociación entre genotipos y fenotipos.

Los marcadores Análogos de Genes de Resistencia o RGAs (*Resistant Genes Analogs*) representan un gran número de secuencias que evidencian una estructura similar a los genes de resistencia (genes R) caracterizados en los genomas vegetales, las cuales han sido definidas a través de la amplificación de ADN genómico utilizando oligonucleótidos degenerados diseñados a partir de los dominios conservados de los genes R. Los dominios predominantes son el LRR (repeticiones ricas en leucina o *leucine-rich repeats*) y el NBS (sitio de unión a nucleótido o *nucleotide binding site*). Otros motivos menos representados en los genes R son el LZ (*leucine zipper*), TIR (*Toll receptor interleukin*) y CC (*coiled coil*), entre otros. La mayoría de los RGAs se organiza en *loci* complejos o *clusters* asociados a genes R conocidos, secuencias homólogas repetidas en forma de tándem, pseudogenes y/o retrotransposones.

Las familias multigénicas más abundantes son del tipo NBS-LRR. De los más de cuarenta genes R conocidos que confieren resistencia a bacteria, hongos, virus y nematodos patógenos de plantas, el 75% codifica para una proteína de tipo NBS-LRR (Kuang et al., 2004; McDowell et al., 1998; Meyers et al., 1998; Parniske et al., 1997; Radwan et al., 2008; Sun et al., 2001).

El girasol cultivado es susceptible a varias enfermedades fúngicas y virales económicamente importantes, dependiendo la severidad de las infecciones, y en gran medida de las condiciones ambientales.

Las enfermedades del girasol son muchas y variables según la región del país. Entre las enfermedades que causan mayores pérdidas económicas en nuestro país se encuentran las de origen fúngico: mildiu cuyo agente causal es *Plasmopara halstedii*, el “Marchitamiento” por *Verticillium dahliae* y la “Podredumbre húmeda del Capítulo” (PHC) causada por *Sclerotinia sclerotiorum*, esta última causa pérdidas del 10% al 20% sobre la producción anual de este cultivo.

Como se mencionó anteriormente, durante las últimas dos décadas el uso de marcadores moleculares de ADN ha sido utilizado exitosamente en el desarrollo de mapas genéticos en girasol, permitiendo la detección de *loci* que controlan características cuantitativas, incluyendo la localización de QTLs asociados con resistencia a mildiu (Radwan et al., 2003; 2004), fomopsis (Alfadil et al., 2007) y podredumbre húmeda del tallo y del capítulo (Gentzbittel et al., 1998; Micic et al., 2004; 2005a y b).

La resistencia a *Sclerotinia sclerotiorum* en girasol ha sido caracterizada como un carácter de base genética compleja y altamente dependiente de las condiciones ambientales. Se han detectado varios *loci* de carácter cuantitativo (QTL) en distintos genotipos, órganos vegetales y condiciones ambientales de crecimiento (Mestries et al., 1998, Gentzbittel et al., 1998, Bert et al., 2004). Estudios recientes han identificado varios QTLs asociados a la resistencia *S. Sclerotiorum* tanto en hojas de girasol, como en tallo y capítulo (Bert et al. 2002; 2004, Micic et al. 2004; 2005a y b). En Argentina se han localizado QTLs para resistencia a la PHC en líneas de buen comportamiento local los cuales han sido localizados fundamentalmente en el grupo de ligamiento 10 del mapa genético de girasol (Maringolo, 2007).

Los primeros estudios genéticos de resistencia a mildiu demostraron que genes dominantes mayores, llamados *Pl* controlan la resistencia a diferentes razas de *P. halstedii*. Hasta la actualidad, once genes *Pl* han sido descritos (Rahim et al., 2002) en especies cultivadas de girasol (Vranceanu & Stoenescu, 1970) y silvestres de *Helianthus* (Miller & Gulla 1991). *Pl*₆, *Pl*₇ y *Pl*₈, encontrados en especies silvestres de *Helianthus*, confieren resistencia a casi todas las razas de *P. halstedii* (Bert et al., 2001; Bouzidi et al., 2002). *Pl*₆ ha sido identificado en especies silvestres de *H. annuus*, mientras que *Pl*₇ proviene de *H. praecox* y *Pl*₈ de *H. argophyllus*. Veare et al. (1997) mostraron que el *locus Pl*₆ ha sido descrito como un *locus* complejo con al menos dos

regiones genéticamente distintas, las cuales confieren resistencia a diferentes razas de *Plasmopara*. El locus *Pl₅* confiere resistencia a un amplio rango de razas de mildiu, pero es susceptible a la raza 330, mientras que el locus *Pl₈* confiere resistencia a todas las razas conocidas de *P. halstedii*.

Los genes *Pl* de resistencia están agrupados en *clusters* sobre diferentes grupos de ligamiento. Al momento del inicio de la presente tesis, se habían descrito al menos tres *clusters* diferentes de genes *Pl* en girasol, considerando su localización y su estructura: los loci *Pl₁₁*, *Pl₁₂*, *Pl₁₆* y *Pl₁₇*, situados sobre el grupo de ligamiento 8 en el mapa genético propuesto por Tang et al. (2003), pertenecientes todos a las familias de los genes de resistencia de tipo TIR, NBS, LRR; el locus *Pl_{15/18}* situado en el grupo de ligamiento 13 del mapa genético de Tang et al. (2003) (Radwan et al., 2003), pertenecientes a la familia non-TIR, NBS, LRR y, además, fue detectada una nueva fuente de resistencia a *P. halstedii* proveniente de *H. argophyllus*, la cual ha sido mapeada en el grupo de ligamiento 1. Este último locus fue denominado *Pl_{Arg}* (Duble et al., 2004). Recientemente, Radwan et al. (2008) identificaron 14 *clusters* de RGAs, cuatro de los cuales habían sido previamente identificados (Gedil et al., 2001; Plocik et al., 2004; Radwan et al., 2003). De los 14 *clusters*, 8 pertenecen a la familia de genes de resistencia de tipo TIR, NBS, LRR y 6 pertenecen a la familia non-TIR, NBS, LRR. Los 14 *clusters* contienen 196 secuencias únicas de tipo NBS-LRR, de las cuales han sido mapeados 167 loci en los 17 grupos de ligamiento.

1.2.2 Métodos de genotipificación de los marcadores moleculares

Existe una gran cantidad de métodos para detectar y caracterizar los marcadores basados en la técnica PCR, como es el caso de los sistemas descritos en las secciones anteriores. La elección del método dependerá de la envergadura del proyecto para proyectos a pequeña escala, son útiles los procedimientos manuales donde los productos amplificados son separados por electroforesis en geles de agarosa o poliacrilamida, y revelados por medio de diferentes métodos de detección (bromuro de etidio, marcación radiactiva del oligonucleótido con γP^{32} y/o con oligonucleótidos fluorescentes). En estudios genéticos a gran escala es recomendable aplicar métodos de genotipificación de alta procesividad, basados en el uso de analizadores automáticos de ADN y

oligonucleótidos fluorescentes, los cuales son relativamente sencillos de implementar y ofrecen la posibilidad de mediano a alto procesamiento de muestras. Entre las ventajas de este método se destacan la alta precisión en la adjudicación de los alelos, la alta procesividad, la posibilidad de automatización del proceso y programas de análisis de datos fácilmente utilizables (Chen & Sullivan, 2003).

Al momento del inicio de este trabajo, se habían utilizado principalmente métodos semiautomáticos para la genotipificación de marcadores microsatélites en girasol, utilizando geles de poliacrilamida y amplicones fluorescentes con el secuenciador ABI Prism 377 DNA Sequencer (Applied Biosystem, Perkin Elmer, Foster City, CA) (Tang et al., 2002; 2003; Yu et al., 2003).

Las técnicas comúnmente usadas para el análisis de segregación de los marcadores SNP se basan en la detección de heteroduplex y son: SSCP (*Single-Stranded Conformational Polymorphism*), DHPLC (*Denaturing High-Performance Liquid Chromatography*), CAPs (*Cleaved Amplified Polymorphic Sequence*) y la técnica de resecuenciación de genes candidatos a partir de colecciones de ESTs o STSs (*Sequence-Tagged-Site*). Hasta el momento, en girasol ha sido utilizada la técnica de DHPLC para la genotipificación de marcadores SNP (Lai et al., 2005). La técnica de resecuenciación de genes ha sido usada para el desarrollo de SNP (Fusari et al., 2008; Kolkman et al., 2004; 2007; Liu & Burke, 2006).

1.2.2.1 Errores en el proceso de genotipificación. Causas y consecuencias

El conjunto ideal de datos de marcadores moleculares para análisis de ligamiento no debe contener datos perdidos ni errores de genotipificación, y los marcadores deben segregarse en la proporción esperada, lo cual dependerá del tipo de población en estudio. Sin embargo, en la práctica, cada conjunto de datos obtenido por genotipificación contiene algún tipo de error. Los errores de genotipificación son definidos como las diferencias observadas sobre un mismo genotipo caracterizado en dos o más análisis independientes (Bonin et al., 2004). En estudios de mapeo, porcentajes de errores del 1% pueden originar un orden incorrecto de los marcadores y aumentar la longitud del mapa genético (Buetow, 1991; Hackett & Broadfoot, 2003; Lincoln & Lander, 1992).

El objetivo de este apartado es examinar las causas y consecuencias de los errores en el proceso de genotipificación, para así reducir su ocurrencia e interpretar sus efectos sobre los resultados del análisis. Todos los tipos de marcadores moleculares son posibles blancos de este tipo de errores, pero sólo se hará referencia a los marcadores usados en la presente tesis: SSR, SNP e InDels.

Pompanón et al. (2005) clasificaron a los tipos de errores de Genotipificación en cuatro categorías:

- Errores ligados a las variaciones en la secuencia de ADN: este tipo de errores puede presentarse por fallas en la amplificación de un alelo determinado debido a sustituciones, inserciones o deleciones sobre la secuencia complementaria a uno de los oligonucleótidos utilizados en la reacción de PCR. Generalmente corresponden a mutaciones cercanas al extremo 3' del oligonucleótido. En el análisis de microsatélites el caso más común es el representado por los alelos nulos (Callen et al., 1993; Paetkau & Strobeck, 1995).

Por otro lado una inserción o deleción que esté cerca de un microsatélite puede generar homoplasia de tamaño (alelos del mismo tamaño, pero que no comparten un ancestro en común), lo que llevaría a la anotación de dos alelos diferentes como un solo alelo.

- Errores debidos a la baja cantidad o calidad de ADN: la baja cantidad y/o calidad de ADN en la mezcla de reacción, debida al uso de soluciones muy diluidas o degradadas, promueve errores de genotipificación debido al fenómeno llamado *allelic dropout* (uno de los dos alelos no amplifica) y la amplificación de falsos alelos (Taberlet et al., 1996). También incrementa el riesgo de contaminación (Taberlet et al., 1996), ya que en condiciones en las cuales el número de moléculas de templado es bajo, es mayor la probabilidad de amplificación de las moléculas contaminantes presentes.
- Errores debidos a artefactos bioquímicos: en la etapa final de la PCR, durante el paso de elongación, la *Taq* polimerasa tiene tendencia a adicionar un nucleótido, usualmente adenina, al extremo 3' de la cadena sintetizada (Brownstein et al., 1996; Magnuson et al., 1996). Esta adenina adicionada crea una banda o pico espurio sobre el gel o el cromatograma.

- Errores debidos a factores humanos: según los casos analizados en el trabajo de Pompanón et al. (2005), el error humano fue el más frecuente. Parte de los errores de este tipo se deben a errores en la anotación manual de geles o autorradiografías, por ejemplo: al analizar autorradiografías leerlas en la orientación opuesta. Pueden presentarse errores en la anotación con métodos automáticos, ya que los microsatélites suelen ser difíciles de analizar debido a la presencia de bandas en escalera o *stutter bands* (Ginot et al., 1996; Harker 2001), por ejemplo, puede producirse un patrón ambiguo cuando dos alelos de un heterocigoto difieren en tamaño por una sola unidad de repetición, causando que las bandas *stutter* se solapen produciendo bandas adyacentes de similar intensidad. Por supuesto, el riesgo de errores en las anotaciones depende de la calidad de los datos y de los recaudos necesarios durante el análisis e interpretación.

En los estudios de ligamiento y asociación estos errores enmascaran la verdadera segregación de los alelos, y los efectos sobre los resultados pueden ser serios, incluso para porcentaje de errores bajos (por ejemplo < 3%). En estudios de ligamiento, los errores de genotipificación pueden distorsionar las frecuencias de los haplotipos (Kirk & Cardon, 2002) conduciendo a un aumento ficticio de la longitud del mapa genético (Goldstein et al, 1997; Hackett & Broadfoot, 2003).

En estudios de asociación, se ha observado que un porcentaje de error del 3% puede impactar seriamente sobre los análisis de desequilibrio de ligamiento (Akey et al., 2001). En estos estudios, donde los eventos de recombinación son menos frecuentes, generalmente los errores afectan a los genotipos no-recombinantes, que son evaluados erróneamente como recombinantes, disminuyendo el poder de detectar asociaciones (Douglas et al., 2000). Sin embargo, porcentajes bajos de error pueden ser tolerables en situaciones que no involucran alelos raros, como en estudios de análisis de QTL (Pompanón et al., 2005).

1.2.3 Mapeo Genético

Uno de los principales usos de los marcadores moleculares en la agricultura ha sido la construcción de mapas de ligamiento para diversos cultivos. Este tipo de mapas es muy

útil en el desarrollo de diferentes estrategias de mejoramiento, pudiendo seleccionar marcadores de mejor calidad y contar con una mayor probabilidad de marcadores polimórficos en un importante intervalo cromosómico (Somers et al., 2004). De esta manera, el mejoramiento molecular es más efectivo si se cuenta con mapas saturados de marcadores moleculares.

Los mapas de ligamiento son utilizados para la identificación de regiones cromosómicas que contienen los genes controladores de características simples o cualitativas (controladas por uno o pocos genes) y caracteres cuantitativos, poligénicos o complejos. Se han desarrollado mapas independientes en diversas especies; en estos casos, sería muy útil contar con un mapa que unifique los mapas existentes para integrar la información de la segregación de los marcadores de diferentes cruzamientos, y así permitir un importante marco en la producción e intercambio de información genética entre miembros de la comunidad científica.

Uno de los factores importante en el desarrollo de un mapa genético es el contar con una población de mapeo adecuada para el estudio que se desea realizar. Las poblaciones de mapeo consisten de individuos de una especie, o, en algunos casos, de cruzamientos entre especies relacionadas, donde los padres difieren en las características que serán estudiadas. Son generalmente utilizadas para identificar factores genéticos y para determinar las distancias de recombinación entre *loci* (Scheinder et al., 2005).

Las características a evaluar en una población de mapeo necesitan ser polimórficas entre las líneas parentales.

Para la construcción de una población de mapeo hay que tener en cuenta el modo de reproducción de la planta a utilizar. En plantas autopolinizables, si es posible generar líneas puras o están disponibles, las poblaciones de mapeo posibles de ser utilizadas son: F₂, Líneas Recombinantes Endocriadas (RILs), poblaciones de retrocruzamientos (*backcross population*), líneas introgresadas y líneas doble haploide. En plantas con polinización cruzada, generalmente son utilizadas plantas de parentales heterocigotos, tales como F₁ y líneas retrocruzadas (*backcross lines*), con diferentes grados de heterocigosidad.

Para el desarrollo de este trabajo se uso una población de tipo RILs (el girasol es una especie autopolinizable). Contar con una población de líneas recombinates

endocriadas facilita el análisis, debido, entre otras razones que se detallan a continuación, a la disponibilidad de una fuente eterna de material vegetal.

Las líneas recombinantes endocriadas o RILs (*Recombinant Inbred lines*) se obtienen de sucesivos pasos de autofecundación a partir de una F2 hasta la generación F8, donde la mayoría de sus *loci* están en homocigosis (Figura 1.1). A lo largo de cada cromosoma alternan bloques de alelos derivados de cada padre. Una vez establecida una línea recombinante, no hay posibilidad de futuras segregaciones en la progenie de cada línea, debido a que la recombinación no puede cambiar la constitución genética de las RILs. Esta es una de sus principales ventajas, ya que estas líneas constituyen una fuente permanente que puede ser indefinidamente replicada. La segunda ventaja es que las líneas sufren varias rondas de meiosis antes de alcanzar la homocigosis, de esta manera, el grado de recombinación es mayor comparado a las poblaciones F2 y así los mapas desarrollados con estas poblaciones presentan mayor resolución que los mapas desarrollados con poblaciones F2 (Burr & Burr, 1991), es decir, los mapas desarrollados con estas poblaciones permiten determinar mejor la posición de cada marcador en un mapa de ligamiento.

Debido a los sucesivos eventos meióticos ocurridos hasta alcanzar la homocigosis, los *loci* ligados tienen una gran probabilidad de volver a recombinar. Este efecto aumenta la sensibilidad o el poder para diferenciar clases genotípicas. Así, las poblaciones recombinantes endocriadas obtenidas por autofecundación tienen un número aproximado del doble de recombinaciones entre marcadores ligados con respecto a una población obtenida con un solo evento meiótico (Haldane & Waddington, 1931). De esta manera, la generación F1 tiene un 100% de heterocigosis, la F2 tiene un 50% de heterocigosis y así sucesivamente hasta la F8, que presenta un 0,78% de heterocigosis.

Niveles de autofecundación de la población	Porcentaje de homocigosidad de cada <i>locus</i> dentro de las líneas	Porcentaje de heterocigosidad de cada <i>locus</i> dentro de las líneas
F1	0	100
F2	50	50
F3	75	25
F4	87,5	12,5
F5	93,75	6,25
F6	96,87	3,12
F7	98,43	1,56
F8	99,21	0,78

Para las líneas RIL obtenidas por autofecundación, la fracción de recombinación observada (R) se correlaciona con la fracción de recombinación de un solo evento meiótico (r) por la ecuación: $r = \frac{R}{2-2R}$ desarrollada por Hadane y Waddington (1931).

De la misma manera las distancias genéticas entre marcadores iguales o menores a 12,5 cM en una población RIL generada por autofecundación serán calculadas con mayor precisión que en una F2 o en una retrocruza con el mismo número de individuos (Taylor, 1978).

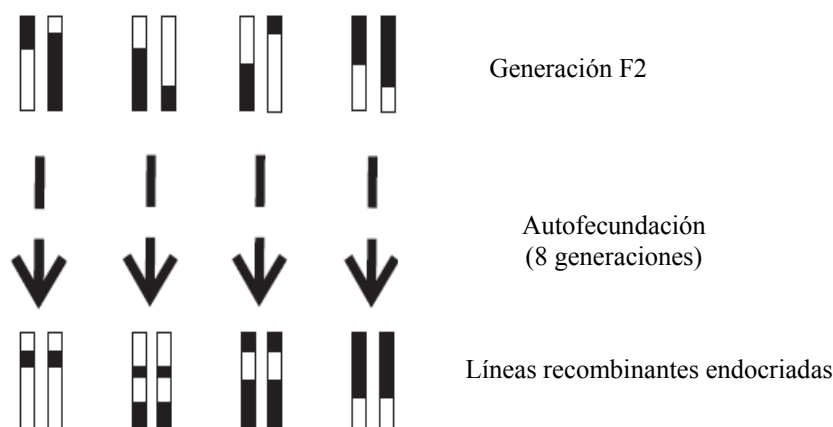


Figura 1.1 Generación de líneas recombinantes endocriadas (RILs). (Schneider, 2005).

Entre las poblaciones de mapeo más relevantes utilizadas en girasol se pueden citar:

a) *Poblaciones de líneas recombinantes*

Las poblaciones de líneas recombinantes endocriadas públicas más utilizadas para desarrollar mapas con diferente nivel de saturación derivan de los cruzamientos entre las siguientes líneas públicas:

- RHA280 x RHA801: la línea RHA280 posee la característica de ser confitero (Fick et al., 1974) y RHA801 es una línea restauradora de la esterilidad citoplasmática (Roath et al., 1981) que posee la característica de ser resistente a mildiu.
- PHA x PHB: ambas líneas son mantenedoras de la esterilidad citoplasmática (Desarrollada por Pioneer Hi-Bred Internacional Inc. (Johnston Iowa).
- RHA266 x PAC2: la línea RHA266 posee la característica de ser resistente a mildiu (Gentzbittel et al., 1998) fue desarrollada por el departamento de agricultura de los Estados Unidos a partir del cruzamiento entre *H. annuus* salvaje x Peredovik. La línea PAC2 es resistente a *S. sclerotiorum* (Gentzbittel et al., 1998) y fue desarrollada por el Instituto Nacional de Investigaciones Agronómicas de Francia, a partir del cruzamiento entre *H. petiolaris* x HA61 (línea USDA). La población derivada de este cruzamiento, así como la generación de las RILs correspondientes, fueron obtenidas por el ENSAT-INRA, Francia.

Estas poblaciones fueron utilizadas para desarrollar varios mapas genéticos de ligamiento en girasol (Al Charrani et al., 2004; Flores Berrios et al., 2000; Hu, 2006; Kiani et al., 2007a; Tang et al., 2003; Tang et al., 2003; Yu et al., 2003).

b) *Poblaciones de F2 usadas en girasol*

Las poblaciones F2 más utilizadas para desarrollar mapas genéticos derivan del cruzamiento de las siguientes líneas públicas:

- ZENB8 x HA89: líneas no restauradora de la esterilidad citoplasmática (León et al., 2000). Poseen la característica de ser susceptible a mildiu.

- HA370 x HA372: líneas públicas mantenedoras de la esterilidad citoplasmática masculina (Miller & Gulla, 1991)

1.2.3.1 Construcción de mapas de ligamiento

Construir un mapa de ligamiento consiste en ordenar los marcadores de acuerdo a sus posiciones relativas en los cromosomas, indicando las distancias genéticas relativas entre ellos y definiendo los grupos de ligamiento. Este ordenamiento se realiza en función de la frecuencia de recombinación o intercambio de material genético entre cromosomas homólogos durante la meiosis. La recombinación puede ocurrir por segregación independiente de loci no ligados o por *crossing-over* entre *loci* ligados, generando nuevas combinaciones de genes. El porcentaje de recombinación se calcula en función de la cantidad de progenie recombinante y su valor se utiliza para estimar la distancia genética entre *loci* (Collard et al., 2005).

La construcción de un mapa genético genera información básica sobre la estructura y la organización de un genoma. Existen tres etapas fundamentales en la construcción de un mapa genético de ligamiento:

La primera etapa consiste en la obtención de una población de mapeo tal como se describió en el punto anterior.

La segunda se basa en la identificación de polimorfismos entre los padres contrastantes. Los marcadores a ser utilizados son seleccionados en base a la disponibilidad de marcadores para la especie en estudio, el objetivo del trabajo y la disponibilidad económica al momento de desarrollar la investigación. Posteriormente a la identificación de los polimorfismos en los parentales se inicia el análisis en los genotipos de la población, proceso denominado genotipificación. Un análisis de los posibles errores durante este proceso se describió en la sección 1.2.2.1.

La tercera etapa consiste en el análisis de ligamiento entre los marcadores, el cual se basa en la frecuencia de recombinación. Para los genes situados muy cerca en el mismo par cromosómico, el ligamiento físico de las combinaciones de los alelos parentales hace imposible la segregación independiente dando lugar, por lo tanto, a frecuencias de recombinantes significativamente menores al 50%. Cuanto mayor es la distancia entre genes ligados, mayor es la probabilidad de que tenga lugar un entrecruzamiento entre

cromátidas no hermanas dentro de la región comprendida entre los genes y, por lo tanto, mayor la proporción de recombinantes que se obtienen. Así, se puede obtener una medida de la distancia entre los genes mediante la determinación de la frecuencia de recombinantes. De hecho, podemos definir una unidad de mapa genético (m.u.) como la distancia entre los genes para la que uno de cada 100 productos de la meiosis es recombinante. Dicho de otra manera, una frecuencia de recombinación (RF) de 0,01 (1%) se define como 1 m.u. (la unidad de mapa también se llama, centimorgan [cM], en honor Thomas Hunt Morgan). Sin embargo, las distancias genéticas expresadas en frecuencia de recombinación no son aditivas esto es debido a que al aumentar la distancia entre los *loci* aumenta la probabilidad de ocurrencia de dobles entrecruzamientos, con lo cual no son detectados los recombinantes y, por lo tanto, se subestima la distancia real. Por este motivo los mapas se construyen de a pequeños intervalos.

Para poder calcular las distancia de mapa, se necesita una función matemática, ya que la frecuencia de ocurrencia de entrecruzamientos y la frecuencia de recombinación no se relacionan de manera lineal, a mayor distancia aumenta el número de quiasmas producidos, y el porcentaje de recombinación hasta un valor máximo de 50%. A distancias menores a 10 cM los valores de recombinación igualan a las distancias de mapa, pero a distancias mayores a 10 cM esta relación no puede aplicarse. En general, se usan dos funciones de mapeo: la función de mapeo de Kosambi, que asume que la ocurrencia de un evento de recombinación influencia la ocurrencia de un evento adyacente; y la función de mapeo de Haldane, la cual asume la ausencia de interferencia entre eventos de recombinación.

Si asumimos que “d” es la distancia de mapa y “r” es la frecuencia de recombinación entre dos marcadores, las dos funciones de mapa son:

$$d = -1/2 \ln (1 - 2r) \text{ (Haldane)}$$

$$d = 1/4 \ln \frac{1+2r}{1-2r} \text{ (Kosambi)}$$

Un test estadístico muy usado basado en criterios de máxima verosimilitud es el denominado *test LOD score* para ligamiento, que compara dos conjuntos de frecuencias,

aquellas esperadas con una frecuencia de recombinación = r , igual a su estimador de máxima verosimilitud, y aquellas esperadas si la recombinación fuera = 0,5 (hipótesis nula, los *loci* son independientes) (Van den Berg et al., 1997). El LOD score es el logaritmo decimal de la razón entre la verosimilitud de la hipótesis de ligamiento versus la hipótesis de independencia:

$$LOD = \log_{10} \frac{eL(r)}{eL(r_0)_{10}}$$

(Barnard, 1947)

El valor LOD expresa la probabilidad de que dos marcadores estén ligados respecto de que segreguen independientemente. Un valor de LOD 4 significa que la probabilidad de que dos marcadores estén ligados es 10.000:1 respecto de la probabilidad de segregación independiente, ya que:

$$LOD = \log_{10} \frac{\text{Probabilidad de estar ligados}}{\text{Probabilidad de segregar independientemente}}$$

Probabilidad de segregar independientemente

Dado que los marcadores pertenecientes a un cromosoma mapean juntos en un grupo de ligamiento, si se cuenta con suficiente cantidad de marcadores, el número final de grupos de ligamiento que se encuentren corresponderá al número básico de cromosomas de la especie (Jones et al., 1997). Una dificultad asociada con obtener igual número de grupos de ligamiento y de cromosomas es que los marcadores polimórficos detectados no están homogéneamente distribuidos a lo largo del cromosoma, y por lo tanto en algunas regiones se encuentran agrupados y en otras están ausentes. Además, en ciertas regiones del genoma denominadas puntos calientes (*hot spots*), los entrecruzamientos ocurren con alta frecuencia, mientras que en los centrómeros y los telómeros existe baja frecuencia de recombinación. Esta es la razón por la cual las distancias genéticas obtenidas por mapeo genético no se corresponden exactamente con las distancias físicas existentes entre los *loci* de un cromosoma. La precisión en la estimación de las distancias genéticas y en la determinación del orden de los marcadores

está directamente relacionada al número de individuos evaluados en la población de mapeo. Poblaciones de menos de 50 individuos proveen resoluciones de mapeo muy bajas para ser de utilidad (Young, 2000).

Los mapas genéticos no brindan información sobre qué grupos de ligamiento corresponden a cada cromosoma o cómo se relacionan los marcadores del grupo de ligamiento con la estructura física del cromosoma, por lo tanto una vez obtenido un mapa de ligamiento es interesante compararlo a través del anclado de marcadores de sitio único con el cariotipo para asignar los grupos de ligamiento a cada cromosoma.

1.2.3.1.1 Mapas genéticos de girasol

A la fecha de inicio del presente trabajo de tesis habían sido publicados distintos mapas de ligamiento genético para girasol, basados en polimorfismos de longitud de fragmentos de restricción del ADN o RFLP (Berry et al., 1996; Gentzbittel et al., 1995; 1999; Jan et al., 1998), en polimorfismos amplificados al azar o RAPD (Riesberg et al., 1993; 1998), y en polimorfismos de longitud de fragmentos de ADN o AFLP (Al-Chaarani et al., 2002; Flores Berrios et al., 2000; Gedil et al., 2001; Peerbolte & Peleman, 1996). Sin embargo, la técnica de RFLP es muy laboriosa para usarla de rutina como marcador molecular, en tanto que los marcadores RAPD y AFLP, si bien tienen muchas ventajas, como ser abundantes y relativamente fáciles de producir, no requieren conocimientos previos de las regiones que se analizan, etc., son mayormente dominantes, a menudo no son específicos, tienen baja reproducibilidad y su utilidad es limitada para conducir estudios sinténicos y/o para referenciar mapas genéticos de ligamiento. La disponibilidad de microsatélites a comienzos de la década de 2000 hizo posible comenzar a integrar y/o referenciar los mapas existentes y/o en desarrollo para girasol. Estos marcadores fueron usados para el desarrollo de diferentes mapas de ligamiento utilizando diferentes poblaciones F₂ y RILs, los cuales se describen a continuación:

Flores Berrios et al. (2000) construyeron un mapa genético con 264 AFLP distribuidos en 18 grupos de ligamiento (GL), utilizando 99 RILs provenientes del cruzamiento entre las líneas parentales RHA266 x PAC2. El mapa cubre una longitud

de 2.558 cM y fue optimizado con el agregado de marcadores AFLP por Al-Chaarani et al. (2002).

El mapa descrito por Al-Chaarani et al. (2004) fue construido sobre 123 RILs provenientes del cruzamiento entre RHA266 x PAC2. Es el mapa más largo reportado en la literatura e incluye 371 marcadores AFLPs y 38 SSR (30 seleccionados de la colección CRSx y 8 de la colección ORSx), cubre una longitud de 2.915,9 cM con una densidad de marcadores de 7,9 cM. Sin embargo, el número de grupos de ligamiento descrito fue de 21, mayor que el número haploide de cromosomas de la especie ($n=17$). La resolución de este mapa mejoró al incorporar microsatélites, ya que si bien los AFLPs son considerados marcadores eficientes en la construcción de mapas genéticos debido a su alto nivel de polimorfismo producido por reacción de PCR, estos marcadores necesitan ser complementados con marcadores anclados para generar un mapa unificado entre diferentes mapas existentes. Asimismo, la utilización de diferentes marcadores permite cubrir diferentes zonas del genoma del girasol.

Tang et al. (2002) desarrollaron el primer mapa genético de SSR utilizando 459 marcadores. Este mapa cubre una longitud de 1.368,3 cM con una densidad media de 3,1 cM y fue construido usando 94 RILs provenientes del cruzamiento entre RHA280 y RHA801. Los marcadores se encuentran distribuidos en 17 GL. Estos autores referenciaron tres mapas construidos a partir de tres poblaciones derivadas del cruzamiento entre germoplasmas contrastantes (RHA280 x RHA281, PHA x PHB y HA370 x HA372), utilizando 701 SSR y 89 RFLPs o InDels (Yu et al., 2003). En el mismo trabajo, estos autores también integraron parcialmente el mapa público de SSR construido con la población derivada del cruzamiento de HA370 x HA372 con el mapa de RFLP construido por Jan et al. (1998). Finalmente publicaron un mapa compuesto que integra 657 *loci* SSR con una densidad media de 2,2 cM por *locus* y una longitud de 1.423 cM (Tang et al., 2003). Si bien este es el mapa más saturado descrito hasta el presente, presenta *gaps* (regiones sin cubrir) más largos de 30 cM sobre 4 grupos de ligamiento, a saber: GL 2 (31,3 cM), GL 4 (36,4 cM), GL6 (32,6 cM) y GL 13 (30,7 cM). En dicho mapa se seleccionaron 78 SSR, considerando su posición en el mapa, el nivel de heterocigosidad, el rango de la longitud de los alelos y la calidad en la genotipificación, desarrollaron una reacción de PCR múltiple con el fin de aumentar la capacidad del procesamiento de datos. Posteriormente, Zhang et al. (2005)

seleccionaron un conjunto de 78 marcadores SSR los cuales fueron usados para ensayar la variabilidad genética en un grupo de 124 líneas endocriadas de girasol. Este estudio permitió definir un conjunto de SSR útiles para la identificación de variedades.

Estos mapas sirvieron de base para mapear *loci* de características cuantitativas (Al-Chaarani et al., 2002; Alfadil et al., 2007; Bert et al., 2001; Darvishzadeh et al., 2007; Ebrahimi et al., 2008; Flores Berrios et al., 1999a y b; Hervé et al., 2001; Kiani et al., 2007a y b; 2008; Lawson et al., 1998; León et al., 1995; 1996; 2000; 2001; 2003; Lu et al., 1999; Maestries et al., 1998; Maringolo, 2007; Mokrani et al., 2002; Pérez-Vich et al., 2001).

El primer mapa funcional de girasol basado en el mapeo de genes que se expresan fue descrito por Lai et al. (2005) y fue construido a partir del cruzamiento entre RHA280 y RHA801. Contiene 243 marcadores funcionales de tipo EST y 196 SSR que constituyen el marco de referencia del mapa descrito por Tang et al. (2002). Los marcadores EST derivan de 44.000 transcritos secuenciados a partir de los genotipos RHA280 y RHA801 y depositadas en bases de datos públicas. Los polimorfismos fueron primero detectados *in silico*, y luego la genotipificación se realizó mediante la detección de heteroduplex utilizando la técnica de DHPLC. El mapa resultante cubre una distancia de 1.349,3 cM e incluye ESTs con funciones candidatas para distintas características, tales como, desarrollo, transporte celular, metabolismo, defensa y tolerancia a estreses bióticos y abióticos.

1.2.4 Mapeo físico

Los marcadores moleculares combinados con la citogenética resultan de gran utilidad para entender la organización cromosómica y la distribución física de las secuencias. El gran potencial de las técnicas de hibridación *in situ* resulta de combinar información acerca de la morfología nuclear o cromosómica con la información molecular de la estructura de las secuencias. En el caso de FISH (*Fluorescent in Situ Hybridization*), una de sus aplicaciones es la obtención de un mapa físico, funcional y estructural del genoma utilizando secuencias de distinto origen como sondas (Poggio & Naranjo, 2005).

1.2.4.1 Características de los genomas de las plantas

El genoma de las plantas consiste mayoritariamente en ADN repetitivo, incluyendo secuencias correspondientes a ADN ribosomal y secuencias con similitud a retrotransposones (Bennetzen, 1993; Grandbastien, 1998; San Miguel y Bennetzen, 1998).

Las regiones organizadoras nucleolares (NOR), responsables de la transcripción del ARN ribosomal (18S-5.8S-26S-28S y 5S), forman una estructura cromosómica conspicua, una constricción secundaria que puede ser identificada por tinción con plata y/o por hibridación *in situ*, siendo la técnica de FISH (*Fluorescenc In Situ Hybridization*) la más utilizada debido a su mayor especificidad.

Los genes ribosomales son de copia múltiple, siendo el número de copias muy variable en el genoma de las plantas. En muchas especies, los genes de ARNr se encuentran en un solo sitio, sobre un par de cromosomas homólogos, sin embargo, lo más frecuente es que se encuentren distribuidos en varios cromosomas. Las regiones organizadoras nucleolares (NOR) a menudo se localizan en los extremos de los cromosomas (Sumner, 2003).

Los transposones están generalmente dispersos a lo largo de los cromosomas de las plantas, pero muchos están asociados con una región particular. Su dispersión cromosómica sugiere que los elementos transponibles juegan un rol importante en la evolución, en la estructura genómica y en la función de los genes. La habilidad relativa de las especies de plantas durante la evolución, para inactivar o remover estos elementos o para tolerar su presencia ha creado poblaciones de transposones particulares para cada genoma (Bennetzen, 2000).

La actividad de los retrotransposones puede cambiar la estructura genómica completa. Debido a su dispersión y alto número de copias en el genoma de las plantas, los retrotransposones sirven como sitios de recombinaciones ectópicas o desiguales. Las recombinaciones ectópicas entre retrotransposones de la misma familia que están en directa orientación sobre el mismo cromosoma pueden causar duplicaciones y deleciones recíprocas, mientras que las recombinaciones desiguales entre elementos sobre el mismo cromosoma que están en orientación opuesta podrían causar inversiones cromosómicas de las secuencias entre los dos retrotransposones. Las recombinaciones

ectópicas entre dos elementos sobre diferentes cromosomas pueden causar translocaciones recíprocas. Los cambios tales como grandes inversiones o translocaciones pueden contribuir al aislamiento reproductivo y a la subsecuente especiación. De esta manera, aunque la mayoría de los elementos transponibles de plantas son inactivos, en un momento dado una actividad ocasional puede tener enormes efectos (Zhang et al., 2004b).

En girasol, Santini et al. (2002) y Natali et al. (2006) usaron como sondas secuencias repetidas derivadas de retroelementos de *H. annuus* de la familia *Ty1/copia-like* y *Ty3/gypsy-like* en experimentos de hibridación *in situ*, y observaron que ambas familias de retroelementos se encontraban dispersas a lo largo de todos los cromosomas en todas las especies de *Helianthus* estudiadas. Sin embargo, las secuencias del retrotrasposón *Ty3/gypsy-like* fueron localizadas preferencialmente en las regiones centroméricas en la mayoría de las especies estudiadas, mientras que la secuencia *Ty1/copia-like* fueron menos representadas o ausentes alrededor del centrómero y más abundante en los extremos de los cromosomas en *H. annuus*. Estas investigaciones sugieren que estas dos familias de secuencias juegan un rol en la evolución del genoma de *Helianthus* y de especies divergentes, evolucionaron de manera independiente compartiendo el mismo fondo genético en especies anuales o perennes y posiblemente adquirieron diferentes funciones en el genoma huésped.

Al utilizar secuencias correspondientes a regiones de ADN repetitivo o de copia única como sondas, podemos aproximarnos aún más a entender la organización cromosómica y la distribución física de las regiones codificantes y no codificantes de los genomas vegetales.

1.2.4.2 Los mapas físicos

Los mapas físicos representan el ordenamiento de los componentes genéticos sobre los cromosomas en términos de distancias físicas. Como se mencionó anteriormente, la frecuencia de recombinación es usada para ordenar genes y marcadores moleculares en grupos de ligamiento en mapas genéticos; como consecuencia, se obtiene un mapa de distancias genéticas (usualmente expresadas en cM), pero la distancia genética no

necesariamente refleja la distancia física, medida en pares de bases, entre los marcadores.

Los mapas físicos permiten determinar no sólo la localización de los marcadores sobre un cromosoma en relación al centrómero y al telómero, sino también detectar fenómenos tales como inserciones, deleciones y translocaciones.

El advenimiento de técnicas de citogenética molecular, como la hibridación *in situ* (FISH: *Fluorescent in Situ Hybridization*, GISH: *Genomic In Situ Hybridization*) y más recientemente la incorporación de otras técnicas como PRINS (*Primed In Situ Labelling*) y C-PRINS (*Cycling-Primed In Situ Labelling*), brindan información adicional a la citogenética clásica permitiendo la integración de los análisis estructurales de los genomas. Estas técnicas combinan información citológica de la morfología nuclear o cromosómica con información molecular de la estructura de las secuencias, y son ampliamente usadas para estudiar la localización física de las secuencias de ADN en el genoma para establecer la correlación entre grupos de ligamiento (determinados por un mapa genético) y cromosomas específicos y para entender la organización genómica y la distribución espacial tridimensional de las secuencias de ADN en interfase y en meiosis. Muchas de las respuestas obtenidas mediante el uso de estas técnicas son difíciles de dilucidar usando otros métodos, por ejemplo, las secuencias repetitivas muestran muchas bandas en geles de electroforesis, con patrones complejos, difíciles de separar, interpretar y asignar a *loci* determinados. Además, los proyectos de secuenciación y análisis de *contigs* de grandes insertos no son capaces de interpretar la extensa y relativamente homogénea distribución de las secuencias repetitivas; y, como se mencionó anteriormente, los mapas de ligamiento nos brindan información sobre dónde ocurren las recombinaciones en el genoma, por lo tanto se obtiene información limitada sobre la localización (Schwarzacher, 2003).

1.2.4.3 La técnica de FISH

La técnica de FISH consiste en el apareamiento de un determinado segmento de ADN o ARN a una secuencia de nucleótidos complementarios dentro de la célula, permitiendo verificar si la célula contiene esa secuencia y observar su exacta localización.

Esta técnica utiliza fluorocromos conjugados a anticuerpos. En un solo preparado pueden ser detectadas simultáneamente varias sondas, cada sonda puede ser visualizada con un color diferente empleando un microscopio de fluorescencia.

Las sondas para FISH son marcadas con biotina o digoxigenina, y estas moléculas no fluorescentes son detectadas mediante fluorocromos conjugados con anticuerpos anti-biotina o anti-digoxigenina, o con estreptoavidina. Uno de los problemas inherente a este método es su alta detección de ruido de fondo producido debido a la unión inespecífica del anticuerpo o la estreptoavidina a la superficie celular. Otro problema es que sustancias de alto peso molecular, tales como los anticuerpos o la estreptoavidina, no penetran bien dentro del citoplasma o en cromosomas condensados. Debido a estos inconvenientes, la señal suele ser muy variable de célula a célula. Esta técnica está siendo utilizada para:

- ✓ Facilitar el mapeo cromosómico.
- ✓ Integrar mapas genéticos y físicos (Cheng et al., 2001).
- ✓ Asociar bandas cromosómicas con datos de secuencias (Jiang & Gill, 1994).
- ✓ Medir *gaps* de los mapas existentes (Jackson et al., 1998).
- ✓ Detección de secuencias de ADN ribosomal, ADN altamente y poco repetido repetitivo (Ambros et al., 1986, Moscone et al., 1996; Mouras & Negrutiu 1989; Mouras et al., 1987).
- ✓ Detección de secuencias de bajo número de copias o de copia única (Lapitan et al., 1997; Wang & Chen., 2005).
- ✓ Detección de transgenes, este tipo de ensayo comenzó en plantas de petunia (Hoopen et al., 1996) y luego se extendió a muchos cereales (Carlson et al., 2001; Pederssen et al., 1997; Salvo-Garrido et al., 2001).

La sensibilidad de la técnica de FISH en plantas es baja, comparada con la observada en animales, incluyendo humanos. Esta diferencia se debe a la presencia de pared celular y citoplasma en los preparados cromosómicos, lo cual inhibe la penetración de la sonda y genera ruido de fondo (*background*) (Jiang & Gill, 1994). Aunque recientemente la sensibilidad de esta técnica en plantas ha sido aumentada y se han publicado algunos artículos en los cuales se utilizan como sondas secuencias

menores a 1kb; los resultados han sido inconsistentes para secuencias de bajo número de copias y la técnica no ha sido aplicada de rutina (Desel et al., 2001; Khurstaleva & Kik, 2001). Una alternativa a este problema es el uso de clones de grandes insertos, tales como BACs, donde el tamaño promedio del inserto va de 120 kb a 300 kb (Sambrook et al., 1989), conteniendo la secuencia de interés como sonda (Jiang et al., 1995; Sadler et al., 2000).

El uso de ADN genómico clonado en Cromosomas Artificiales de Bacterias (BACs) como sondas en experimentos de hibridación *in situ* se denomina BAC-FISH. Como la mayoría de los clones de insertos grandes contienen secuencias repetitivas, es necesario el bloqueo de las secuencias de ADN repetitivo utilizando ADN competidor no marcado, tal como la fracción Cot-1 y/o secuencias repetitivas propias de la especie en estudio, no marcada en la mezcla de hibridación. Sin embargo, no siempre se logra bloquear completamente con Cot-1 u otras fracciones repetitivas, aunque sí se puede mejorar el ruido de fondo (Kim et al., 2002).

El desarrollo de este tipo de técnicas es una herramienta poderosa para relacionar los marcadores y/o genes ubicados en un mapa de ligamiento a los cromosomas correspondientes.

La estrategia se basa en el uso de marcadores de ADN o productos de PCR, tales como STSs y SSR seleccionados de un mapa genético de ligamiento y usados como sondas en librerías genómicas de grandes insertos. Para realizar la búsqueda en las librerías genómicas de manera más eficiente es recomendable utilizar en cada hibridación un gran número de sondas (Chang et al., 2001; Tao et al., 2001). Las sondas usadas pueden ser diseñadas, a partir de secuencias conocidas, como oligonucleótidos superpuestos u *overgos* (*overlapping oligos*) utilizando programas bioinformáticos. Cada *overgo* consiste de un par de oligonucleótidos de 22 a 24 nucleótidos, los cuales se complementan en ocho bases en el extremo 3'. Así, la longitud total de cada par de oligonucleótidos abarca de 36 a 40 bases. El par de oligonucleótidos es marcado radiactivamente por una reacción de rellenado usando la polimerasa Klenow (*Klenow fill-in*), produciendo un fragmento doble cadena de 36 a 40 pb que es usado como sonda para la búsqueda en librería de clones de BACs. Este método tiene varias ventajas; en primer lugar, el par de oligonucleótidos es una secuencia de copia única del genoma, ya que son diseñados en base a la información de la secuencia del marcador, lo cual

minimiza la influencia de secuencias repetitivas; en segundo lugar, las sondas *overgo* tienen alta hibridación específica ya que son de copia única y poseen características para una hibridación perfecta.

Las hibridaciones con *overgo* muestran señales fuertes y bajo ruido de fondo, debido a que las sondas son marcadas por una reacción de llenado de extremos de ADN con nucleótidos marcados radioactivamente utilizando el fragmento Klenow de la polimerasa I (*Klenow fill-in*). Finalmente, estas sondas cortas son lavadas fácilmente de los filtros y las membranas pueden ser rehusadas varias veces (Figura 1.2). Tradicionalmente, la búsqueda en librerías de BACs se realiza utilizando como sondas productos de PCR o fragmentos de ADN clonado. Estas sondas contienen fragmentos de varias cientos de pares de bases de ADN y tienen mayor probabilidad de contener secuencias de ADN repetitivo que pueden causar hibridación cruzada. Esto es especialmente cierto en especies con genomas grandes como girasol. La mayor ventaja de las sondas *overgos* con respecto a las sondas tradicionales para la búsqueda de secuencias únicas en librerías de BACs es que las secuencias repetitivas presentes en un fragmento de ADN usado como sonda puede ser excluidas para el diseño de *overgo* minimizando así la hibridación cruzada. Además de esta ventaja, es posible utilizar un conjunto de hasta cuarenta sondas *overgos* en cada hibridación, lo que las hace el método elegido para estudios a gran escala. En los últimos años se han publicado varios trabajos en los cuales se ha utilizado esta técnica en la construcción de mapas físicos en plantas (Islam-Faridi et al., 2002; Kim et al., 2002; Lapitan et al., 1997; Zhang et al., 2004a).

La técnica BAC-FISH ha sido utilizada para localizar físicamente secuencias específicas de ADN e individualizar cromosomas, para identificar regiones cromosómicas con secuencias repetidas y para la identificación citogenética de marcadores moleculares (Zhang et al., 2004), tanto en plantas con genomas pequeños como arroz (Jiang et al., 1995), algodón (Hanson et al., 1995), sorgo (Gomez et al., 1997), *Arabidopsis* (Fransz et al., 2000, Jackson et al., 2000) y papa (Dong et al., 2000) como en plantas con genomas grandes.

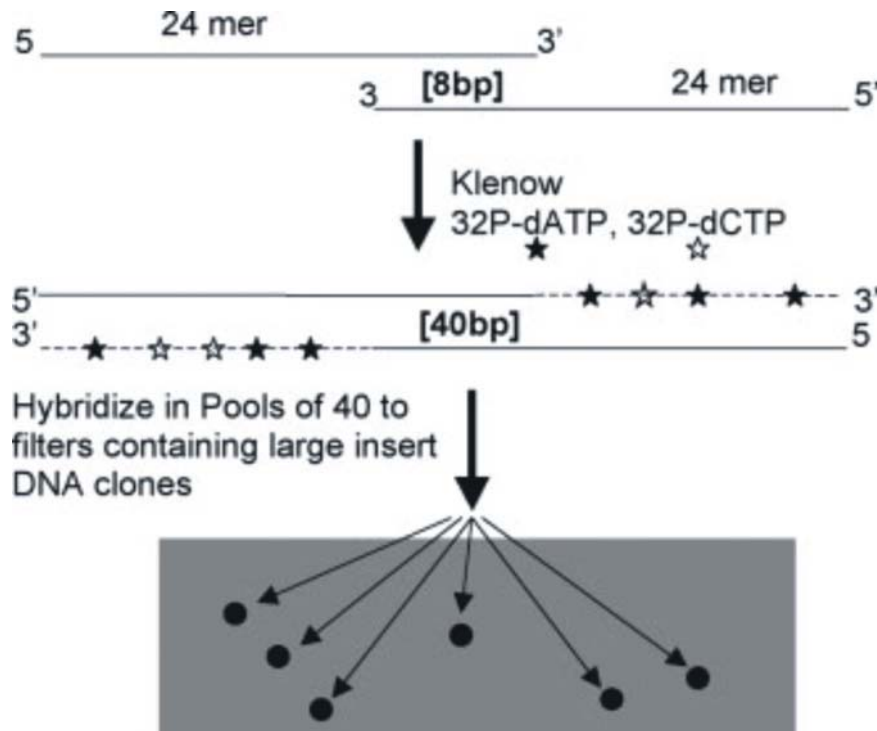


Foto: Meksem et al. (2005)

Figura 1.2 Método para diseñar *overgos*.

En girasol es posible desarrollar la técnica de BAC-FISH debido a la disponibilidad de librerías de cromosomas artificiales de bacterias que posibilitan el aislamiento de genes y/o el estudio de características agronómicas importantes. En total se han descrito cinco librerías. La primera librería de BACs de girasol fue construida utilizando el vector pBeloBAC11 y la enzima *HindIII*, cubre de 4 a 5 veces el tamaño del genoma; el tamaño promedio de los insertos es de 80 kb (Gentzbittel et al., 2002). Una segunda librería de BACs fue construida por Özdemir et al. (2004) usando el mismo vector y la misma enzima, con un tamaño promedio de insertos de 60 kb esta librería cubre 1,9 veces el genoma. Feng et al. (2006) construyeron dos librerías de BACs y BIBACs complementarias. La librería de BACs fue construida con el vector pECBAC1 y la enzima *BamHI*, los clones tienen un tamaño promedio de inserto de 140 kb y cubre cinco veces el tamaño del genoma haploide. La librería de BIBACs fue construida con el vector pCLDO4541 y la enzima *HindIII*, con un tamaño promedio de insertos de 137 kb y equivale a 3,9 veces el tamaño del genoma. Estas dos librerías combinadas cubren aproximadamente 8,9 veces el tamaño del genoma, con lo cual existe una probabilidad mayor al 99% de obtener el clon de interés. Asimismo, Bouzidi

et al. (2006) construyeron una librería de BACs utilizando la enzima *HindIII* y el vector pBeloBAC11, con un tamaño promedio de insertos de 118 kb y equivalente a 5-6 veces el tamaño del genoma haploide.

La quinta librería de BAC de girasol fue desarrollada por la Universidad de Clemson y construida con el vector pIndigoBac536 con un tamaño promedio de insertos de 125 kb equivale a 8,3 veces el tamaño del genoma.

1.2.4.4 La técnica de GISH

La técnica de GISH se basa en el mismo principio que la técnica de FISH, sólo que en este caso la sonda consiste de ADN genómico total de una especie. Esta técnica ha sido usada exitosamente en diversas especies de plantas para evaluar la introgresión de especies silvestres en plantas cultivadas (Benabdelmouna et al., 2003; Shigyo et al., 2003; Wei et al., 2003) así como en estudios evolutivos usando poliploides silvestres (Bennett et al., 1995; Poggio et al., 1999).

1.2.4.5 La técnica de PRINS

La técnica de PRINS se basa en el apareamiento de oligonucleótidos iniciadores no marcados a secuencias complementarias en cromosomas desnaturalizados *in situ*, seguido de la elongación debida a la incorporación, mediante una ADN polimerasa, de nucleótidos por métodos directos o indirectos. Los métodos directos (PRINS directo) se basan en el uso de nucleótidos marcados por fluorescencia, mientras que los indirectos (PRINS indirecto) se basan en la incorporación de nucleótidos marcados con haptenos (biotina o digoxigenina) que serán luego detectados mediante anticuerpos marcados con fluorocromos (Kubálaková et al., 1997). Una variación de esta técnica es el **C-PRINS** (Cycling-PRINS), en la cual intervienen una serie de ciclos termales análogos a la reacción en cadena de la polimerasa.

La técnica de PRINS tiene varias ventajas con respecto a la técnica de FISH convencional, es fácil de realizar, rápida, elimina el paso de marcación de sonda y es más económica debido a que es posible el diseño y la síntesis de oligonucleótidos iniciadores cortos con mínima información sobre la secuencia a amplificar. Estos

oligonucleótidos hibridan más rápida y fácilmente con sus secuencias homólogas en cromatina densamente condensada que las sondas más largas utilizadas para hibridación *in situ*. Además, los iniciadores pueden elongarse hasta secuencias repetidas flanqueantes (en menos de 30 minutos) y, por lo tanto, intensifican la señal del oligonucleótido que hibridó previamente (Menke et al., 1998).

Debido a su rapidez y sensibilidad, la técnica de PRINS y sus variantes están siendo actualmente usadas en numerosas aplicaciones, entre las cuales se encuentran:

- ✓ Mapeo físico de secuencias de ADN repetitivo en cromosomas mitóticos.
- ✓ Mapeo genético en cromosomas politénicos.
- ✓ Individualización de cromosomas humanos.
- ✓ Identificación de marcadores cromosómicos.
- ✓ Análisis de líneas celulares tumorales.
- ✓ Detección de células fetales.
- ✓ Estimación del sexo.
- ✓ Detección de aneuploidías.

(Kubálaková et al., 2001).

1.2.4.6 Avances en citogenética molecular en girasol

A pesar de los enormes avances científicos de la biología estructural (incluyendo la secuenciación nucleotídica de los genomas de varias especies), el conocimiento de las regiones cromosómicas en donde se localizan muchos caracteres de interés teórico y otros importantes para la producción no ha sido suficientemente explorado, reduciendo considerablemente su potencial aplicación a los procesos de mejora genética, siendo uno de los factores determinantes de ello la complejidad taxonómica del género *Helianthus*. Entre los factores que contribuyen a esta complejidad se encuentran el alto número de especies anuales y perennes, los híbridos interespecíficos naturales o artificiales con distintos grados de introgresión por parte de especies silvestres, la poliploidía en

especies perennes y una amplia distribución geográfica que lleva a una gran variación fenotípica.

Las técnicas de hibridación *in situ*, al ser útiles para entender la organización cromosómica y la distribución física de las secuencias repetidas, permiten analizar comparativamente el genoma del girasol cultivado con el de especies silvestres, y estudiar importantes fuentes de genes de resistencia a enfermedades, tolerancia a sequía, androesterilidad citoplasmática, entre otros caracteres. Asimismo, el desarrollo de estas técnicas tiene aplicación en el mapeo físico de caracteres de genes de interés para girasol cultivado, especialmente en la localización de familias multigénicas (Jiang & Gill., 1994).

Los estudios de citogenética molecular preliminares en girasol, utilizando la técnica de GISH, mostraron que la especie salvaje *H. petiolaris* está muy relacionada al girasol cultivado, aunque una hibridación diferencial entre cromosomas podría indicar una divergencia durante el proceso de mejoramiento del girasol cultivado (Paniego et al., 2007; Rocco et al., 2002). Asimismo, los mapas de ligamiento genético y estudios de colinearidad indican la presencia de once rearrreglos (ocho translocaciones y tres inversiones) entre las especies *H. petiolaris* y *H. annuus* (Burke et al., 2004).

Estudios realizados mediante bandeo cromosómico y la técnica de FISH usando ADN ribosomal 45S como sonda permitieron la diferenciación entre especies diploides, tetraploides y hexaploides (Vanzela et al., 2002).

El girasol cultivado posee cromosomas muy semejantes en tamaño y prácticamente indistinguibles entre sí dentro de cada clase morfológica. Si bien han sido reportados varios trabajos en los cuales se intentó identificar y caracterizar a los pares cromosómicos estos trabajos han sido contradictorios. Raicu et al. (1976) y Schrader et al. (1997) identificaron 10 pares metacéntricos, 3 submetacéntricos y 4 acrocéntricos. Asimismo, Al-Allaf & Godward (1979) y Cuéllar et al. (1996) encontraron 4 pares metacéntricos, 8 submetacéntricos y 5 acrocéntricos. Recientemente Ceccarelli et al. (2007) realizaron la caracterización del complemento cromosómico de *H. annuus* utilizando una secuencia repetida en tándem de 368 pb aislada del clon HAG004N15. En el citado trabajo los cromosomas fueron divididos en dos grupos, cuatro pares acrocéntricos y trece pares meta y submetacéntricos. Además hay un debate acerca del

número de pares de cromosomas que tienen satélites, identificándose en los diferentes trabajos 1 (Bohorova & Georgieva-Todorova, 1987); 2 (Kulshreshta & Gupta 1981) o 3 (Ceccarelli et al. 2007; Cuéllar et al., 1996; Raicu et al., 1976; Schrader et al., 1997). Otra discrepancia se presenta en los pares de cromosomas conteniendo ADN ribosomal, Cuéllar et al., (1996) identificaron 3 y Scharer et al., (1997) y Vanzela et al., (2002) reportaron 4. Asimismo Ceccarelli et al., (2007), utilizando la sonda pTa71 de trigo observaron señales en 4 pares de cromosomas, pero sólo en 3 pares fueron intensas.

El presente trabajo se focaliza en los grupos de ligamiento en los que mapean *loci* o QTLs para resistencia a mildiu, PHC de girasol, y/o a estrés hídrico con el fin de encontrar la localización cromosómica de dichos caracteres de interés agronómico.

1.3 Hipótesis de trabajo

La hipótesis que sustenta este trabajo es que es posible localizar genéticamente cualquier secuencia nucleotídica que presente variación genética, sea ésta un marcador molecular anónimo o un gen conocido mediante el análisis del patrón de segregación diferencial con respecto a otro *locus* o conjunto de *loci*.

La disponibilidad de polimorfismos suficientemente distribuidos a nivel cromosómico constituye una herramienta eficaz para ser usada en proyectos de mejoramiento genético, en la localización de caracteres complejos y en la genotipificación o caracterización varietal.

Es posible utilizar secuencias repetitivas de ADN contenidas en BACs para caracterizar el complemento cromosómico del girasol cultivado, y correlacionar grupos de ligamiento de interés agronómico, particularmente aquellos que contienen marcadores asociados a *loci* o QTLs de resistencia a enfermedades con cromosomas específicos a través de técnicas de citogenética molecular tales como FISH y/o PRINS.

1.4 Objetivos

1.4.1 Objetivo general

- Desarrollar un mapa genético de referencia para girasol cultivado, incluyendo marcadores neutros y funcionales y la localización física de grupos de ligamiento que incluyan genes y/o regiones genómicas de interés agronómico a través de herramientas de citogenética molecular.

1.4.2 Objetivos específicos

- Integrar la información obtenida a partir del desarrollo de marcadores moleculares neutros, de tipo microsatélites o SSR (*Simple Sequence Repeats*), en un mapa de ligamiento de AFLPs construido sobre una población de líneas recombinantes endocriadas (RILs), obtenidas del cruzamiento de RHA266 x PAC2 (Dr. Gentzbittel, EMSAT, Francia) y su referencia a otros mapas publicados.
- Identificar, caracterizar y mapear marcadores funcionales de tipo ESTs (*Expressed Sequence Tags*) mediante técnicas de identificación de SNPs (*Single Nucleotide Polimorphism*) y microsatélites contenidos dentro de las secuencias de EST conocidos como ESTs-SSR.
- Contribuir mediante distintas sondas a la caracterización del complemento cromosómico de *H. annuus*.
- Iniciar el desarrollo de un mapa físico que permita definir la correlación cromosómica de grupos de ligamiento de interés, con especial énfasis en aquellos asociados a resistencia a patógenos, por medio de técnicas de citogenética molecular, incluyendo BAC-FISH (*Bacterial Artificial Chromosome-Fluorescent In Situ Hybridization*), PRINS (*Primed In Situ Labeling*) y/o C-PRINS (*Cycling-Primed In Situ Labeling*).

2. MATERIALES Y MÉTODOS

2.1 Material vegetal

Para el análisis y el mapeo genético de los marcadores moleculares se trabajó con una población de mapeo compuesta por 123 líneas recombinantes endocriadas (RILs) derivada del cruzamiento de las líneas puras de girasol RHA266 y PAC2 cedidas por el Dr. Gentzbittel, INP-ENSAT/INRA Centre Biotech. Veg. Francia, en el ámbito de una cooperación académica entre este grupo y el grupo de genómica de girasol del Instituto de Biotecnología, INTA Castelar.

Para la identificación y validación de marcadores funcionales se incluyeron dos líneas puras de girasol HA89 y RHA801 que son parentales de una población de mapeo en desarrollo en la EEA INTA, Balcarce para estudios de estrés biótico. La Tabla 2.1 describe las características de las líneas parentales utilizadas.

Para la preparación de las sondas y los preparados citológicos a ser utilizados en la construcción y caracterización de un mapa físico de girasol cultivado, se utilizó la línea pública HA89 (Tabla 2.1).

Tabla 2.1. Origen de las líneas parentales de girasol utilizadas en el mapeo genético de marcadores moleculares. USDA, *United States Department of Agriculture (USA)*; INRA, *Institut National de la Recherche Agronomique (France)*.

Líneas	Origen	Institución
PAC2	<i>H. petiolaris</i> (restaurador de la fertilidad) x línea USDA (HA61)	INRA
RHA266	<i>H. annuus</i> salvaje x Pederovik	USDA
HA89	VNIIMK 8931 – CM303	USDA
RHA801	Derivada de restaurador compuesto	USDA

Para la obtención del material vegetal y la conservación de la población de mapeo las semillas provenientes de 94 RILs, más las líneas parentales, se cultivaron a campo, bajo protección antipájaro en sombráculo en las instalaciones del Instituto de Biotecnología de INTA Castelar. Asimismo, una copia de la población de mapeo se

mantiene en la EEA INTA, Balcarce. El procedimiento de multiplicación de materiales consistió en la siembra de 10 semillas correspondientes a cada genotipo, los cuales se sembraron en arena estéril para su germinación. Las plántulas de 5 cm se transplantaron de manera individual a macetas conteniendo 40% tierra estéril, 30% turba, 20% estiércol, 10% perlitas y fueron fertilizadas con fertilizante hidrosoluble en polvo Hakaphos (de grado equivalente 18-18-18 (N-P-K), el cual fue aplicado por riego en dosis de 250 ppm dos veces por semana. Para la extracción de ácidos nucleicos se tomaron 4 hojas de plantas de 6 a 8 semanas. Las mismas fueron congeladas inmediatamente en nitrógeno líquido y conservadas en ultrafreezer a -80°C hasta el momento de ser liofilizadas. Una vez completada la cosecha, los materiales se liofilizaron durante 24 horas. Las muestras liofilizadas fueron molidas por abrasión con bolitas de metal recubiertas de cerámica durante 2 minutos en *vortex*.

Para la multiplicación de la población los capítulos se autofecundaron en estadio R5 y se cubrieron con bolsas de papel para prevenir la fecundación cruzada. Las semillas cosechadas se guardaron secas a una temperatura de 4°C .

2.2 Extracción de ADN

Los reactivos utilizados en los siguientes protocolos corresponden a la firma SIGMA (Argentina), excepto en los casos en que se indique otro proveedor.

2.2.1. Extracción por CTAB con el agregado de agentes para extraer polisacáridos

Para la extracción de ADN se utilizó el método de CTAB (bromuro de cetil-metilamonio), modificado por Saghai-Marroof et al. (1984). Aproximadamente 1000 mg de material molido por muestra fueron incubados en 30 ml de *buffer* de extracción (Tabla 2.2) de 60 a 90 minutos a 65°C en agitación.

Tabla 2.2. La tabla detalla la concentración final de los componentes usados para la extracción de ácidos nucleicos

Componentes	[final]
Tris-HCl pH 7,5	100 mM
NaCl	700 mM
EDTA pH 8,0	10 mM
CTAB	1% m/v
2-mercaptoetanol	140 mM
PVP (polivinilpirrolidona)	1% m/v

Luego de la incubación se realizaron extracciones con cloroformo:octanol en proporción 24:1 v/v, la emulsión se agitó durante 10 min y se centrifugó a 3.000 rpm para facilitar la separación de fases. Se recuperó la fase acuosa y el ADN se precipitó con 1 volumen de isopropanol, y se realizaron precipitaciones sucesivas con 76% (v/v) de etanol/0,2 M de NaAcO y 66% (v/v) de etanol/10 mM NH₄AcO. Luego, el ADN se resuspendió en 1 a 4 ml de TE (Tris HCl 10 mM, EDTA 1 mM pH 8,0) por 16 hs en agitación a temperatura ambiente. A continuación, el ADN se trató con 50 µl de RNAsa A (10 mg/ml, Sigma Chem. Co., EUA) por 30 min a 37° C. Finalmente, se adicionó un paso final de purificación para eliminar contaminantes, principalmente polifenoles, mediante el tratamiento con la resina comercial Nucleon PhytoPure Plant DNA Extracción, Kit (Amersham Pharmacia Biotech, RU), 100 µl de resina cada 0,1 g de tejido fresco, por 10 min en agitación y se centrifugó por 5 min a 3.000 rpm. La fase acuosa se conservó en alícuotas de 2 ml a -20° C.

2.3 Cuantificación del ADN de la población RIL

La cuantificación del ADN se realizó por medio de espectrofotometría ultravioleta GeneQuant (Amersham Pharmacia Biotech, RU), asumiendo una equivalencia de DO (260 nm) = 50 µg. Alternativamente, el ADN fue cuantificado utilizando el espectrofotómetro Nanodrop[®] (Thermo Scientific Nanodrop 1000). Paralelamente, se realizó una verificación de la integridad del ADN extraído y de posible contaminación

remanente de ARN mediante electroforesis en geles de agarosa al 1% (m/v) y *buffer* TAE 1X (Tris HCl [40 mM], NaAcO [5 mM], EDTA [0,77 mM]). Para la visualización de los ácidos nucleicos se utilizó tinción con bromuro de etidio (Sambrook et al., 1989).

2.4 Selección de los marcadores microsatélites y análisis de polimorfismo en los parentales

Para realizar los estudios de análisis de ligamiento y poder referenciar los mapas existentes, se seleccionó un grupo de 404 marcadores microsatélites de un conjunto de 550 marcadores los cuales habían sido caracterizados como altamente polimórficos en 16 líneas elite de girasol, desarrollados en el INTA Castelar (HAX) (Paniego et al., 2002). Asimismo, se incluyó un grupo de 256 marcadores microsatélites públicos desarrollados en la Universidad de Oregon (ORSx), seleccionados de un conjunto de 657 marcadores mapeados por Tang et al. (2003) en tres poblaciones de mapeo (RHA280 x RHA281, RHA370 x RHA372 y PHA x PHB) y distribuidos en los 17 grupos de ligamiento del mapa público desarrollado por los citados autores.

Los microsatélites seleccionados fueron amplificados por PCR usando como templado ADN genómico correspondiente a los parentales de la población de mapeo (RHA266 y PAC2) con el fin de identificar los marcadores polimórficos, los cuales luego fueron genotipificados en la progenie.

2.5 Utilización de programas bioinformáticos aplicados al descubrimiento de marcadores moleculares

2.5.1 Estrategia para explotar la base de datos de EST de girasol para detectar marcadores microsatélites

Con el propósito de identificar marcadores funcionales se realizó la búsqueda de SSR contenidos en las secuencias de EST de girasol. Para ello, se usó la base de datos de unigenes de girasol de TIGR (HAGI), actualmente mantenidos y actualizados en el Dana-Farber Cancer Institute (<http://compbio.dfci.harvard.edu/tgi/>), la cual esta representada por 36.741 unigenes o secuencias únicas depositadas al 4 de mayo de

2007, correspondientes a genes predichos a partir del ensamblado de las secuencias de EST de *H. annuus* depositadas en GenBank.

Para la identificación de microsatélites con repeticiones de di, tri, tetra o pentanucleótidos se utilizó el programa CUGI-SSR Server (Universidad de Clemson, EUA). Los parámetros de búsqueda establecidos fueron: longitud de la repetición de 18 bases para los di y trinucleótido y de 20 bases para los tetra y pentanucleótidos.

En este análisis se consideraron las secuencias que contenían, microsatélites perfectos, los cuales se caracterizan por tener un motivo de repetición único y sin interrupciones en la secuencia en tándem, y microsatélites compuestos, con más de un motivo de repetición en la misma secuencia. Se descartaron los llamados microsatélites imperfectos, cuando existen interrupciones de la repetición.

Al momento de estimar las frecuencias fueron consideradas todas las posibles combinaciones de los cuatro nucleótidos. Los motivos de los microsatélites están representados por 4 diferentes dinucleótidos (AT)_n, (AG)_n, (AC)_n, (GC)_n, 10 diferentes trinucleótidos, 33 diferentes tetranucleótidos y 102 motivos diferentes de pentanucleótidos. Así, todos estos motivos pueden estar presentes de diversas formas a partir del mismo conjunto básico o su reverso complementario (Jurka & Pethiyagoda, 1995; Katti et al., 2001). Los diferentes motivos de los di-y trinucleótidos están representados en la Tabla 2.3.

Tabla 2.3. Posibles combinaciones de los di-y trinucleótidos

Motivos dinucleótidos	Motivos trinucleótidos
AT/TA	AAT/ATA/TAA/ATT/TTA/TAT
AG/GA/CT/T	AAG/AGA/GAA/CTT/TTC/TCT
AC/CA/TG/GT	AAC/ACA/CAA/GTT/TTG/TGT
GC/CG	ATG/TGA/GAT/CAT/ATC/TCA
	AGT/GTA/TAG/ACT/CTA/TCA
	AGG/GGA/GAG/CCT/CTC/TCC
	AGC/GCA/CAG/GCT/CTG/TGC
	ACC/CCA/CAC/GGT/GTG/TGG
	ACC/CCA/CAC/GGT/GTG/TGG
	GGC/GCG/CGG/GCC/CCG/CGC

2.5.2 Estrategias para detectar polimorfismos de tipo SNP/InDels en los unigenes de girasol de girasol

Con el fin de desarrollar marcadores funcionales se realizaron ensayos para la identificación de polimorfismo de nucleótido simple (SNP), o por inserciones y deleciones de nucleótidos (InDels) en las secuencias de EST depositadas en bases de datos públicas. Para ello se ensayó una estrategia que permite la identificación de sitios de restricción diferenciales entre las secuencias analizadas para producir marcadores de tipo CAPS (*Cleaved Amplified Polymorphic Sequences*). Se seleccionaron secuencias de ESTs que presentaban similitud a genes de función conocida, especialmente aquellos categorizados como genes candidatos en respuestas a estrés biótico y abiótico, incluyendo factores de transcripción, componentes del sistema de transducción de señales, proteasas, proteínas de defensa, enzimas detoxificantes, etc., (Fernández et al., 2003).

Para el reconocimiento de SNPs se utilizó el programa SNP2CAPS (Thiel et al., 2004), el cual utiliza un algoritmo simple para la detección de sitios de restricción en secuencias alineadas, usando como referencia el conjunto de enzimas de restricción comerciales en la base de datos REBASE (www.biochem.ucl.ac.uk/bsm/dbbrowser/protocol/rebasefrm.html).

2.6 Condiciones de amplificación de fragmentos mediante la reacción en cadena de la polimerasa (PCR)

2.6.1 Condiciones de amplificación de los marcadores moleculares

La amplificación de los microsatélites se realizó por medio de la técnica de reacción en cadena de la polimerasa (PCR) utilizando diferentes termocicladores, PTC100 96U (MJ Research, Watertown, EUA) y Mastercycler ep-gradient Eppendorf (Eppendorf, RU). El volumen final de la reacción fue de 12 µl conteniendo:

Componentes	[final]
ADN	30 ng
Buffer de Taq Polimerasa 10X	1X
MgCl ₂	1,5 mM
Oligonucleótidos	0,25 µM
dNTPs	0,2 mM
Taq Polimerasa	0,75 UE
H ₂ O bidestilada	Llevar a 12 µl

Para la amplificación de los marcadores moleculares se utilizó el método de disminución progresiva de la temperatura de hibridación o *annealing*, llamado *touchdown* en un rango de temperaturas desde 63 a 50° C (Paniego et al., 2002). En los termocicladores mencionados arriba se utilizaron los siguientes programas:

Paso	Temperatura	Tiempo min:seg	Volver Al paso	Repetir #	D[° C]/dT
1	94° C	04:00			
2	94° C	00:30			
3	63° C	00:30			-1
4	72° C	01:00	2	13	
5	94° C	00:30			
6	50° C	00:30			
7	72° C	01:00	5	30	
8	72° C	10:00			
9	4° C	00:00			

Las condiciones de amplificación en los parentales de la población de mapeo fueron ajustadas para cada combinación de oligonucleótidos iniciadores a probar, utilizando gradientes de temperatura y distintas concentraciones de los productos de reacción, para luego utilizar la combinación óptima.

2.6.2 Reacciones de amplificación para uso de sondas específicas

Las reacciones de amplificación se llevaron a cabo en un termociclador, Mastercycler ep-gradient Eppendorf (Eppendorf, RU). El volumen final de cada reacción fue de 25 ó 50 μ l, conteniendo los siguientes componentes:

Componentes	[final]
ADN	de 30 ng a 100 ng
Buffer de Taq Polimerasa 10X	1X
MgCl ₂	de 1,5 mM a 2 mM
Oligonucleótidos	de 0,25 μ M a 0,4 μ M
dNTPs	de 0,1 mM a 0,2 mM
Taq Polimerasa	de 0,4 UE a 1,5 UE
H ₂ O bidestilada	Llevar a 25 o 50 μ l

Todos los reactivos fueron de la firma Invitrogen (Invitrogen, EUA).

Las condiciones de amplificación fueron ajustadas para cada combinación de oligonucleótidos a probar, utilizando distintas concentraciones de los productos de reacción para luego utilizar la combinación óptima.

El termociclador se programó con una desnaturalización inicial de 3 min a 95° C y 35 ciclos cada uno, con una desnaturalización de 30 seg a 94° C, en donde la temperatura de hibridación se estableció de acuerdo a la composición de bases de los oligonucleótidos utilizados, aplicando la siguiente relación para su cálculo $T_m = 2x(A+T)+4x(G+C)-5^\circ C$ durante 30 seg, extensión de 1 min 30 seg a 72° C y una extensión final de 10 min a 72° C.

2.7 Electroforesis y visualización de ADN en geles de agarosa

Los productos de PCR fueron separados por electroforesis en geles de agarosa 1% (p/v), empleándose *buffer* TAE (Tris-Acetato 20 mM, pH 8,5, EDTA 1 mM), tanto para la preparación del gel como para la solución de corrida. La concentración de bromuro de

etidido utilizada fue de 0,3 µg/ml y se adicionó 1/10 vol de solución de siembra (azul de bromofenol 0,25% (p/v) y/o xilen-cianol 0,25% (p/v)). El gel se corrió a voltaje constante entre 1 y 5 V por cm de gel. Para la visualización del ADN en el gel se empleó un transiluminador de luz UV. En todos los casos en los que fue necesario estimar la longitud de los fragmentos de ADN separados mediante electroforesis en geles de agarosa, se empleó como marcador 1 Kb Ladder y/o 1 Kb Ladder Plus (Invitrogen, EUA) según recomendaciones del fabricante.

2.8 Purificación de fragmentos de ADN de geles de agarosa

Una vez corridas las muestras de ADN mediante la técnica de electroforesis en agarosa e identificadas mediante su visualización con luz UV, las bandas correspondientes a los fragmentos esperados se aislaron y se purificaron a partir del gel utilizando el kit comercial Gel Band Purification Kit (Amersham Pharmacia Biotech, Inc., RU) o Qiaex II gel extraction Kit (Qiagen, Alemania), y/o se purificaron directamente a partir de la reacción de PCR utilizando el kit comercial Qiaquick PCR Purification Kit (Qiagen, Alemania).

2.9 Electroforesis y visualización de los fragmentos en geles de poliacrilamida

Los fragmentos de ADN amplificados se separaron por electroforesis en geles de 6% acrilamida-bisacrilamida (19:1) con urea 7 M, *buffer* TBE 0,5X (Tris borato 9 mM, EDTA 0,1 mM, persulfato de amonio 0,05 (m/v) y 25 µl de TEMED [catalizadores]). Se corrieron a 22 mA con un *buffer* de corrida TBE 1X/ TBE 0.5X entre 3 y 4 horas, dependiendo del tamaño del microsatélite. Se adicionó a 5 µl de cada muestra *buffer* de carga (98% de formamida, EDTA 10 mM y colorantes de azul de bromofenol y xilen-cianol [0,5 mg/ml]) y se desnaturalizaron a 95° C por 5 min. Las muestras se colocaron inmediatamente en hielo y se sembraron en el gel.

2.9.1 Métodos manuales

Luego de la corrida, los fragmentos fueron visualizados con nitrato de plata, SYBR®Gold (Molecular Probes, Eugene, EUA) o con marcación con radioisótopos.

La tinción con AgNO_3 se realizó siguiendo las instrucciones del proveedor (Silver Sequence™ DNA Sequencing System, Promega, EUA). El proceso de tinción consistió en la fijación en ácido acético 10% durante 15 min, para remover los restos de *buffer* de electroforesis y la urea, y para evitar la difusión de los pequeños productos de amplificación. Luego, el gel se enjuagó con agua ultrapura para eliminar el ácido acético y la urea junto con el Tris/borato. A continuación, el gel se tiñó durante 30 min en una solución de AgNO_3 0,1% (m/v) y formaldehído 0,056% (m/v). Posteriormente, se lavó durante 3 seg en agua fría para remover el exceso de plata e inmediatamente se reveló con una solución alcalina de carbonato de sodio 3% (m/v), formaldehído 0,056% (m/v) y tiosulfato de sodio 0,2 mg/ml previamente refrigerada entre 4 -10° C. En estas condiciones de pH, los iones de la plata depositados sobre los ácidos nucleicos se reducen a iones metálicos por acción del formaldehído. El tiempo de revelado de las bandas de ADN se controló visualmente hasta obtener el nivel de tinción y contraste adecuado para cada gel, lo cual aproximadamente se corresponde con un tiempo de 15 min. La reacción de tinción se detuvo por inmersión del gel en ácido acético al 10%. La lectura del gel se realizó usando un transiluminador de luz blanca.

Otra técnica de tinción manual evaluada fue la tinción con SYBR® Gold (Molecular Probes, EUA) la cual se realizó siguiendo las instrucciones del proveedor. Para la preparación del reactivo se partió de una concentración inicial de 10.000X y se diluyó hasta una solución 1X en *buffer* TBE (89 mM Tris base, 89 mM ácido acético, 1 mM EDTA, pH 8,0), los geles fueron incubados durante 30 min con 350 ml de solución diluida (1X) a temperatura ambiente, en oscuridad y con agitación continua. La solución de tinción fue almacenada en oscuridad a 4° C y pudo ser rehusada de manera eficiente hasta 4 veces en un período de tiempo máximo de 30 días. La visualización y lectura de las bandas de ADN se realizó en un transiluminador de luz UV (FotoDyne, EUA y/o Biorad, EUA).

El tamaño de los alelos correspondiente a cada marcador se estableció en referencia al marcador de tamaño molecular de 25 pb (Invitrogen, EUA).

Para la marcación con métodos radioactivos se utilizó el radioisótopo $\gamma\text{-P}^{32}$ (NEN, PerkinElmer, EUA). El oligonucleótido sentido se marcó siguiendo el protocolo descrito a continuación para un volumen final de 1 μl por reacción: 0,8 μCi de ATP $\gamma\text{-P}^{32}$; 1X *buffer* de *Polynucleotide Kinase T4*; 0,3 U de *Polynucleotide Kinase T4* y 0,8

pmoles de oligonucleótido sentido. La reacción se realizó en un termociclador *MJ-Research Inc.* PT100 (Waltham, EUA) durante 30 minutos a 37° C, seguido de un ciclo de 5 minutos a 70° C con la finalidad de inactivar la enzima T4. Una vez finalizado el marcado se mantuvo la sonda a -20° C hasta el momento de realizar la reacción de PCR, la cual, junto con el programa de ciclado, se realizaron de la misma manera que lo explicado en el apartado 2.6.1.

Los productos amplificados fueron separados por electroforesis en geles de poliacrilamida desnaturalizante al 6%. Una vez finalizada la corrida electroforética, los geles fueron transferidos a papel tipo Whatman y cubiertos con film antiadhesivo (Resinite™, EUA). Los mismos fueron secados en un secador de geles (Gel-Dryer-583 BioRad, EUA) a una temperatura de 80° C durante 30 minutos y en condiciones de vacío. Una vez secos, los geles se colocaron en placas autorradiográficas de alta sensibilidad CP-BU (AGFA, Argentina) a -80° C o a temperatura ambiente entre 72 y 120 horas, dependiendo de la intensidad de la señal. Para el revelado de las películas se utilizó el revelador G-150 (AGFA, Bélgica) y las mismas fueron fijadas durante un minuto con G-334 (AGFA, Bélgica). Una vez terminado el proceso de revelado, las películas fueron secadas a temperatura ambiente para su posterior lectura. Las autorradiografías fueron analizadas usando un transiluminador de luz blanca.

2.9.2 Métodos automáticos

Para la automatización del proceso de genotipificación se establecieron técnicas basadas en electroforesis capilar (EC) usando oligonucleótidos marcados con grupos fluorescentes. Los fragmentos se evaluaron por EC utilizando un analizador de ADN (ABI 3130xl, Applied Biosystems, EUA). Los oligonucleótidos sentido se marcaron con los siguientes fluoróforos:

Nombre del fluoróforo	Color emitido	Emisión de fluorescencia
6-FAM	azul	535
HEX	verde	560
NED	amarillo	585
ROX	rojo	604

El filtro utilizado en el equipo reserva el color rojo (ROX) para el estándar interno de peso molecular el cual acompaña la corrida electroforética. Para los fragmentos de 50 a 500 pb, el estándar de peso molecular fue GeneScan ROX500 (-250), tiene 15 fragmentos de: 35, 50, 75, 100, 139, 150, 160, 200, 300, 340, 350, 400, 450, 490 y 500 pb. El fragmento de 35 pb debe corregirse de manera manual en el análisis, ya que migra anormalmente y distorsiona los resultados de la curva estándar (Figura 2.1). Cuando el tamaño del fragmento amplificado fue mayor o igual a 500 pb, el estándar de peso molecular utilizado fue GeneScan ROX 2500.

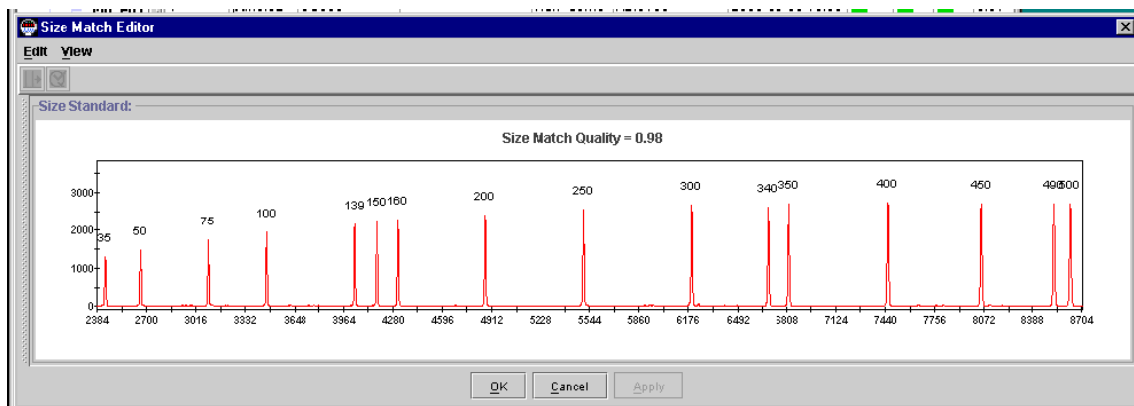


Figura 2.1 Fragmentos del estándar de peso molecular.

2.9.2.1 Marcación directa

Para la marcación directa los oligonucleótidos sentido se utilizaron los siguiente fluoróforos 6-FAM, HEX y NED.

Para explotar al máximo las posibilidades del sistema, ampliar la capacidad de procesamiento de muestras y hacerlo más económico, se tuvieron en cuenta dos aspectos importantes para definir la marcación de los oligonucleótidos y la preparación de las muestras para la electroforesis para una colección de microsatélites conocidos, de manera tal de reunir en cada panel un grupo de hasta seis marcadores compatibles, para aumentar la eficiencia de discriminación entre alelos en una sola corrida. Las consideraciones tomadas fueron:

- a) Se usaron los tres colores de fluoróforos disponibles (HEX, 6-FAM, NED) para el diseño de los oligonucleótidos a ensayar.

- b) Se combinaron los colores con los diferentes tamaños en pares de bases (pb) de los fragmentos amplificados, evitando la superposición de fragmentos del mismo tamaño y del mismo color.

La compatibilidad de los paneles diseñados fue comprobada experimentalmente a partir de ensayos de co-electroforesis. En todos los casos, antes de preparar los paneles de co-electroforesis, los fragmentos se evaluaron en geles de agarosa para ajustar las diluciones correspondientes, ya que es muy importante probar que la intensidad de la señal para los distintos fluorocromos se encuentre dentro del rango de fluorescencia adecuado (entre 200 y 4.000 RFU) y que no haya superposiciones de tamaños de fragmentos marcados con el mismo fluoróforo.

En este proceso se tuvieron en cuenta las distintas intensidades de emisión que presentan los fluoróforos utilizados. La señal emitida por 6-FAM es más alta que la emitida por HEX, y ésta que la emitida por NED.

Para el protocolo de preparación de las muestras para su resolución por EC se realizó una optimización de la corrida con pocas muestras en una primera etapa, y luego, en una segunda etapa, se realizó el ensayo con el total de las muestras a analizar.

Los fragmentos resueltos por EC fueron analizados de manera automática utilizando el programa GeneMapper® 4.0 (Applied Biosystem, EUA).

2.9.2.2 Marcación indirecta

Se evaluó un método alternativo de marcado indirecto descrito por Zhang et al. (2003), el cual implica el agregado de una secuencia universal en el extremo 5' del oligonucleótido sentido de hasta 19 nucleótidos, generalmente correspondiente a la del oligonucleótido universal M13. La reacción de amplificación se llevó adelante con tres oligonucleótidos, el sentido modificado, el antisentido y el universal M13 marcado con una modificación fluorescente en el extremo 5' (Figura 2.2). La ventaja de esta metodología es que solamente se necesita disponer de distintas versiones de un sólo oligonucleótido marcado con los fluoróforos 6-FAM, HEX y NED, con lo cual se abaratan los costos. Para llevar adelante estos ensayos se seleccionó un grupo de 139 microsatélites con diferente calidad de amplificación en reacciones de PCR, cuyos oligonucleótidos sentido fueron modificados incorporando una secuencia de 19

nucleótidos de M13. Las condiciones de reacción fueron las estándares descritas en el apartado 2.6.1 excepto por el agregado en la mezcla de reacción 0,02 μM de oligonucleótido M13 marcado.

Una vez obtenidos los fragmentos fueron analizados siguiendo el mismo procedimiento descrito para los fragmentos obtenidos por marca directa (ver apartado 2.9.2.1).

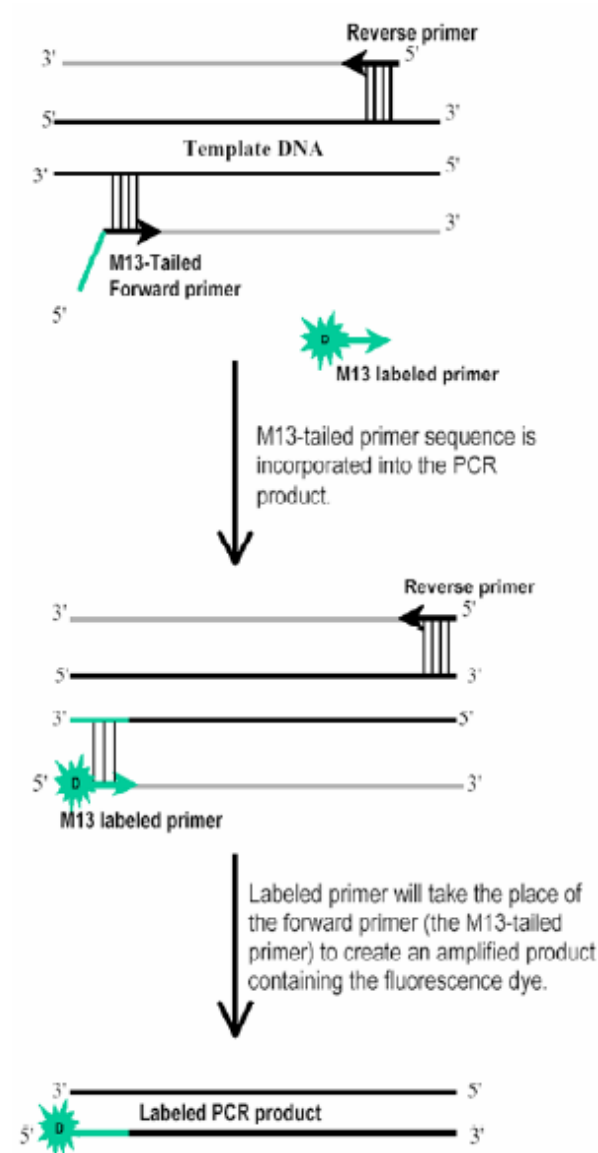


Figura 2.2 Ejemplificación del método de marcado indirecto (Vallumen, 2006).

2.10 Análisis de microsatélites usando el programa GeneMapper® 4.0 (Applied Biosystem, EUA)

El primer paso en el análisis de microsatélites usando el programa GeneMapper® 4.0 (Applied Biosystem, EUA) es la creación de un proyecto dentro del cual se define un grupo de paneles llamado “kit”, donde cada “panel” corresponde al grupo de marcadores que componen el conjunto de co-electroforesis definido previamente a la corrida por EC. Para cada uno de estos marcadores se completa una tabla que contiene el nombre, la longitud del fragmento en pb, el color del fluoróforo y el tipo de motivo de repetición.

A continuación se seleccionan los parámetros de análisis para el proyecto. Para esto, inicialmente se define el método de análisis y el estándar de peso molecular. A partir de allí, y trabajando sobre un “panel” determinado, se corrobora que la asignación automática de tamaños a los picos del estándar de peso molecular sea correcta; en el caso de encontrar discrepancias, las asignaciones se corrigen manualmente y se reanalizan los datos antes de continuar el análisis. El programa brinda, a su vez, indicadores de calidad de las secuencias (SQ), visibles como etiquetas de colores asignadas a diferentes parámetros, tal como lo indica la Figura 2.3.

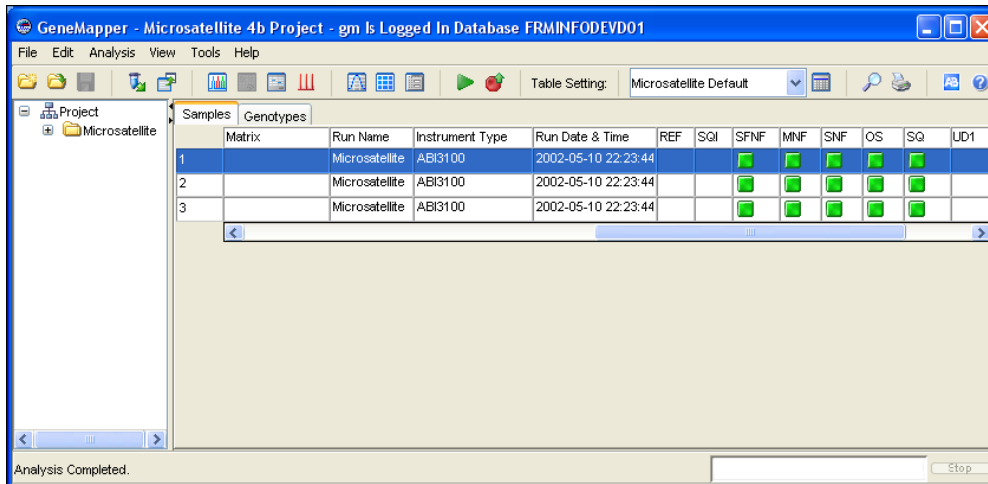


Figura 2.3 Parámetros de análisis del proyecto utilizando el programa GenMapper[®] 4.0 (Applied Biosystem, EUA). El programa asigna etiquetas de calidad de las secuencias (SQ), las cuales son determinadas por los siguientes parámetros:

SFNF (*Sample File Not Found*): indica que no se encuentra la muestra.

MNF (*Matriz Not Found*): indica que no se encuentra la matriz.

SNF (*Size Standart Not Found*): indica que no se encuentra el estándar de peso molecular.

OS (*Offscale*): indica que la señal esta fuera de la escala establecida.

Referencias: verde: buena calidad de la corrida; amarillo: se debe realizar un chequeo manual; rojo: corrida de baja calidad.

Luego, se identifican los alelos correspondientes a cada marcador y se define un “bin” (región del espectro definida para registrar la señal de un alelo) por alelo dentro del proyecto creado inicialmente. Cada bin entonces representa un alelo del marcador con dos atributos: el tamaño del fragmento en pb y el color del fluoróforo con el cual se marcó el oligonucleótido sentido correspondiente al marcador analizado (Figura 2.4 a y 2.4 b).

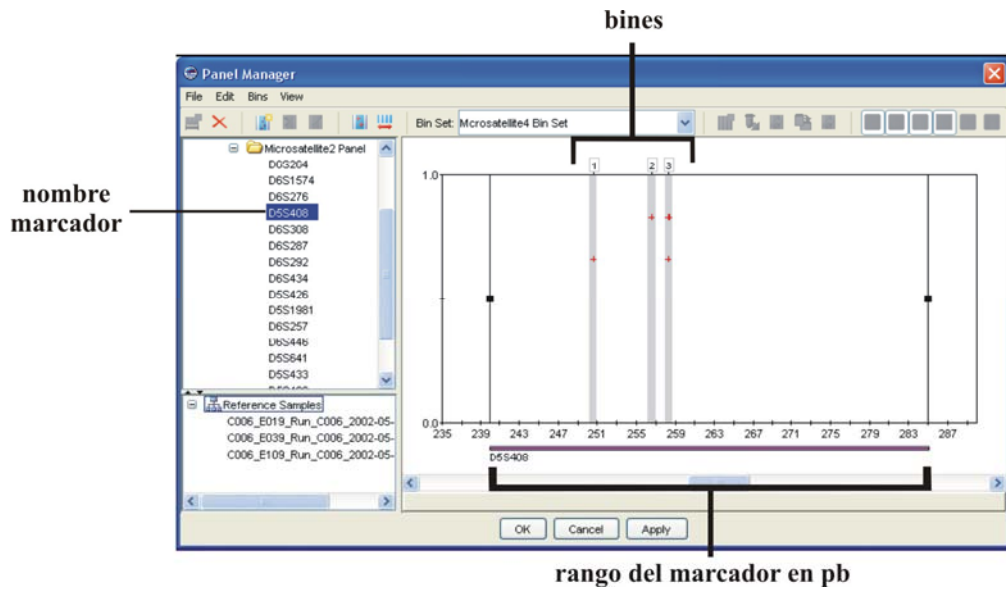


Figura 2.4a Alelos indicados con los bins y el rango en pb del marcador seleccionado.

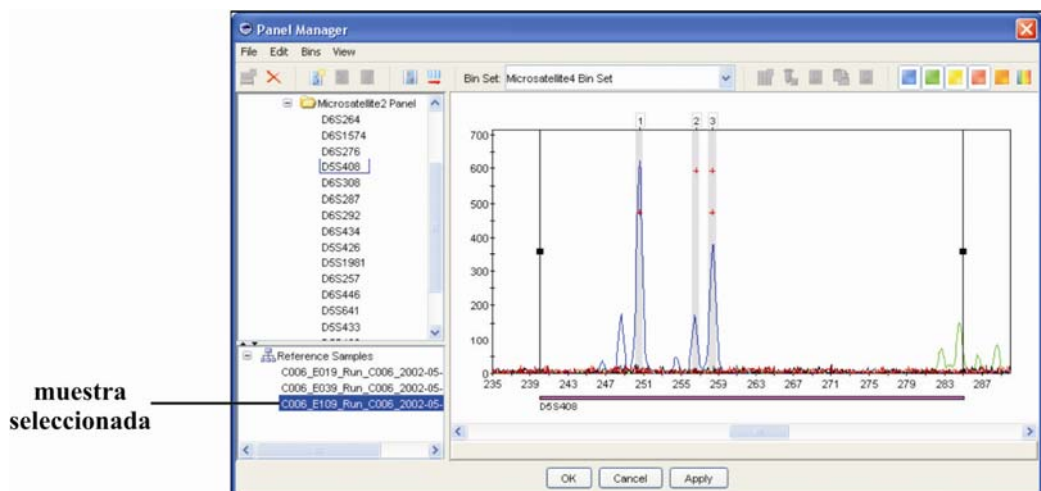


Figura 2.4b Alelos indicados como picos para la muestra seleccionada y el rango en pb del marcador seleccionado.

Para examinar los resultados y evaluar de manera conjunta los genotipos analizados en el proyecto, se utiliza la función Genotype Plot del programa, la cual permite visualizar los resultados de hasta ocho individuos simultáneamente. En el ejemplo se visualizan los gráficos correspondientes a dos muestras (Figura 2.5).

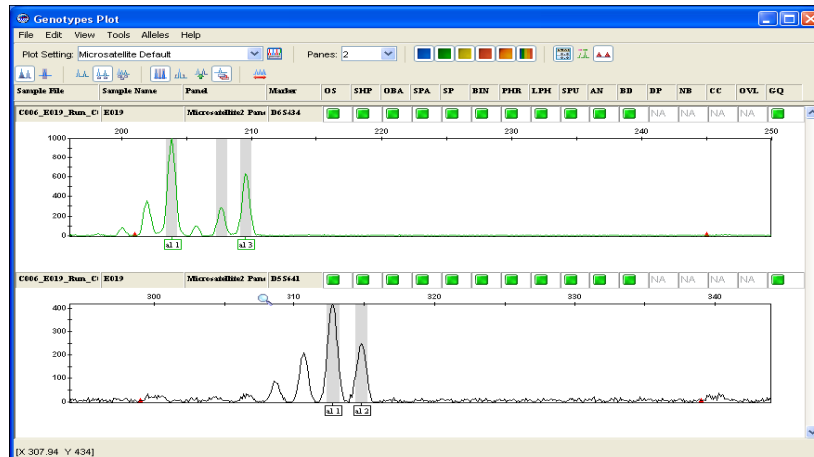


Figura 2.5 Visualización simultánea de dos genotipos en el Genotype Plot. En la parte superior del gráfico se observa una grilla horizontal donde figuran los valores de calidad (QV) de los genotipos, ellos son:

OS (*Offscale*): un pico está fuera de la escala del rango del marcador.

SHP (*Sharp Peak*): conjunto de picos más largos y angostos que el conjunto de picos analizados, sólo se utiliza en dinucleótidos.

OBA (*One Basepair Allele*): sólo se utiliza en humanos.

SPA (*Single Peak Artifact*): muestra un sólo pico debido a errores de la electroforesis.

SP (*Split Peak*): corrimiento del pico en menos de 0,25 pb sólo se utiliza en dinucleótidos.

BIN (*Out of bin allele*): aparece cuando el ancho del pico se encuentra fuera del rango establecido. El rango más utilizado es del 20%.

PHR (*Peak Height Ratio*): dos alelos están presentes y la proporción entre el alelo de mayor tamaño con respecto al de menor tamaño es menor a cierto nivel. Generalmente es 50%.

LPH (*Low Peak Height*): se muestra cuando los alelos son de menor tamaño que el valor especificado, los valores estándares son 200 para homocigotos y 100 para heterocigotos.

SPU (*Spectral Pull-Up*): indica que la fluorescencia del marcador esta fuera de escala lo cual provoca la emisión de otros colores del espectro.

AN (*Allele Number*): el número de alelos excede el máximo permitido para el organismo o no se encuentran alelos. El número de alelos se especifica en el método de análisis.

BD (*Broad Peak*): indica que el ancho del pico es más ancho que el especificado en el método de análisis.

DP (*Double Peak*): sólo se utiliza en el análisis de SNP por SNaPshot.

NB (*Narrow Bin*): indica que un pico está a 0,5 pb del bin, pero no necesariamente en el bin.

CC (*Control Concordance*): la muestra control no define exactamente al alelo.

OLV (*Overlap*): sólo se utiliza en humanos.

Una vez considerados los criterios de calidad descritos para cada corrida y, en los casos necesarios, habiendo realizado las ediciones correspondientes, los genotipos asignados por el programa pueden ser aceptados como confiables para continuar con el análisis de los mismos. Las tablas se exportaron como archivos de texto en el formato requerido para su posterior análisis con programas de mapeo.

2.11 Caracterización de los marcadores funcionales

2.11.1 Análisis de los marcadores EST-SSR

De los EST-SSR identificados, fueron seleccionados 127 para los cuales se sintetizaron oligonucleótidos específicos y se amplificaron en 4 líneas puras de girasol (Tabla 2.1). En primer lugar, fueron seleccionadas secuencias con repeticiones mayores o iguales a 20 pb, favoreciendo repeticiones dinucleótidos, seguidas por trinucleótidos, tetra y pentanucleótidos y que, además, presentaban distinto grado de similitud con secuencias de genes de función conocida, tales como aquellas relacionadas al metabolismo y/o estructura celular, transporte, proteínas de almacenamiento, mecanismos de respuesta a estrés biótico y abiótico, factores de transcripción, componentes del sistema de transducción de señales, etc. En una segunda etapa, se amplió el rango de longitud de la repetición a mayor o igual a 12 bases nucleotídicas.

Los fragmentos obtenidos por PCR se resolvieron por tinción con nitrato de plata (ver apartado 2.9.1) y todos aquellos que mostraron polimorfismo entre las líneas parentales RHA266 y PAC2, fueron genotipificados en 94 líneas recombinantes de la progenie RIL mediante la técnica de genotipificación automática descrita en el apartado 2.9.2.1.

2.11.2 Análisis de los SNPs por medio de la técnica de CAPS

Las secuencias seleccionadas de acuerdo al procedimiento descrito en el apartado 2.5.2, fueron amplificadas por PCR, el volumen final de la reacción de PCR fue de 50 μ l, conteniendo: 45 ng de ADN, 1X *Buffer* taq DNA polimerasa (Invitrogen, Argentina), CL_2Mg [1,5 mM], dNTPs [0,1mM], taq DNA polimerasa 0,5 UE y oligonucleótidos [0,3 mM]. Los productos de PCR obtenidos se resolvieron mediante electroforesis en geles de agarosa al 1%, teñidos con bromuro de etidio. Aquellos que presentaban una única banda se purificaron con el sistema comercial Exosap-it® (USB, EUA), y los productos que presentaban dos o más bandas fueron purificados por medio de elusión de la banda mayoritaria utilizando el kit comercial Quiaex II (Qiagen, Alemania). Luego, los productos purificados fueron cuantificados y secuenciados en ambos sentidos. Las secuencias obtenidas fueron editadas eliminando regiones de baja calidad y/o contaminantes, se les asignaron valores de calidad usando el algoritmo Phred

(www.phrap.com/phred) y se confirmó la identidad del gen candidato buscado por comparaciones en bases de datos públicas. En total, se secuenciaron 25 secuencias de ESTs en ambos padres de la población de mapeo.

Para validar los polimorfismos reconocidos *in silico*, se usaron las enzimas de restricción correspondientes para digerir los productos de amplificación obtenidos a partir de las líneas parentales de la población de mapeo. Los productos de digestión se analizaron por medio de electroforesis en geles de agarosa 2%, agarosa Metaphor (FMC bioproduct, EUA), poliacrilamida al 6% y 8% y se visualizaron con bromuro de etidio, SYBR[®]Gold y/o nitrato de plata, como se describió anteriormente.

2.11.3 Identificación de SNP e InDels mediante la técnica de resecuenciación de genes candidatos a partir de colecciones de ESTs.

Los marcadores InDels genotificados fueron identificados por la Lic. Corina Fusari, en un trabajo complementario a este proyecto (Fusari et al., 2008).

2.12 Análisis estadístico de los datos obtenidos

2.12.1 Análisis de segregación de los alelos en la población estudiada

Doscientos veinticuatro marcadores informativos fueron genotificados en las líneas parentales y en 94 RILs de la población RHA266 x PAC2.

Se construyó una matriz de genotipos para las RILs analizadas, los alelos derivados del parental RHA266 fueron identificados como alelo “A”, y los alelos derivados del parental PAC2 fueron identificados como alelo “B”. Los datos fueron analizados mediante la detección de diferencias en los alelos por *locus*.

Independientemente del método de genotificación utilizado, los análisis fueron realizados en forma independiente por dos operadores.

La segregación de cada alelo del marcador en la progenie se evaluó utilizando el programa GQMOL (Schuster & Cruz, 2004) disponible en: (www.ufv.br/dbg/gqmol/gqmol.htm) mediante el test FDR (*False Discovery Rate*), el cual propone controlar la razón de los falsos descubrimientos. Así, FDR es la proporción de hipótesis nulas rechazadas $H_0 =$ “La segregación de los marcadores se

ajusta a la relación 1:1”, que es la proporción de falsos negativos (Benjamín & Höcheberg, 1995) considerando un nivel de significancia global del 5%.

Los datos obtenidos del análisis de la segregación de los alelos entre los individuos de la población estudiada fueron incorporados a una matriz de datos de AFLP disponible para la población de trabajo (Gentzbittel et al., 1999) y cedida a nuestro grupo en el marco de la colaboración establecida.

Para una población de líneas recombinantes endocriadas la segregación esperada es 1:1. Debido a que el análisis del conjunto de marcadores implica una prueba múltiple y para el caso de SSR ligados, las pruebas no son independientes, se utilizó el procedimiento FDR (Benjamin & Höcheberg, 1995), este método propone controlar la razón de los falsos descubrimientos, siendo FDR la proporción de H_0 verdaderas entre las H_0 rechazadas; que es la proporción de falsos negativos (Q) (Benjamín & Yekutieli, 2005). Para determinar el punto de corte en test múltiples, mediante el control de FDR, son necesarios los siguientes pasos:

- a) Obtener el valor de chi-cuadrado H_0 para cada marcador, para una segregación 1:1 y el valor de probabilidad (P) asociado a éste.
- b) Ordenar los valores de $P_{(i)}$, asociado a cada $H_0(i)$ de las pruebas obtenido en el paso anterior.
- c) Definir el nivel de significancia global (q) a ser utilizado para controlar la FDR, siendo $q = N \times P_{(i)} / i$, donde N es el número de pruebas realizadas, $P_{(i)}$ es la probabilidad de test de chi-cuadrado e i es el número de hipótesis dado.
- d) Determinar el mayor i para el cual (a) $q^* \leq N \times P_{(i)} / i$, siendo q^* el valor de significancia del experimento.
- e) Las H_0 con valor de P crítico mayor a la P_i que corresponda a la expresión (a), deben ser rechazadas.

2.12.2 Utilización de herramientas bioinformáticas para el desarrollo del mapa genético

El uso de un gran número de marcadores en un mapa de ligamiento acarrea gran complejidad en el análisis de los datos observados. Esta tarea es facilitada con el uso de programas informáticos. En este trabajo se utilizaron los programas MapMaker 3.0

(Lander et al., 1987) y Carthagene 0.999 (Schiex & Gaspin, 1997). El programa MapChart 2.1 (Voorrips, 2002) fue usado para la presentación gráfica de los grupos de ligamiento y la posición de los marcadores en el mapa.

En la construcción del mapa se utilizó el programa Carthagene 0.999 (Schiex y & Gaspin., 1997) para definir la distribución de los marcadores en los distintos grupos de ligamiento y el orden de los *loci* dentro de cada grupo. La función de mapeo Kosambi (Kosambi, 1944) fue usada para calcular las distancias de mapa en cM a partir de las frecuencias de recombinación. Las frecuencias de recombinaciones estimadas en las poblaciones de mapeo RILs fueron corregidas por el factor que considera meiosis múltiples $r = R / (2-2R)$, donde $R = n_r / (n_r + n_n)$ es la frecuencia de recombinación y n_r y n_n son el número de genotipos recombinantes y no recombinantes, respectivamente (Haldane & Waddington, 1931). Los *loci* fueron agrupados dentro de cada grupo usando un valor mínimo de LOD (*likelihood odds*) score de 4,0 y una frecuencia de recombinación máxima de 35. Una vez definidos los grupos de ligamiento que componen el mapa, el orden dentro de cada grupo de ligamiento fue analizado de manera individual usando el programa Mapmaker 3.0 (Lander et al., 1987). Este programa también se utilizó para incluir marcadores no ligados en cada grupo de ligamiento a valores de LOD 4, disminuyendo para ello el valor hasta LOD 3.

2.12.3 Distribución de los marcadores moleculares en el mapa genético

La distribución de los marcadores dentro de los grupos de ligamiento fue estimada utilizando el método estadístico llamado test de rachas (*runs test*), el cual es una herramienta útil para determinar si una serie de ítems o de eventos son producidos al azar (Daniel, 1978).

De esta manera, para evaluar si los marcadores SSRs se distribuyen al azar, cada GL fue dividido en rachas (intervalos) de acuerdo al orden de aparición de los marcadores. Fueron consideradas dos categorías: n : número de observaciones; n_1 = número de observaciones de SSRs y n_2 = número de observaciones de AFLPs. Siendo en todos los casos H_0 : “Los marcadores se distribuyen al azar en los grupos de ligamiento”. De esta manera, se obtiene un valor r = número de rachas y valores críticos del test estadístico; si r es inferior o igual al valor crítico menor o mayor, o igual al valor crítico mayor, se

rechaza la H_0 para un nivel de confianza de 0,05. Con valores de n_1 y n_2 mayores a 20 no se pueden usar las tablas desarrolladas por Swed & Eisenhart (1943) para probar la hipótesis, sino que deben considerarse como una distribución normal según la siguiente ecuación:

$$z = \frac{r - \left\{ \left[\frac{2n_1n_2}{n_1 + n_2} \right] + 1 \right\}}{\sqrt{\frac{2n_1n_2(2n_1n_2 - n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - n_1)}}}$$

Luego, se comparó el valor obtenido de z con el valor de z tabulado para un valor de significancia de $\alpha = 0,05$.

2.13 Preparación de sondas de copia única y de ADN repetitivo para hibridación *in situ*

Se prepararon sondas de ADN repetitivo, ADN ribosomal 26S, y sondas específicas diseñadas a partir de genes relacionados con mecanismos de defensa a patógenos de importancia agronómica para girasol, como *Sclerotinia sclerotiorum*, de secuencias correspondientes a Análogos de Genes de Resistencia a patógenos (RGA), descritos para *Plasmopara halstedii*, y de microsatélites asociados a QLTs de interés en distintos mapas genéticos.

Las sondas correspondientes a ADN repetitivo se utilizaron para optimizar la técnica FISH en girasol, así como para la caracterización del complemento cromosómico. Las sondas específicas fueron evaluadas para la localización física de regiones de interés y para ensayar las técnicas de PRINS (*Primed in situ labeling*) y C-PRINS (*Cycling-Primed in situ labeling*) como herramienta para la integración del mapa genético y físico de girasol.

481 TTGACCTTAATGCCGTTTACCAATAATAAGAAGTCAAATCTAGATTTGTCAAAGCTTTG
 541 GTAAATAAGTCGGCACGTTGGTCATCGGTGTGGACCTTAACAACATCGATTAGCCTTTTC
 601 TCAAAGCAATCACGTATGAAGTGATATTTGATTTTCGATGTGTTTGGTCTTTGAATGCTGC
 661 ACAGGATTTCTAGTGATATCTAAAGCAGCAGAATTATCAACGTAAATAGGAGTAGTTAGG
 721 AATTCAAACCGTAGTCCCGCAATTGTTGTTG

Figura 2.9 Secuencia completa del retrotransposón de girasol perteneciente a la familia *copia-like*, AN AJ00967. En negrita se muestra el fragmento amplificado y usado como sonda. En rojo se indican los oligonucleótidos utilizados para amplificar el fragmento (Tabla 2.3).

Tabla 2.4 Regiones amplificadas con similitud a retrotransposones. La tabla detalla el nombre, tamaño y la secuencia de los oligonucleótidos seleccionados.

Nombre de la secuencia amplificada	Tamaño de la secuencia amplificada	Secuencia del oligonucleótido sentido	Secuencia del oligonucleótido antisentido	Referencias
SSR-Ha785	239 pb	5' GTCTCGGGATCCAAGATTGA 3'	5' TCACGACGGTTGTAAAACGA 3'	Paniego et al., (2002)
Retro1 (<i>gypsy-like</i>)	865 pb	5' TGATGCAATTCTGCAACGAT 3'	5' AGGAGCTGGTAGCAGTGGAA 3'	Tang et al., (2006)
Retro2 (<i>gypsy-like</i>)	847 pb	5' AGCTCATTGTCAGCTCCTGT 3'	5' GATGCCCTCTTGCATTA 3'	Tang et al., (2006)
Retro3 (<i>gypsy-like</i>)	538 pb	5' AGGGCATTCAAATGGCTATG 3'	5' GTCTCATCCGGAAGATCCAA 3'	Tang et al., (2006)
Retro4 (<i>gypsy-like</i>)	969 pb	5' TCGGTCCCTTGGTGAGATAC 3'	5' ACGGGTCAAATCAGTTCAGG 3'	Tang et al., (2006)
Rep1 (<i>copia-like</i>)	170 pb	5' TCTCAGAACCTCGGCAATCT 3'	5' GGCAGCAAAGAGAAAATG 3'	Santini et al., (2002)

2.13.1.3. Utilización de secuencias clonadas en Cromosomas Artificiales de Bacterias (BACs) conteniendo ADN repetitivo como sondas para la caracterización del complemento cromosómico

Para la identificación del complemento cromosómico se utilizó la técnica de BAC-FISH. El clon de BAC utilizado como sonda en experimentos de hibridación *in situ*, (BAC772), fue aislado de una genoteca de BACs para *H. annuus* (BAC: HA HBa. www.genome.clemson.edu) utilizando como sonda la secuencia (SSR- Ha785). El clon BAC772 presenta similitud a una secuencia altamente repetitiva de *H. annuus*, (AN AJ009965) (ver Figura 2.7). Este clon fue amplificado y purificado utilizando el kit

comercial Qiagen® Large-Construct Kit (Qiagen, Alemania), previo a su utilización como sonda en experimentos de BAC-FISH

La genoteca BAC usada en este estudio fue construida por la Universidad de Clemson a partir del cultivar HA383 de *H. annuus*. Contiene 202.752 clones ordenados en 528 placas distribuidos en 11 filtros, y representa 8,3 veces el tamaño del genoma. Los clones están distribuidos por duplicado en la genoteca, con un orden determinado como se observa en la Figura 2.10.

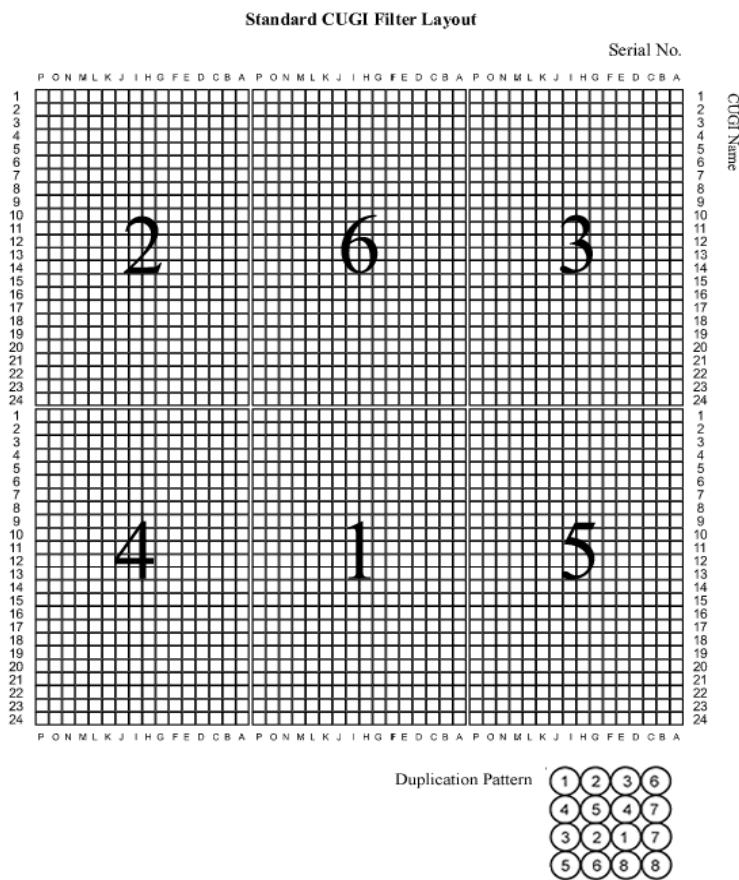


Figura 2.10 Distribución de los arreglos por duplicado del total de los clones que componen la genoteca.

2.13.2 Preparación de sondas de copia única

2.13.2.1 Preparación de sondas a partir de RGAs

Se amplificaron seis sondas específicas, las cuales fueron diseñadas por Radwan et al. (2003; 2004) a partir de los llamados genes *Pl*, los cuales controlan la resistencia a *P. halstedii*, agente causal de mildiu en girasol, a partir de secuencias con similitud a RGAs de las clases non-TIR-NBS-LRR (Radwan et al., 2003) y CC-NBS-LRR (Radwan et al., 2004). Los oligonucleótidos denominados HaNTP3, HaNTP5, HaNTP6 fueron diseñados a partir de secuencias de RGAS de girasol descritas por Gedil et al. (2001a), pertenecientes a la clase non-TIR-NBS-LRR y depositadas en la base de datos del Genbank AN: AF272766, AF272768, AF272769 (Tabla 2.4). Las reacciones de amplificación se llevaron a cabo con 50 ng de ADN, 0,2 mM de cada dNTP, 0,4 UE de *Taq* ADN polimerasa (Advantage 2, Clontech, EUA), 1X *Buffer* de *Taq* ADN polimerasa y 0,4 mM de cada oligonucleótido. Mediante el siguiente programa: 95° C 3 min de desnaturalización inicial, 35 ciclos a 94° C 10 seg, 58° C por 30 seg y 72° C durante 1 min 30 seg.

Los oligonucleótidos llamados Ha-P1, Ha-P2 y Ha-P3 fueron diseñados por Radwan et al. (2004) a partir de la secuencias de RGAs de girasol Ha-NTIR11g y Ha-NTIR3A (AN: AY490793 y AY490791 respectivamente) (Tabla 2.5). Las reacciones de amplificación se llevaron a cabo con 50 ng de ADN, 0,2 mM de cada dNTP, 0,4 UE de *Taq* ADN polimerasa (Advantage 2, Clontech, EUA), 1X *Buffer* de *Taq* ADN polimerasa y 0,4 mM de cada oligonucleótido. Mediante el siguiente programa: 95° C 3 min de desnaturalización inicial, 35 ciclos a 95° C 30 seg, 60° C por 30 seg y 72° C durante 2 min.

Los fragmentos amplificados fueron purificados a partir del gel utilizando el *kit* comercial Qiaex II (Qiagen, Alemania) y clonadas en un T vector (pGemT easy vector, promega, EUA). La reacción de ligación se realizó en un volumen final de 10 μ l conteniendo: 2X *buffer* de “rápida ligación”, vector pGemT (50 ng), producto de PCR (en una relación 3:1 inserto: vector), T4 DNA ligasa 3 UE. La ligación se realizó por un período de 72 hs a 4° C. El producto de ligación se utilizó para transformar células competentes DH5 α por el método químico (Sambrook et al., 1989), y los clones recombinantes se seleccionaron por observación de la actividad del producto del gen β -

galactosidasa en presencia de X-gal e IPTG (Sambrook et al., 1989). Los clones recombinantes obtenidos se almacenaron en microtubos tipo Eppendorf de 5 ml con 500 μ l de glicerol 50% y con ampicilina (100 μ g/ml) a -80° C para su correcta conservación. El análisis de los mismos se realizó a través de la observación del patrón de restricción generado después de electroforesis del ADN plasmídico previamente digerido con la enzima de restricción *EcoRI* en un volumen final de reacción de 20 μ l conteniendo 10 unidades de enzima de restricción *EcoRI* (New England Biolabs, EUA) *buffer* 10X, durante 2 horas a 37° C. Alternativamente, se realizó una PCR utilizando oligonucleótidos complementarios a las secuencias promotoras T7 y SP6 presentes en el plásmido, en un volumen final de la reacción de 25 μ l, conteniendo: de 50 a 200 ng de ADN plasmídico, 1X *buffer* de Taq polimerasa 10X, 1 UE de Taq polimerasa (Life Technologies, EUA), 2 mM de $MgCl_2$, 0,1 mM de dNTPs, 0,4 μ M de oligonucleótido T7 y 0,4 μ M de oligonucleótido SP6. Los productos de digestión (o amplificación) se analizaron por medio de electroforesis en geles de agarosa al 1%, revelando las bandas con bromuro de etidio. Los productos obtenidos fueron purificados utilizando el sistema comercial prep-A-Gene-DNA Purification System (Biorad, EUA). Estos clones fueron secuenciados utilizando el servicio de secuenciación del Instituto de Biotecnología, INTA Castelar, utilizando los oligonucleótidos T7 y SP6. El análisis de las secuencias generadas se realizó utilizando el programa de comparación y alineación múltiple Bioedit (Hall, 1999). Se realizó el análisis de las secuencias obtenidas por comparación contra secuencias presentes en bases de datos públicas (Genbank) por medio de los programas: BLASTN y BLASTX (Altschul et al., 1990). Se seleccionaron 4 clones (Ha-P2, Ha-P3, HaNTP3 y HaNTP6 de 1.200, 1.150, 250 y 270 pb respectivamente) para ser usados como sondas para hibridación *in situ*.

Tabla 2.5. Secuencia de los oligonucleótidos con similitud a genes de resistencia seleccionados para su utilización como sondas en hibridación *in situ*.

Nombre del oligonucleótido	Subfamilia NBS-LRR	Secuencia del oligonucleótido sentido	Secuencia del oligonucleótido antisentido
HaNTP3	non-TIR	5' GAATATTGTATAACGATACACGAG 3'	5' TTCCAGTAGCCCTAGAATGAAATG 3'
HaNTP5	non-TIR	5' GACATTTATATAACGACGCACAAG 3'	5' TTCCATGGGCACATGAACGAAATG 3'
HaNTP6	non-TIR	5' GACTATTGTACAATGAAAAGCAAG 3'	5' TTCCAGGAGCACATGCATGAAATG 3'
Ha-P1	CC (<i>coiled coils</i>)	5' GCCCAAATTGAAAGAAAGGTGTG 3'	5' GGCGAAATTGGTTCCCCTGAGTCG 3'
Ha-P2	CC (<i>coiled coils</i>)	5' AATCTTGAGTCATTACCCGAGC 3'	5' CAGCGTCTCTGGTAGATCGTTCACC 3'
Ha-P3	CC (<i>coiled coils</i>)	5' TAGTTAACCATGGCTGAAACCGCTG 3'	5' TTTGAAAGATAAGTTCGCCTCTCG 3'

2.13.2.2 Preparación de sondas correspondientes a marcadores SSR de localización única en el genoma de girasol

2.13.2.2.1 Diseño de sondas tipo overgo (*overlapping oligonucleotide probes*)

Las sondas *overgo* están compuestas de dos oligonucleótidos sintéticos que contienen una región complementaria de 8 pb (<http://tree.caltech.edu/protocols/picture/overgo.jpg>). Los dos oligonucleótidos de 24 pb que representan al overgo se solapan en una región de 8 pb dejando dos regiones de 16 pb libres a cada lado, las cuales son rellenadas con nucleótidos marcados usando la ADN polimerasa Klenow (Promega, EUA). Ambas cadenas de ADN son marcadas con [α - 32 P]-dCTP y [α - 32 P]-dATP (NEN, Perkin Elmer, EUA), de esta manera, se genera una sonda de 38-44 pb con alta actividad específica (Gardiner et al., 2004).

Una consideración importante en el diseño de las sondas overgo es la eliminación de cualquier secuencia repetitiva. Las secuencias de tipo overgo fueron examinadas manualmente para eliminar cualquier posible resto de *hairpins* y para comprobar que contengan al menos diez sitios de marcado para [32 P]-dATP y [32 P]-dCTP. Si en las 8 pb de la región que se solapa se encontraban menos de tres G-C, la longitud del solapamiento se incrementó a 9 pb para permitir la asociación estable entre los dos oligonucleótidos durante los pasos de marcado *in vitro*.

Las sondas fueron diseñadas a partir de marcadores SSR, EST-SSR e InDels anclados al mapa genético de referencia, usando los programas Overgo 1.02i (Cai et al., 1998, <http://mouse-genome.bcm.tmc.edu/webovergo/OvergoInput.asp>) y OligoSpawn (<http://oligospawn.ucr.edu/>). Se diseñaron treinta y dos sondas pertenecientes a los grupos de ligamiento 9, 10 y 16.

2.13.2.2.2 Marcación de sondas

Las sondas *overgo* fueron marcados con [α - 32 P]-dATP y [α - 32 P]-dCTP, siguiendo método descrito por J. D. McPherson (<http://www.tree.caltech.edu>), con pequeñas modificaciones.

Para tal fin, se desnaturalizó la mezcla de oligonucleótidos conteniendo 10 pmol de cada uno a 80° C durante 5 minutos, inmediatamente se colocó la mezcla a 37° C durante 10 minutos y se almacenó en hielo. Se realizó una reacción para cada par de oligonucleótidos utilizados; el volumen final de la reacción fue de 10 μ l, conteniendo:

Componentes	[final]
<i>Buffer</i> OLB (<i>Overgo Labelling Buffer</i>) 5X	1X
BSA (Albúmina de Suero Bovino)	0,1 mg/ml
Oligonucleótidos	10 pmol c/u
[α - 32 P]-dATP*	10 μ Ci/ μ l
[α - 32 P]-dCTP*	10 μ Ci/ μ l
<i>Klenow fragment</i> (Invitrogen, EUA)	2 UE
H ₂ O	Hasta 10 μ l

A continuación, la reacción se incubó a 37° C durante 1 hora. La remoción de los nucleótidos no incorporados se realizó utilizando el kit comercial “QiAquick® Nucleotide Removal Kit” (Qiagen, Alemania).

Las sondas se desnaturalizaron agregando 1 volumen de NaOH 0,4 N durante 10 minutos a temperatura ambiente.

La incorporación de deoxinucleótidos presentó un rango de 10 a 80%, de esta manera la actividad específica de la sonda overgo fue entre $0,5-2,5 \times 10^6$ dpm/pmol.

2.13.2.2.3 Hibridación, lavados y exposición de membranas

Los filtros de BACs fueron individualmente prehibridados en 15 ml de *Buffer* PAES (Acido polianetolesulfónico: 0,1% (p/v), pirofosfato de sodio 0,01% (p/v), SDS 1% y SSPE 2,5X) (Sambrook et al., 1989) de 4 a 16 horas a 58° C en un horno de hibridación con rotación continua (Biometra, Alemania), en botellas de hibridación.

Luego de la prehibridación, se realizó la hibridación utilizando 10 ml de solución PAES. Se adicionaron entre 10 y 12 sondas marcadas radioactivamente, previamente desnaturalizadas a 90° C durante 10 minutos, al centro de cada botella conteniendo cada una un filtro de alta densidad de clones de BAC. La hibridación se realizó a la misma temperatura que la prehibridación durante 16 horas.

Posteriormente, los filtros se lavaron a la misma temperatura de hibridación con 20 ml de solución 2X SSC /0,1% SDS durante 5 minutos en la misma botella. Luego fueron retirados de la botella y lavados en bandeja durante 15 minutos adicionales con 200 ml de la misma solución. La radioactividad final en cada membrana fue entre 20 y 50 cuentas/min, evaluadas con un contador Geiger. Cuando el nivel de radioactividad detectado en las membranas superaba a 200 cuentas/min se aplicaron lavados adicionales con solución 1X SSC /0,1X SDS durante 15 minutos a la misma temperatura de hibridación (Sambrook et al., 1989).

A continuación, los filtros de BAC se expusieron durante tres días en una pantalla de captación de imagen (Fuji film, Japón) y los patrones de hibridación fueron obtenidos usando el escáner Typhoon trio (Amersham Biosciences, RU) mediante el programa ImagenQuant (Amersham Biosciences, RU). Luego, las membranas se expusieron a placas autoradiográficas de alta sensibilidad (Biomax MS, Kodak) de 7 a 21 días según la actividad específica de las sondas y la intensidad de la señal esperada.

2.13.2.2.4 Identificación de clones positivos y confirmación por PCR

La identificación de los clones BAC positivos se realizó utilizando un transiluminador y el escáner de imágenes Typhoon trio (Amersham Biosciences, RU), seleccionándose clones por duplicado y de distinta intensidad. Los clones de BACs seleccionados fueron purificados utilizando el kit comercial NucleoBond® Xtra Midi Plus (Machery-Nagel, Alemania). La confirmación de los clones positivos se realizó por PCR, usando los oligonucleótidos específicos para cada microsatélite y utilizando como templado cada clon de BAC seleccionado.

2.13.2.2.5 BAC-Southern blot

Para evaluar la presencia de secuencias repetitivas en los clones BAC seleccionados, se realizó la digestión de 10 µg de ADN de cada uno de los clones durante 5 horas con la enzima *EcoRI*, utilizando 2,5 UE enzima/µg de ADN. Los productos de la digestión fueron sometidos a una electroforesis en geles de agarosa 0,8 % (m/v, en *Buffer* TAE 1X a 0,6 mA/cm) a un voltaje constante de 20 V por 17 hs. El ADN fue transferido durante 16 hs a una membrana Hybond N+ (Amersham Pharmacia Biotech, Inc.) siguiendo las instrucciones del proveedor. Luego se inmovilizó el ADN transferido a la membrana mediante tratamiento con luz UV (120.000 µjoules/cm²) y horneado a 80° C por 120 min.

Las membranas fueron hibridadas con las siguientes sondas de ADN repetitivo: (descriptas en el apartado 2.13.1.2):

a) La sonda SSR-Ha785, con similitud a una secuencia altamente repetitiva (AN AJ009965) b) la sonda Rep1, derivada de la secuencia AN AJ009967 de girasol, con similitud a retrotransposones de la familia *copia-like* (Santini et al., 2002) y c) la sonda Retro2, derivada de la secuencia AN DQ229838, con similitud a retrotransposones de la familia *gypsy-like* (Tang et al., 2006).

Estos fragmentos de ADN fueron purificados con el kit comercial Quiaex II gel extraction kit (Qiagen, Alemania) y marcados con [α^{32} P]-dCTP (actividad específica Ci/µg, New England Nuclear, EUA), utilizando el kit de marcación Prime-a Gene® Labeling System (Promega, EUA), el volumen final de la reacción fue de 50 µl conteniendo:

Componentes	[final]
ADN	100 ng
<i>Buffer</i> Klenow 5X	1X
dNTPs (dATP, dGTP, dTTP)	0,03 mM
[α - ³² P*]-dCTP	0,02 mCi
BSA	1X
Fragmento Klenow de la ADN Polimerasa	1 UE

La reacción se incubó 1 h a 37° C y luego se detuvo por inactivación de la enzima durante 10 min con NaOH 0,4 N. Durante este paso también se desnaturalizó la sonda marcada para poder incorporarla al *buffer* de prehibridación e iniciar la etapa de hibridación.

Las membranas se prehibridaron a 65° C durante 4 horas con una solución PAES y la hibridación se realizó durante 16 horas a 65° C, en un horno de hibridación con rotación continua, como se describiera en el apartado 2.13.2.2.3.

Posteriormente, las membranas se lavaron y se expusieron en una pantalla de captación de imagen (Fuji film, Japón), los patrones de hibridación fueron obtenidos usando el escáner Typhoon trio (Amersham Biosciences).

2.14 Preparados citológicos, hibridación mediante la técnica FISH y amplificación de regiones específicas por PRINS/C-PRINS

2.14.1 Obtención de preparados citológicos

2.14.1.1 Obtención y fijación del material vegetal

Se germinaron semillas de girasol en placas de Petri conteniendo algodón humedecido y papel de filtro hasta que las raíces alcanzaron una longitud de entre 1 y 2 cm (aproximadamente de 3 a 5 días). A continuación, las raíces se trataron con colchicina 0,05% con el fin de arrestar a los cromosomas en metafase mitótica, durante 2 hs a temperatura ambiente, y luego fueron fijadas en una solución 3:1 de alcohol absoluto: ácido acético glacial a 4° C durante un período de 48 hs.

Las raíces fijadas fueron lavadas con agua destilada durante 10 min y luego colocadas en *buffer* ácido cítrico-citrato de sodio 1X durante 10 min. A continuación fueron incubadas en solución enzimática (celulasa 2% (p/v) y pectinasa 20% (v/v) a 37° C durante 3 horas y colocadas en *buffer* ácido cítrico-citrato de sodio 1X durante 24 hs a 4° C.

2.14.1.2 Técnica de aplastado (squash)

El ápice radical de 2 a 3 mm de longitud se colocó sobre un portaobjetos con una gota de ácido acético 45%, se disgregó suavemente el material con una pinza, se colocó un cubreobjetos y se observó al microscopio de contraste de fase. En caso de observarse restos de citoplasma en las células, el portaobjetos se expuso al calor cuidadosamente a fin de eliminarlos. Una vez libre de citoplasma se procedió a la separación de los cromosomas por aplastado o *squash* para obtener preparados con células dispuestas en un solo plano. Los preparados se colocaron en hielo seco durante 60 min, se eliminaron los cubreobjetos, se secaron a temperatura ambiente y se conservaron a -20° C hasta el momento de su utilización.

2.14.2 Hibridación *in situ* (FISH)

La técnica de FISH fue realizada según Cuadrado & Jouve (1994). Los preparados se incubaron en alcohol: ácido acético 3:1 durante 10 min y luego se realizaron dos lavados de 10 min cada uno en alcohol absoluto.

2.14.2.1 Acondicionamiento de los preparados para FISH

Los preparados citológicos fueron pretratados con el fin de reducir la hibridación inespecífica. Estos pasos también ayudan a que la sonda y los reactivos de la detección penetren hasta el ADN, manteniendo la estabilidad de las secuencias blanco frente a los distintos tratamientos. Para ello, a cada portaobjeto se lo trató con 200 µl de RNAsa [100 µg/ml], y se lo cubrió con un cubreobjeto de plástico. Los preparados fueron incubados 1 h a 37° C en cámara húmeda.

Los preparados se lavaron en una solución 2X SSC a temperatura ambiente 3 veces durante de 5 min en agitación. A continuación, se incubaron en una solución de paraformaldehído al 4% (p/v) a temperatura ambiente durante 10 min en agitación, y se lavaron en 2X SSC a temperatura ambiente 3 veces durante 5 min en agitación.

Luego, se procedió a su deshidratación, para ello, los preparados se incubaron sucesivamente en distintas soluciones de alcohol (70%, 90% y 100%) durante 3 min cada uno a temperatura ambiente. Los preparados se dejaron secar a temperatura ambiente durante 2 hs.

2.14.2.2 Marcado de sondas para FISH

Las sondas se marcaron por la técnica de *random priming* o iniciación con oligonucleótidos al azar, con digoxigenina 11-dUTP o mediante *nick translation* o desplazamiento de mella con biotina 14-dUTP.

2.14.2.3. Hibridación

Las reacciones de hibridación se llevaron a cabo en un volumen final de 30 μ l, conteniendo los siguientes componentes:

Componentes	[final]
Formamida	50 %
Sulfato de dextrano	10 %
SSC 20X	2 X
Sonda	25-200 ng
DNA salmón [10 mg/ml]	0,3 mg/ml
DNA competidor (bloqueante)	2-100 X
SDS al 10 %	0,33 mg/ml
H ₂ O	Llevar a 30 μ l

Todos los clones de BAC, conteniendo la secuencia de interés, fueron hibridados en primera instancia sin utilizar ADN competidor, a fin de estimar la cantidad de ADN competidor necesaria para bloquear las señales repetitivas, la cual varió de 10 a 80 veces con respecto a la cantidad de sonda. Como ADN competidor se utilizaron sondas de ADN repetitivo de girasol sin marcar; a saber: SSR-Ha785 y/o Rep1 teniendo en cuenta el resultado de los ensayos de BAC-*Southern blot*.

La desnaturalización de la mezcla de hibridación se realizó a 70° C durante 15 min. A continuación, la mezcla se incubó en hielo durante 5 min y se centrifugó durante 10 seg. Se colocaron 30 µl de la mezcla de hibridación a cada preparado, los cuales habían sido previamente deshidratados y secados al aire. Luego, se colocó un cubreobjeto de plástico a cada preparado.

Los preparados fueron colocados en un termociclador (Eppendorf, Mastercycler, Alemania) utilizando el siguiente programa: 75° C durante 7 min; 55° C durante 30 seg; 45° C durante 1 min; 42° C durante 2 min; 40° C durante 5 min; 38° C durante 5 min y 37° C por tiempo indefinido. A continuación las muestras fueron incubadas en cámaras húmedas a 37° C en estufa, durante 16 horas.

2.14.2.4 Lavados post-hibridación

Para la remoción de la sonda libre, o sonda hibridada de manera inespecífica, se realizaron los siguientes lavados de manera secuencial con agitación:

- 1) 2X SSC, durante 10 min a 42° C
- 2) Formamida al 20% (v/v) en 0,1X SSC, durante 10 min a 42° C
- 3) 0,1X SSC, durante 5 min a 42° C
- 4) 2X SSC, durante 5 min a 42° C
- 5) 4X SSC /tween 0,2%, durante 5 min a 42° C
- 6) 4X SSC /tween 0,2%, durante 5 min a temperatura ambiente, dos veces.

2.14.2.5 *Detección visual de los sitios de hibridación*

Como bloqueante de los sitios donde se pueden unir los anticuerpos utilizados (anti-biotina y/o anti-digoxigenina) se utilizó de BSA al 5% (p/v) en solución 4X SSC/ tween 0,2%. Se añadieron 100 μ l de esta solución por preparado y se cubrió con cubreobjetos de plástico. Los preparados fueron incubados durante 5 min a temperatura ambiente. Posteriormente, se quitaron los cubreobjetos y se drenó la solución en exceso.

Para la detección de sondas marcadas con digoxigenina y/o biotina se adicionaron 50 μ l anti-digoxigenina FITC (isotiocianato de fluoresceína) en BSA al 5% (p/v) (1:50) y/o estreptoavidina-cy3 (1:300), respectivamente, a cada preparación. A continuación se colocó un cubreobjeto de plástico a cada preparado, los cuales, fueron incubados a 37° C durante una hora en cámara húmeda. Luego, fueron lavados tres veces en 4X SSC/ tween 0,2% durante 10 min.

A fin de realizar una contra tinción de los cromosomas se adicionaron 75 μ l de DAPI (4'6- diamidino-2-phenylindole) [2 μ g/ μ l], se colocó un cubreobjeto de plástico a cada preparado y se incubaron durante 15 min a temperatura ambiente. A continuación, los preparados se lavaron en 4X SSC/ tween 0,2%, se agregaron 30 μ l de medio de montaje (Vectashield, Vector laboratories, EUA) y se cubrieron con cubreobjetos de vidrio. Los preparados fueron conservados en oscuridad a 2° C hasta su observación.

2.14.2.6 *Obtención y análisis de imágenes*

Las células fueron observadas en un microscopio de epifluorescencia (Leica DMLB, Alemania), las fotos fueron tomadas con una cámara Leica DFC 350 FX y las imágenes se analizaron con el programa Adobe Photoshop CS2.

Los cariogramas fueron confeccionados utilizando el programa MicroMeasure 3.01 (<http://www.colostate.edu/Depts/Biology/Micromeasure>).

2.14.3 *Primed in situ labelling (PRINS)*

La hibridación por la técnica PRINS fue realizada según Menke et al. (1998) y desarrollada como parte de la tesis de Licenciatura de la Lic. Carolina Díaz (Díaz Quijano, 2007), dirigido por el Dr. Eduardo Greizerstein y la Dra. Lidia Poggio en el

Laboratorio de Citogenética y Evolución de la Facultad de Ciencias Exactas y Naturales de la Universidad Nacional de Buenos Aires, utilizando los oligonucleótidos diseñados y/o desarrollados en la presente tesis.

2.14.4 *Cycling-Primed in situ labelling (C-PRINS)*

La hibridación por la técnica C-PRINS fue realizada según Kubaláková et al. (2001) con modificaciones, según se detalla a continuación.

2.14.4.1 *Preparación de la mezcla de extensión*

La mezcla de extensión se preparó para un volumen final de 50 μ l, conteniendo:

Componentes	[final]
<i>Buffer de Taq Polimeras 10X</i>	1X
MgCl ₂	2,5 mM
dATPs	100 μ M
dCTP	100 μ M
dGTP	100 μ M
dTTP	34 μ M
Digoxigenina-11-dUTP o biotina-16-dUTP (<i>Roche Applied Science</i>)	8 μ M
Oligonucleótidos	0,8 μ M
Taq Polimerasa (Invitrogen, EUA)	2 UE
H ₂ O bidestilada	Llevar a 50 μ l

Los preparados se sellaron colocando un marco (Frame-sealTM Incubation Chambers, Bio-Rad, EUA). En la zona delimitada por el marco se agregaron 50 μ l de la mezcla de reacción y se colocó un cubreobjeto de plástico.

Los preparados se colocaron en un termociclador (Eppendorf Mastercycler, Alemania) siguiendo el programa detallado a continuación: 91° C durante 4 min, 35

ciclos de 91° C durante 45 seg, 55-60° C durante 45 seg y 72° C durante 45 seg y la extensión final a 72° C por 5 min.

2.14.4.2. Método indirecto de detección de la reacción

La reacción se detuvo adicionando 100 µl de *Stop buffer* (2,923 g de NaCl, 1,861 g de Na₂EDTA en 100 ml de agua destilada, pH 8,0). Los preparados se incubaron en estufa durante 2 min a 70° C. Se retiraron los marcos y a cada preparado se le adicionaron 100 µl de BSA y se colocó un cubreobjeto de plástico. Los preparados fueron incubados durante 30 min a 37° C en 50 µl de anti-Dig fluoresceína diluida en *buffer* de bloqueo (BSA); la solución que se utilizó fue 1/40 anti-dig en BSA al 5%.

Finalmente, los preparados fueron lavados 3 veces en 4X SSC/tween 0,2% durante 5 min a temperatura ambiente, y fueron contra teñidos con 75 µl de DAPI (4'6-diamino-2-phenylondole) [2 µg/ml], se colocaron cubreobjetos de plástico y se incubaron a temperatura ambiente durante 15 min. Luego los preparados se lavaron brevemente en 4X SSC/tween 0,2%, se agregaron 30 µl de medio de montaje (Vectashield, Vector Laboratories, EUA) y se les colocó un cubreobjetos de vidrio. Los preparados se conservaron en oscuridad a 2° C.

La obtención y el análisis de las imágenes se realizaron como se describió en el apartado 2.14.2.6.

3. RESULTADOS

3.1 Población de mapeo: caracterización y evaluación genotípica

De las 123 RILs que componen la población, 94 se seleccionaron para llevar adelante el trabajo de genotipificación y mapeo, dado que presentaban un poder germinativo alto y la extracción de ácidos nucleicos posibilitó la obtención de ADN de buena cantidad y calidad (Figura 3.1). En la multiplicación de las líneas recombinantes endocriadas derivadas del cruzamiento de RHA266 x PAC2 se observó una fuerte depresión por endogamia. El índice de germinación observado fue del 79%.



Figura 3.1 Multiplicación de las 124 líneas recombinantes endocriadas en las instalaciones del Instituto de Biotecnología CICVyA INTA Castelar.

El ADN de girasol fue extraído y purificado por el método de CTAB. Dado el alto contenido de polisacáridos que coprecipitan con el ADN, fue necesario incluir algunas modificaciones al método, como el agregado de PVP (polivinilpirrolidona) en una proporción de 100 μ l por cada 0,1 g de tejido fresco al *buffer* de extracción, y un paso adicional de purificación utilizando una resina comercial (Nucleon PhytoPure Plant

DNA Extraction; Amersham Pharmacia Biotech, RU). Este protocolo permitió aislar ADN genómico en cantidad y calidad adecuada para llevar adelante el trabajo de mapeo genético.

3.2 Desarrollo de marcadores funcionales

El desarrollo de marcadores funcionales se realizó a partir de secuencias de unigenes disponibles en bases de datos públicas (Índice de Genes de Girasol, HAGI, <http://compbio.dfci.harvard.edu/tgi/>) a través de la identificación de repeticiones en tándem, microsatélites contenidos dentro de secuencias que se expresan (EST-SSR), polimorfismo de un sólo nucleótido (SNP) o por inserciones y deleciones de nucleótidos (InDels). A continuación se detallan los resultados obtenidos con las diferentes técnicas empleadas.

3.2.1. Búsqueda y caracterización de SSR en bases de datos de unigenes y análisis en cuatro genotipos

La búsqueda de motivos de microsatélites contenidos en secuencias de unigenes de *H. annuus* descriptos en el *Plant Gene Index*, Dana-Farber Cancer Institute (<http://www.dfci.harvard.edu/>), utilizando el programa CUGI-SSR Server permitió detectar 1.134 motivos de repetición compuestos por di-, tri-, tetra y pentanucleótidos en las 36.741 secuencias analizadas, considerando sólo aquellos SSR con una longitud mínima de la repetición de 18 bases para di- y trinucleótido y de 20 bases para tetra- y pentanucleótidos. La Figura 3.2 muestra las frecuencias de SSRs agrupadas por el tipo de SSR (di-, tri-, tetra- y pentanucleótido).

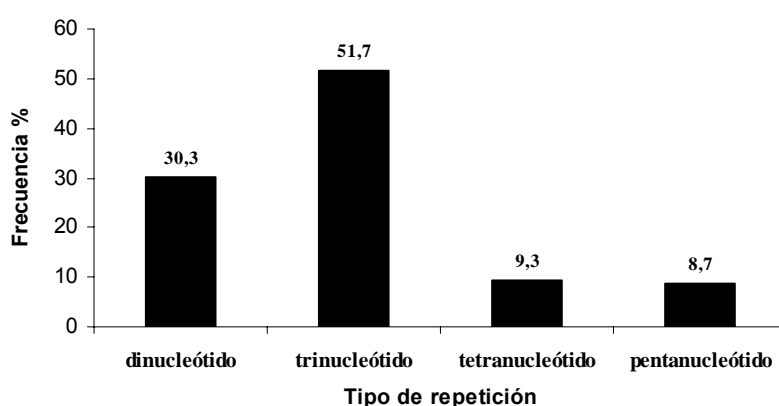


Figura 3.2 Distribución de los diferentes tipos de repetición de EST-SSR presentes en unigenes de girasol.

De los 1.134 EST-SSR detectados, el 51,65% presentó motivos de trinucleótidos, el 30,28% de dinucleótidos, el 9,32% de tetranucleótidos y el 8,74% de pentanucleótidos (Tabla 3.1).

Ciento cuarenta y seis unigenes contenían más de una repetición (12,9%).

Tabla 3.1 Frecuencia de SSR en la base de datos de EST de girasol.

Tipo de repetición	Cantidad de SSRs	Longitud total (bp)
Dinucleótido	367=30,28%	265144=29,98%
Trinucleótido	626=51,65%	462236=52,27%
Tetranucleótido	113=9,32%	81340=9,2%
Pentanucleótido	106=8,74%	75622=8,55%
Total	1134=100%	884342=100%

Como se explicó en Materiales y Métodos apartado 2.5.1 los motivos de los microsatélites están representados por 4 diferentes dinucleótidos, 10 diferentes trinucleótidos, 33 diferentes tetranucleótidos y 102 motivos diferentes de

pentanucleótidos. Al momento de estimar las frecuencias fueron consideradas todas las posibles combinaciones de los cuatro nucleótidos.

En la base de datos de unigenes de girasol (HAGI) se identificaron 281 EST con el motivo GA/CT, este fue el motivo más abundante identificado en un 23,2% de las secuencias analizadas, seguido por el motivo trinucleótido AAG/CTT presente en 144 secuencias (11,9%). La Figura 3.3 resume la distribución de frecuencias de SSR derivados de EST, basada en el tipo de motivo presente en la secuencia.

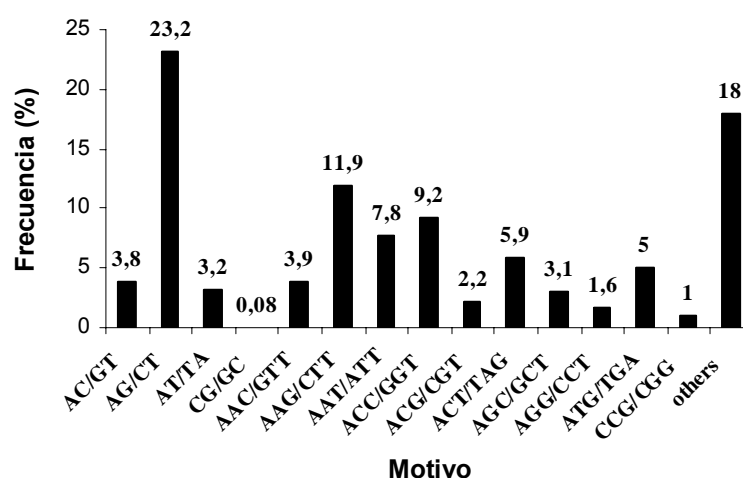


Figura 3.3 Distribución de la frecuencia de los motivos de los EST-SSR presentes en unigenes de girasol.

Un motivo dinucleótido puede representar múltiples codones dependiendo del marco de lectura, y traducirse en diferentes aminoácidos. Por ejemplo, el motivo GA/CT puede representar los codones GAG, AGA, UCU y CUC en el ARNm y traducirse en los aminoácidos Arg, Ala, Leu, respectivamente. Ala y Leu están presentes en alta frecuencia en proteínas, 8% y 10%, respectivamente (Lewin, 1994). Esta podría ser una de las razones por las cuales el motivo GA/CT está presente en alta frecuencia en las colecciones de EST.

Con el fin de asignar una función molecular probable a los marcadores derivados de EST-SSR, las secuencias correspondientes fueron analizadas usando el programa

[Blast2Go](http://www.blast2go.de/) (<http://www.blast2go.de/>), el cual asigna una clasificación a las secuencias según Gene Ontology (GO) (Tabla 3.2).

Tabla 3.2 Clasificación según Gene Ontology (GO) de los EST de girasol conteniendo microsatélites genotificados en la población de mapeo. La tabla muestra además el tipo de repetición y la localización en el mapa genético de ligamiento desarrollado en el presente trabajo.

Nombre EST (TIGR)	Clasificación según Gene Ontology	Tipo de repetición	GL
TC21354	GO: 0004601 <i>peroxidase activity</i>	trinucleótido	-
TC20625	GO: 0005488 <i>binding</i>	trinucleótido	4
TC26869	GO: 0005198 <i>structural molecule activity</i>	trinucleótido	16
TC23078	GO: 0003824 <i>catalytic activity</i>	trinucleótido	-
TC24992	GO: 0005198 <i>structural molecule activity</i>	trinucleótido	8
TC26323	Sin anotación GO	dinucleótido	17
TC23602	GO: 0003824 <i>catalytic activity</i>	tetranucleótido	8
TC21542	GO: 0003824 <i>catalytic activity</i>	trinucleótido	-
TC24343	Sin anotación GO	tetranucleótido	16
TC18194	GO: 0003824 <i>catalytic activity</i>	dinucleótido	3
TC25054	GO: 0005215 <i>transporter activity</i>	pentanucleótido	16
TC21990	GO: 0003824 <i>catalytic activit</i>	dinucleótido	-
TC21689	GO: 0005488 <i>binding</i>	trinucleótido	4
TC22997	Sin anotación GO	trinucleótido	-
TC15366	GO: 0003824 <i>catalytic activit</i>	InDel	10
TC18146	GO: 0003677 <i>DNA binding</i>	InDel	10
TC23762	GO: 0003824 <i>catalytic activit</i>	InDel	14

Para desarrollar marcadores a partir de los EST-SSR identificados en los unigenes de girasol se seleccionaron 127 secuencias, las cuales fueron amplificadas en los genotipos RHA266, PAC2, HA89 y RHA80, resueltos y visualizados mediante la técnica de geles de poliacrilamida y tinción con nitrato de plata. Los resultados obtenidos fueron los siguientes: 74 EST-SSR mostraron patrones de amplificación claros, 47 fallaron en la obtención de productos de amplificación o generaron productos de amplificación ambiguos, complejos o difíciles de resolver y 6 oligonucleótidos amplificaron alelos nulos (0/1 o 1/0) (Tabla 3.3). Dentro del conjunto de marcadores que resultaron funcionales, 17 fueron polimórficos entre los genotipos analizados y 57, monomórficos.

Tabla 3.3 Resultados obtenidos de la evaluación de 127 EST- SSR en cuatro líneas de girasol RHA266, PAC2, HA89 y RHA801

Líneas analizadas	
RHA266, PAC2, HA89 y RHA801	
Polimórficos	13%
Monomórficos	45%
Falló PCR	37%
Alelos nulos	5%

Un ejemplo de los fragmentos obtenidos mediante geles de poliacrilamida 6% teñidos con nitrato de plata se puede ver en la Figura 3.4, la cual muestra el patrón de amplificación obtenido para un conjunto de 7 marcadores sobre los 4 genotipos analizados.

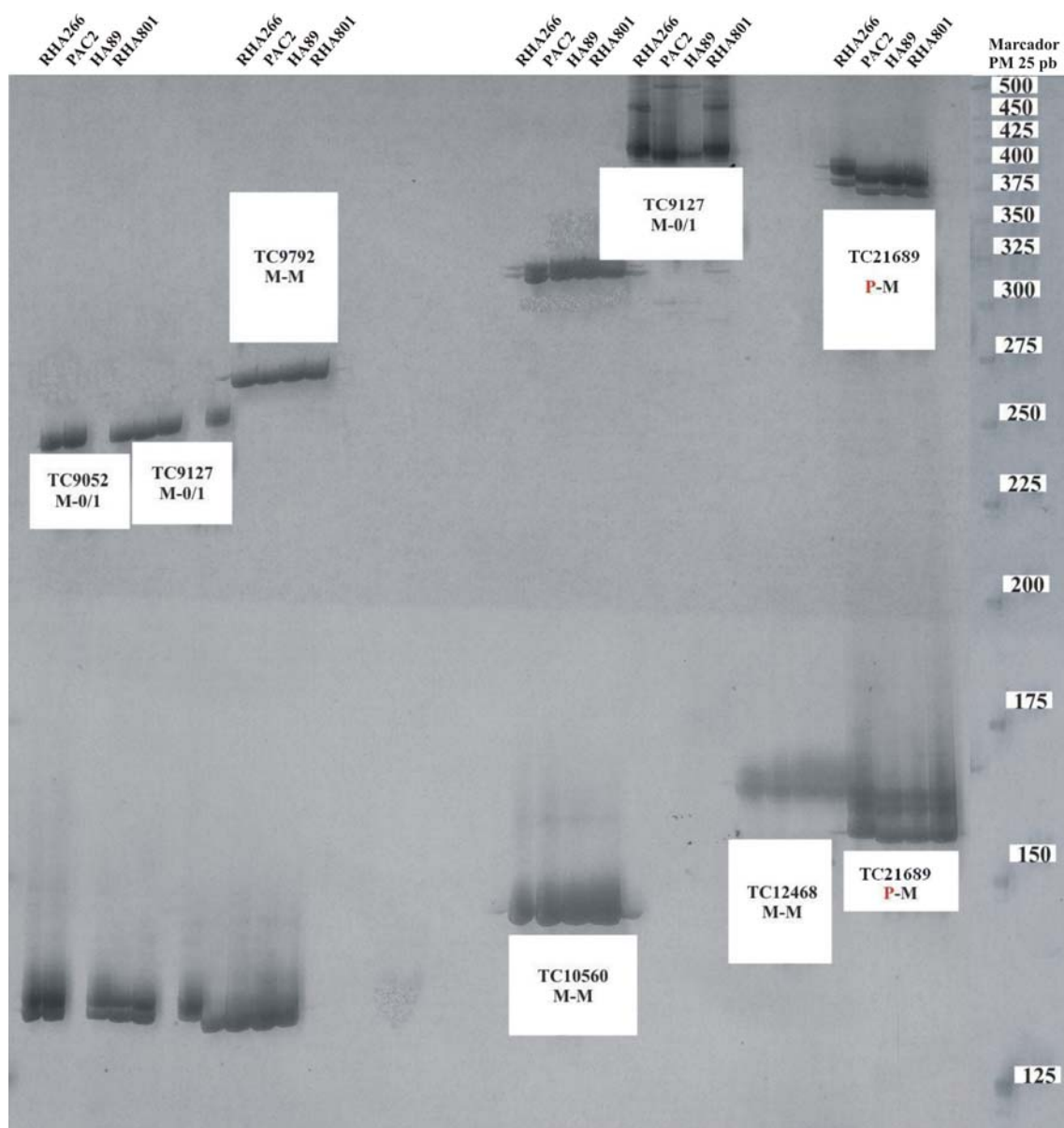


Figura 3.4 Patrón de amplificación de 7 marcadores EST-SSR sobre distintos genotipos de girasol. Los fragmentos fueron resueltos en un gel de poliacrilamida 6% desnaturalizante. Para cada marcador la primera calle corresponde al genotipo RHA266, la segunda al genotipo parental PAC2, la tercera al genotipo parental HA89 y la cuarta al genotipo parental RHA801. Referencias: P: polimórfico, M: monomórfico y 0/1 o 1/0: alelo nulo. N/A: falló PCR.

3.2.2 Análisis de marcadores funcionales mediante CAPS

Para evaluar polimorfismos de tipo SNPs mediante la técnica de CAPS se seleccionaron 25 secuencias de ESTs a partir de la colección local, las cuales presentaban similitud con

secuencias de genes de función conocida (Fernández et al., 2003) un ejemplo se muestra en la Tabla 3.4.

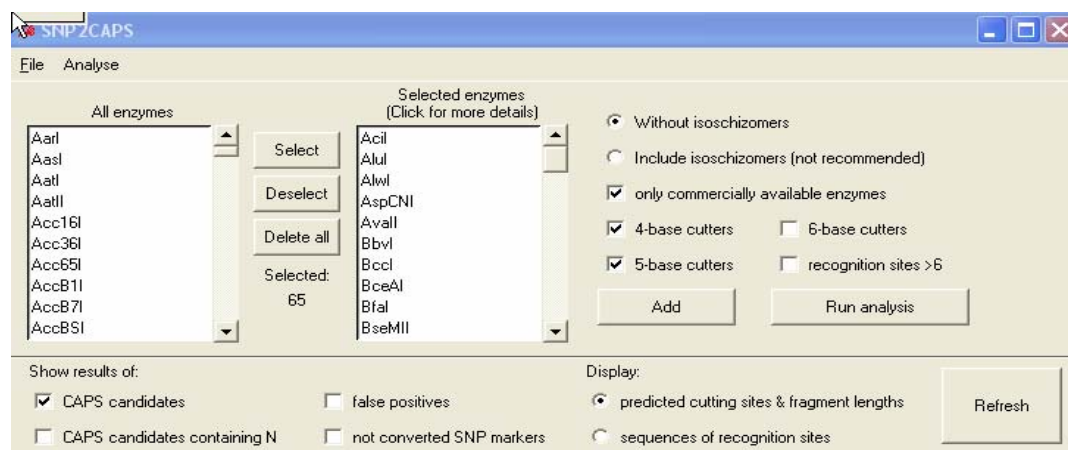
Tabla 3.4 Ejemplos de regiones candidatas seleccionadas para el mapeo de EST

EST ID	pb	Genbank AN	Función putativa (BLASTX)	Organismo	e-valor
Defensa/estrés					
EF562	690	BU671932	<i>germin-like protein</i>	<i>Oryza sativa</i>	$8,0 \times 10^{-41}$
EF332	908	BU671904	<i>Fructosil transferase</i>	<i>Helianthus tuberosus</i>	$1,00 \times 10^{-59}$
EF421	436	Z49063	<i>Polygalacturonase inhibitor factor</i>	<i>Actinidia deliciosa</i>	$3,00 \times 10^{-42}$
EF502	346	BU671910	<i>Lipid transfer protein</i>	<i>Gossypium hirsutum</i>	$2,00 \times 10^{-14}$
F201	243	BQ789917	<i>Auxin-responsive protein</i>	<i>Arabidopsis thaliana</i>	$2,00 \times 10^{-25}$
F205	276	BU671946	<i>leucine zipper protein</i>	<i>Arabidopsis thaliana</i>	$3,00 \times 10^{-40}$
Señalización					
F211	356	BU671951	<i>serine/threonine protein phosphatase PP2A-3 catalytic subunit protein</i>	<i>Oryza sativa subsp. Indica</i>	$2,00 \times 10^{-25}$
F241	287	AJ003197	<i>adenine nucleotide translocator protein</i>	<i>Lupinus albus</i>	$4,00 \times 10^{-33}$
Proteasas e inhibidores de proteasas					
F232	544	BU671929	<i>putative zinc protease protein</i>	<i>Arabidopsis thaliana</i>	$9,00 \times 10^{-62}$
Oxido-reducción					
F221	268	M25528	<i>ferredoxin-NADP+ reductase protein</i>	<i>Mesembryanthemum crystallinum</i>	$8,00 \times 10^{-26}$
Pared celular					
F226	308	U81049	<i>actin protein.</i>	<i>Pisum sativum</i>	$1,00 \times 10^{-16}$
F228	538	BU671932	<i>calreticulin protein</i>	<i>Nicotiana glumbaginifolia</i>	$1,00 \times 10^{-21}$

Las secuencias seleccionadas fueron amplificadas en los genotipos RHA266 y PAC2 y secuenciadas en ambos sentidos. Las secuencias obtenidas fueron editadas eliminando regiones de baja calidad y/o contaminantes, se les asignaron valores de calidad usando el algoritmo Phred y se confirmó la identidad del gen candidato comparando la secuencia en bases de datos públicas de proteínas, para lo cual se utilizó el programa BLASTX

(Altschul et al., 1990). De las 50 secuencias ensayadas, 46 fueron de buena calidad, con una lectura promedio de 550 bases, las secuencias fueron editadas utilizando el programa Bioedit (<http://www.mbio.ncsu.edu/BioEdit/BioEdit.html>). Para el reconocimiento de SNPs se utilizó el programa SNP2CAPS. El programa identificó *in silico* 21 polimorfismos a partir de 23 secuencias confiables. La Figura 3.5a muestra un ejemplo de la utilización de dicho programa con el marcador EF421 (Genbank AN: Z49063). En este caso, el programa predijo que existen dos sitios de restricción para la enzima *TaqI* en el parental HA1 (RHA266), de 200 pb, a las 31 y 46 pb, con lo cual los fragmentos esperados son tres: 154, 15 y 31 pb y que existen tres sitios de restricción para la misma enzima con el parental HA2 (PAC2), de 210 pb. De esta manera, los fragmentos definidos son cuatro: 101, 63, 15 y 31 pb.

Los polimorfismos identificados *in silico* fueron evaluados experimentalmente utilizando las enzimas de restricción correspondientes. Sólo 2 de los 21 polimorfismos detectados en los parentales de la población de mapeo pudieron desarrollarse como marcadores CAPs (F205 y F234 cuyos números de acceso son GenBank AN BU671882 y BU671930 Figura 3.5b).



The screenshot shows the SNP2CAPS software interface. The 'Selected enzymes' list includes: Acil, Alul, Alwl, AspCNI, Avall, Bbvl, Bccl, BceAl, Bfal, and BseMII. The 'Show results of' section has checkboxes for 'CAPS candidates' (checked), 'CAPS candidates containing N' (unchecked), 'false positives' (unchecked), and 'not converted SNP markers' (unchecked). The 'Display' section has radio buttons for 'predicted cutting sites & fragment lengths' (selected) and 'sequences of recognition sites' (unchecked). A 'Refresh' button is present.

```

>EF421_HA1
CAGGAATGGGCCCGCTTAATTCATTCCGGTCGAGGTGAAGGCTTTCGAGATTTGGAAGTGTAGCGAGCTCTGGCGGGATGGAGC
>EF421_HA2
CAGGAATGGGCCCGCTTAATTCATTCCGGTCGAGGTGAAGGCTTTCGAGATTTGGAAGTGTAGCGAGCTCTGGCGGGATGGAGC

Predicted CAPS candidates
-----
EF421      TaqI      200                31, 46                154, 31, 15      HA1
EF421      TaqI      210                31, 46, 109          101, 63, 31, 15  HA2
  
```

Figura 3.5a Vista de la pantalla de resultados del programa SNP2CAPS para el alineamiento de las secuencias correspondientes al fragmento EF421.



Figura 3.5b Gel de poliacrilamida 6% desnaturalizante, teñido con SYBR[®] Gold (Molecular Probes, EUA) mostrando el polimorfismo entre los parentales de la población de mapeo RHA266 x PAC2 de los marcadores F205 y F234.

3.3 Análisis de marcadores utilizando diferentes estrategias de genotipificación

Durante el desarrollo de esta tesis se ensayaron diferentes técnicas de genotipificación con el fin de optimizar un procedimiento que fuera sencillo de implementar, reproducible y automatizable de acuerdo a las posibilidades disponibles en el laboratorio. Inicialmente se utilizó la técnica de amplificación por PCR con oligonucleótidos fríos, la resolución en geles de poliacrilamida y la visualización mediante tinción con nitrato de plata. Con la utilización de esta técnica los resultados fueron satisfactorios. Sin embargo, la técnica de visualización implementada fue muy demandante de tiempo y muy sensible a la calidad del agua y a los reactivos utilizados. Con el propósito de simplificar el paso de visualización, se reemplazó la tinción con AgNO_3 por la tinción con SYBR[®] Gold (Molecular Probes, EUA), este método es más rápido y ágil, pero sólo fue eficiente en aquellos marcadores que presentaban una buena amplificación y bandas nítidas.

Con el fin de aumentar la sensibilidad y la procesividad del sistema de detección, se ensayó el método de amplificación usando nucleótidos marcados con radiactivo y detección por autorradiografía como método alternativo para reducir el tiempo de cada ensayo. Los resultados obtenidos fueron satisfactorios, pero el tiempo empleado fue similar al empleado con la tinción con plata y, teniendo en cuenta los riesgos del operador, fue discontinuado.

Finalmente, se desarrolló una técnica de genotipificación automática mediante electroforesis capilar usando un analizador automático de fragmentos (ABI3130XL Applied Biosystems, EUA) y oligonucleótidos fluorescentes. Para explorar al máximo

las posibilidades del sistema, ampliar la capacidad de procesamiento de muestras y hacerlo más económico, las muestras para la electroforesis fueron preparadas utilizando los tres fluoróforos disponibles (HEX, 6-FAM y NED), combinando los colores con los diferentes tamaños en pb de los fragmentos amplificados, evitando la superposición de fragmentos del mismo tamaño y del mismo color. Se reunieron en cada panel un grupo de hasta seis marcadores compatibles.

Antes de preparar los paneles de co-electroforesis, los fragmentos se evaluaron en geles de agarosa para ajustar las diluciones correspondientes, a fin de que la intensidad de la señal para los distintos fluorocromos se encuentre dentro del rango de fluorescencia adecuado, entre 200 a 4.000 RFU. En la mayoría de los marcadores, la intensidad de la señal de fluorescencia se encontró dentro del rango adecuado, y para aquellos fragmentos donde se identificaron intensidades fuera del rango esperado se realizó una nueva dilución de la muestra.

Con esta técnica se lograron resultados consistentes a los obtenidos previamente usando métodos manuales, con la ventaja de un aumento en la calidad del genotipado y una reducción considerable del tiempo de análisis de cada marcador.

Si bien el método de genotipificación automática utilizando oligonucleótidos fluorescentes resultó exitoso y se observó un aumento en la calidad del genotipado comparado con los métodos colorimétricos, tales como nitrato de plata, los costos de los oligonucleótidos marcados son altos y, por este motivo, se ensayó un método alternativo de marcado indirecto, el cual implicó el agregado de 19 nucleótidos de la secuencia universal M13 al extremo 5' del oligonucleótido sentido. De esta manera la reacción de amplificación se llevó adelante con tres oligonucleótidos; el sentido modificado, el antisentido y el universal M13 marcado con una modificación fluorescente en el extremo 5'. Esta metodología se llevó a cabo con una selección de 139 microsatélites. Los resultados obtenidos no fueron satisfactorios, observándose baja eficiencia y baja reproducibilidad en la amplificación. Debido a esto, se resolvió proseguir con la técnica de marcado directo.

3.4 Análisis de marcadores moleculares

3.4.1 Análisis de marcadores microsatélites genómicos

Con el fin de referenciar los mapas existentes para girasol cultivado, se seleccionó un grupo de 404 marcadores microsatélites de un conjunto de 550 marcadores hipervariables (Paniego et al., 2002). Paralelamente, se incluyó un grupo de 256 marcadores microsatélites públicos desarrollados en la Universidad de Oregon (ORSx) seleccionados de un conjunto de 657 marcadores mapeados por Tang et al. (2003) en tres poblaciones de mapeo (RHA280 x RHA281, RHA370 x RHA372 y PHA x PHB) y distribuidos en los 17 grupos de ligamiento del mapa público desarrollado por Tang et al. (2003). En total se analizaron 594 SSR (279 HAx y 315 ORSx) en los parentales de la población de mapeo RHA266 x PAC2, de los cuales resultaron informativos 224 SSR (101 HAx y 123 ORSx), 241 fueron monomórficos, 89 amplificaron alelos nulos (0/1 ó 1/0) y 40 presentaron productos de amplificación inespecíficos, complejos, difíciles de resolver o no pudieron ser amplificados. Los porcentajes se resumen en la Tabla 3.5.

Tabla 3.5 Resultado del análisis de 594 marcadores SSR en los parentales RHA266 y PAC2.

RHA266 x PAC2	
Polimórficos	37,7%
Monomórficos	40,6%
Falló PCR	6,7%
Alelos nulos	15%

Los 224 marcadores polimórficos detectados en las líneas parentales fueron genotipificados en 94 RILs mediante los métodos descritos en la sección 3.3.

Con la utilización de geles de poliacrilamida y visualización mediante tinción con nitrato de plata se analizaron en total 135 SSR en los parentales de la población de mapeo de los cuales resultaron informativos 52 HAx y 44 ORSx. La Figura 3.6 muestra un ejemplo de análisis de los parentales de la población de mapeo mediante geles de

poliacrilamida teñidos con AgNO_3 . La Figura 3.7 muestra un ejemplo de genotipificación de la población de mapeo mediante geles de poliacrilamida teñidos con AgNO_3 .

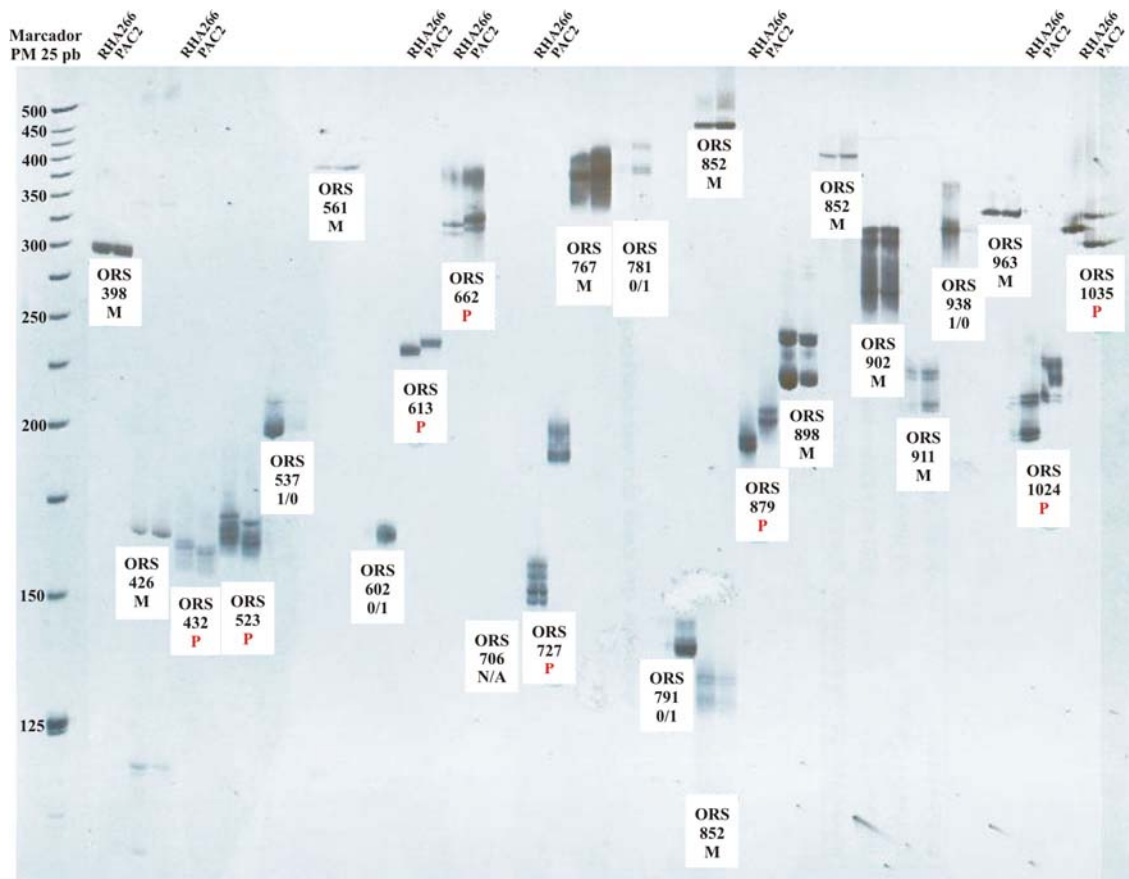


Figura 3.6 Identificación de polimorfismos de SSRs genómicos en los genotipos RHA266 y PAC2. Electroforesis de fragmentos de amplificación de microsatélites en girasol en gel de poliacrilamida 6% desnaturizante provenientes de la amplificación de 25 marcadores microsatélites en los genotipos parentales, RHA266 y PAC2. En la primera calle se muestran el marcadores de peso molecular (PM) estándar de 25 pb (Invitrogen, EUA). Para cada marcador la primera calle corresponde al genotipo RHA266 y la segunda al genotipo parental PAC2. Referencias: P: polimórfico, M: monomórfico y 0/1 o 1/0: alelo nulo. N/A: falló la PCR.

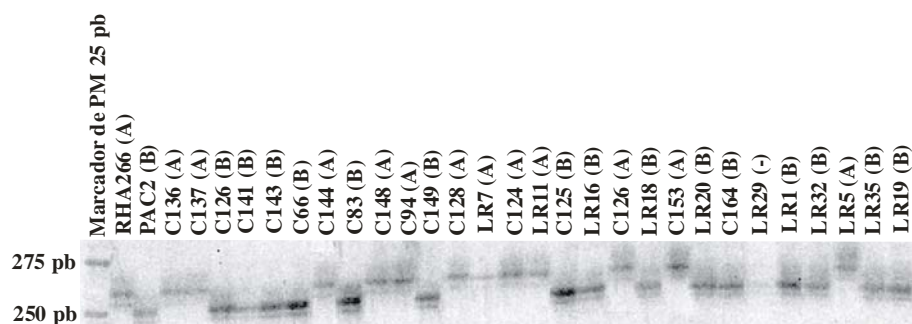


Figura 3.7 Fotografía de un gel de poliacrilamida 6% teñido con nitrato de plata mostrando la segregación de los alelos del marcador microsatélite ORS510 en los parentales (RHA266 y PAC2) y en 28 líneas recombinantes de la progenie.

Con la utilización de nucleótidos marcados con radiactivo se analizaron en total 214 SSR en los parentales de la población de mapeo, de los cuales resultaron informativos 26 HAx y 32 ORSx (Figura 3.8).

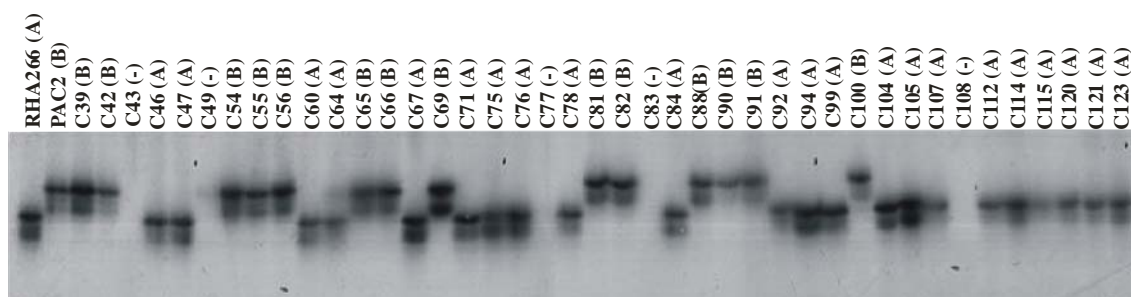


Figura 3.8 Fotografía de un gel de poliacrilamida 6% mostrando la segregación del marcador microsatélites HA3204 utilizando métodos radioactivos (γ -P³²) en los parentales de la población de mapeo (RHA266 y PAC2) y en 46 RILs.

Mediante EC fueron genotipificados 70 marcadores (23 HAx y 47 ORSx). Además, la mayoría de los marcadores genotipificados de manera manual, fueron re-genotipificados en forma automática con el propósito de completar los datos faltantes y corroborar el tamaño de los alelos. A fin de preparar las diluciones óptimas para la corrida por electroforesis capilar los fragmentos se evaluaron en geles de agarosa 1,5% teñidos con bromuro de etidio. En la Figura 3.9a se observa el producto de amplificación de los marcadores HA2605 y HA503, utilizando el oligonucleótido sentido marcado con

el fluoróforo HEX (cuyo color emitido es el verde) y 6-FAM (cuyo color emitido es el azul) respectivamente. En la Figura 3.9b se observa el producto de amplificación de los marcadores HA3886, ORS432 y HA3319, utilizando el oligonucleótido sentido marcado con el fluoróforo NED (cuyo color emitido es el amarillo).

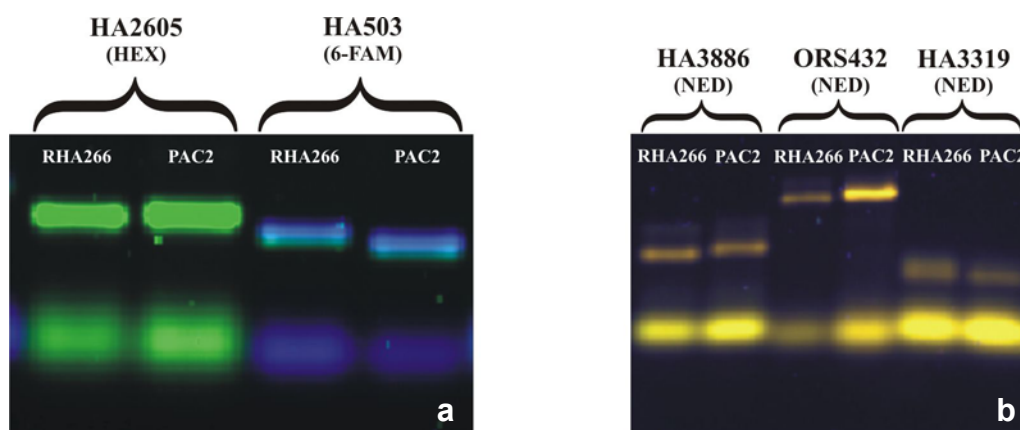


Figura 3.9 Fotografía de un gel de agarosa 1,5% mostrando el producto de amplificación de los marcadores microsatélites. **3.9a** HA2605 y HA503 utilizando los fluoróforos HEX y 6-FAM. **3.9b** HA3886, ORS432 y HA3319 utilizando el fluoróforo NED.

Con el propósito de definir los alelos correspondientes de cada marcador se definió para cada alelo la región del espectro en la cual se registra su señal, llamada “bin”. Como ejemplo se puede observar las Figuras 3.10a y b, las cuales muestran los electroferogramas correspondientes al marcador microsatélite ORS297 en las líneas recombinantes C100 y C139 las cuales presentan un alelo de 226 pb y 222 pb respectivamente, (fluoróforo utilizado HEX). El parental RHA266 presenta un alelo de 226 pb (indicado con el bin A) y el parental PAC2 un alelo de 222 pb (indicado con el bin B).

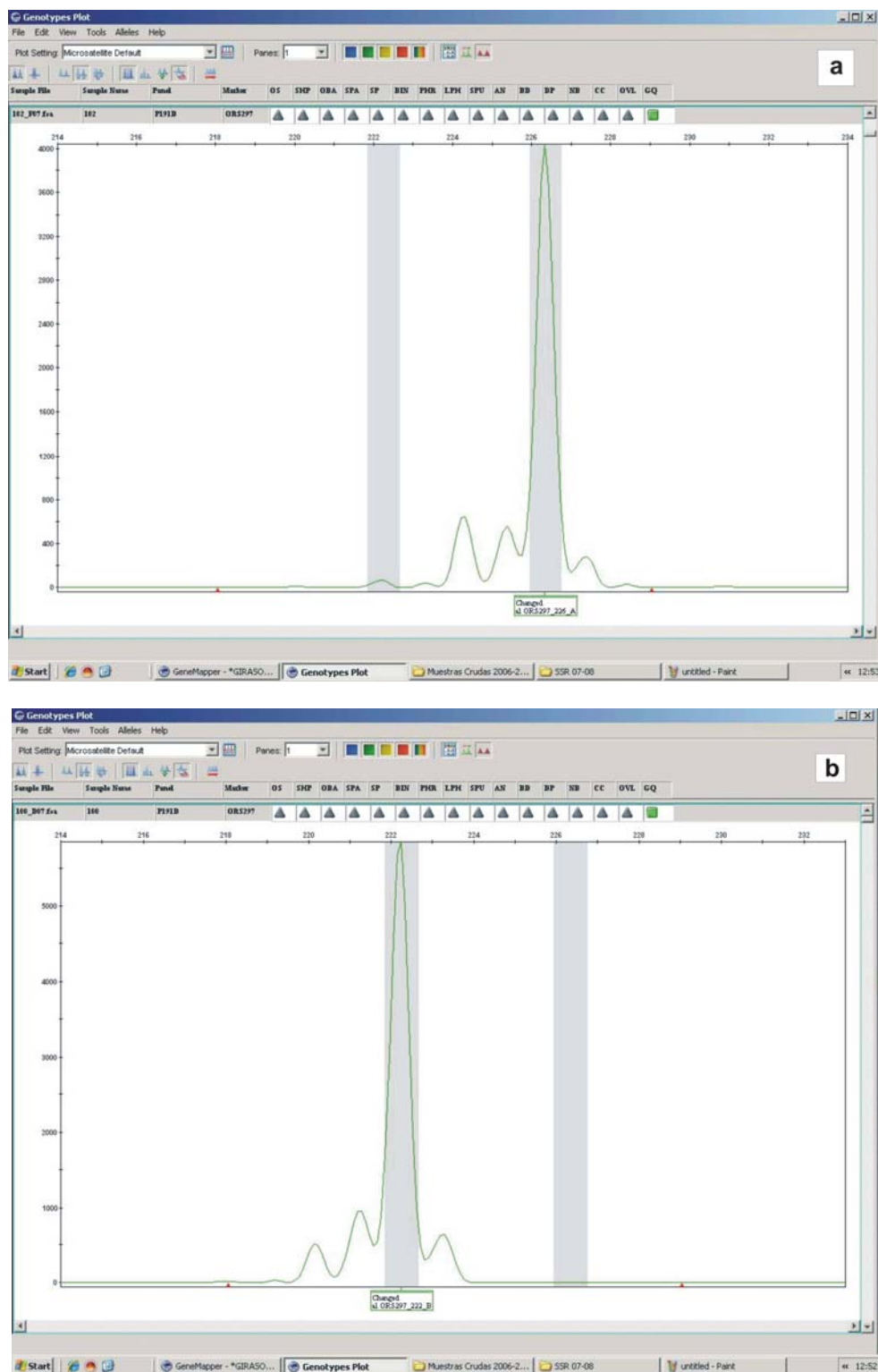


Figura 3.10 Electroferogramas del marcador microsatélite ORS297, obtenido sobre las líneas recombinantes C100 y C139 utilizando el secuenciador automático de capilares Applied Biosystems 3130XL. **3.10a** El pico muestra un alelo de 226 pb derivado de la segregación del parental RHA266 (A). **3.10b** El pico muestra un alelo de 222 pb derivado de la segregación del parental PAC2 (B).

En los casos donde fue posible, los paneles para la EC fueron preparados utilizando los tres fluoróforos disponibles (HEX, 6-FAM y NED) combinando los colores con los diferentes tamaños en pb de los fragmentos amplificados, evitando la superposición de fragmentos del mismo tamaño y color. La tabla 3.6 muestra un ejemplo de co-electroforesis de seis marcadores compatibles. En la Figura 3.11a se observan los datos luego de la EC de seis marcadores que conforman un panel descritos en la Tabla 3.6, a saber: ORS316 y ORS483 (6-FAM, azul), ORS297 y ORS1085 (HEX, verde), y ORS243 y ORS301 (NED, negro) sobre la línea recombinante C125. En la figura 3.11b se definieron los alelos correspondientes de cada marcador sobre la línea recombinante C125.

En la Figura 3.12 se muestra la co-electroforesis de los marcadores descritos en la Tabla 3.6 y la visualización simultánea de tres genotipos RILs (LR46, C121 y C138 en el Genotype Plot).

Tabla 3.6 La tabla muestra los datos de la EC de seis marcadores que conforman un panel.

Marcador	Fluoróforo	Alelo A	Alelo B
ORS316	6-FAM (azul)	182 pb	178 pb
ORS483	6-FAM (azul)	267 pb	269 pb
ORS297	HEX (verde)	226 pb	222 pb
ORS1085	HEX (verde)	281 pb	277 pb
ORS243	NED (negro)	173 pb	169 pb
ORS301	NED (negro)	235 pb	237 pb

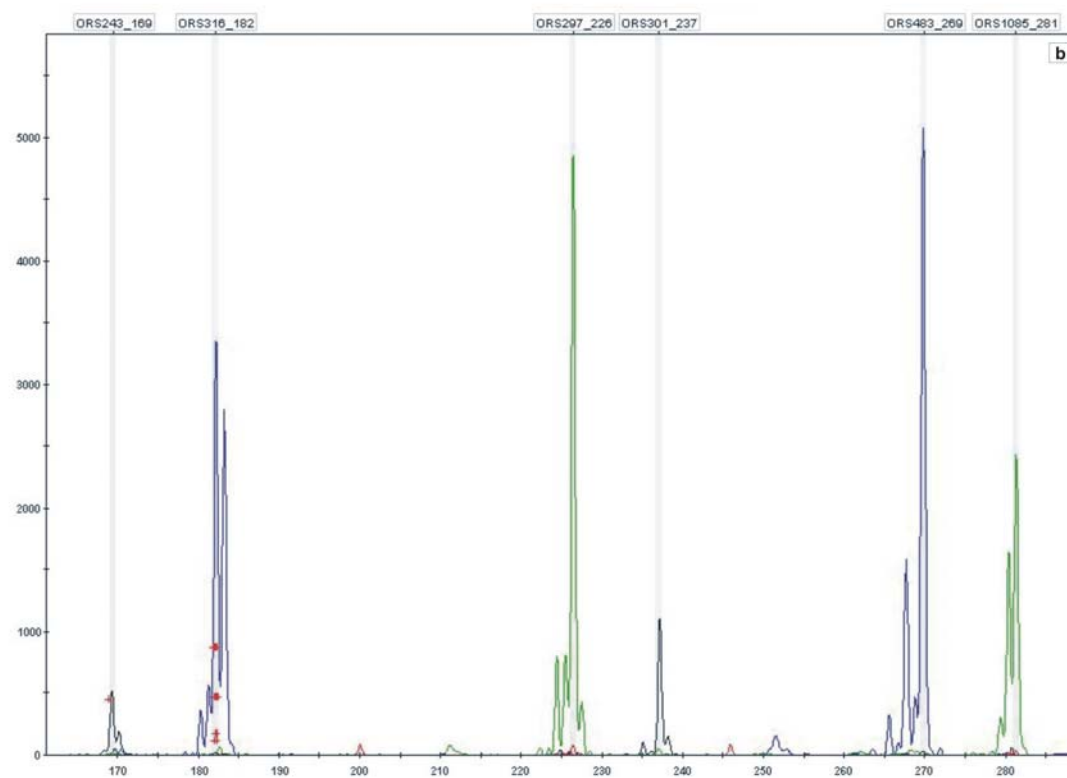
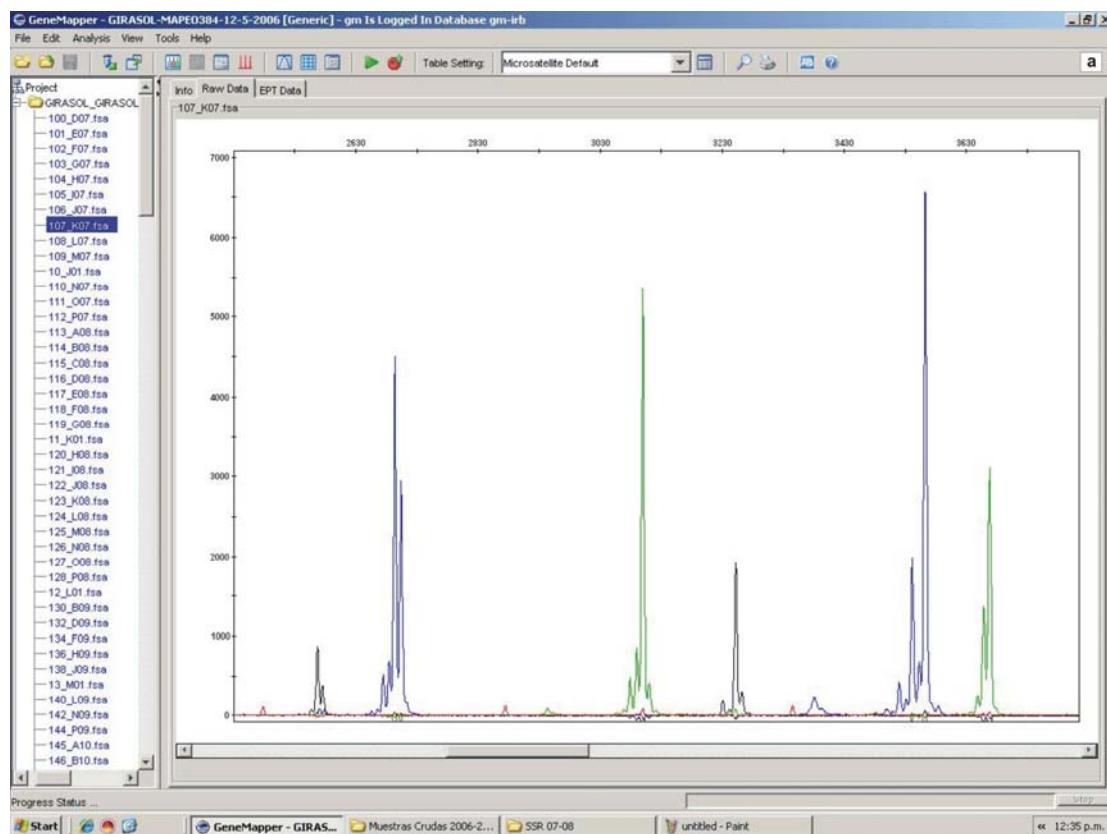


Figura 3.11a y b Co-electroforesis de seis marcadores microsatélites que conforman un panel descrito en la Tabla 3.6 sobre la línea recombinante C125 en el Genotype Plot.

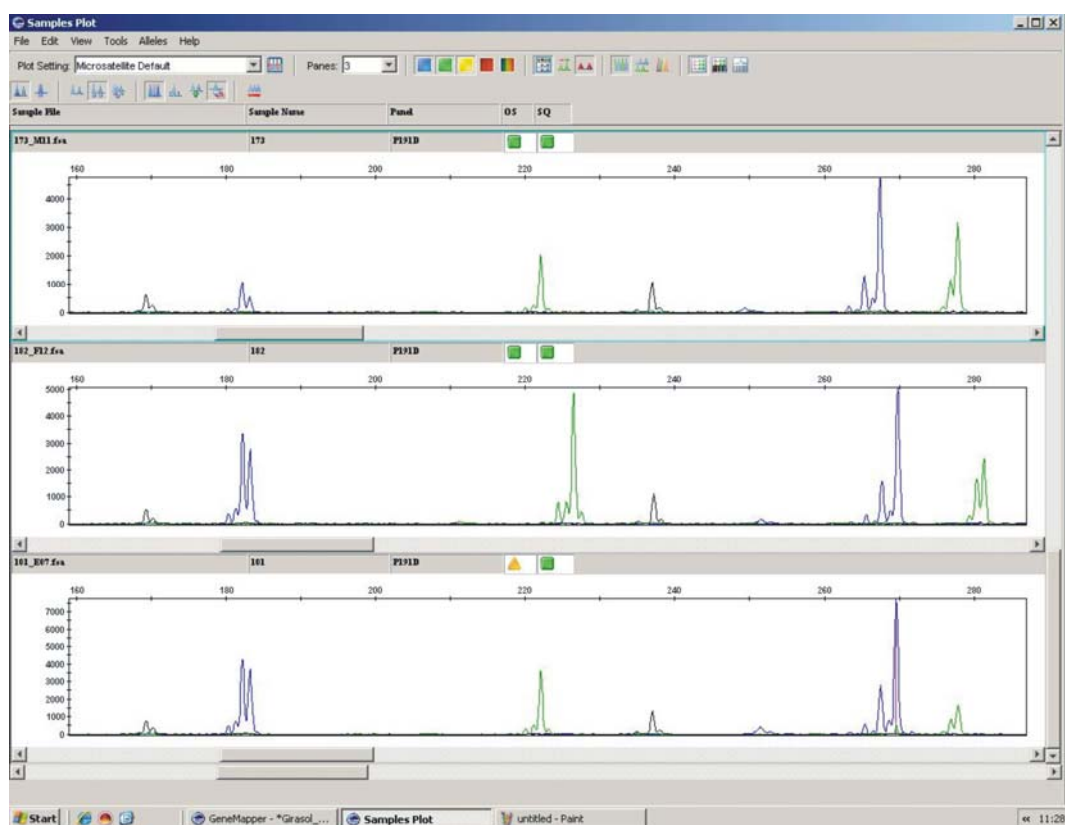


Figura 3.12 Visualización simultánea de tres genotipos en el Genotype Plot. Se observa la co-electroforesis de seis marcadores microsatélites ORS316 y ORS483 (6-FAM, azul), ORS297 y ORS1085 (HEX, verde), y ORS243 y ORS301 (NED, negro). En la parte superior, central e inferior de la pantalla se observa el análisis de los fragmentos utilizando las líneas LR46, C121 y C138 respectivamente.

Independientemente del método de genotipificación empleado los alelos observados en las líneas parentales para los SSR desarrollados en el INTA (HAX) y en la Universidad de Oregon (ORSx) coincidieron en el 92% de los casos con los alelos definidos previamente en la línea de girasol HA89 (Paniego et al., 2002) y HA370 (Tang et al., 2002). Así, en la mayoría de las líneas analizadas se observaron los mismos alelos y en muy pocos casos fueron identificados alelos nuevos. (ver Tabla anexa).

La información completa de todos los marcadores se puede consultar en la Tabla anexa.

Del total de *loci* analizados, en un porcentaje cercano al 0,9 % de las 94 líneas analizadas fueron heterocigotas, el cual se aproxima al valor de heterocigosidad esperado (0,78%).

3.4.2 Análisis de marcadores EST-SSR

Los marcadores polimórficos entre los genotipos PAC2 y RHA266 fueron genotipificados en 94 RILS de la progenie mediante la técnica de electroforesis capilar usando oligonucleótidos fluorescentes.

Previo al paso de genotipificación mediante EC, los fragmentos fueron evaluados en geles de agarosa 1,5% a fin de realizar las diluciones correspondientes. Un ejemplo se puede observar en la Figura 3.13 la cual muestra el producto de amplificación de los marcadores TC18194 (marcado con el fluoróforo 6-FAM) y TC22997 (marcado con el fluoróforo HEX) en los genotipos RHA266, PAC2, HA89 y RHA801.

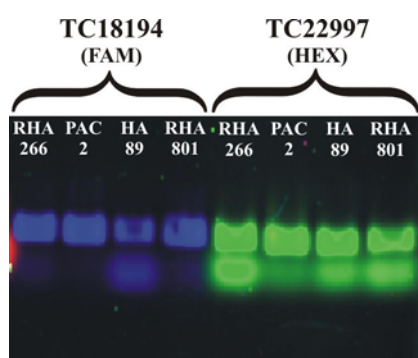


Figura 3.13 Fotografía de un gel de agarosa 1,5% mostrando el producto de amplificación de los marcadores TC18194 (6-FAM) y TC22997 (HEX). Para cada marcador la primera calle corresponde al genotipo RHA266, la segunda al genotipo PAC2, la tercera al genotipo HA89 y la cuarta al genotipo RHA801.

El análisis de los fragmentos correspondientes a los marcadores funcionales se realizó usando el programa GeneMapper® 4.0 (Applied Biosystem, EUA). En la Figura 3.14 se puede observar simultáneamente la amplificación de las líneas C82, C115 y LR29 para el marcador TC11140 (actualmente incluido en el *contig* TC24992) en el Genotype Plot.

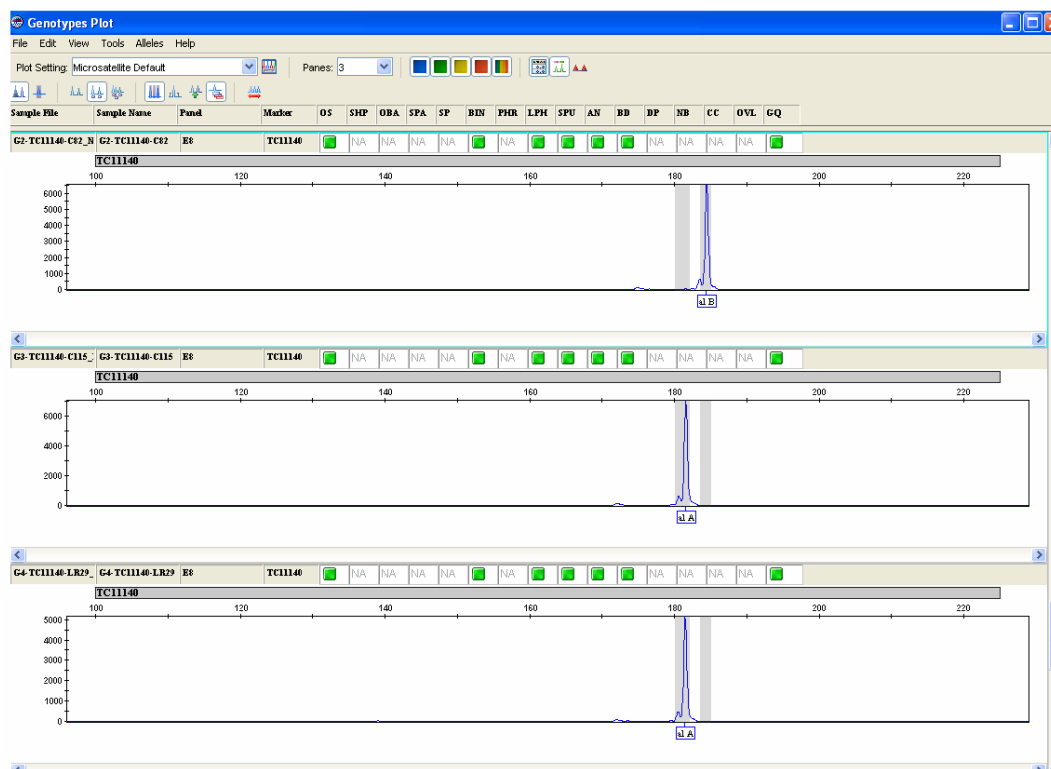


Figura 3.14 Visualización simultánea correspondiente a la amplificación de tres líneas para el marcador TC11140. El alelo parental RHA266 (A) tiene un tamaño de 184 pb y el alelo parental PAC2 (B) tiene un tamaño de 180 pb. En la parte superior, central e inferior se observa el producto de amplificación del marcador TC11140 en las líneas C82, C115 y LR29.

3.5 Construcción del mapa genético de ligamiento

Para evaluar la selección inicial de marcadores, el sistema de mapeo, el tamaño muestral, el tratamiento estadístico y la utilidad de los programas computacionales disponibles, los datos se incorporaron al mapa de referencia a distintos intervalos de tiempo, lo que permitió realizar ajustes de la selección de marcadores y de la metodología.

El análisis de segregación de los 584 marcadores que componen la matriz de datos se realizó utilizando el programa GQMOL (Schuter & Cruz, 2004; WWW.ufv.br/dbg/gqmol/gqmol.htm) y el test estadístico FDR (*False Discovery Rate*) (Benajmin & Hocheberg, 1995), considerándose una significancia global del experimento del 5%. Sólo el marcador HA765 presentó segregación distorsionada y fue descartado de futuros análisis.

Como se describió en Materiales y Métodos, para la confección del mapa genético de ligamiento se utilizó el programa Carthagene 0.999 (Schiex & Gaspin, 1997), estableciendo un porcentaje de recombinación máximo de 35% y un LOD mínimo de 4.

Los grupos de ligamiento obtenidos con este programa fueron analizados de manera independiente utilizando el programa Mapmaker 3.0 (Lander et al., 1987) con el fin de corroborar la posición de los marcadores dentro del grupo y de testear la inclusión de marcadores no ligados bajo los parámetros de mapeo establecidos inicialmente. Para este propósito, los nuevos marcadores se analizaron bajo el porcentaje de recombinación máximo de 35% y se disminuyó el valor de LOD hasta un mínimo de 3. De esta manera, el mapa genético quedó conformado por un total de 547 marcadores (231 SSR (81 SSR/HAx, 113 SSR/ORSx y 34 SSR/cartisol), 9 EST-SSR, 3 InDels y 304 AFLPs) localizados en 17 grupos de ligamiento. De los 547 marcadores, 209 fueron incorporados al mapa a partir de este trabajo, de los cuales 172 fueron incluidos con LOD 4 y el resto con LOD 3.

La longitud total estimada del mapa genético fue de 1.942,3 cM con una densidad media de 3,6 cM por *locus*. El grupo de ligamiento 10 fue el más denso (2,9 cM/*locus*), mientras que el grupo de ligamiento 17 fue el menos denso (4,9 cM/*locus*). La longitud de los grupos fluctúa de 71,9 (GL 7) a 203 cM (GL 10) y contienen de 20 (GL 13) hasta 70 (GL 10) marcadores. La región de mayor longitud sin cubrir de marcadores (*gaps*) fue de 13,2 cM y se observó en el GL 17 (ver Tabla 3.7 y Figura 3.15).

- El **GL 1** incluyó 21 marcadores, 8 marcadores SSR (6 ORSx y 2 HAx) y 13 marcadores AFLP, con una longitud promedio de 83,3 cM y una densidad de 3,9 cM/*locus*. Los marcadores ORS53, ORS662, ORS803 y ORS509 se correlacionaron con el GL 1 del mapa compuesto de Yu et al. (2003) y el marcador ORS365 correlacionó con el mismo GL del mapa de Tang et al. (2003).
- El **GL 2** incluyó 29 marcadores, 15 SSR (11 ORSx y 4 HAx) y 14 AFLP, con una longitud de 112,5 cM y una densidad 3,8 cM/*locus*. Los marcadores ORS423, ORS1035, ORS1065, ORS1194, ORS1045, ORS925, ORS653, ORS229 y ORS342 se correlacionaron con el GL 2 del mapa de Yu et al. (2003). El marcador ORS342 fue incluido con LOD 3.
- El **GL 3** incluyó 30 marcadores, 13 SSR (3 ORSx, 3 SSx, 7 HAx) y 16 AFLP, con una longitud de 92,9 cM y una densidad media de 3 cM/*locus*. Los marcadores ORS718, ORS657 y ORS432 se correlacionaron con el GL 3 del

mapa de Yu et al. (2003). Se incorporó el marcador microsatélite SSL303, el cual estaba presente en la matriz y no había sido incluido al mapa genético previamente por no encontrarse ligado a los marcadores disponibles. Los marcadores HA920, HA3938 y SSL303 fueron incluidos con LOD 3.

- El **GL 4** incluyó 29 marcadores, 7 SSR (3 ORSx, 4 HAx), 2 EST-SSR y 20 AFLP, con una longitud de 95,9 cM y una densidad media de 3,3 cM/*locus*. Los marcadores ORS620 y ORS1068 se correlacionaron con el GL 4 del mapa descrito por Yu et al. (2003).
- El **GL 5** incluyó 27 marcadores, 11 SSR (6 ORSx, 1 SSx y 4 HAx) y 16 AFLP, con una longitud de 103,5 cM y una densidad media de 3,8 cM/*locus*. Los marcadores ORS1024 y ORS433 se correlacionaron con el GL5 del mapa de Yu et al. (2003), el marcador y el ORS1159 correlacionaron con el GL 5 del mapa de Tang et al. (2003). Se incorporó el marcador microsatélite SSL231, el cual estaba presente en la matriz y no había sido incluido al mapa genético de ligamiento previamente. Los marcadores HA140 y SSL231 fueron incluidos con LOD 3.
- El **GL 6** incluyó 21 marcadores, 8 SSR (5 ORSx, 1 SSx y 2 HAx) y 13 AFLP, con una longitud de 83,6 cM y una densidad media de 3,9 cM/*locus*. Los marcadores ORS483, ORS381 y ORS1233 se correlacionaron con el GL 6 del mapa de Tang et al. (2003), el marcador ORS381 se correlacionó con el mapa de Yu et al. (2003).
- El **GL 7** incluyó 23 marcadores, 8 SSR (4 ORSx y 4 HAx) y 15 AFLP, con una longitud de 71,9 cM y una densidad media de 3,1 cM/*locus*. El marcador ORS331 se correlacionó con el mapa de Tang et al. (2002) y el marcador ORS1041 se correlacionó con el mapa de Yu et al. (2003).
- El **GL 8** incluyó 34 marcadores, 17 SSR (7 ORSx, 3 SSx y 7 HAx), 2 EST-SSR y 15 AFLP, con una longitud de 112 cM y una densidad media de 3,2 cM/*locus*. Los marcadores ORS1043, ORS243 y ORS894 se correlacionaron con el GL 8 del mapa de Tang et al. (2002) y los marcadores ORS418 y ORS894 se correlacionaron con el mapa de Yu et al. (2003). Los marcadores HA3513, ORS1043, TC23602, TC24992, HA911, ORS243, ORS418_1, ORS329, HA3581, HA3278, HA362, HA3971, ORS894_2 fueron incluidos con LOD 3.

- El **GL 9** incluyó 42 marcadores, 16 SSR (8 ORSx, 3 SSx y 5 HAx) y 26 AFLP con una longitud de 141,7 cM y una densidad media de 3,3 cM/*locus*. Los marcadores ORS805, ORS887, ORS510 y ORS1265 se correlacionaron con el GL9 del mapa de Tang et al. (2002) y los marcadores ORS1127 y ORS428 se correlacionaron con el mapa de Tang et al. (2003). Los marcadores ORS1009 y HA2564 fueron incluidos con LOD 3.
- El **GL 10** incluyó 70 marcadores, 23 SSR (9 ORSx, 6 SSx y 8 HAx), 2 EST-SSR y 45 AFLP, con una longitud de 203 cM y una densidad media de 2,9 cM/*locus*. Los marcadores ORS889, ORS878, ORS591 y ORS691 se correlacionaron con el GL 10 del mapa de Tang et al. (2002). Los marcadores ORS380, ORS437, ORS1144 y ORS691 se correlacionaron con el GL10 del mapa de Tang et al. (2003) y el marcador ORS78 se correlacionó con el mapa de Yu et al. (2003).
- El **GL 11** incluyó 21 marcadores, 7 SSR (4 ORSx, 1 SSx y 2 HAx) y 14 AFLP, con una longitud de 77,5 cM y una densidad media de 3,6 cM/*locus*. Los marcadores ORS733, ORS354, ORS1146 y ORS5 se correlacionaron con el GL 11 del mapa de Yu et al. (2003).
- El **GL 12** incluyó 31 marcadores, 15 SSR (6 ORSx, 4 SSx y 5 HAx) y 16 AFLP, con una longitud de 109,1 cM y una densidad media de 3,5 cM/*locus*. Los marcadores ORS671, ORS679, ORS123 y ORS1085 se correlacionaron con el GL 12 del mapa descrito por Tang et al. (2002).
- El **GL 13** incluyó 20 marcadores, 10 SSR (3 ORSx, 1 SSx y 6 HAx) y 10 AFLP, con una longitud de 89,3 cM y una densidad media de 4,4 cM/*locus*. Los marcadores ORS630 y ORS511 se correlacionaron con el GL 13 del mapa descrito por Tang et al. (2002) y el marcador ORS316 se correlacionó con el GL 13 del mapa de Yu et al. (2003). Los marcadores ORS630, ORS511, ORS316 y HA4208 fueron incluidos con LOD 3.
- El **GL 14** incluyó 39 marcadores, 20 SSR (8 ORSx, 8 HAx, 4 SSx), 1 INDEL y 18 AFLP, con una longitud de 153,4 cM y una densidad media de 3,9 cM/*locus*. Los marcadores ORS578, ORS391 y ORS1086 se correlacionaron con el GL 14 del mapa descrito por Tang et al. (2002) y los marcadores ORS1152 y ORS301

se correlacionaron con el GL 14 del mapa de Yu et al. (2003). Los marcadores ORS1128 y ORS578 fueron incluidos con LOD 3.

- El **GL 15** incluyó 24 marcadores, 13 SSR (7 ORSx, 2 SSx y 4 HAx) y 11 AFLP, con una longitud de 89,2 cM y una densidad media de 3,7 cM/*locus*. Los marcadores ORS8 y ORS401 se correlacionaron con el GL 15 del mapa descrito por Tang et al. (2003), los marcadores ORS121 y ORS420 se correlacionan con el GL 15 del mapa descrito por Yu et al. (2003) y los marcadores ORS1242, ORS499 y ORS687 se correlacionaron con el mapa descrito por Tang et al. (2002).
- El **GL 16** incluyó 51 marcadores, 28 SSR (18 ORSx, 3SSx, 1 iuBx y 6 HAx), 3 EST-SSR y 20 AFLP, con una longitud de 178,7 cM y una densidad media de 3,5 cM/*locus*. Los marcadores ORS899, ORS303, ORS126, ORS656, ORS303, ORS788 y ORS485 se correlacionaron con el GL 16 del mapa descrito por Tang et al. (2002). Los marcadores ORS128 y ORS993 se correlacionaron con el GL 16 del mapa descrito por Yu et al. (2003) y los marcadores ORS455 y ORS407 se correlacionaron con el GL 16 del mapa descrito por Tang et al. (2003). Los marcadores ORS5_2, ORS126_2, ORS788, ORS993, ORS495_1, ORS470, HA360 y ORS495_2 fueron incluidos con LOD 3.
- El **GL 17** incluyó 35 marcadores, 12 SSR (8 ORSx, 1 SSx y 3 HAx), 1 EST-SSR y 22 AFLP, con una longitud de 173,4 cM y una densidad media de 4,9 cM/*locus*. Los marcadores ORS169, ORS297, ORS677, ORS1097, ORS1247 se correlacionaron con el GL 17 del mapa descrito por Tang et al. (2002). Los marcadores ORS31_3, ORS297, TC26323 y HA3971 fueron incluidos con LOD 3.

Treinta y dos marcadores SSR (20 SSR/HAx y 12 SSR/ORSx) y 5 EST-SSR no pudieron ser incluidos en el mapa de ligamiento.

La evaluación de los marcadores de distintas fuentes, los programas de mapeo y los parámetros estadísticos aplicados resultaron adecuados, ya que en general los marcadores se distribuyen de manera amplia y coherente para aquellos marcadores localizados previamente en otros mapas de ligamiento (Tang et al., 2002; 2003; Yu et al., 2003).

Tabla 3.7 Distribución y densidad de los marcadores en cada grupo de ligamiento.

Grupo de ligamiento	AFLP	SSR	EST-SSR	InDels	Número de loci	Longitud (cM)	Densidad (cM/locus)
1	13	8	-	-	21	83,3	3,9
2	14	15	-	-	29	112,5	3,8
3	16	13	1	-	30	92,9	3,0
4	20	7	2	-	29	95,9	3,3
5	16	11	-	-	27	103,5	3,8
6	13	8	-	-	21	83,6	3,9
7	15	8	-	-	23	71,9	3,1
8	15	17	2	-	34	112,0	3,2
9	26	16	-	-	42	141,7	3,3
10	45	23	-	2	70	203,0	2,9
11	14	7	-	-	21	77,5	3,6
12	16	15	-	-	31	109,1	3,5
13	10	10	-	-	20	89,3	4,4
14	18	20	-	1	39	153,4	3,9
15	11	13	-	-	24	89,2	3,7
16	20	28	3	-	51	178,7	3,5
17	22	12	1	-	35	173,4	4,9
Total	304	231	9	3	547	1.942,3	3,6

Mapa genético de ligamiento

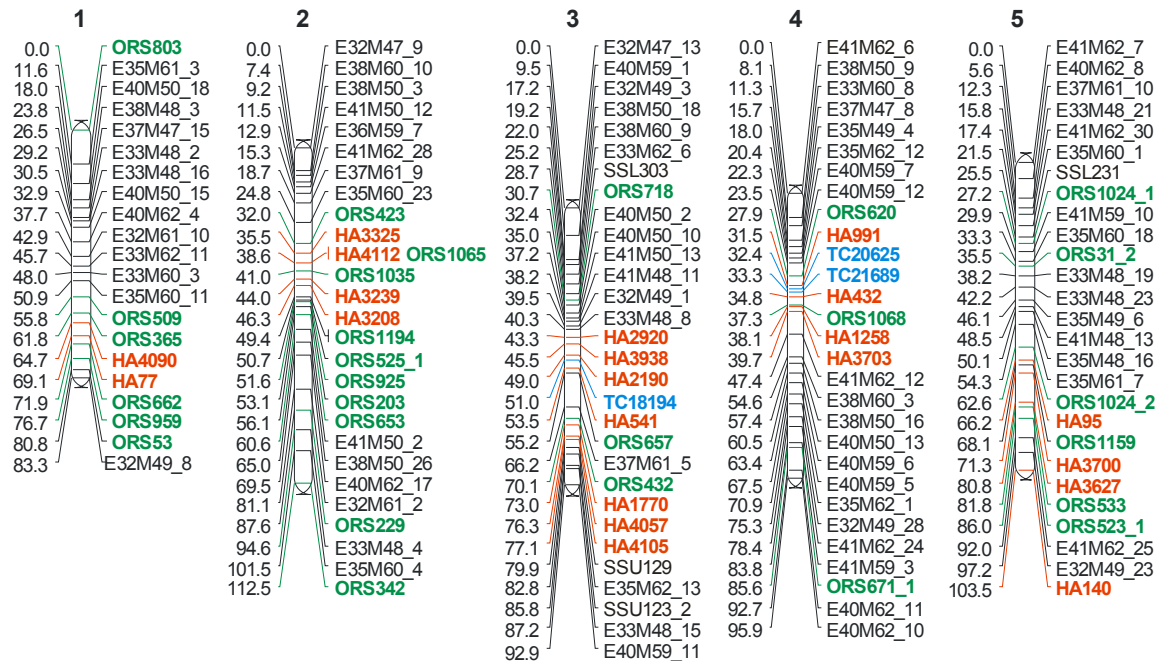


Figura 3.15 Mapa de ligamiento de SSRs, EST-SSR, EST-InDels y AFLPs sobre una población de líneas recombinantes del cruzamiento de RHA266 x PAC2. Los grupos de ligamiento descriptos fueron numerados en base a la nomenclatura del mapa de microsatélites descrito por Tang et al., (2003). Las distancias de mapa (en cM) entre los *loci* están indicadas sobre el lado izquierdo de cada grupo de ligamiento. En rojo: SSR desarrollados por el INTA (HAx), en verde: SSR desarrollados en la Universidad de Oregon (ORSx), en celeste: EST-SSR, en violeta: EST-InDels y en negro: AFLPs (ExMx) y SSR (iubx, SSLx y SSUx) desarrollados por el grupo Cartisol y genotipificados por Al-Chaarani et al. (2004).

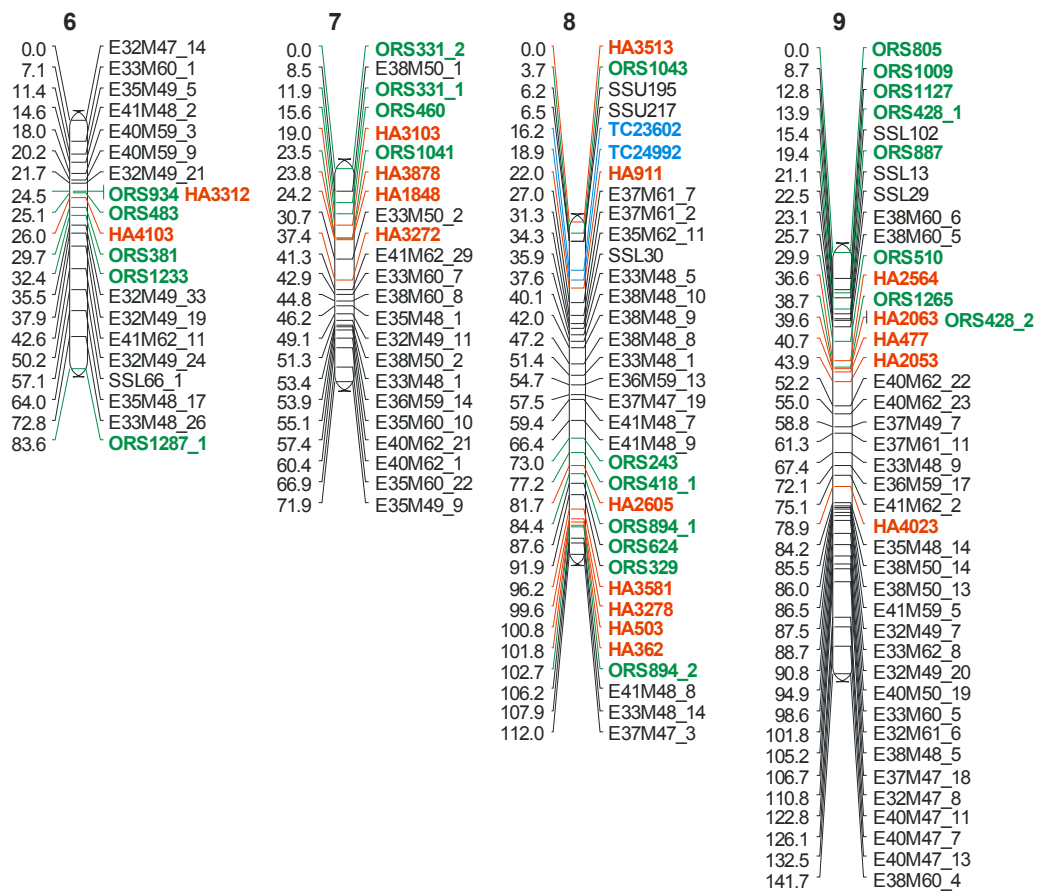


Figura 3.15 Continuación

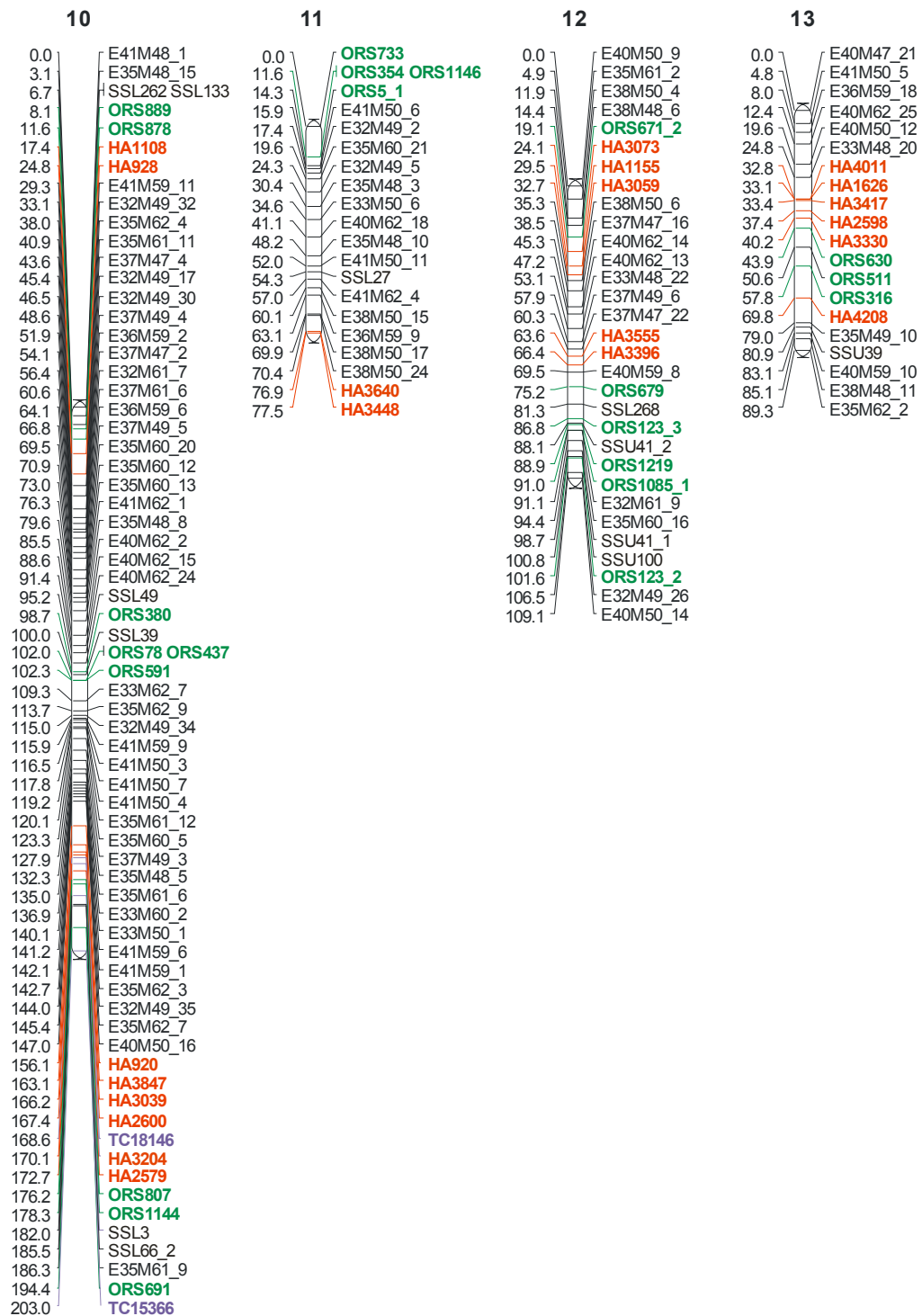


Figura 3.15 Continuación

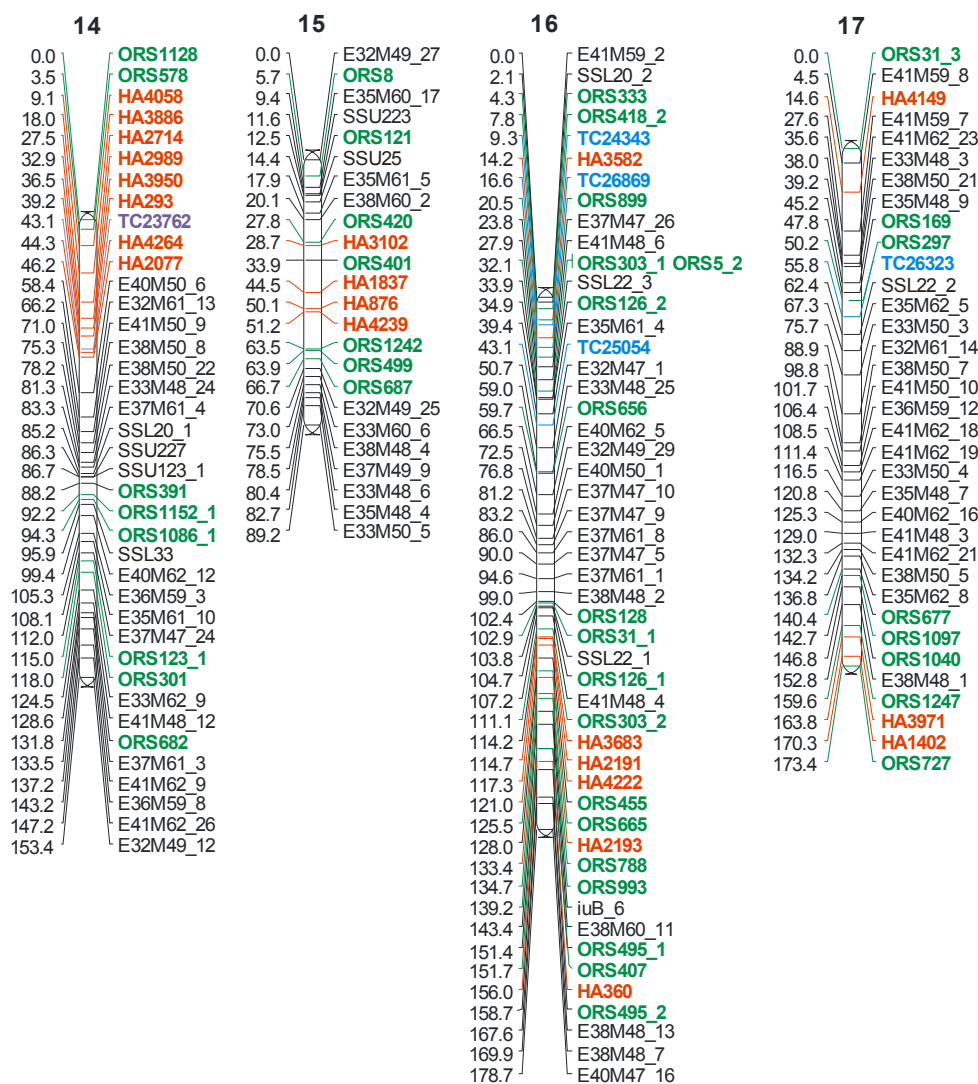


Figura 3.15 Continuación

3.5.1 Distribución de los marcadores moleculares

El test de rachas, utilizado para analizar la distribución de los marcadores SSR y AFLPs, no presentó evidencias para rechazar la H_0 (“Los marcadores se distribuyen al azar en los grupos de ligamiento”) en los grupos de ligamiento 5, 6, 7 y 11, con un nivel de confianza de 95%, lo que indicaría que en los mencionados GL los marcadores se distribuirían al azar. Sin embargo, en los grupos de ligamiento restantes a saber: GL 1, 2, 3, 4, 8, 9, 10, 12, 13, 14, 15, 16 y 17 se rechaza la H_0 , lo que indicaría que en los mencionados GL los marcadores no se distribuirían al azar sino que se encontrarían agrupados en distintas regiones a lo largo del grupo de ligamiento, alternando agrupaciones de microsatélites con conglomerados de marcadores de tipo AFLPs.

3.6 Contribución mediante diversas sondas a la caracterización del complemento cromosómico de *Helianthus annuus* utilizando la técnica de FISH

Con el fin de caracterizar el complemento cromosómico de *H. annuus* se diseñaron y/o amplificaron diversas sondas correspondientes a secuencias ribosomales y a secuencias repetitivas de diferente origen. Este trabajo fue realizado en colaboración con el Laboratorio de Citogenética de la Facultad de Ciencias Exactas y Naturales de la UBA, bajo la dirección de la Dra. Lidia Poggio y el Dr. Eduardo Greizerstein. Los resultados de esta sección están parcialmente documentados en la tesis de licenciatura de Díaz Quijano (2007).

3.6.1 Sondas derivadas de la región ribosomal 26S

Se amplificó un fragmento de 232 pb con similitud a ADN ribosomal 26S, denominado EF235 (GenBank, AN: BU671882) obtenido a partir de la colección local de ESTs (Fernández et al., 2003) (ver Materiales y Métodos, Figura 2.13.1).

En los ensayos de hibridación mediante la técnica FISH, se observó que la sonda EF235 hibridó en tres pares cromosómicos revelando la presencia de 6 zonas de genes ribosomales. Como se observa en las Figuras 3.16a y b las señales de hibridación coinciden con los satélites. Se observan además, diferencias en el tamaño de las señales, lo que podría deberse a polimorfismos en el número de copias de dichas secuencias en cada uno de los cromosomas. Esta sonda permitió identificar tres de los diecisiete pares cromosómicos y, asimismo, ha permitido la identificación de regiones ribosomales en otras especies *H. resinosus*, *amaranthus* y *allium cepa*, con mejor definición que la sonda de ADN ribosomal pTa71 de trigo, de 9 Kb, (Gerlach & Bedbrook, 1979) utilizada generalmente y reportada previamente para girasol (Greizerstein, comunicación personal).

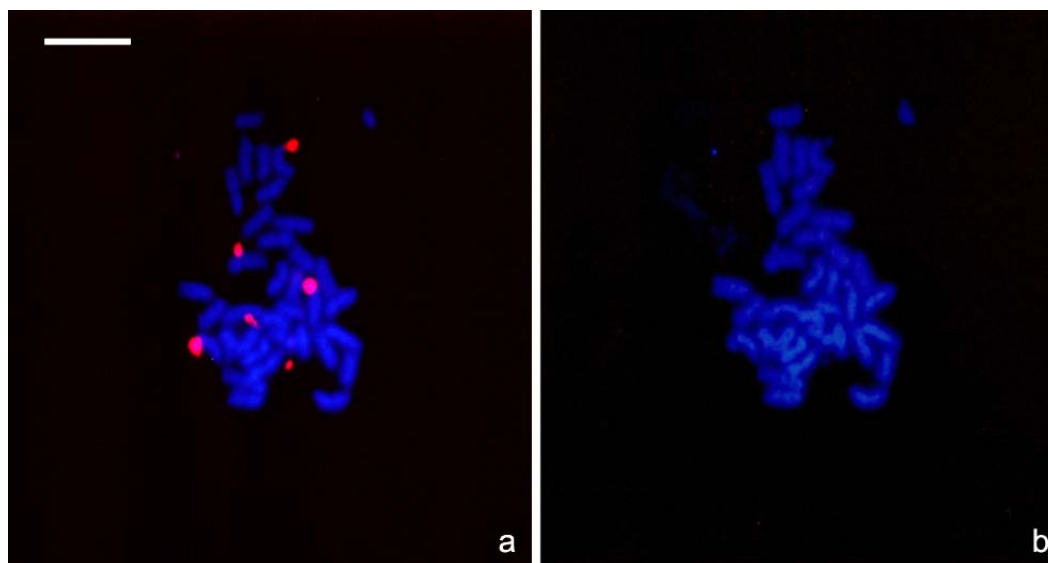


Figura 3.16a Metafases mitóticas hibridadas con sonda EF235 correspondiente a la región ribosomal 26S. Se observan seis señales de hibridación detectadas con Cy3 (roja). **3.16b** Misma célula contratinción con DAPI. La barra indica 10 μm .

3.6.2 Sonda derivada de una secuencia altamente repetitiva de girasol

La sonda denominada **SSR-Ha785** tiene un tamaño de 239 pb y presenta similitud en bases de datos públicas de nucleótidos con una región altamente repetitiva de girasol denominada HAN9965 (Genbank AN: AJ009965) (ver Materiales y Métodos Tabla 2.3 y Figura 2.7). En los ensayos de hibridación mediante FISH mostró una señal de hibridación dispersa sobre todos los cromosomas (Díaz Quijano, 2007), pudiéndose identificar sólo algunos pares de cromosomas.

3.6.3 Utilización de un clon de BAC con alto contenido de ADN repetitivo como sonda en experimentos de FISH

El clon de BAC denominado BAC772, conteniendo secuencias repetitivas con similitud a la secuencia SSR-Ha785 (ver sección Materiales y Métodos, Figura 2.13.1.3) fue aislado de una genoteca de BACs para girasol (Clemson, CUGI HA_HBa) y utilizado como sonda en ensayos de hibridación *in situ*. Las hibridaciones BAC-FISH revelaron un patrón de hibridación menos disperso al observado mediante FISH utilizando la misma sonda (SSR-Ha785) aislada y marcada individualmente.

Las bandas-FISH observadas al hibridar con el clon BAC772 conjuntamente con el análisis de la morfología y tamaño de los cromosomas permitió identificar los 17 cromosomas presentes en girasol en base al análisis sobre varias células (Figura 3.17 y 3.18). Se confeccionó un cariograma (Figura 3.19a y b) y su correspondiente idiograma ilustrado en la figura 3.19c. Se identificaron 11 pares metacéntricos, 2 pares submetacéntricos y 4 pares subtelocéntricos. Las bandas-FISH permitieron discriminar entre pares de cromosomas de igual morfología y tamaño.

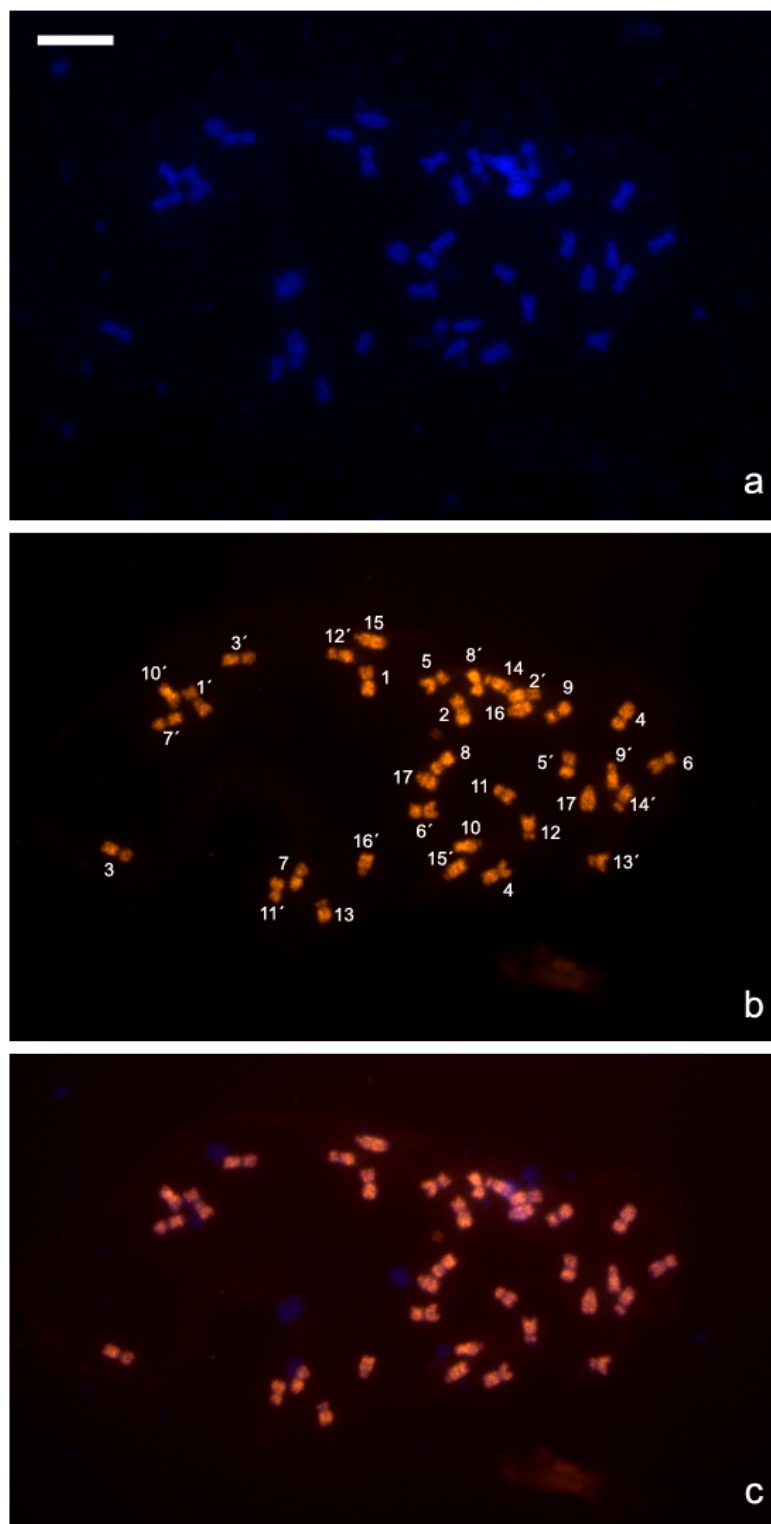


Figura 3.17a-c Metafase mitótica: hibridación utilizando el clon BAC772. **3.17a** Tinción con DAPI. **3.17b** Hibridación detectada con estreptoavidina-Cy3. Los números indican los miembros de cada par cromosómico. **3.17c** Doble exposición de la misma célula. La barra indica 10 μm .

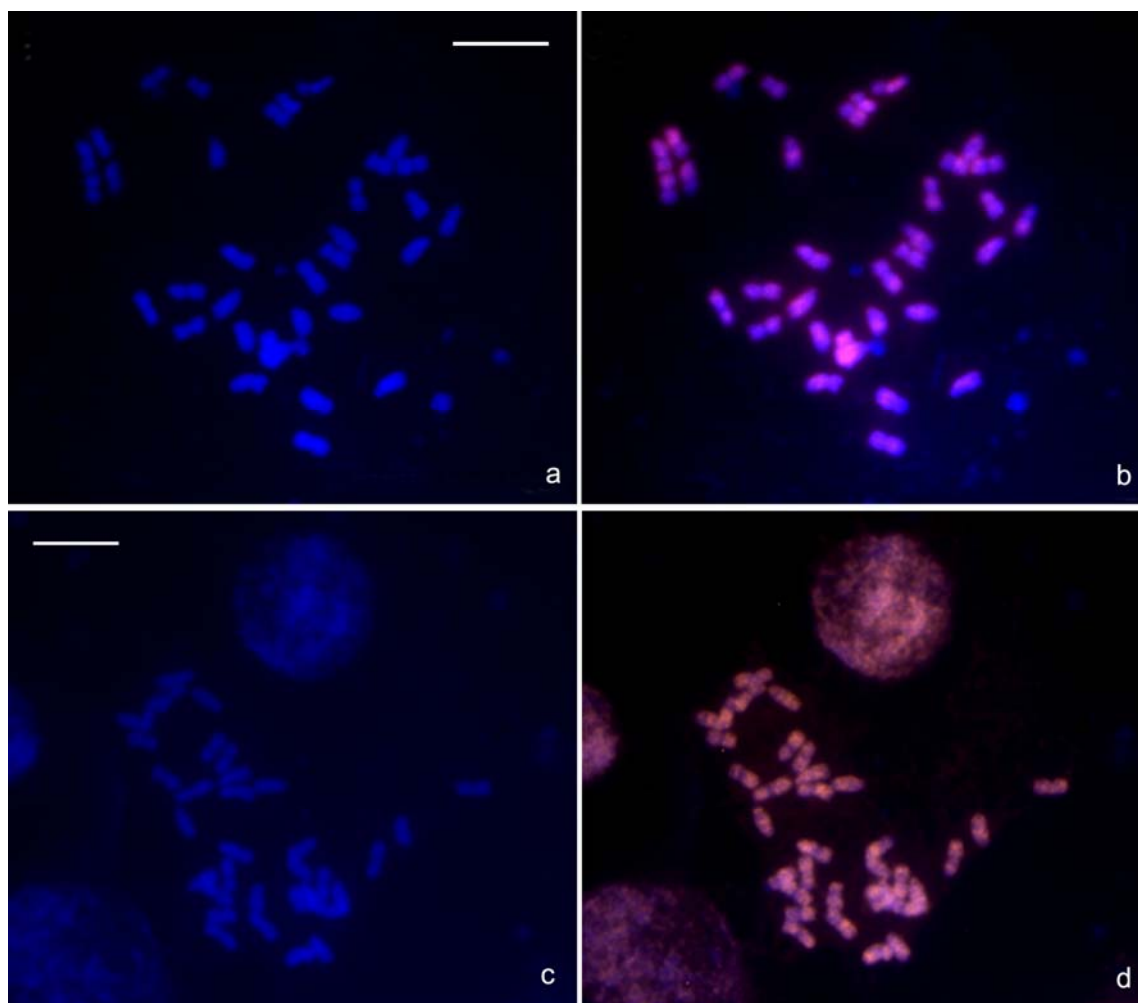
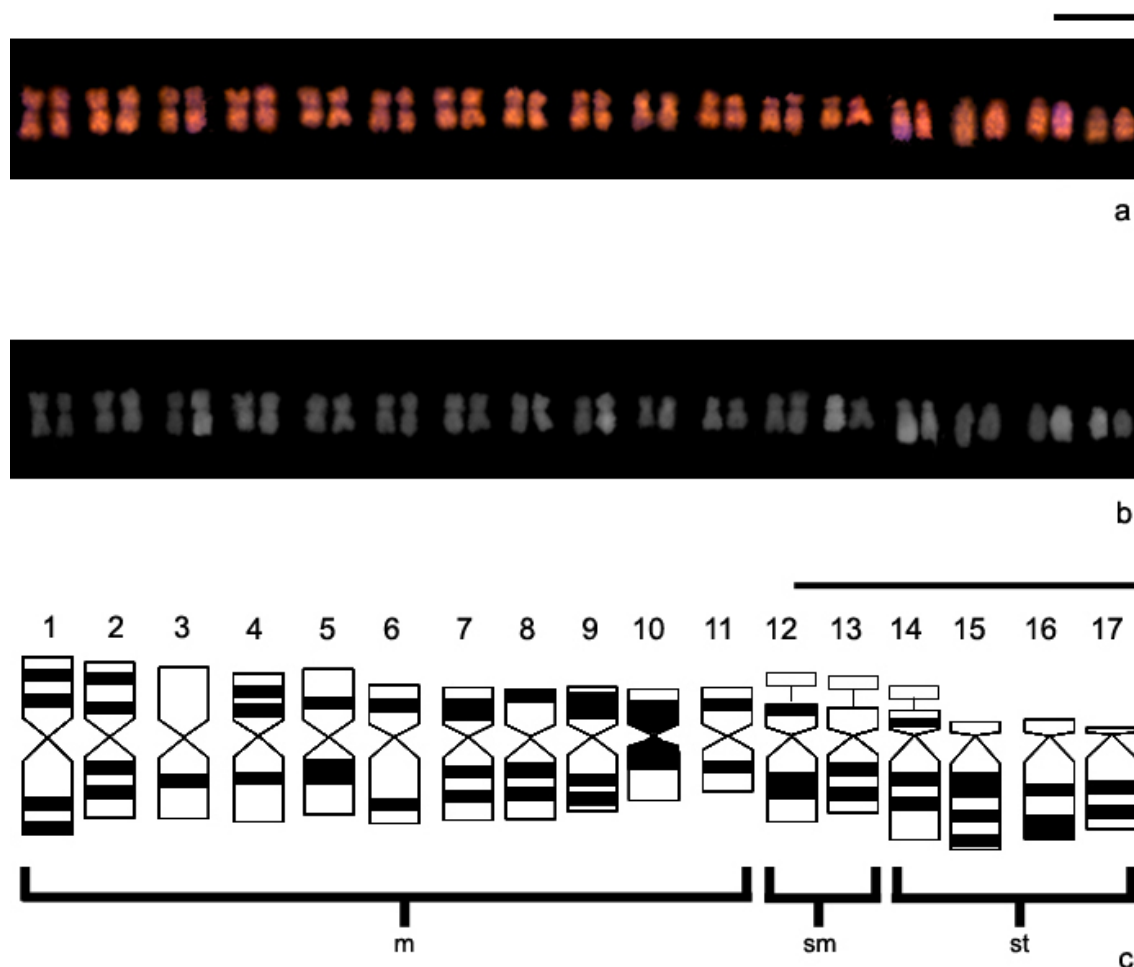


Figura 3.18a-d Metafases mitóticas de la línea HA89 de girasol. **3.18 b y d** Hibridación utilizando como sonda el clon BAC772 detectado con estreptoavidina-Cy3. **3.18 a y c** Contratación con DAPI. La barra indica 10 μm .



Tomado de Díaz Quijano (2007), con modificaciones

Figuras 3.19a y b Cariograma de girasol realizado a partir de la célula de la figura 3.17. **3.19c** Idiograma correspondiente al cariograma. m= cromosomas metacéntricos, sm= cromosomas submetacéntricos, st= cromosomas subtelo-céntricos. La barra indica 10 µm.

A fin de caracterizar el contenido de las secuencias repetitivas presentes en el clon de BAC usado como sonda, se secuenciaron los extremos del mismo y se realizaron ensayos de BAC-*Southern-blot*. El análisis comparativo de las secuencias de los extremos del BAC repetitivo con bases de datos públicas mediante el programa BLASTX reveló similitud a un retrotransposón de arroz de la familia *Ty1-copia-like*. Los ensayos de BAC-*Southern-blot* se llevaron a cabo utilizando las siguientes sondas: SSR- Ha785 (Paniego et al., 2002) con similitud a secuencias altamente repetitivas de girasol (GenBank AN: AJ009965); Rep1 con similitud a retrotransposones de la familia *copia-like* (GenBank AN: AJ009967) (Santini et al., 2002); y Retro3 con similitud a retrotransposones de la familia *gypsy-like* (GenBank AN: DQ229838) (Tang et al.,

2006). Las tres sondas revelaron señales de hibridación, siendo la sonda SSR-Ha785 la que presentó señales de mayor intensidad, lo que posiblemente indica un número de copias mayor de estas secuencias (ver Figura 3.20).

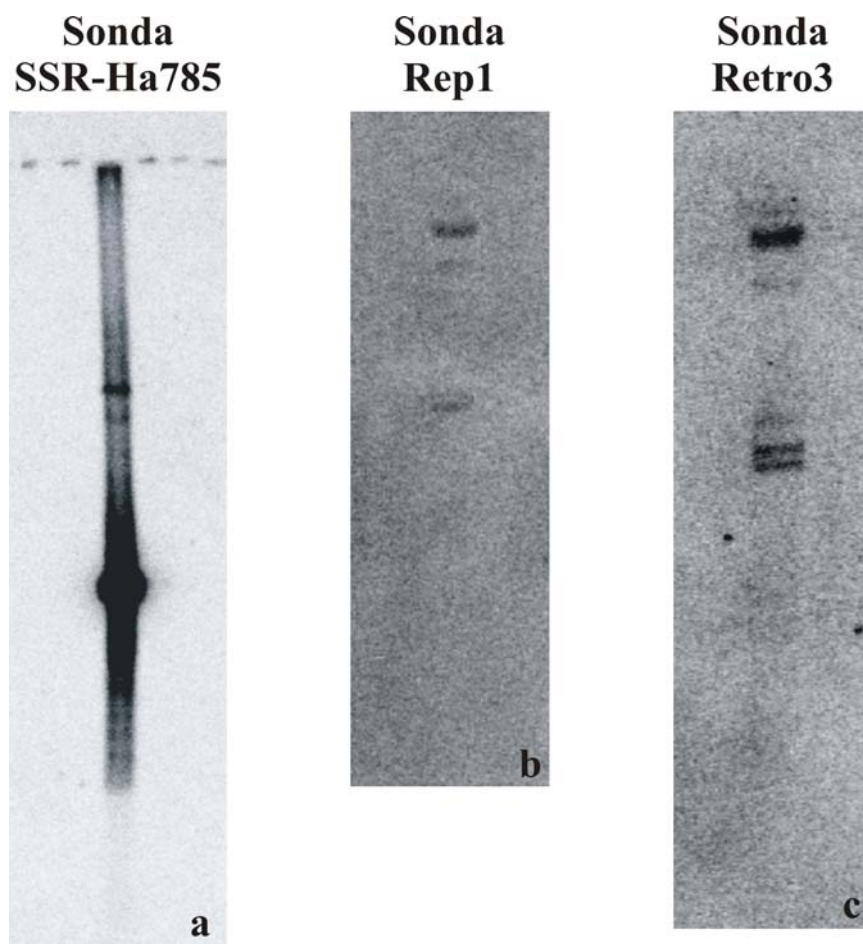


Figura 3.20 BAC- Southern blot del clon repetitivo BAC772 utilizando como sondas tres secuencias repetitivas de girasol. **3.20a** La secuencia SSR- Ha785 (Paniego et al., 2002) con similitud a secuencias altamente repetitivas de girasol. **3.20b** La secuencia Rep1 con similitud a retrotransposones de la familia *copia-like* (Santini et al., 2002). **3.20c** La secuencia Retro3 con similitud a retrotransposones de la familia *gypsy-like* (Tang et al., 2006).

3.6.4 Contribución mediante diversas sondas repetitivas a la caracterización del complemento cromosómico de *H. annuus* utilizando la técnica de PRINS

En forma simultánea al desarrollo de técnicas FISH para caracterizar el complemento cromosómico de girasol, se evaluó la aplicabilidad de la técnica de PRINS. En primer lugar se amplificaron regiones correspondientes a la secuencia Retro2 utilizando oligonucleótidos específicos, (ver Materiales y Métodos, apartado 2.13.1.2b). Los

oligonucleótidos Retro2 amplifican un fragmento de 844 pb del retrotransposon *Ty3 gypsy-like* (Tang et al., 2006). En la Figura 3.21 se evidencian señales de amplificación de esta secuencia en distintos cromosomas. En este ensayo se utilizó Biotina-16-dUTP como nucleótido marcado y se reveló con estreptoavidina-Cy3 (rojo). Aunque se observaron señales de hibridación discretas en algunos cromosomas, esta sonda no permitió identificar el complemento cromosómico completo.

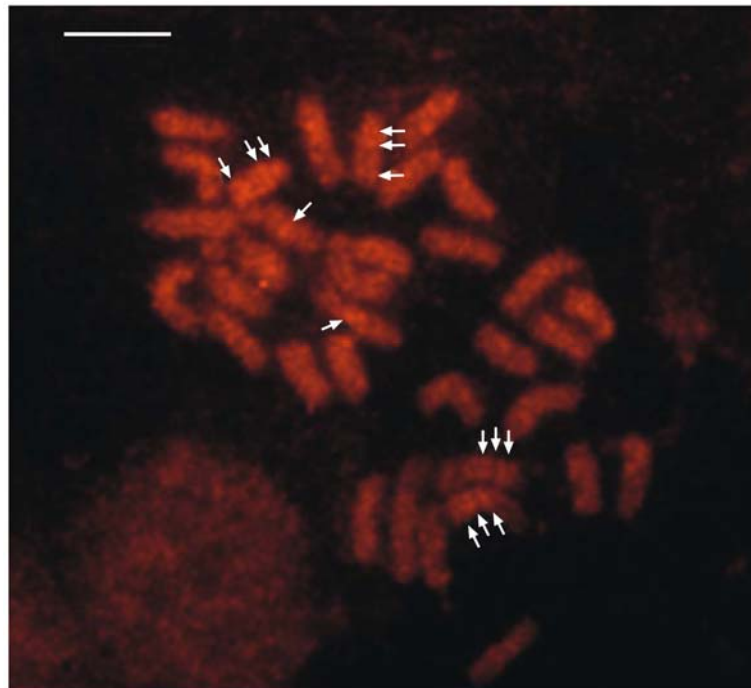


Figura 3.21 Metafase mitótica de hibridación mediante PRINS utilizando el oligonucleótido Retro2. Las flechas señalan bandas que corresponden a la extensión del oligonucleótido. Fluorocromo utilizado para la detección: Cy3. La barra indica 10 μm .

3.7 Localización de secuencias de copia única mediante BAC-FISH, PRINS y C-PRINS

3.7.1 Utilización de clones de BAC conteniendo microsatélites como sonda en experimentos de hibridación in situ

Con el fin de relacionar los grupos de ligamiento con los cromosomas correspondientes, se realizaron ensayos aplicando la técnica de BAC-FISH, usando como sondas clones de BAC seleccionados a partir de microsatélites de localización única en el mapa de girasol, asociados a QTLs de interés.

Para esta etapa se seleccionó un conjunto de 32 marcadores microsatélites distribuidos en los grupos de ligamiento 9, 10 y 16. Esta selección se basó en resultados de un trabajo paralelo a esta tesis, en el cual se identificaron distintos QTLs de resistencia al patógeno de girasol *S. sclerotiorum* en el GL 10 (Maringolo, 2007) y QTLs que están siendo confirmados en los GL 9 y 16 (Maringolo, comunicación personal).

Los marcadores seleccionados fueron usados como sondas, en grupos de entre diez y doce, para analizar una genoteca de BAC de girasol representada en once filtros de alta densidad de clones de BACs (BAC: HA_HBa, www.genome.clemson.edu). Las sondas fueron diseñadas como overgo utilizando los programas bioinformáticos Overgo 1.02i (Cai et al., 1998, <http://mouse-genome.bcm.tmc.edu/webovergo/OvergoInput.asp>) y/u OligoSpawn (<http://oligospawn.ucr.edu/>) y marcadas radioactivamente como se describe en Materiales y Métodos, produciendo un fragmento doble cadena de 36 a 40 pb.

Cada grupo de sondas overgo, conteniendo entre 10 y 12 sondas, fue hibridado en una primera etapa en dos filtros de BACs. Como resultado de las hibridaciones se obtuvieron más de 70 clones positivos de variada intensidad (ver Figura 3.22) y se seleccionaron al azar 35 clones de distinta intensidad de la señal. Con el fin de identificar qué marcador estaba incluido en cada clon de BAC, se realizaron reacciones de PCR de cada clon con cada uno de los oligonucleótidos específicos para cada fragmento. Se obtuvieron seis clones positivos, dos de los cuales pertenecían al mismo microsatélite, distribuidos en los tres grupos de ligamiento seleccionados inicialmente. Los clones positivos fueron denominados BAC1, BAC2, BAC3, BAC4, BAC5, BAC6 y fueron identificados con las sondas de los fragmentos HA928, HA928, HA4222, HA2063, ORS805 y HA2600 respectivamente (ver Tabla 3.8).

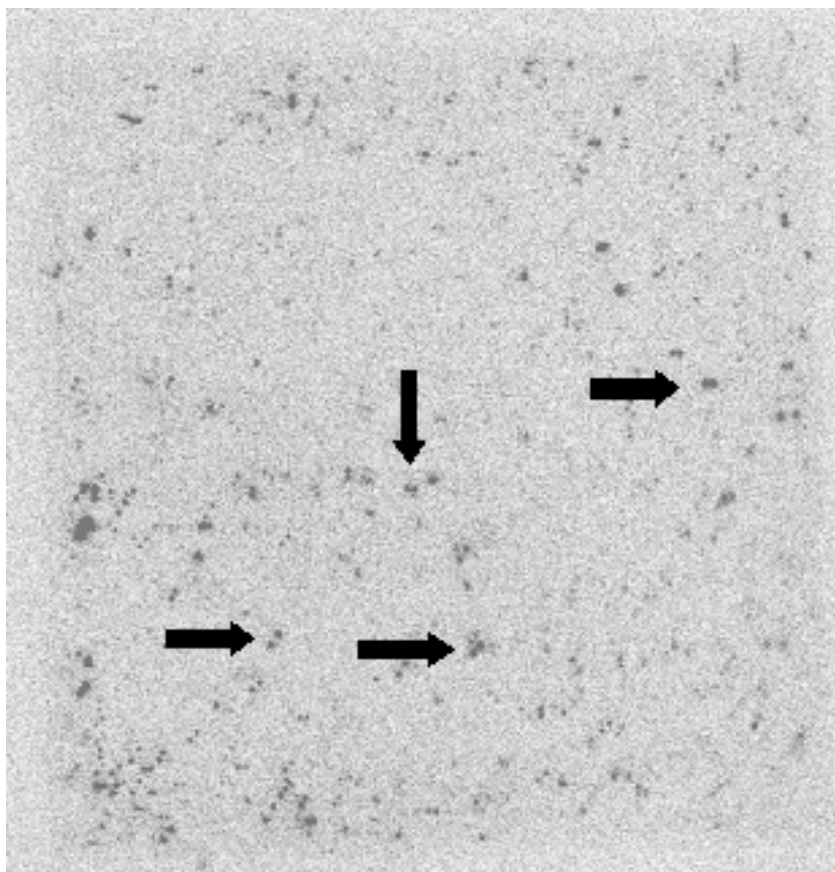


Figura 3.22 Autorradiografía de un filtro de BAC de alta densidad (Clemson, CUGI, HA_HBa), hibridado con un grupo de diez sondas overgo. Las flechas indican ejemplos de clones positivos con señales duplicadas.

Tabla 3.8 Nombre, tipo y motivo de la repetición del marcador microsatélite incluido en el clon de BAC.

Nombre clon de BAC	Marcador microsatélite	Tipo de repetición	Motivo de la repetición	GL
BAC1	HA928	Compuesto (di-trinucleótido)	GT-ATT	10
BAC2	HA928	Compuesto (di-trinucleótido)	GT-ATT	10
BAC3	HA4222	Trinucleótido	ATT	16
BAC4	HA2063	Compuesto (di-dinucleótido)	GA-GT	9
BAC5	ORS805	Dinucleótido	GA	9
BAC6	HA2600	Dinucleótido	GA	10

A fin de realizar las hibridaciones mediante la técnica de BAC-FISH los seis clones de BACs fueron marcados por la técnica de *nick translation* con biotina-14-dUTP y revelados con estreptoavidina-Cy3 y/o por la técnica de *random priming* con digoxigenina 11-dUTP y revelados con FITC.

En los ensayos de hibridación *in situ* utilizando como sonda los seis clones de BACs se observaron en todas las células señales de hibridación dispersa. Sin embargo, se observó un patrón de dispersión distinto para cada uno. Un ejemplo de estos ensayos puede verse en la Figura 3.23, al utilizar como sonda el clon de BAC3 marcado con digoxigenina 11-dUTP en cromosomas prometafásicos, se observan señales de hibridación de manera dispersa sobre la mayoría de los cromosomas existiendo diferencias entre cromosomas con respecto a la intensidad y distribución. Los resultados de la hibridación utilizando como sonda los clones de BACs sugerirían que los mismos podrían contener además de las secuencias de localización única en el genoma, correspondientes a los SSR, secuencias repetitivas. La presencia de estas regiones repetitivas podrían explicar los patrones de hibridación dispersos obtenidos.

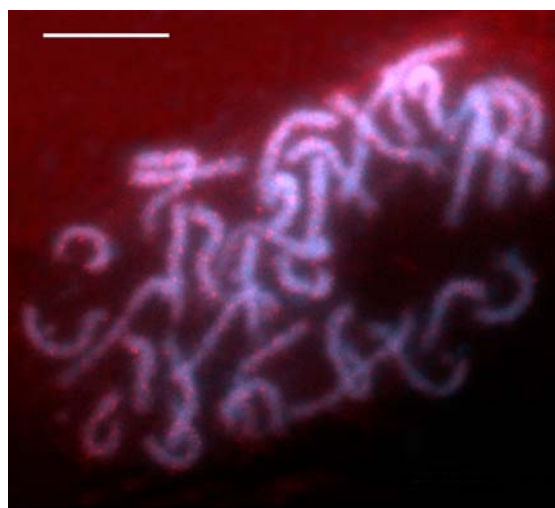


Figura 3.23 Prometafase mitótica: se observan señales de localización dispersa en ensayo de BAC-FISH utilizando como sonda el clon de BAC3 marcado con digoxigenina 11-dUTP y detectado con FITC. Contratinción con DAPI. La barra indica 10 μm .

Con el fin de evaluar la presencia de secuencias repetitivas presentes en los clones de BACs seleccionados para FISH, se realizaron ensayos de BACs-*Southern blot* utilizando como sondas las tres secuencias de ADN repetitivo descritas en el apartado 3.6.3. Estos ensayos permitieron confirmar la presencia de elementos repetitivos en todos los BACs seleccionados, los cuales mayoritariamente hibridaron positivamente con la sonda derivada del fragmento SSR-Ha785 siendo los clones de BACs 2 y 5 los que presentaron mayor número de copias de elementos repetitivos (Figura 3.24a). La sonda derivada de Rep1, con similitud a retrotransposones de la familia *copia-like* GenBank AN: AJ009967 (Santini et al., 2002), reveló señales de hibridación en los clones 1 y 6 (Figura 3.24b). No se observaron señales de hibridación al utilizar como sondas los fragmentos denominados Retro2 y Retro3 con similitud a retrotransposones de la familia *gypsy-like* GenBank AN: DQ229838 (Tang et al., 2006).

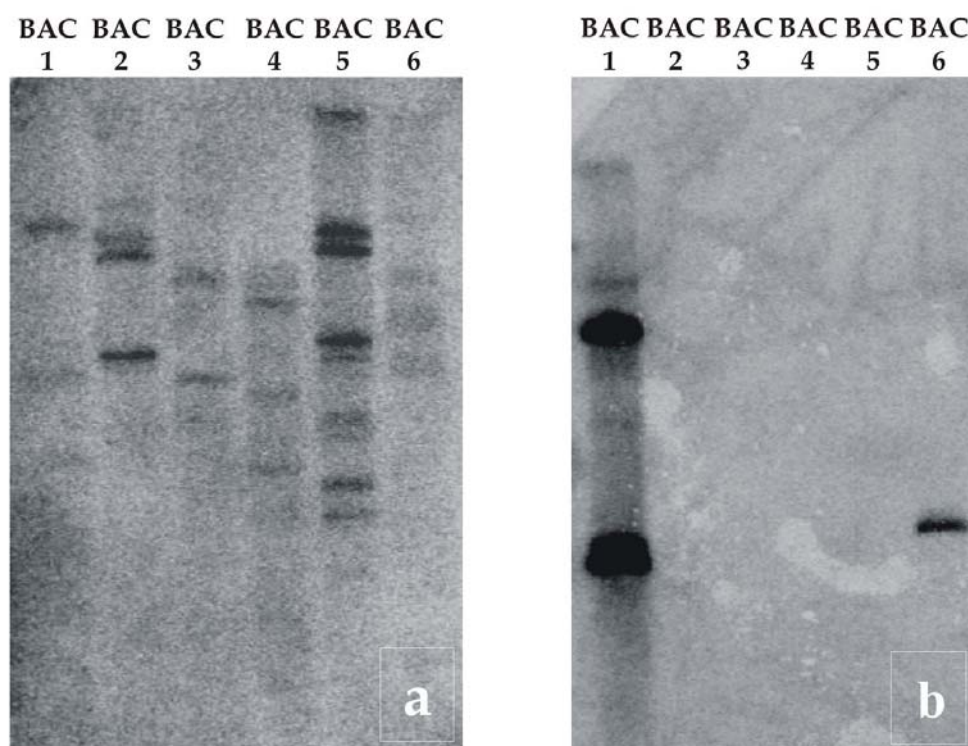


Figura 3.24 BAC- *Southern blot* de los 6 clones BAC seleccionados. **3.24a** Señales de hibridación observada en todos los clones de BACs al utilizar como sonda el fragmento SSR-Ha785 (Paniego et al., 2002) con similitud a secuencias altamente repetitivas de girasol. **3.24b** Señales de hibridación observada en los clones de BACs 1 y 6 al utilizar como sonda el fragmento Rep1, con similitud a retrotransposones de la familia *copia-like* AN AJ009967 (Santini et al., 2002).

La información derivada de estas hibridaciones fue importante para el diseño de los experimentos de bloqueo de preparados cromosómicos para la detección de secuencias únicas por la técnica BAC-FISH. Como se mencionó anteriormente, esta técnica puede resultar limitada por la presencia de señales inespecíficas debidas a hibridaciones de secuencias altamente repetitivas contenidas en la sonda.

Teniendo en cuenta los resultados del ensayo de BAC-*Southern blot*, los clones de BACs fueron hibridados utilizando como ADN bloqueante secuencias repetitivas descritas para girasol sin marcar. Para ello, se ajustaron las concentraciones de ADN bloqueante para cada BAC en particular y se realizaron varios ensayos utilizando de 10 a 80 veces de ADN bloqueador con respecto a la cantidad de ADN usado como sonda. En la Figura 3.25 pueden observarse los resultados obtenidos utilizando 20 veces de ADN bloqueador SSR-HA785 con respecto a la concentración de sonda BAC3, el cual contiene al microsatélite HA4222 anclado en el GL 16. En estas condiciones se visualizan dos señales de hibridación intensa, correspondientes a la localización única de este fragmento en un par de cromosomas supuestamente homólogos. Este resultado fue reproducible en siete células distribuidas en dos preparados.

En la Figura 3.26 se observan los resultados obtenidos utilizando 15 veces de ADN bloqueante SSR-HA785 con respecto a la concentración de sonda, y 10 veces de ADN bloqueante Rep1 en un ensayo de FISH utilizando como sonda el BAC6, el cual contiene al microsatélite HA2600, anclado en el GL 10. Si bien este resultado necesita ser confirmado se observan dos señales teloméricas en un par de cromosomas metacéntricos de tamaño similar. Este resultado fue reproducible en distintas células de un preparado.

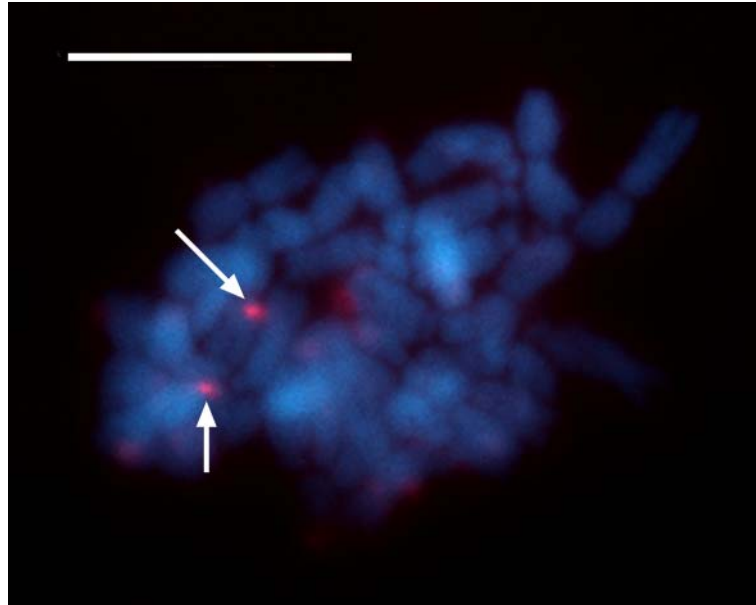


Figura 3.25 Metafase mitótica: las flechas indican señales de hibridación en cromosomas homólogos en un ensayo de BAC-FISH utilizando como sonda el clon de BAC3 conteniendo al microsatélite HA4222 (GL 16). La contratinción fue realizada con DAPI. Relación ADN marcado: ADN bloqueante 1:20. La barra indica 10 μm .

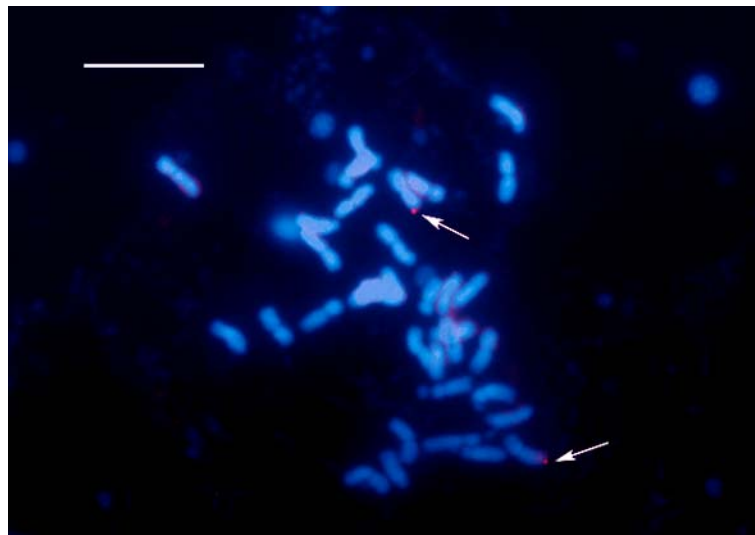


Figura 3.26 Metafase mitótica: las flechas indican dos señales teloméricas en cromosomas homólogos en un ensayo de BAC-FISH utilizando como sonda el clon de BAC6 conteniendo al microsatélite HA2600 (GL 10). La contratinción fue realizada con DAPI. Relación ADN marcado: ADN bloqueante (SSR-Ha785) 1:15 y relación ADN marcado: ADN bloqueante (Rep1) 1:10. La barra indica 10 μm .

3.7.2 Localización de secuencias de bajo número de copias o de copia única mediante PRINS y C-PRINS (Cycling-Primed in situ labelling)

3.7.2.1 Localización de secuencias microsatélites

En paralelo, y como ensayo complementario a la técnica de BAC-FISH, se probaron las técnicas de PRINS y C-PRINS. Para ello, se seleccionaron seis microsatélites incluidos en el mapa genético descrito en el apartado 3.5. En esta selección se tuvo en cuenta que el marcador hubiera presentado al menos una de las siguientes condiciones: muy buena lectura en el electroferograma al realizarse la genotipificación (picos bien definidos), que amplifique un solo alelo, que estén preasociados a QTLs de interés y/o tamaño de la región amplificada (ver Tabla 3.9).

Tabla 3.9 La tabla muestra tamaño, tipo de repetición, asociación a QTLs y el grupo de ligamiento al cual están anclados los microsatélites seleccionados para poner a punto las técnicas de PRINS y C-PRINS.

Nombre del SSR	Tamaño	Motivo	Función asociada	GL	Referencias
HA3102	229pb	GA	QTL de resistencia a Phoma	15	Afadil et al. (2007)
HA3103	184pb	GA	QTL de resistencia a estrés hídrico	7	Kiani et al. (2007)
ORS297	226pb	TG	QTL de resistencia a estrés hídrico	17	Kiani et al. (2007)
ORS805	279pb	CCA/AG	QTL de resistencia a Phoma	9	Afadil et al. (2007)
ORS1085	281pb	CT	QTL de resistencia a estrés hídrico	12	Kiani et al. (2007)
ORS1247	339pb	CT	QTL de resistencia a <i>S. sclerotiorum</i>	17	Maringolo, comunicación personal

Mediante la técnica de PRINS, utilizando los oligonucleótidos correspondientes a la región flanqueante de los microsatélites seleccionados se detectaron para todos los casos señales de hibridación inespecíficas y baja eficiencia en la reproducibilidad. Debido a la dificultad en la puesta a punto de la técnica y a la baja sensibilidad de la misma para detectar secuencias de copia única y con el propósito de disponer de una técnica de citogenética molecular altamente sensible y específica para la detección de secuencias de

bajo número de copia o de copia única, se optimizó la técnica C-PRINS (ver Materiales y Métodos, sección 2.14.1).

En una primera etapa se utilizaron los oligonucleótidos correspondientes a la región flanqueante de los marcadores microsatélites ORS1247, anclado en el GL 17, de 339pb, el cual presenta asociación a QTLs de resistencia a *S. sclerotiorum* (Maringolo, comunicación personal) y el microsatélite HA3103, anclado en el GL 7, el cual presenta asociación a QTLs de resistencia a estrés hídrico (Kiani et al., 2007).

La Figura 3.27 muestra la extensión mediante C-PRINS utilizando oligonucleótidos que amplifican al microsatélite ORS1247. Se observaron dos señales de hibridación localizadas en los satélites de un par de cromosomas homólogos. Estos resultados fueron reproducibles en distintas células de un preparado.

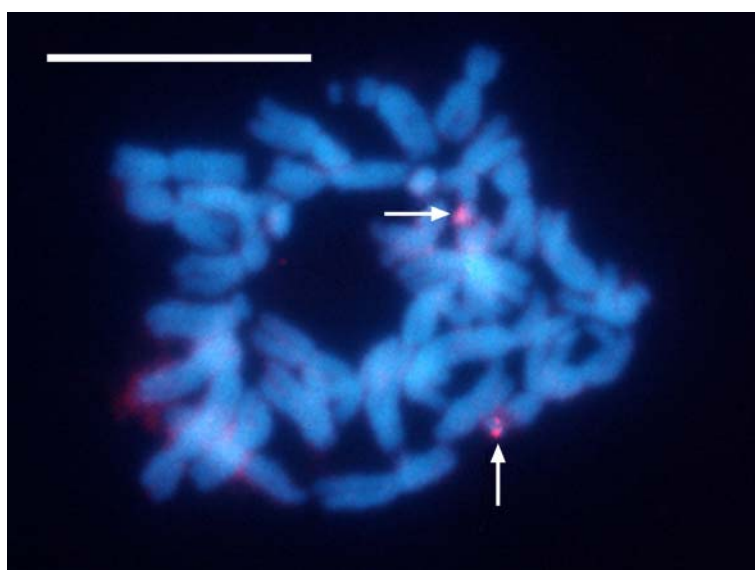


Figura 3.27 Metafase mitótica: las flechas indican la localización del microsatélite ORS1247 usando C-PRINS. La barra indica 10 μm .

La Figura 3.28a muestra la extensión mediante C-PRINS, utilizando oligonucleótidos que amplifican al microsatélite HA3103 en cromosomas metafásicos; se observa una señal de hibridación localizada en la región media de un cromosoma. La Figura 3.28b muestra la extensión del mismo marcador en un núcleo interfásico. Esto

indicaría que en la línea HA89 existiría una mutación en el sitio de hibridación del oligonucleótido en el cromosoma homólogo. Estos resultados fueron reproducibles en 10 células distribuidas en dos preparados.

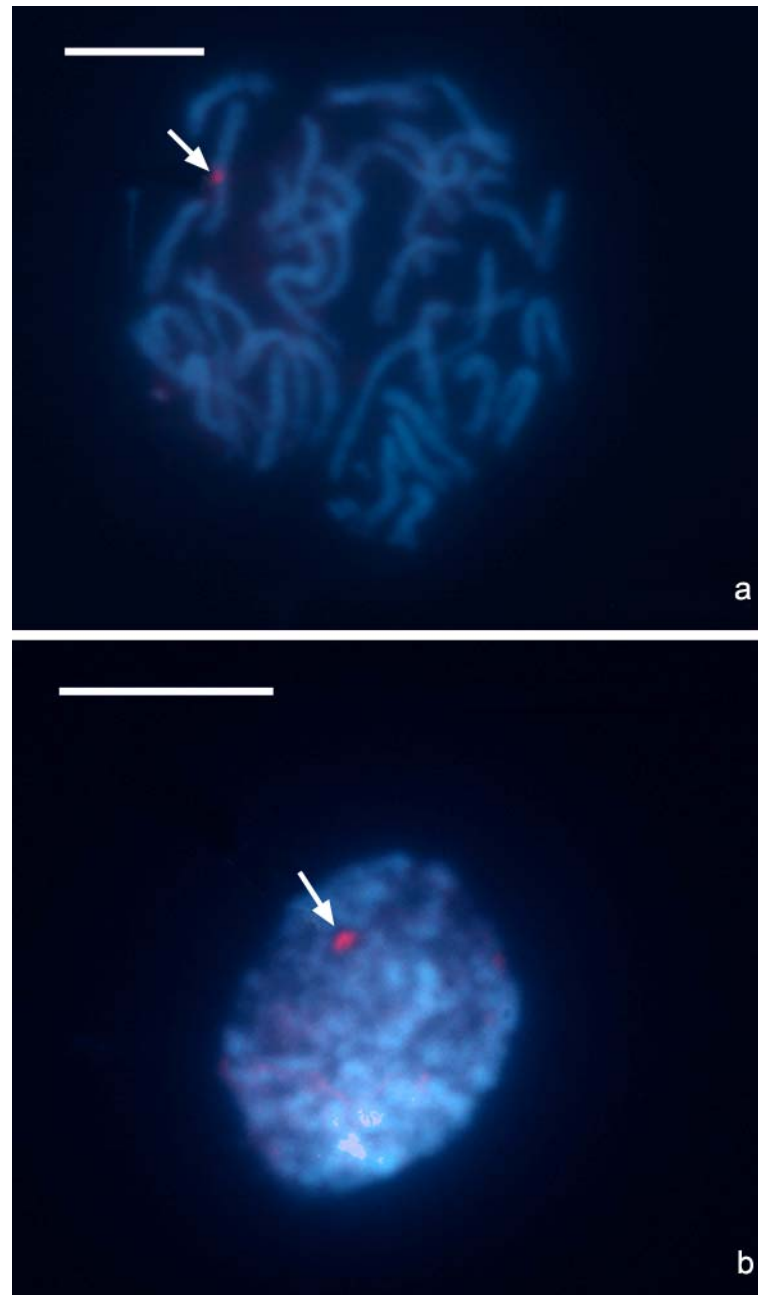


Figura 3.28 a: Metafase mitótica **b:** núcleo interfásico. Las flechas señalan la localización del microsatélite HA3103 mediante C-PRINS. La barra indica 10 μm .

3.7.2.2 Localización de secuencias de RGAs (Análogos de genes de resistencia)

Se amplificaron y se clonaron cuatro fragmentos que presentaban similitud a RGAs en base de datos públicas de girasol, denominados HaNTP3 de 250 pb y HaNTP6 de 270 pb, pertenecientes a la familia de los genes R de tipo non-TIR-NBS-LRR, caracterizados y mapeados por Radwan et al. (2003) en el GL 13 del mapa de Yu et al. (2003) y los fragmentos denominados HaP2 de 1.200 pb, HaP3 de 1.150 pb, pertenecientes a la familia de los genes R de tipo CC-NBS-LRR caracterizados y mapeados por Radwan et al. (2004) en el GL13 del mapa de Yu et al. (2003).

Se realizaron varios ensayos de hibridación mediante FISH, utilizando como sonda el producto de amplificación de los fragmentos HaP2 de 1.200 pb y HaP3 de 1.150 pb y mediante la técnica de PRINS, utilizando el oligonucleótido HaP2. En ningún caso se observó señal de hibridación con dichas sondas. Si se considera que se trata de secuencias de bajo número de copias, y atendiendo al resultado positivo obtenido con la técnica C-PRINS para la localización de marcadores únicos en el genoma, la optimización de la técnica C-PRINS para la detección de RGAs en girasol puede considerarse como una alternativa a explorar.

4. DISCUSIÓN

Este trabajo se focaliza en el desarrollo de un mapa genético saturado de referencia y la localización de secuencias de interés en un mapa físico. Contar con un mapa de estas características que integre la información de los mapas existentes para girasol, el desarrollo de marcadores moleculares y la localización cromosómica de secuencias de interés mediante la elaboración de un mapa físico, constituyen herramientas genómicas valiosas tanto para profundizar el análisis genómico de regiones agrónomicamente importantes y el análisis comparativo de genomas de especies relacionadas, como para su aplicación en programas de selección asistida con marcadores moleculares y/o el clonado molecular de genes de interés.

4.1 Población de mapeo

En el desarrollo de esta tesis se utilizó una población de líneas recombinantes endocriadas provenientes del cruzamiento de las líneas puras de girasol RHA266 (resistente a mildiu) y PAC2 (resistente a *S. sclerotium*). Esta población ha sido utilizada en numerosos trabajos de mapeo genético y mapeo de QTLs (Al-Chaarani et al., 2002; 2004; Alfadil et al., 2007; Ebrahimi et al., 2008; Flores Berrios et al., 1999a; 2000; Herve et al., 2001; Hu et al., 2006; Kiani et al., 2007a y b; Kiani; 2008). Algunos estudios sostienen la ventaja de la utilización de RILs para el desarrollo de mapas genéticos y la detección de QTLs (Austin & Lee, 1996), debido a que las RILs son genotipos homocigotas y pueden ser propagadas sin futuras segregaciones. Además, sufren meiosis múltiples antes de alcanzar la homocigosidad, por lo tanto, los genes ligados tienen gran probabilidad de recombinar y, de esta manera, los efectos pleiotrópicos pueden ser detectados (Burr & Burr, 1991). Estos efectos incrementan el poder de evaluar diferencias entre clases genotípicas. Así, los mapas desarrollados con estas poblaciones presentan mayor resolución que los mapas desarrollados con poblaciones F2.

El grado de precisión en la medición de las distancias genéticas y la determinación del orden de los marcadores están directamente relacionados al número de individuos estudiados en la población de mapeo. Idealmente, una población de mapeo para la

construcción de mapas de ligamiento debe contener como mínimo 50 individuos (Collard et al., 2005). La población utilizada en este trabajo consiste de 123 líneas recombinantes, por lo que puede ser considerada dentro de los parámetros admisibles para la obtención de un mapa confiable.

4.2 Desarrollo de marcadores moleculares funcionales en girasol

El desarrollo de marcadores funcionales se realizó a partir de 36.741 secuencias de unigenes disponibles en bases de datos públicas (Índice de Genes de Girasol, <http://compbio.dfci.harvard.edu/tgi/>), a través de la identificación de microsatélites contenidos dentro de secuencias que se expresan (EST-SSR), polimorfismos de un solo nucleótido (SNP) o por inserciones y deleciones de nucleótidos (InDels).

En la búsqueda de los marcadores funciones de tipo EST-SSR fueron consideradas las secuencias que presentaron más de un motivo microsatélite, llamados microsatélites compuestos, pero no aquellas donde la repetición está interrumpida, llamadas microsatélites imperfectos, ya que esto hubiese dificultado la interpretación de los resultados. Además, según lo reportado en otras especies, este tipo de SSR genera bajo nivel de polimorfismo, presumiblemente debido a que el estadio mutacional inicial genera menos posibilidades para el deslizamiento de la polimerasa (Smulders et al., 1997).

El proceso de identificación de microsatélites en secuencias de unigenes disponibles en bases de datos públicas se inicia con la definición de los criterios de identificación de los tipos de repetición, los cuales se ajustan en gran medida al propósito final del estudio. En este caso, como el propósito fue identificar marcadores informativos, se utilizaron criterios restrictivos fundamentalmente en relación al largo total de las repeticiones, considerando que en estudios previos la probabilidad de encontrar polimorfismos aumenta de manera proporcional a la longitud del SSRs (Cho et al., 2000; La Rota et al., 2005; Temnykh et al., 2001). Los criterios utilizados incluyeron una longitud mínima de 18 bases para los di- y trinucleótidos, y de 20 bases para los tetra- y pentanucleótidos, identificándose un total de 1.134 secuencias. Niveles similares de astringencia fueron reportados por Kantety et al. (2002) y Gao et al. (2003). En una segunda etapa, se amplió el rango de longitud de la repetición a un valor mayor o igual a

12 nucleótidos, con el propósito de incluir mayor cantidad de secuencias candidatas relacionadas con características de interés agronómico, identificándose en este caso 4.002 secuencias. Este considerable aumento en la identificación de EST-SSR coincide con lo observado por otros autores al utilizar como criterio de búsqueda longitudes de repetición entre 10 y 12 nucleótidos.

Como resultado de la búsqueda de motivos de microsatélites contenidos en secuencias de unigenes de *H. annuus*, fueron identificados SSR en más del 3% de las secuencias analizadas, bajo el criterio de 18 bases para los di- y trinucleótidos, y de 20 bases para los tetra- y pentanucleótidos. Este resultado es consistente con lo encontrado por otros autores, los cuales describen porcentajes de EST-SSR entre 1,5% y 4,7% de las secuencias analizadas en cebada, maíz, arroz, sorgo y trigo (Gao et al., 2003; Kantety et al., 2002). Sin embargo, en especies como el trigo y la pimienta, han sido detectadas frecuencias mayores al 11% (Nicot et al., 2004; Yi et al., 2006). Kumpatla & Mukhopadhyay (2005) analizaron EST-SSR en 49 dicotiledóneas incluyendo al girasol y describieron porcentajes que varían entre 2,65% y 16,82%. Para girasol en particular, analizaron 60.007 secuencias de ESTs depositadas en el GenBank (<http://www.ncbi.nlm.nih.gov>) y encontraron una frecuencia de ESTs-SSRs de 6,18%. En dicho estudio, el criterio para la identificación de SSR fue menos astringente que el utilizado en este trabajo, incluyendo entre los motivos de repetición a los mononucleótidos y tomando el valor de 10 nucleótidos como el largo mínimo de la repetición. Dichos autores concluyen que los mononucleótidos fueron los más frecuentes, seguidos por los dinucleótidos y los trinucleótidos. Por otro lado, Pashley et al. (2006) buscaron EST-SSR a partir de secuencias de EST de girasol depositadas en la base de datos CGPDB (<http://cgpdb.ucdavis.edu>), con un criterio de selección de una repetición mínima de 10 unidades para los dinucleótidos y 12 para los tri- y tetranucleótidos. En esta colección, el tipo de repetición más abundante fue el de trinucleótidos, seguido por el de tetranucleótidos y luego el de dinucleótidos, siendo ATG/CAT el motivo trinucleótido más abundante. Una selección de 188 secuencias fueron utilizadas para diseñar oligonucleótidos y amplificarlos en 12 especies silvestres de *H. annuus*, con el propósito de analizar la transferibilidad de los mismos. Como resultado de este estudio se identificó un conjunto 48 EST-SSR funcionales. Recientemente, Heesacker et al. (2008) desarrollaron EST-SSR a partir de una

colección de 17.904 unigenes obtenidos a partir del ensamblado de 89.225 ESTs depositados en el GenBank (<http://www.ncbi.nlm.nih.gov/dbEST/>) provenientes de *H. annuus*, *H. argophyllus* y *H. paradoxus*. Encontraron que el 10,9% de los 17.904 unigenes de *H. annuus* contenían SSRs, adoptando como criterio que la longitud de la repetición fuera mayor o igual de 10 pb para dinucleótidos ($n \geq 5$). El tipo de repetición más frecuente fue el de dinucleótidos, seguido por el de trinucleótidos y, por último el de tetranucleótidos. Para este análisis el criterio de búsqueda utilizado se basó en un número mínimo de repeticiones (n) considerando los valores de $n \geq 4$ y $n \geq 5$.

Las diferencias en los resultados obtenidos entre estos trabajos en general pueden ser debidas a los diferentes conjuntos de datos utilizados para conducir las búsquedas (ESTs y/o unigenes provenientes de distintos procesos de ensamblado), y a las herramientas y criterios de selección utilizados en la búsqueda. En particular, las diferencias observadas en este estudio respecto a las observaciones de otros autores pueden explicarse en que sólo se consideraron microsatélites perfectos con una longitud mayor o igual a 20 pb y no se incluyeron mononucleótidos.

En la base de datos de unigenes de girasol (HAGI) según los procedimientos aplicados en este trabajo, los trinucleótidos fueron el tipo de repetición más abundante, representada por el 51,7% de las secuencias, seguido por los dinucleótidos (30,3%), luego por los tetranucleótidos (9,3%) y, por último, por los pentanucleótidos con 8,7%. Estos resultados podrían estar reflejando el hecho de que el número de repeticiones de los trinucleótidos puede cambiar sin alterar el marco de lectura del ARNm, y que el número de posibles combinaciones de los trinucleótidos es mayor a la de los dinucleótidos. Esta dominancia de SSR trinucleótidos puede ser explicada, además, sobre la base de la supresión de SSRs no trinucleótidos en regiones codificantes, debido al riesgo de mutaciones en el marco de lectura que puede ocurrir cuando aquellos microsatélites cambian en una unidad (Rungis et al., 2004).

El motivo dinucleótido más abundante encontrado en la base de datos analizada fue GA/CT, presente en el 23,2% de las secuencias. Un dinucleótido puede representar múltiples codones, dependiendo del marco de lectura, y ser traducido en diferentes aminoácidos. Por ejemplo, el motivo GA/CT puede representar los codones GAG, AGA, UCU y CUC en una población de ARNm, y traducirse en los aminoácidos Arg, Glu, Ala y Leu, respectivamente. Ala y Leu están presentes con alta frecuencia en las

proteínas, 8% y 10% respectivamente (Lewin 1994). Esta podría ser una de las razones por las que el motivo GA/CT está presente en alta frecuencia en las colecciones de ESTs. Investigaciones previas demostraron que el motivo AG fue el motivo dinucleótido más frecuente en ESTs de plantas (Kumpatla & Mukhopadhyay, 2005; Morgante et al., 2002; Parida et al., 2006; Varshney et al., 2002, Yi et al., 2006). El motivo trinucleótido más frecuente en girasol encontrado en este trabajo fue el AAG/CTT (11,9%), seguido por el ACC/GGT (9,2%) y por el AAT/ATT (7,8%). Según el estudio desarrollado por Kumpatla & Mukhopadhyay (2005), el motivo AAG/CTT es el motivo más abundante en dicotiledóneas. Sin embargo, estos autores reportaron que para el girasol el motivo trinucleótido más frecuente fue GGT/ACC.

4.3 Evaluación de diferentes estrategias de genotipificación

Con las diferentes técnicas de genotipificación utilizadas durante el desarrollo de este trabajo se obtuvieron resultados confiables y reproducibles. Sin embargo, la utilización de las técnicas de genotipificación manuales, mediante geles de poliacrilamida y la visualización con nitrato de plata, con SYBR[®]Gold o con nucleótidos marcados con radiactivo y detección por autorradiografía fueron muy demandantes de tiempo. La genotipificación de los marcadores moleculares alcanzó un alto nivel de procesividad con el desarrollo de la técnica de electroforesis capilar utilizando tres fluoróforos (6-FAM, HEX y NED) en el diseño de los oligonucleótidos. Esa estrategia permitió reunir en cada panel de 2 a 6 marcadores compatibles, combinado los colores con los diferentes tamaños en pb de los fragmentos amplificados, de manera de prevenir la superposición de fragmentos del mismo tamaño y del mismo color. Con esta metodología se obtuvo una matriz de datos para mapeo con una mínima cantidad de datos perdidos.

A fin de disminuir los errores derivados del proceso de genotipificación se emplearon distintas estrategias: a) para errores humanos: todos los análisis, tanto manuales como automáticos, fueron leídos por dos operadores; b) para disminuir los errores debidos a la baja cantidad y calidad del ADN se prepararon diluciones de las muestras y fueron chequeadas por agarosa 1% antes de la amplificación de los fragmentos. Además, a fin de preparar las diluciones óptimas para la corrida por EC, los

fragmentos se evaluaron en geles de agarosa 1,5%. Los resultados de este trabajo indican que estas medidas fueron eficaces en el sentido de que sólo un marcador presentó segregación distorsionada, los resultados de la genotipificación manual se reprodujeron en la genotipificación automática y, finalmente, en la obtención de un mapa genético robusto en el cual la co-localización de marcadores respecto de otros mapas públicos permitió referenciar los grupos de ligamiento, unificando la nomenclatura con los mapas disponibles.

Con el método de marcado indirecto se observó baja eficiencia en la amplificación y en la reproducibilidad del mismo en girasol, dificultando su aplicación en el mapeo de microsátélites. Estos resultados posiblemente se relacionan con la complejidad que introduce en la reacción de PCR, el agregado de tres oligonucleótidos, lo cual exige que las temperaturas de hibridación de los oligonucleótidos sean lo más similares posibles y que se ajusten a las relaciones entre los oligonucleótidos no marcados respecto al oligonucleótido universal M13 marcado y el ciclado de la PCR. Estas demandas del sistema hacen que no todos los estudios referidos a la técnica coincidan en la aplicabilidad de la misma (Missiaggia & Grattapaglia, 2006; Zhou et al., 2002). Dado que la única ventaja de este método es el costo de producción del dato, y en esta experiencia el procedimiento fue exitoso sólo para la amplificación de un número bajo de *loci*, rindiendo en la mayoría de los casos ensayados un porcentaje alto de datos faltantes, se optó por el método de marcado directo.

Independientemente del método de genotipificación empleado los alelos observados en las líneas parentales para los SSR desarrollados en el INTA (HAX) y en la Universidad de Oregon (ORSx) coincidieron en el 92% de los casos con los alelos definidos previamente. Así, en la mayoría de las líneas analizadas se observaron los mismos alelos y en muy pocos casos fueron identificados alelos nuevos, lo cual puede ser atribuido a errores en la estimación del tamaño molecular durante la genotipificación por métodos manuales, en algunos casos podría estar asociada a la inestabilidad propia de este tipo de marcadores y/o diferentes procesos de recombinación o mutación entre los distintos genotipos; que conducen a la aparición de nuevos alelos; no obstante no se realizaron estudios tendientes a confirmar esta última causa.

4.4 Análisis de polimorfismo de marcadores neutros y funcionales

En el análisis de los marcadores genómicos en los parentales de la población de mapeo (RHA266 x PAC2) se observó un 37,7% de marcadores polimórficos, un 40,6% de monomórficos, un 15% de alelos nulos y en un 6,7% falló la reacción de amplificación para ambos alelos. Estos resultados coinciden con lo reportado por Tang et al. (2002), quienes encuentran un nivel de polimorfismo del 41,2%, entre los genotipos RHA280 x RHA801. Para la construcción de un mapa genético es fundamental que existan polimorfismos entre los genotipo parentales. Este factor no puede ser subestimado porque la ausencia de polimorfismos hace imposible el análisis de segregación y el mapeo de ligamiento (Young et al., 1994). Los resultados descritos respecto del nivel de polimorfismo observado entre parentales de distintas poblaciones de mapeo para un conjunto representativo de la colección de SSR públicos de girasol permiten prever el número de marcadores SSR que deberán analizarse al inicio de un trabajo de mapeo en girasol cultivado. El porcentaje de alelos nulos en el análisis de los SSR genómicos fue del 15%, coincidiendo este resultado con un estudio reportado por Tang et al. (2002), en el cual se observó un porcentaje de 13,7% en uno o más de los *loci* analizados.

La validación de los SSR génicos en cuatro genotipos de girasol (RHA266, PAC2, HA89 y RHA801) resultó en la amplificación de 74 secuencias de un total de 127 analizadas. El 13% de los *loci* amplificados fueron polimórficos, el 45%, monomórficos, el 5% alelos nulos y en el 37% restante no se observó producto de amplificación o se generaron productos de amplificación inespecíficos, complejos o difíciles de resolver. El porcentaje de polimorfismo observado coincide con el reportado por Heesacker et al. (2008), los cuales aseguran que menos del 10% de los *loci* que se transcriben pueden ser mapeados genéticamente usando SSR en girasol y con lo reportado en otras especies (Eujayl et al., 2004; Fraser et al., 2004; Varshney et al., 2005b). Las posibles explicaciones de las fallas en la amplificación incluyen: a) uno o ambos oligonucleótidos se extienden a través de un sitio de procesamiento de ARN o *splicing*; b) presencia de intrones, InDels o sustituciones en secuencias de ADN genómico; y c) desarrollo de los oligonucleótidos a partir de secuencias con alto número de indeterminaciones o errores de ensamblado. De esta manera, la calidad de las secuencias de EST depositadas en bases de datos, y los procesos de curado, edición y ensamblado aplicados a estas secuencias en la creación de bases de datos secundarias,

como la de unigenes, son muy importantes para el diseño de marcadores moleculares (Varshney et al., 2005b).

La ocurrencia de alelos nulos puede producirse por pequeñas diferencias en las regiones flanqueantes debidas a mutaciones puntuales (InDels o sustituciones) en el sitio de pegado de los oligonucleótidos (Lehman et al., 1996), que pueden conducir a un débil pegado de los oligonucleótidos, especialmente en la región 3' y/o a deleciones del microsatélite en un *locus* específico (Callen et al., 1993). Fueron observados alelos nulos en EST-SSR en arroz (Cho et al., 2000), en kiwi (Fraser et al., 2004) y en trigo (Eujayl et al., 2001; Gupta et al., 2003), entre otras especies de importancia agronómica. En algunas especies, tales como trigo y *Picea*, se ha encontrado una menor frecuencia de alelos nulos en los marcadores EST-SSR con respecto a los SSR genómicos (Leigh et al., 2003; Rungis et al., 2004). Los resultado del presente trabajo en girasol son congruentes con los reportados por estos autores, habiéndose detectado una frecuencia de 5-7% en EST-SSR y de 15% en SSR genómicos.

En general, se han observado porcentajes de amplificación tanto para SSR genómicos como para SSR génicos similares a los encontrados en diferentes estudios (Cordeiro et al., 2001; Gupta et al., 2003; Kota et al., 2001; Saha et al., 2004; Thiel et al., 2003, Yu et al., 2004).

El método de CAPS, basado en la identificación de polimorfismos de nucleótido simple por digestión con endonucleasas de restricción de productos de PCR específicos, resultó exitoso para la detección virtual de SNP. Sin embargo, no fue posible confirmar experimentalmente la mayoría de esos polimorfismos entre los parentales de la población de líneas recombinantes. Cabe aclarar que el programa SNP2CAPS no evalúa la utilidad técnica del patrón de restricción sugerido, por lo que se realizó una inspección visual de los resultados a fin de confirmar que el sitio de restricción sugerido no se encontrara en el extremo de los fragmentos de PCR, o muy cercano a un segundo sitio de restricción como para permitir la resolución de los fragmentos en un gel de agarosa. Además, se realizaron optimizaciones de las condiciones de la reacción de PCR, incluyendo ajustes del número de ciclos, de la temperatura de hibridación y de la concentración de oligonucleótidos, y de la reacción de digestión probando distintas concentraciones de los productos de PCR y de la enzima (en todos los casos se utilizaron controles para las condiciones enzimáticas usadas). No obstante estos

recaudos, sólo pudieron ser evaluadas experimentalmente dos secuencias polimórficas, las cuales, a su vez, diferían del tamaño esperado e inferido por el programa.

4.5 Mapa genético de ligamiento

Con el propósito de referenciar los mapas de ligamiento existentes para girasol, se seleccionó un grupo de 256 marcadores microsatélites públicos desarrollados en la Universidad de Oregon de un conjunto de 657 marcadores mapeados en tres poblaciones de mapeo por Tang et al. (2002; 2003), distribuidos en los 17 grupos de ligamiento. Asimismo, se incluyeron 404 marcadores microsatélites desarrollados en el INTA. Para el análisis de polimorfismo entre las líneas parentales fueron analizados 594 SSR genómicos (278 HAx y 315 ORSx) y 127 EST-SSR, de los cuales resultaron polimórficos y fueron genotipificados en la progenie de RILs 224 SSR (101 HAx y 123 ORSx) y 14 EST-SSR. Ochenta y siete marcadores SSR (ORSx) correlacionaron con los GL de mapas publicados (36 con el mapa de Yu et al. (2003); 34 con el mapa de Tang et al. (2002) y 17 con el mapa de Tang et al. (2003). La selección de los marcadores microsatélites para referenciar los mapas de ligamiento existentes para girasol resultó adecuada, ya que los grupos de ligamiento y el orden de los marcadores en cada GL se correlacionaron con aquellos marcadores localizados previamente en tres mapas de ligamiento (Paniego et al., 2006; Tang et al., 2002; 2003; Yu et al., 2003).

De los 584 marcadores que componen la matriz de datos, sólo el marcador HA765 presentó distorsión en la segregación de los alelos al ser analizado con la prueba estadística FDR. Debido a que este tipo de datos influye negativamente en la construcción del mapa genético (Zhang et al., 2002) y en la detección de QTLs, ocasionando asociaciones espurias, se eliminó este marcador de la matriz de datos. Este tipo de distorsiones han sido reportadas previamente en este y en otros cultivos (León et al., 2000; Mutengwa et al., 2005; Vogl & Xu, 2000; Winter et al., 2000) y pueden deberse a errores en la genotipificación de los marcadores, a artefactos de la PCR o a algún tipo de selección que haya ocurrido durante el desarrollo de las poblaciones (León et al., 2000).

De los 224 marcadores microsatélites genotipificados, 37 (32 SSR y 5 EST-SSR) no fueron incluidos en ningún grupo de ligamiento por falta de asociación a otros

marcadores. Esta falta de ligamiento podría ser debida a que algunos de los marcadores se encuentran en posiciones teloméricas, y la baja densidad de marcadores en esta zona ocasiona distancias muy grandes que no permiten la detección de ligamiento (Hu 2006). Algunos de los marcadores que no fueron incluidos en este mapa posiblemente podrán ser incluidos cuando se adicione más marcadores al análisis.

El mapa desarrollado en el presente trabajo a partir de una matriz de datos que incluye 304 marcadores AFLP, 231 SSR, 9 EST-SSR y 3 InDels, cubre una longitud total de 1.942,3 cM y una distancia media entre marcadores de 3,6 cM. Fue construido con 123 RILs provenientes del cruzamiento entre RHA266 y PAC2. La región de mayor longitud sin cubrir por marcadores (*gaps*) fue de 13,2 cM, y se observó en el GL 17. La longitud de los GL varió entre 71,9 cM en el GL 7 y 203 cM en el GL 10. El mapa presentado en este trabajo representa una versión avanzada del mapa previo publicado por Al-Charani et al. (2004), dicho mapa incluye 371 marcadores AFLP y 38 SSR, cubre una longitud de 2.915,9 cM a una densidad media entre marcadores de 7,9 cM. Este mapa presenta regiones de baja cobertura o *gaps* de marcadores de más de 35 cM y un número total de 21 GL, mayor al número haploide de cromosomas de la especie ($n=17$). La saturación y la integración de marcadores de distinta naturaleza, fundamentalmente AFLPs y SSR, realizada en este trabajo de tesis sobre el mapa desarrollado por Al-Charani ha permitido reducir el número de GL a 17, disminuyendo notablemente la extensión de las regiones de baja cobertura existentes.

Por su parte, Tang et al. (2002) desarrollaron un mapa genético utilizando 459 marcadores SSR. Este mapa cubre una longitud de 1.368,3 cM con una densidad media de 3,1 cM, y fue construido con 94 RILs provenientes del cruzamiento entre RHA280 y RHA801. Los marcadores se encuentran distribuidos en 17 GL. Estos autores referenciaron tres mapas construidos a partir de tres poblaciones derivadas del cruzamiento entre germoplasmas contrastantes (RHA280 x RHA281, PHA x PHB y HA370 x HA372), utilizando 701 SSR y 89 RFLPs o InDels (Yu et al., 2003). Finalmente, publicaron un mapa compuesto que integra 657 *loci* SSR con una densidad media de 2,2 cM por *locus* y una longitud de 1.423 cM (Tang et al., 2003). Sin embargo, a pesar de la densidad del mismo, este mapa presenta en varios GL intervalos o *gaps* de más de 30 cM sin cubrir. Las diferencias en la longitud del mapa y en el

tamaño de los *gaps* de este mapa con respecto al desarrollado en el presente trabajo pueden ser atribuidas a los distintos tipos de marcadores moleculares utilizados.

En este trabajo se observa una amplia distribución de los marcadores SSR en los 17 GL, coincidiendo con el número haploide de la especie. Si bien los marcadores mapeados (AFLP y SSR) se observan homogéneamente distribuidos en algunas regiones del genoma y agrupados en otras, el fenómeno de formación de *clusters* fue mínimo, ya que no obstaculizó el desarrollo del mapa genético. Esta observación coincide con lo reportado por Flores Berrios et al. (2000) utilizando la misma población que en este trabajo y los mismos marcadores AFLPs, y con lo descrito en el mapa genético de girasol desarrollado por Tang et al. (2002). Estos autores reportaron que los marcadores microsátélites se encontraron distribuidos formando un denso *cluster* en los GL 4, 10 y 17 y un moderado agrupamiento en los GL 2, 3, 5, 13 y 15. Además, el presente mapa y/o los marcadores utilizados en este trabajo han sido utilizados para realizar diferentes estudios de caracterización de QTLs para características de importancia agronómica en girasol (Al-Chaarani et al., 2002; Alfadil et al., 2007; Bert et al., 2001; Darvishzadeh et al., 2007; Ebrahimi et al., 2008; Flores Berrios et al., 1999a y Flores Berrios et al., 1999b; Hervé et al., 2001; Kiani et al., 2007a; 2007b; 2008; Lawson et al., 1998; León et al., 1995; 1996; 2000; 2001; 2003; Lu et al., 1999; Maringolo, 2007; Mestries et al., 1998; Mokrani et al., 2002; Pérez-Vich et al., 2004). El fenómeno de formación de agrupamientos o *clusters* ha sido atribuido a la frecuente asociación de los SSR con retrotransposones y con otros elementos de ADN repetitivo, a la alta densidad de dichos elementos en regiones centroméricas y a la reducida recombinación en las regiones centroméricas (Areshechenkova & Ganal, 1999; Cuadrado & Schwarzacher, 1998; Künzel et al., 2000; Ramsay et al., 1999; 2000). Este fenómeno se ha observado en varias especies, tales como algunas coníferas y cebada (Paglia et al., 1998; Ramsay et al., 2000), pero no en otras como soja y arroz (Cregan et al., 1999b; Temnykh et al., 2000). En girasol, estudios de hibridación *in situ*, usando como sondas secuencias repetidas derivadas de retroelementos de *H. annuus*, permitieron la detección de secuencias de las familias de retroelementos *Ty1/copia-like* y *Ty3/gypsy-like* dispersas a lo largo de todos los cromosomas en todas las especies estudiadas. Sin embargo, las secuencias del retrotrasposón *Ty3/gypsy-like* fueron localizadas preferencialmente en las regiones centroméricas en la mayoría de las

especies estudiadas, mientras que las secuencias *Ty1/copia-like* fueron menos representadas o ausentes alrededor del centrómero y más abundantes en los extremos de los cromosomas en *H. annuus*.

La evaluación de los marcadores de distintas fuentes, los programas de mapeo y los parámetros estadísticos aplicados resultaron adecuados, ya que en general los marcadores se distribuyen de manera amplia y coherente para aquellos marcadores localizados previamente en otros mapas de ligamiento.

El desarrollo de mapas genéticos es el primer paso hacia la detección de factores que controlan la expresión de características de importancia agronómica. La identificación de marcadores o genes ligados a QTLs que controlan características de importancia agronómica, permite su utilización en programas de mejoramiento asistido por marcadores, facilitando la selección de genotipos superiores en forma eficiente, reduciendo el tiempo y el espacio necesarios. En este contexto la información generada por el mapa de ligamiento presentado en este trabajo, podría ser útil para la identificación y validación de diversos QTLs. De esta manera este mapa sirve como punto de partida a futuros trabajos del grupo de investigación, tendientes a la identificación de regiones que controlan características de importancia agronómica y el análisis comparativo de genomas de especies relacionadas.

4.6 Caracterización del complemento cromosómico de *H. annuus*

En el presente trabajo se presenta la caracterización del complemento cromosómico de *H. annuus* mediante la utilización de diversas sondas, tanto ribosomales como repetitivas, de diferente origen. Se identificaron 11 pares metacéntricos, 2 pares submetacéntricos y 4 pares subtlocéntricos. La utilización de la sonda SSR-HA785 con similitud a secuencias repetitivas de girasol en ensayos de hibridación *in situ* mediante FISH permitió identificar algunos pares de cromosomas pero no todo el complemento. Por otra parte, la aplicación de la técnica de PRINS utilizando oligonucleótidos específicos que amplifican secuencias repetitivas de girasol con similitud al retrotransposon *Ty3-gypsy-like*, permitió observar señales de hibridación en varios cromosomas pero tampoco permitió identificar el complemento cromosómico completo. De las estrategias ensayadas la que permitió obtener un cariograma y su correspondiente

idiograma fue la técnica de BAC-FISH, utilizando como sonda un clon de BAC con alto contenido de secuencias repetitivas. Las señales de hibridación consideradas para la confección del cariograma para algunos cromosomas fueron nítidas, permitiendo la identificación de los 17 pares cromosómicos que conforman el complemento. Recientemente, Ceccarelli et al. (2007) realizaron la caracterización del complemento cromosómico de *H. annuus* utilizando una secuencia repetida en tándem de 368 pb; los cromosomas fueron subdivididos en dos grupos, 13 pares meta y submetacéntricos y 4 pares subtlocéntricos. Además mostraron un idiograma en el cual representaron la ubicación de dichas secuencias en cada uno de los 17 pares de cromosomas del girasol cultivado. Tanto en el trabajo de Ceccarelli et al. (2007) como en este trabajo, la utilización de una única sonda permitió la identificación del complemento cromosómico total para la especie, siendo la sonda reportada en el presente trabajo el primer reporte de utilización de una sonda BAC para esta finalidad, con alta reproducibilidad y disponibilidad local para una futura identificación física de regiones de interés agronómico.

La utilización de la secuencia EF235 (AN BU671882), con similitud a un ADN ribosomal 26S, obtenida a partir de la colección local de ESTs de girasol (Fernández et al., 2003), ha permitido observar la presencia de seis señales de hibridación de distinta intensidad y tamaño en satélites de 3 pares cromosómicos; lo que sugeriría variación del número de copias de la región 26S. De acuerdo a Appels & Honeticutt (1986) y Pedersen & Linde-Laursen (1994), los *clusters* de genes de ADNr repetidos en tándem pueden aparecer en más de un cromosoma, el número de los sitios y copias de las repeticiones pueden incluso variar dentro de cada complemento cromosómico haploide. Los sitios de ADNr adicionales pueden ser originados por la delección parcial o total de dichos *loci*.

Vanzela et al. (2002), utilizando la sonda pTa71 de 9 kb conteniendo las secuencias ribosomales 18S, 5.8S y 26S de *Triticum aestivum* (Gerlach & Bedbrook, 1979), en *Helianthus* obtuvieron como resultado seis señales de hibridación consistentes con las regiones organizadoras nucleolares que portaban los satélites. Por su parte, Cuellar et al. (1996; 1999) informaron la presencia de seis señales de ADNr en *H. annuus* variedad *macrocarpus*, mientras que Schrader et al. (1997) reportaron ocho señales en el mismo material. Ceccarelli et al. (2007), utilizando la misma sonda, observaron ocho señales de

hibridación, de las cuales sólo detectan señales de hibridación intensa en satélites de 3 pares cromosómicos. En todos los casos, los autores informan que no todas las señales presentan la misma intensidad. Con la sonda EF235 utilizada en este trabajo se obtuvieron excelentes señales tanto en girasol (*H. annuus* y *H. resinosus*) como en otras especies, entre ellas amaranto y cebolla (Greizerstein, comunicación personal). Esta sonda presenta ventajas respecto a la sonda pTa71 de 9Kb de trigo, generalmente utilizada para detectar este tipo de secuencias, que derivan de su menor tamaño y posibilidad de amplificación por PCR, lo que la señala como una buena opción para la detección de regiones de ADNr en distintos materiales.

4.7 Localización de secuencias repetitivas, de bajo número de copias y de copia única mediante las técnicas de BAC-FISH, PRINS o C-PINS

Con el fin de establecer la correlación del mapa genético con el mapa físico de girasol, se seleccionó inicialmente un conjunto de marcadores microsátélites de localización única en el mapa de girasol y representativo de los grupos de ligamiento de interés 9, 10, 16 y 17. En el GL10 se han localizado QTLs para resistencia al patógeno fúngico *S. sclerotiorum*, agente causal de la Podredumbre Húmeda del Capítulo (PHC) de girasol (Maringolo 2007). Asimismo, se incluyeron marcadores SSR localizados en los GL 9, 16 y 17 que contienen QTLs adicionales para este carácter, recientemente detectados (Maringolo, comunicación personal). Por otra parte se seleccionaron SSR de los GL 7, 12 y 17 asociados a QTLs para el carácter estrés hídrico (Kiani et al., 2007a) y marcadores localizados en el GL 15, en el cual Alfadil et al. (2007) encontraron QTLs asociados con resistencia a fomosis.

La estrategia de búsqueda de clones de BACs conteniendo marcadores en GL de interés, utilizando sondas de tipo *overgo*, permitió el relevamiento de filtros de alta densidad en forma procesiva, aumentando la eficiencia en la detección de clones positivos respecto de estrategias de hibridación con sondas específicas individuales diseñadas para cada secuencia blanco. Esta estrategia ha sido usada exitosamente en varias especies de plantas, tales como maíz, arroz, maní e incluso girasol (Chen et al., 2002; Feng et al., 2006; Gardiner et al., 2004; Yüksel & Paterson 2005). En el presente trabajo aplicando esta estrategia se identificaron 6 clones BAC positivos, que contienen 5 marcadores de localización única que mapean en 3 de los GL de interés: GL 10

(HA928, HA2600), GL 16 (HA4222) y GL 9 (HA2063; ORS8059). La implementación de la técnica de BAC-FISH utilizando estos clones de BACs como sondas en hibridación *in situ* reveló un patrón de hibridación disperso indicando la posible presencia de elementos repetitivos en estos clones de BACs. Los resultados del ensayo de BAC- *Southern blot* permitieron confirmar la presencia de elementos repetitivos en todos los clones de BACs utilizados, observándose un patrón de distribución diferente para cada clon de BAC, así como también diferencias en intensidad de la señal para distintas secuencias repetitivas evaluadas (secuencias repetitivas reportadas previamente en girasol y secuencias con similitud a retrotransposones de tipo *copia-like* y *gypsy-like*). Este resultado confirma lo reportado previamente por diferentes autores, los cuales coinciden en que secuencias repetidas con similitud a retrotransposones estarían altamente distribuidas en el genoma de girasol (Natali et al., 2006; Tang et al., 2006). Teniendo en cuenta los resultados del ensayo de BAC-*Southern blot*, los clones de BACs fueron hibridados utilizando como ADN bloqueante secuencias de ADN repetitivo de girasol sin marcar y fueron ajustadas las concentraciones de ADN bloqueante para cada BAC en particular. En un primera etapa se seleccionaron los BACs con bajo contenido de secuencias repetitivas, debido a que la sensibilidad de la técnica de FISH es mayor cuando no se usa ADN bloqueante (Kim et al., 2002). Nuestro estudio mostró que la técnica de BAC-FISH permitió la identificación de marcadores microsatélites anclados en GLs, en los cuales se han reportado QTLs asociados con resistencia a *S. sclerotiorum* en cromosomas metafásicos, el microsatélite HA4222 anclado en el GL 16 y el microsatélite HA2600 anclado en el GL 10. Estos resultados confirman que la técnica de BAC-FISH es una poderosa técnica para la identificación citogenética de marcadores moleculares de localización única en plantas con genomas grandes tales como girasol. La técnica de BAC-FISH ha sido utilizada para localizar físicamente secuencias específicas de ADN e individualizar cromosomas, y para la identificación citogenética de marcadores moleculares en otras especies tales como sorgo, maíz, arroz, algodón, *Arabidopsis* y papa (Dong et al., 2000; Fransz et al., 2000; Gomez et al., 1997; Hanson et al., 1995; Islam-Faridi et al., 2002; Kim et al., 2002). Si bien en algunas especies, tales como sorgo (Kim et al., 2002) no es necesario el bloqueo de las secuencias repetitivas para la localización de determinados marcadores, en la mayoría es necesario el bloqueo en algunos clones de BACs.

Debido a la necesidad de encontrar un método más simple para marcar con fluorescencia secuencias de ADN en cromosomas mitóticos, se ensayaron los métodos de PRINS y de C-PRINS. No hay reportes previos de utilización de la técnica de PRINS en estudios de citogenética molecular en girasol. La técnica de PRINS se ensayó tanto para la localización de secuencias de copia única, microsatélites incluidos en el mapa genético de ligamiento y secuencias de bajo número de copias con similitud a RGAS, como para la caracterización del complemento cromosómico, utilizando oligonucleótidos que amplifican secuencias repetitivas. Mediante la técnica de PRINS se amplificó una región correspondiente a una secuencia con similitud a retrotransposones de la familia *gypsy-like* de girasol, denominada Retro2, utilizando oligonucleótidos específicos, y se obtuvieron señales de amplificación en distintos cromosomas. Estos resultados muestran que la técnica de PRINS permite la localización de secuencias repetitivas, en menor tiempo de trabajo, comparada con la técnica de FISH. Sin embargo, la utilización de esta técnica, por el momento, no ha permitido la ubicación de secuencias de copia única, tales como SSR o de bajo número de copias como los RGAs, posiblemente debido a la dificultad en la puesta a punto de la técnica, a la dificultad de obtener preparados citológicos libres de citoplasma, a la baja sensibilidad de la técnica y/o al riesgo de que la alta señal de ruido de fondo enmascare el resultado. Este resultado coincide con lo reportado por Menke et al. (1998), quienes informaron que la técnica de PRINS es menos sensible que la de FISH y no es conveniente para la detección de secuencias de bajo número de copias. Una alternativa interesante es la puesta a punto del método de PRINS directo; el uso de este método de marcado eliminaría las señales de fondo, pero posiblemente también debilitaría la intensidad de la señal (Kubaláková et al., 2001).

En este trabajo se utilizó la técnica de C-PRINS indirecto como alternativa a la de PRINS, con el fin de detectar secuencias de copia única en girasol, microsatélites de localización génica conocida en los GL de girasol. Por el momento se han detectado señales de hibridación con la utilización de oligonucleótidos que amplifican a los marcadores microsatélites HA3103, anclado en el GL 7, el cual presenta asociación a QTLs de resistencia a estrés hídrico (Kiani et al., 2007) y el microsatélite ORS1247, anclado en el GL 17, de 339 pb, el cual presenta asociación a QTLs de resistencia a *S. sclerotiorum*. Con la utilización de la técnica de C-PRINS en lugar de la de PRINS se

observaron señales de mayor intensidad, posiblemente debido a que al aumentar el número de ciclos de la PCR hay mayor concentración de moléculas marcadas amplificadas en esa localización. En este trabajo se ha demostrado que el método de C-PRINS indirecto es un método útil para detectar secuencias de copia única en girasol. Estos resultados coinciden con lo reportado por Kubaláková et al. (2001), los cuales encontraron que la técnica de C-PRINS es una variante más sensible que la técnica de PRINS y más útil para la detección de secuencias de bajo número de copias.

Las técnicas de FISH y de PRINS, utilizando como sonda secuencias de RGAs, no permitieron hasta el momento la localización de las mismas. Para poner a punto ambas técnicas se amplificaron y clonaron cuatro fragmentos que presentaban similitud a RGAs en base de datos públicas de girasol, denominados HaNTP3 y HaNTP6, pertenecientes a la familia de los genes R de tipo non-TIR-NBS-LRR, caracterizados a partir de los llamados genes *Pl* que controlan la resistencia a *P. halstedii*, agente causal de mildiu en girasol y mapeados en el GL 13 del mapa de Yu et al. (2003) por Radwan et al. (2003) y los fragmentos denominados HaP2, HaP3, pertenecientes a la familia de los genes R de tipo CC-NBS-LRR, caracterizados y mapeados por Radwan et al. (2004) en el GL 13 del mapa de Yu et al. (2003). Se realizaron varios ensayos de hibridación mediante la técnica de FISH utilizando dichos RGAs, y mediante la técnica de PRINS utilizando el oligonucleótido Ha-P2. Dado que en ningún caso se observó señal de hibridación, se propone la utilización del método de C-PRINS, validado para secuencias únicas en este trabajo, para detectar este tipo de secuencias en el genoma de girasol, así como también la utilización de tiramida para la amplificación de la señal. El método *Tyramide-FISH* amplifica la señal mediante el depósito de tiramida conjugada a un fluoróforo; esta técnica ha sido adaptada a plantas para detectar secuencias en cromosomas mitóticos de *Allium cepa* (Khrustaleva & Kik, 2001).

En este trabajo se comprobó que la técnica de C-PRINS es una estrategia alternativa y/o complementaria a la técnica de BAC-FISH para la localización de secuencias de copia única en el genoma de girasol. La técnica de C-PRINS tiene la ventaja de ser fácil de aplicar, más rápida, más económica y más específica (evita la detección de ADN repetitivo) que la técnica de BAC-FISH. Asimismo, si bien la técnica de BAC-FISH es más utilizada y difundida que la de C-PRINS (Cheng et al., 2001; Islam-Faridi et al., 2002; Kim et al., 2002; Kulikova et al., 2001; Wang et al., 2006) la presencia de

grandes cantidades de elementos repetitivos en determinados clones de BACs imposibilita la localización de secuencias únicas presentes en dichos clones (Kim et al., 2002; Guerra, comunicación personal) y requiere la optimización del bloqueo de esas secuencias.

5. CONCLUSIONES

- Este trabajo implicó la búsqueda y el desarrollo de marcadores moleculares, el establecimiento y la evaluación de distintas técnicas de genotipificación para su adaptación a procesos de mediano desempeño para el análisis masivo de marcadores moleculares, el mapeo genético, la caracterización del complemento cromosómico y la localización física de regiones genómicas de interés agronómico.
- El desarrollo de marcadores moleculares funcionales se condujo a partir de la identificación *in silico* de EST-SSR, SNPs e InDels presentes en colecciones de transcritos o EST, y la evaluación experimental de subconjuntos seleccionados.
- La búsqueda de EST-SSRs se realizó sobre una colección de 36.741 unigenes. El 3% de las secuencias analizadas presentaron uno o más motivos de repetición. El tipo de repetición más abundante fueron los trinucleótidos, representada por el 51,7%, seguida por los dinucleótidos con el 30,3%, luego los tetranucleótidos con el 9,3% y, por último, los pentanucleótidos con el 8,7%. El tipo de motivo más abundante fue el GA/CT, encontrado en el 23,2% de las secuencias, seguido por el motivo AAG/CTT (11,9%). La validación se realizó en cuatro genotipos de girasol, de un total de 127 secuencias seleccionadas el 55% de los *loci* pudieron ser amplificados, el 13% resultaron polimórficos, el 45% monomórficos, el 5% alelos nulos y el 37% restante falló la PCR. En la población de mapeo fueron genotipificados 14 EST-SSR, de los cuales 9 pudieron anclarse al mapa genético de ligamiento. Por lo tanto, la base de datos de EST de girasol puede ser considerada una valiosa fuente de marcadores microsatélites para el desarrollo de mapas genéticos.
- La búsqueda de SNPs e InDels se realizó mediante la técnica de CAPS. Para ello, se seleccionaron 25 secuencias de ESTs a partir de la colección local, las cuales se amplificaron y secuenciaron en los parentales de la población de mapeo, resultando en 23 secuencias de buena calidad. En el análisis *in silico* se

detectaron 21 polimorfismos, los cuales fueron evaluados experimentalmente utilizando las enzimas de restricción correspondientes. Sólo 2 de los 21 polimorfismos detectados en los parentales de la población de mapeo pudieron desarrollarse como marcadores CAPS. De esta manera, el método de CAPS resultó exitoso para la detección virtual de SNP, sin embargo, no fue posible confirmar experimentalmente la mayoría de esos polimorfismos.

- Las diferentes técnicas de genotipificación utilizadas durante el desarrollo de este trabajo resultaron confiables y reproducibles. Sin embargo, la utilización de las técnicas de genotipificación manuales, mediante geles de poliacrilamida y la visualización con nitrato de plata, SYBR[®]Gold o con nucleótidos marcados con radiactivo y detección por autorradiografía, fueron muy demandantes de tiempo. Con la utilización del método de marcado indirecto de los fragmentos utilizando el oligonucleótido universal M13, se observó baja eficiencia en la amplificación y baja reproducibilidad, dificultando su aplicación en el mapeo de microsátélites. Finalmente, se optó por la utilización de la técnica de marcado directo para la genotipificación de los fragmentos, debido a que se observó un alto nivel de procesividad con el desarrollo de la misma.
- Se analizaron 746 marcadores moleculares (594 SSR, 127 EST-SSR y 25 SNPs-InDels) en los parentales de la población de mapeo, de los cuales 241 (224 SSR, 14 EST-SSR y 3 SNPs-InDels) fueron polimórficos y se genotipificaron en las 94 RILs utilizadas en este estudio.
- Doscientos cuarenta y un marcadores de tipo SSR, EST-SSR y SNPs-InDels se incorporaron en una matriz de datos de mapeo conteniendo 343 AFLPs y 38 SSR, de los cuales 197 SSR, 14 EST-SSR y 3 InDels pudieron anclarse al mapa genético de ligamiento.
- Se generó un mapa genético de ligamiento de referencia que incluye 304 marcadores AFLP, 231 SSR, 9 EST-SSR y 3 InDels, en 123 RILs con una longitud total de 1.942,3 cM, una distancia media entre marcadores de 3,6 cM y una región de mayor longitud sin cubrir por marcadores (*gaps*) de 13,2 cM. Ochenta y siete marcadores SSR correlacionaron con los GLs de tres mapas

públicos de girasol (Tang et al., 2002, 2003; Yu et al., 2004), lo que permitió referenciar los grupos de ligamiento unificando los mapas disponibles.

- Este mapa genético de referencia saturado ofrece un marco en donde seguir ubicando nuevos marcadores disponibles, permite orientar la búsqueda de regiones candidatas para la localización de polimorfismos asociados a caracteres agronómicos de interés y conducir estudios de asociación genotipo-fenotipo, los cuales pueden servir de base para conducir estudios de caracterización de QTLs y el desarrollo de nuevas líneas e híbridos mediante selección asistida por marcadores (MAS). Además, el desarrollo de este tipo de mapas contribuye a la investigación y a la actualización de la información de dominio público. Los futuros trabajos del grupo de investigación deben basarse en las posibilidades que brinda el mapa genético generado.
- En este trabajo se optimizaron distintas técnicas de citogenética molecular, tanto para la caracterización del complemento cromosómico de *H. annuus* como para la localización de secuencias repetitivas de copia única.
- La utilización de una secuencia con similitud a un ADN ribosomal 26S, obtenida a partir de la colección local de ESTs de girasol (Fernández et al., 2003), ha permitido observar la presencia de 6 señales de hibridación de distinta intensidad y tamaño.
- La técnica de BAC-FISH permitió la caracterización del complemento cromosómico mediante la utilización de un clon de BAC, conteniendo secuencias repetitivas, como sonda y la localización física en cromosomas metafásicos de dos marcadores microsatélites anclados en los GL asociados a QTLs para resistencia a *S. sclerotiorum* (microsatélites HA2600 y HA4222 localizados en el GL 10 y 16 respectivamente). Estos resultados confirman que la técnica de BAC-FISH es una poderosa técnica para la identificación citogenética de marcadores moleculares de localización única en girasol, revelando ser una herramienta útil y complementaria de los procesos de mejoramiento molecular en girasol.
- Los resultados de este trabajo muestran que la técnica de PRINS permite la localización de secuencias repetitivas permitiendo ahorrar tiempo de trabajo,

pero por el momento no ha permitido la ubicación de secuencias de copia única, tales como SSR o secuencias con similitud a RGAs.

- Mediante el método de C-PRINS se han detectado señales de hibridación en cromosomas metafásicos con la utilización de oligonucleótidos que amplifican a los marcadores microsatélites HA3103, anclado en el GL 7, el cual presenta asociación a QTLs de resistencia a estrés hídrico (Kiani y et al., 2007) y el microsatélite ORS1247, anclado en el GL 17, el cual presenta asociación a QTLs de resistencia a *S. sclerotiorum*. En este trabajo se ha demostrado que el método de C-PRINS indirecto es un método útil para detectar secuencias de copia única en girasol.

Lic. Paola M. Talia

Dra. Ruth Heinz

Dra. Norma Paniago

6. REFERENCIAS BIBLIOGRÁFICAS

- AKEY, J. M., K. ZHANG, M. M. XIONG, P. DORIS & L. JIN, 2001 The effect that genotyping errors have on the robustness of common linkage-disequilibrium measures. *Am. J. Hum. Genet.* **68**: 1447-1456.
- AL-ALLAF, S. & GODWARD, M.B.E. 1979 Karyotype analysis of four varieties of *Helianthus annuus* L. *Cytologia* **44**: 319-323.
- AL-CHAARANI, G. R., L. GENTZBITTEL, X. Q. HUANG & A. SARRAFI, 2004 Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *Theor Appl Genet* **109**: 1353-1360.
- AL-CHAARANI, G. R., A. ROUSTAEI, L. GENTZBITTEL, L. MOKRANI, G. BARRAULT *et al.*, 2002 A QTL analysis of sunflower partial resistance to downy mildew (*Plasmopara halstedii*) and black stem (*Phoma macdonaldii*) by the use of recombinant inbred lines (RILs). *Theor Appl Genet* **104**: 490-496.
- ALFADIL, A. T., S. POORMOHAMMAD KIANI, G. DECHAMP-GUILLAUME, L. GENTZBITTEL & A. SARRAFI, 2007 QTL mapping of partial resistance to *Phoma* basal stem and root necrosis in sunflower (*Helianthus annuus* L.). *Plant Science* **172**: 815-823
- ALIGNAN, M., T. HEWEZI, M. PETITPREZ, G. DECHAMP-GUILLAUME & L. GENTZBITTEL, 2006 A cDNA microarray approach to decipher sunflower (*Helianthus annuus*) responses to the necrotrophic fungus *Phoma macdonaldii*. *New Phytol* **170**: 523-536.
- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS & D. J. LIPMAN, 1990 Basic local alignment search tool *J Mol Bio* **215**: 208-218.
- AMBROS, P. F., M. A. MATZKE & A. J. M. MATZKE, 1986 Detection of a 17-kb unique sequence (T-DNA) in plant chromosomes by in situ hybridization. *EMBO J.* **5**: 2073–2077.
- APPELS, R., & R. L. HONEYCUTT, 1986 Evolution over a billion years, pp. 81–125 in *DNA Systematics II. Plant DNA*. CRC Press, Boca Raton, F.L.
- ARESHECHENKOVA, T., & M. W. GANAL, 1999 Long tomato microsatellites are predominantly associated with centromeric regions. *Genome* **42**: 536-544.
- AUSTIN, D. F., & M. LEE, 1996 Comparative mapping in F2:3 and F6:7 generations of quantitative trait loci for grain yield and yield components in maize. *Theor Appl Genet* **92**: 817–826.
- BAACK, E., K. WHITNEY & L. RIESEBERG, 2005 Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytol* **167**: 623–630.

- BATLEY, J., G. BARKER, H. O'SULLIVAN, K. J. EDWARDS & D. EDWARDS, 2003 Mining for Single Nucleotide Polymorphisms and Insertions/Deletions in Maize Expressed Sequence Tag Data. *Plant Physiology* **132**: 84–91.
- BECKER, J., & M. HEUN, 1995 Barley microsatellites: allele variation and mapping. *Plant Mol. Biol.* **27**: 835-845.
- BENABDELMOUNA, A., G. GUERITAINE, M. ABIRACHED-DARMENCY & H. DARMENCY, 2003 Genome discrimination in progeny of interspecific hybrids between *Brassica napus* and *Raphanus raphanistrum*. *Genome* **46**: 469-472.
- BENJAMIN, Y., & Y. HOCHEBERG, 1995 Controlling the False Discovery Rate: a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistics Society* **57**: 289-300.
- BENJAMIN, Y., & D. YEKUTIELLI, 2005 Quantitative Trait Loci Analysis Using the False Discovery Rate. *Genetics* **171**: 783-790.
- BENNETZEN, J. L., 2000 Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 251-269.
- BENNETZEN, J. L., S. P.S., A. D. CRESSE & M. HENDRICKX, 1993 Specificity and regulation of the Mutator transposable element system in maize. *Crit. Rev. Plant Sci* **12**: 67-95.
- BENNETT, M. D., 1995 The development and use of genomic in situ hybridization (GISH) as a new tool in plant biosystematics. Bransham PE, Bennett MD (eds) *Kew Chromosome Conference IV*, Royal Botanical Gardens, Kew, UK: 167-183.
- BARNARD, G. A., 1947 Statistical inference. *J. R. Statist. Soc. B* **11**: 115-139.
- BERRY, S. T., A. J. LEON, P. CHALLIS, C. LIVIN, R. JONES et al., 1996 Construction of a high density, composite RFLP linkage map for cultivated sunflower *Helianthus annuus*, pp. 1150–1160 in 14th International Sunflower Conference, Beijing, China.
- BERRY, S. T., A. J. LEON, C. C. HANFREY, P. CHALLIS, A. BURKHOLZ *et al.*, 1995 Molecular marker analysis of *Helianthus annuus* L.: 2. Construction of an RFLP linkage map for cultivated sunflower. *Theor. Appl. Genet.* **91**: 195-199.
- BERT, P. F., D. TOURVIEILLE DE LABROUHE, P. NICOLAS, J. PHILIPPON, S. MOUZEYAR *et al.*, 2001 Identification of a second linkage group carrying genes controlling resistance to downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). *Theor Appl Genet* **103**: 992–997.
- BLAKE, S. F., 1918 A revision of the genus *Viguiera*. *Contr. Gray Herb* **54**: 1-205.
- BOHOROVA, N. & GEORGIEVA-TODOROVA, Y. 1987 Hybridization between *Helianthus annuus* L. ($2n = 34$) and *H. hirsutus* Raf. ($2n = 68$). *Genet. Breed.* **20**: 205-210.
- BONIN, A., E. BELLEMAIN, P. BRONKEN EIDSEN, F. POMPANON, C. BROCHMANN et al., 2004 How to track and assess genotyping errors in population genetics studies. *Molecular Ecology* **13**: 3261–3273.
- BOUZIDI, M. F., S. BADAoui, F. CAMBON, F. VEAR, D. TOURVIEILLE DE LABROUHE *et al.*, 2002 Molecular analysis of a major locus for resistance to downy

- mildew in sunflower with specific PCR-based markers. *Theor Appl Genet* **104**: 592-600.
- BOUZIDI, M. F., J. FRANCHÉL, Q. TAO, K. STORMO, A. MRAZ *et al.*, 2006 A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions. *Theor. Appl. Genet.* **113**: 81-89.
- BRINKMANN, B., M. KLINTSCHAR, F. NEUHUBER, J. HUHNE & B. ROLF, 1998 Mutation rate in human microsatellites: influence of the structure and length of the tandem repeat. *Am J Hum Genet* **62**: 1408-1415.
- BROWNSTEIN, M. J., J. D. CARPTEN & J. R. SMITH, 1996 Modulation of non-templated nucleotide addition by Taq DNA polymerase: primer modifications that facilitate genotyping. *BioTechniques* **20**: 1004-1010.
- BUETOW, K. H., 1991 Influence of aberrant observations on highresolution linkage analysis outcomes. *American Journal of Human Genetics* **49**: 985-994.
- BURKE, J. M., Z. LAI, M. SALMASO, T. NAKAZATO, S. TANG *et al.*, 2004 Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. *Genetics* **167**: 449-457.
- BURR, B., & F. BURR, 1991 Recombinant inbred lines for molecular mapping in maize: theoretical and practical considerations. *Trends Genet* **7**: 55-60.
- CAI, W. W., J. RENEKER, C. W. CHOW, M. VAISHNAY & A. BRADLEY, 1998 An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. *Genomics* **54**: 387-397.
- CALLEN, D. F., A. D. THOMPSON, Y. SHEN, H. A. PHILLIPS, R. I. RICHARDS *et al.*, 1993 Incidence and origin of 'null' alleles in the (AC)_n microsatellite markers. *Am. J. Hum. Genet.* **52**: 922-927.
- CARDLEY, L., L. RAMSAY, D. MILBOURNE, M. MACAULAY, D. MRSHALL *et al.*, 2000 Computational and experimental characterization of physically clustered simple sequence repeat in plants. *Genetics* **156**: 847-854.
- CARLSON, A. R., J. LETARTE, J. CHEN & K. J. KASHA, 2001 Visual screening of microspore-derived transgenic barley (*Hordeum vulgare* L.) with green-fluorescent protein. *Plant Cell Rep* **20**: 331-337.
- CECCARELLI, M., V. SARRI, L. NATALI, T. GIORDANI, A. CAVALLINI *et al.*, 2007 Characterization of the chromosome complement of *Helianthus annuus* by in situ hybridization of a tandemly repeated DNA sequence. *Genome* **50**: 429-434.
- COLLAR, B. C. Y., M. Z. Z. JAHUFER, J. B. BROUWER & E. C. PANG, An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for cropimprovement: The Basic concepts. *Euphytica* **142**: 169-196.
- CORDEIRO, G. M., R. CASU, C. L. MCINTYRE, J. M. MANNERS & R. J. HENRY, 2001 Microsatellite markers from sugarcane (*Saccharum* spp.) ESTs cross transferable to *Helianthus* and sorghum. *Plant Sci* **160**: 1115-1123.
- CREGAN, P. B., T. JARVIK, A. BUSH, R. C. SHOEMAKER, K. G. LARK *et al.*, 1999 An integrated genetic linkage map of the soybean genome. *Crop Science* **39**: 1464-1490.

- CUADRADO, A., & N. JOUVE, 1994 Mapping and organization of highly repeated DNA sequences by means of simultaneous and sequential FISH and C-banding in 6X-triticales. *Chromosome Res* **2**: 331-338.
- CUADRADO, A., & T. SCHWARZACHER, 1998 The chromosomal organization of simple sequence repeats in wheat and rye chromosomes. *Chromosoma* **107**: 587-594.
- CUÉLLAR, T., E. BELHASEN, B. FERNANDEZ-CALVIN, J. ORELLANA & J. L. BELLA, 1996 Chromosomal differentiation in *Helianthus annuus* var. *macrocarpus*: heterocromatin characterization and rDNA location. *Heredity* **76**: 586-591.
- CUÉLLAR, T., J. ORELLANA, E. BELHASSEN & J. L. BELLA, 1999 Chromosomal characterization and physical mapping of the 5S and the 18S-5.8S-25S ribosomal DNA in *Helianthus argophyllus*, with new data from *Helianthus annuus*. *Genome* **42**: 110-115.
- CUBERO, J. L., 2003 *Introducción a la mejora genética vegetal*. Ediciones Mundi-Prensa España.
- CHAGNÉ, D., J. BATLEY, D. EDWARDS & J. W. FORSTER, 2007 Single Nucleotide Polymorphism Genotyping in Plants, pp. 77-94 in *Association Mapping in Plants*.
- CHAKRABORTY, R., M. KIMMEL, D. N. STRIVERS, L. J. DAVISON & R. DEKA, 1997 Relative mutation rates at di- tri- and tetranucleotide microsatellite loci. *Proc Natl Acad Sci USA* **94**: 1041-1046.
- CHANG, Y. L., Q. TAO, C. SCHEURING, K. DING, K. MEKSEM *et al.*, 2001 An integrated map of *Arabidopsis thaliana* for functional análisis of its genome sequence. *Genetics* **159**: 1231-1242.
- CHEN, M. S., G. PRESTING, W. B. BARBAZUK, J. L. GOICOCHEA, B. BLACKMON *et al.*, 2002 An integrated physical and genetic map of the rice genome. *The Plant Cell* **14**: 537-545.
- CHEN, X., & P. F. SULLIVAN, 2003 Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *Pharmacogenomics J* **3**: 77-96.
- CHENG, Z., PRESTING, G., ROBIN BUELL, C., WING, R.A. & JIANG, J. 2001 High-Resolution Pachytene Chromosome Mapping of Bacterial Artificial Chromosomes Anchored by Genetic Markers Reveals the Centromere Location and the Distribution of Genetic Recombination Along Chromosome 10 of Rice. *Genetics* **157**: 1749-1757
- CHING, A., K. S. CALDWELL, M. JUNG, M. DOLAN, O. S. SMITH *et al.*, 2002 SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet* **3**: 1-14.
- CHO, Y. G., ISHII, T., TEMNYKH, S., CHEN, X., LIPOVICH, L., MCCOUCH, S. R., PARK, W. D., AYRES, N. & CARTINHO, S. 2000 Diversity of microsatellites derived from genomic libraries and GenBank sequences in rice (*Oryza sativa* L.). *Theor. Appl. Genet.* **100**: 713-722.
- DANIEL, W. W., 1978 *Applied Nonparametric statistics*. Houghton Mifflin complay Boston USA. pp 53-58.

- DARVISHZADEH, R., S. POORMOHAMMAD KIANI, G. DECHAMP-GUILLAUME, L. GENTZBITTEL & A. SARRAFI, 2007 Quantitative trait loci associated with isolate specific and isolate nonspecific partial resistance to *Phoma macdonaldii* in sunflower. *Plant Pathology* **56**: 855–861
- DEHEMER, K. J., & W. FRIEDT, 1998 Evaluation of different microsatellite motifs for analyzing genetic relationships in cultivated sunflower (*Helianthus annuus* L.). *Plant Breed* **1**: 39-42.
- DESEL, C., C. JUNG, D. G. CAI, M. KLEINE & T. SCHMIDT, 2001 High-resolution mapping of YACs and the single-copy gene HS1 (pro)-1 on Beta vulgaris chromosome by multi-colour fluorescence in situ hybridization. *Plant Mol. Biol.* **45**: 113–122.
- DIAZ QUIJANO, C. M., 2007 Caracterización estructural de secuencias de ADN repetitivo de girasol por técnicas de hibridación “in situ”, pp. 1-60 in *Departamento de Ecología, Genética y Evolución*. Facultad de Ciencias Exactas y Naturales Universidad de Buenos Aires, Buenos Aires.
- DIERINGER, D., & C. SCHLÖTTERER, 2003 Two distinct modes of microsatellite mutation processes: evidence from the complete genomic sequences of nine species. *Genome Res* **13**: 2242–2251.
- DONG, F., J. SONG, S. K. NAESS, J. P. HELGESON, C. GEBHARDT *et al.*, 2000 Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. *Theor. Appl. Genet.* **101**: 1001–1007.
- DOUGLAS, J. A., M. BOEHNKE & K. LANGE, 2000 A multipoint method for detecting genotyping errors and mutations in sibling-pair linkage data. *Am. J. Hum. Genet.* **66**: 1287-1297.
- DUBLE, C. M., V. HAHN, S. J. KNAPP & E. BAUER, 2004 PIArg from *Helianthus argophyllus* is unlinked to other known downy mildew resistance genes in sunflower. *Theor Appl Genet* **109**: 1083–1086.
- EBRAHIMI, A., P. MAURY, BERGER., S. POORMOHAMMAD KIANI, A. NABIPOUR *et al.*, 2008 QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes. *Genome* **8**: 599-615.
- EDWARDS, D., J. W. FORSTER, D. CHAGNÉ & J. BATLEY, 2007 WHAT ARE SNPs?, in *Association Mapping in Plants*, edited by N. ORAGUZIE, E. RIKKERINK, S. GARDINER and H. DE SILVA, Auckland, New Zealand pp. 41-52
- ELLEGREN, H., 2000 Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet* **16**: 551-558.
- EUJAYL, I., M. K. SLEDGE, L. WANG, G. D. MAY, K. CHEKHOVSKIY *et al.*, 2004 *Medicago truncatula* EST-SSRs reveal cross-species genetic markers for *Medicago* spp. *Theor Appl Genet* **108**: 414-422.
- EUJAYL, I., M. SORRELLS, M. BAUM, P. WOLTERS & W. POWELL, 2002 Isolation of EST-derived microsatellite markers for genotyping the A and B genomes of wheat. *Theor. Appl. Genet.* **104**: 399–407.

- EUJAYL, I., M. ZORREÍIS, M. BAUM, P. WOLTERS & W. POWELL, 2001 Assessment of genotypic variation among cultivated durum wheat based on EST-SSRS and genomic SSRS. *Euphytica* **119**: 39–43.
- FEN, J., & G. G. JAN, 2008 Introgression and molecular tagging of Rf (4), a new male fertility restoration gene from wild sunflower *Helianthus maximiliani* L. *Theor Appl Genet* **117**: 241-249.
- FENG, J., B. A. VICK, M. K. LEE, H. B. ZHANG & C. C. JAN, 2006 Construction of BAC and BIBAC libraries from sunflower and identification of linkage group-specific clones by overgo hybridization. *Theor Appl Genet* **113**: 23-32.
- FERNÁNDEZ, P., J. DI RIENZO, L. FERNANDEZ, H. E. HOPP, N. PANIEGO *et al.*, 2008 Transcriptomic identification of candidate genes involved in sunflower responses to chilling and salt stresses based on cDNA microarray analysis. *BMC Plant Biol* **8**: 11.
- FERNÁNDEZ, P., N. PANIEGO, S. LEW, H. E. HOPP & R. A. HEINZ, 2003 Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. *BMC Genomics* **4**: 40.
- FICK, G. N., & D. E. ZIMMER, 1976 Yield stability of sunflower hybrids and open pollinated varieties. Proc. 7th Int.Sunflower Conf. Krasnadar, Russia Int. Sunflower Assoc., Paris, France: 253-258.
- FICK, G. N., D. E. ZIMMER & M. L. KINMAN, 1974 Registration of six sunflower parental lines. *Crop Sci* **14**: 912.
- FLORES BERRIOS, E., L. GENTZBITTEL, G. ALIBERT, Y. GRIVEAU, A. BERVILLÉ *et al.*, 1999a Genetic control of in vitro organogenesis in recombinant inbred lines of sunflower (*Helianthus annuus* L.). *Plant Breed.* **118**: 359-361.
- FLORES BERRIOS, E., L. GENTZBITTEL, H. KAYYAL, G. ALIBERT & A. SARRAFI, 2000 AFLP mapping of QTLs for in vitro organogenesis traits using recombinant inbred lines in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* **101**: 1299-1306.
- FLORES BERRIOS, E., L. GENTZBITTEL, H. SERIEYS, G. ALIBERT & A. SARRAFI, 1999b Influence of genotype and gelling agents on in vivo regeneration by organogenesis in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* **101**: 1299-1306.
- FRANSZ, P. F., S. ARMSTRONG, J. H. JONG, L. D. PARNELL & C. VAN DRUNEN, 2000 Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: structural organization of heterochromatic knob and centromere region. *Cell* **100**: 367–376.
- FRASER, L. G., C. F. HARVEY, R. N. CROWHURST & H. N. DE SILVA, 2004 EST-derived microsatellites from *Actinidia* species and their potential for mapping. *Theor Appl Genet* **108**: 1010–1016.
- FUSARI, C. M., V. V. LIA, H. E. HOPP, R. A. HEINZ & N. B. PANIEGO, 2008 Identification of Single Nucleotide Polymorphisms and analysis of Linkage Disequilibrium in sunflower elite inbred lines using the candidate gene approach. *BMC Plant Biology* **8**: 7.
- GAO, L., J. TANG, H. LI & J. JIA, 2003 Analysis of microsatellites in major crops assessed by computational and experimental approaches. *Molecular Breeding* **12**: 245-261.

- GARDINER, J., S. SCHROEDER, M. L. POLACCO, H. SANCHEZ-VILLEDA, Z. FANG *et al.*, 2004 Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. *Plant Physiol* **134**: 1317-1326.
- GEDIL, M. A., 1999 Marker development, genome mapping, and cloning of candidate disease resistance genes in sunflower, *Helianthus annuus L.* Ph.D. Thesis. Oregon state Univ., Corvallis, OR.
- GEDIL, M. A., M. B. SLABAUGH, S. BERRY, R. JOHNSON, R. MICHELMORE *et al.*, 2001a Candidate disease resistance genes in sunflower cloned using conserved nucleotide-binding site motifs: genetic mapping and linkage to the downy mildew resistance gene P11. *Genome* **44**: 205-212.
- GEDIL, M. A., C. WYE, S. BERRY, B. SEGERS, J. PELEMAN *et al.*, 2001b An integrated restriction fragment length polymorphism--amplified fragment length polymorphism linkage map for cultivated sunflower. *Genome* **44**: 213-221.
- GENTZBITTEL, L., A. ABBOTT, J. P. GALAUD, L. GEORGI, F. FABRE *et al.*, 2002 A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. *Mol Genet Genomics* **266**: 979-987.
- GENTZBITTEL, L., E. MESTRIES, S. MOUZEYAR, F. MAZEYRAT, S. BADAOUI *et al.*, 1999 A composite map of expressed sequences and phenotypic traits of the sunflower *Helianthus annuus L.* genome. *Theor. Appl. Genet.* **99**: 218-234.
- GENTZBITTEL, L., S. MOUZEYAR, S. BADAOUI, F. MESTRIES, F. VEAR *et al.*, 1998 Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus L.* *Theor. Appl. Genet.* **96**: 519-525.
- GENTZBITTEL, L., F. VEAR, Y. X. ZHANG, A. BERVILLÉ & P. NICOLAS, 1995 Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus L.*). *Theor. Appl. Genet.* **90**: 1079-1086.
- GERLACH & BEDBROOK, 1979 Cloning and characterization of ribosomal RNA genes from wheat and barley. *Nucl. Acids Res* **7**: 1869-1885.
- GINOT, F., I. BORDELAIS, S. NGUYEN & G. GYAPAY, 1996 Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research* **24**: 540-541.
- GOLDSTEIN, D. B., & A. G. CLARK, 1995 Microsatellite variation in North American populations of *Drosophila melanogaster*. *Nucleic Acids Res* **23**: 3882-3886.
- GOLDSTEIN, D., & C. SCHLÖTTERER, 1999 *Microsatellites: Evolution and Applications*. Oxford Univ. Press, Oxford.
- GOLDSTEIN, D. R., H. Y. ZHAO & T. P. SPEED, 1997 The effects of genotyping errors and interference on estimation of genetic distance *Hum. Hered* **47**: 86-100.
- GOMEZ, M. I., M. N. ISLAM-FARIDI, S. S. WOO, K. F. SCHERTZ, D. CZESCHIN *et al.*, 1997 FISH of a maize sh2-selected sorghum BAC to chromosomes of *Sorghum bicolor*. *Genome* **40**: 475-478.

- GRANDBASTEIN, A. M., 1998 Activation of *plat* retrotransposons under stress conditions. *Trends Plant Sci* **3**: 181-187.
- GRIVET, L., J. C. GLASZMANN, M. VINCENTZ, F. DA SILVA & P. ARRUDA, 2003 ESTs as a source for sequence polymorphism discovery in sugarcane: example of *Adh* genes. *Theor. Appl. Genet.* **106**: 190–197.
- GUPTA, P. K., J. K. ROY & M. PRASAD, 2001 Single nucleotide polymorphisms: A new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. *CURRENT SCIENCE* **80**: 524-535.
- GUPTA, P. K., S. RUSTGI, S. SHARMA, R. SINGH, K. N. *et al.*, 2003 Transferable EST-SSR markers for the study of polymorphism and genetic diversity in bread wheat. *Mol Genet Genomics* **270**: 315–323.
- GUPTA, P. K., R. K. VARSHNEY & M. PRASAD, 2002 *Molecular Markers: Principles and methodology*. *Techniques in Crop Improvement* 9-54.
- HACKETT, C. A., & L. B. BROADFOOT, 2003 Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90**: 33-38.
- HACKAUF, B., & P. WEHLING, 2002 Identification of microsatellite polymorphisms in an expressed portion of the rye genome. *Plant Breed* **121**: 17-25.
- HALDANE, J. B. S., & C. H. WADDINGTON, 1931 Inbreeding and linkage. *Genetics* **16**: 357-374.
- HALL, T. A., 1999 BioEdit: a user-friendly biological sequence alignment editor and análisis program for Windows. *Nucl. Acids Symp* **41**: 95-98.
- HANSON, R. E., M. S. ZWICK, S. CHOI, M. N. ISLAM-FARIDI, T. D. MCKNIGHT *et al.*, 1995 Fluorescent in situ hybridization of a bacterial artificial chromosome. *Genome* **38**: 646–651.
- HARKER, N., 2001 Collection, reporting and storage of microsatellite genotype data, pp. 251–264 in *Plant Genotyping: the DNA Fingerprinting of Plants*, edited by R. J. HENRY, UK & New York, USA.
- HEESACKER, A., V. KISHORE, W. GAO, S. TANG, J. KOLKMAN *et al.*, 2008 SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *Theor Appl Genet.*
- HEISER, C. B., 1976 The sunflowers, pp. 198 in *Evolution of Crop Plants*. University of Oklahoma Press. Ed. N.W. Simmonds. pp 36-38
- HEISER, C. B., & D. M. SMITH, 1955 New chromosome number in *Helianthus* and related. *Proc. Nat. Acad. Sci. USA* **64**: 250-253.
- HERVE, D., F. FABRE, E. FLORES BERRIOS, N. LEROUX, G. R. AL-CHAARANI *et al.*, 2001 QTL analysis of photosynthesis and water status traits in sunflower (*Helianthus annuus* L.) under greenhouse conditions. *J Exp. Bot*
- HEWEZI, T., M. PETITPREZ & L. GENTZBITTEL, 2006 Primary metabolic pathways and signal transduction in sunflower (*Helianthus annuus* L.): comparison of transcriptional

- profiling in leaves and immature embryos using cDNA microarrays. *Planta* **223**: 948-964.
- HOOPEN, R. T., T. P. ROBBINS, P. F. FRANZ, B. M. MONTIJN, O. OUD et al., 1996 Localization of T-DNA insertions in *Petunia* by fluorescence in situ hybridization: physical evidence for suppression of recombination. *Plant Cell* **8**: 823-830.
- HU, J., 2006 Defining the sunflower (*Helianthus annuus* L.) linkage group ends with the Arabidopsis-type telomere sequence repeat-derived markers. *Chromosome Res* **14**: 535-548.
- INNAN, H., R. TERAUCHI & T. MIYASHITA, 1997 Microsatellite polymorphism in natural populations of the wild plant *Arabidopsis thaliana*. *Genetics* **146**: 1441-1452.
- ISLAM-FARIDI, M. N., K. L. CHILDS, P. E. KLEIN, G. HODNETT & M. A. MENZ, 2002 A molecular cytogenetic map of sorghum chromosome 1: fluorescence in situ hybridization analysis with mapped bacterial artificial chromosomes. *Genetics* **161**: 345-353.
- JACKSON, S. A., Z. CHENG, M. L. WANG, H. M. GOODMAN & J. JIANG, 2000 Comparative fluorescence in situ hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. *Genetics* **156**: 833-838.
- JACKSON, S., M. L. WANG, H. M. GOODMAN & J. JIANG, 1998 Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. *Genome* **41**: 566-577.
- JAN, C. C., B. A. VICK, J. F. MILLER, A. L. KAHLER & E. T. BUTLER, 1998 Construction of an RFLP linkage map for cultivated sunflower. *Theor Appl Genet* **96**: 15-22.
- JIANG, J., & B. S. GILL, 1994 Nonisotopic in situ hybridization and plant genome mapping: the first 10 years. *Genome* **37**: 717-725.
- JIANG, J., B. S. GILL, G. L. WANG, P. C. RONALD & D. C. WARD, 1995 Metaphase and interphase fluorescence in situ hybridization mapping of rice genome with bacterial artificial chromosomes. *Proc Natl Acad Sci USA* **92**: 4487-4491.
- JONES, N., H. OUGHAM & H. THOMAS 1997 Markers and mapping: we are all geneticists now. *New Phytol.* **137**:165-177.
- JURKA, J., & C. PETHIYGODA, 1995. Simple repetitive DNA sequences from primates: compilation and analysis. *J. Mol. Evol.* **40**: 120-126.
- KANTETY, R. V., M. LA ROTA, D. E. MATTHEWS & M. E. SORRELLS, 2002 Data mining for simple sequence repeats in expressed sequence tags from barley, maize, rice, sorghum and wheat. *Plant Mol Biol* **48**: 501-510.
- KATTI, M.V., RANJEKAR, P.K., & V.S.GUPTA, 2001. Differential distribution of simple sequence repeats in eukaryotic genome sequences. *Mol. Biol. Evol.* **18**: 1161-1167.
- KESSELI, R. V., & R. W. MICHELMORE, 1997 The compositae: systematically fascinating but specifically neglected. *Genome mapping in Plants*: 179-191

- KHRUSTALEVA, L., & C. KIK, 2001 Localization of single-copy TDNA insertion in transgenic shallots (*Allium cepa*) by using ultra-sensitive FISH with tyramide signal amplification. *Plant J* **25**: 699–707.
- KIANI, P. S., P. GRIEU, P. MAURY, T. HEWEZI, L. GENTZBITTEL *et al.*, 2007b Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* **114**: 193–207.
- KIANI, P., P. MAURY, A. SARRAFI & P. GRIEU, 2008 QTL analysis of chlorophyll fluorescence parameters in sunflower (*Helianthus annuus* L.) under well-watered and water-stressed conditions. *Plant Science* **175**: 565-573.
- KIANI, S. P., P. TALIA, P. MAURY, P. GRIEU, R. HEINZ *et al.*, 2007a Genic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under water treatments. *Plant Science* **172**: 773-787.
- KIM, J. S., K. L. CHILDS, M. N. ISLAM-FARIDI, M. A. MENZ & R. R. KLEIN, 2002 Integrated karyotyping of sorghum by in situ hybridization of landed BACs. *Genome* **45**: 402–412.
- KINMAN, M. L., 1970 New developments in the USDA and state experimental station breeding programmes. *Proc. 4th Int. Sunflower Conf. Int. Sunflower Int.sunflower Assoc., Toowoomba, Qld, Australia*: 181-183.
- KIRK, K. M., & L. R. CARDON, 2002 The impact of genotyping error on haplotype reconstruction and frequency estimation. *Eur. J. Hum. Genet.* **10**: 616-622.
- KOEBNER, 2004 Marker assisted selection in the cereals: the dream and the reality. *Cereal genomics* **10**: 317-331.
- KOLKMAN, J. M., S. T. BERRY, A. J. LEON, M. B. SLABAUGH, S. TANG *et al.*, 2007 Single Nucleotide Polymorphisms and Linkage Disequilibrium in Sunflower. *Genetics*.
- KOLKMAN, J., M. B. SLABAUGH, J. BRUNIARD, S. BERRY, S. B. BUSHMAN *et al.*, 2004 Acetohydroxyacid synthase mutations conferring resistance to imidazolinone or sulfonylurea herbicides in sunflower. *Theor Appl Genet* **109**: 1147-1155.
- KOSAMBI, D. D., 1944 The estimation of a map distance from recombination values. *Ann. Eugen.* **12**: 172–175.
- KOTA, R., R. VARSHNEY, T. THIEL & K. DEHMER, 2001 Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas* **135**: 145–151.
- KUANG, H., W. S.S., B. C. MEYERS, E. NEVO & R. W. MICHELMORE, 2004 Multiple genetic processes result in heterogeneous rates of evolution within the major cluster disease resistance genes in lettuce. *Plant Cell* **16**: 2870–2894.
- KUBALÁKOVÁ, M., M. A. LYSÁK, J. VRÁNA, H. SIMKOVÁ, J. CÍHALÍKOVÁ *et al.*, 2000 Rapid identification and determination of purity of flow-sorted plant chromosomes using C-PRINS. *Cytometry* **41**: 102–108.
- KUBALÁKOVÁ, M., J. MACAS & J. DOLEZEL, 1997 Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. *Theor Appl Genet* **94**: 758–763.

- KUBALÁKOVÁ, M., J. VRÁNA, J. CÍHALÍKOVÁ, M. A. LYSÁK & J. DOLE, 2001 Localisation of DNA sequences on plant chromosomes using PRINS and C-PRINS. *Methods in Cell Science* **23**: 71-82.
- KULIKOVA, O., G. GUALTIERI, R. GEURTS, D.-J. KIM, D. COOK et al., 2001 Integration of the FISH pachytene and genetic maps of *Medicago truncatula*. *Plant J.* **27**: 49–58.
- KULSHRESHTA, V.B. & GUPTA, P.R. 1981 Cytogenetic studies in the genus *Helinathus* I. I. Karyological studies in twelvespecies. *Cytologia (Tokyo)* **46**: 279-289.
- KUMPATLA, S., & S. MUKHOPADHYAY, 2005 Mining and survey of simple sequence repeats in expressed sequence tags of dicotyledonous species. *Genome* **48**: 985-998.
- KÜNZEL, G., L. KORZUN & A. MEISTER, 2000 Cytologically intergrated physical restriction fragment length polymorphism maps for the barley genome based on translocation breakpoints *Genetics* **154**: 397-412.
- KWOK, P. Y., 2001 Methods for genotyping single nucleotide polymorphisms. *Annu Rev Genomics Hum Genet* **2**: 235-258.
- LA ROTA, M., R. V. KANTETY, J. YU & M. E. ZORREÉIS, 2005 Nonrandom distribution and frequencies of genomics and EST-derived microsatellite markers in rice, wheat, and barley. *BMC Genomics* **6**: 23
- LAI, Z., T. NAKAZATO, M. SALMASO, J. M. BURKE, S. TANG et al., 2005 Extensive chromosomal repatterning and the evolution of sterility barriers in hybrid sunflower species. *Genetics* **171**: 291-303.
- LANDER, E. S., P. GREEN, J. ABRAHAMSON, A. BARLOW, M. J. DALY et al., 1987 MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics* **1**: 174-181.
- LAPITAN, N. L. V., S. E. BROWN, W. KENNARD, J. L. STEPHENS & D. L. KNUDSON, 1997 FISH physical mapping with barley BAC clones. *Plant J.* **11**: 149–156.
- LAWSON, W. R., K. C. GOULTER, R. J. HENRY, G. A. KONG & J. K. KOCHMAN, 1998 Marker-assisted selection for two rust resistance genes in sunflower. *Mol Breed* **4**: 227-234.
- LECLERQ, P., 1969 Une sterilité male cytoplasmique chez le tournesol. *Ann. Amerlior Plantes* **19**: 99-106.
- LEE, C. E., & T. MITCHELL-OLDS, 2006 Preface to the special issue: ecological and evolutionary genomics of populations in nature. *Mol Ecol* **15**: 1193-1196.
- LEE, J. M., S. H. NAHM, Y. M. KIM & B. D. KIM, 2004 Characterization and molecular genetic mapping of microsatellite loci in pepper. *Theor Appl Genet* **108**: 619-627.
- LEHMAN, T., W. A. HAWLEY, L. KAMAU, D. FONTENILLE & F. H. SIMARD, 1996 Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellite and allozyme loci. *Heredity* **77**: 192-208.
- LEIGH, F., V. LEA, LAW J, P. WOLTERS, POWELL W *et al.*, 2003 Assessment of EST- and genomic microsatellite markers for variety discrimination and genetic diversity studies in wheat. *Euphytica* **133**: 359–366.

- LEÓN, A. J., F. H. ANDRADE & M. LEE, 2003 Genetic analysis of seedoil concentration across generations and environments in sunflower (*Helianthus annuus* L.). *Crop Sci* **43**: 135-140.
- LEÓN, A. J., F. H. ANDRADE & M. LEE, 2000 Genetic mapping of factors affecting quantitative variation for flowering in sunflower (*Helianthus annuus* L.). *Crop Sci* **40**: 404-407.
- LEÓN, A. J., M. LEE & F. H. ANDRADE, 2001 Quantitative trait loci for growing degree days to flowering and photoperiod response in sunflower (*Helianthus annuus* L.). *Theor. Appl. Genet.* **102**: 497-503.
- LEÓN, A. J., M. LEE, G. K. RUFENER, S. T. BERRY & R. P. MOWERS, 1996 Genetic mapping of a locus (*hyp*) affecting seed hypodermis color in sunflower. *Crop Sci* **36**: 1666-1668.
- LEON, A. J., M. LEE, G. K. RUFENER, S. T. BERRY & R. P. MOWERS, 1995 Use of RFLP markers for genetics linkage analysis of oil percentage in sunflower seed. *Crop Sci* **35**: 558-564.
- LEWIN, B., 1994 *Genes*. Oxford University Press, New York.
- LINCOLN, S. E., & E. S. LANDER, 1992 Systematic detection of errors in genetic linkage data. *Genomics* **14**: 604-610.
- LIU, A., & J. M. BURKE, 2006 Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics* **173**: 321-330.
- LU, Y. H., G. GAGNE, B. GREZES-BESSET & P. BLANCHARD, 1999 Integration of a molecular linkage group containing broomrape resistance gene *Or5* into an RFLP in sunflower. *Genome* **42**: 453-456.
- MAGNUSON, V. L. et al., 1996 Substrate nucleotide-determined non-templates addition of adenine by Taq DNA polymerase: implications for PCR-based genotyping and cloning. *BioTechniques* **21**: 700-709
- MARINGOLO, C. A., 2007 Regiones cromosómicas asociadas a resistencia a podredumbre húmeda del capítulo de girasol (*Sclerotinia sclerotiorum* (lib.) de bary). Tesis de maestría. Facultad de Ciencias Agrarias. Universidad Nacional de Mar del Plata.
- MCDOWELL, J. M., M. DHANDAYDHAM, T. A. LONG, M. G. AARTS, S. GOFF *et al.*, 1998 Intragenic recombination and diversifying selection contribute to the evolution of downy mildew resistance at the RPP8 locus of *Arabidopsis*. *Plant Cell* **10**: 1861-1874.
- MCMURRAY, C. T., 1995 Mechanisms of DNA expansion. *Chromosoma* **104**: 2-13.
- MEKSEM, K., H. ISHIHARA & T. JESSE, 2005 Integration of Physical and Genetic Maps, pp. 215-231 in *The Handbook of Plant Genome Mapping*, edited by K. MEKSEM and G. KAHL.
- MENKE, M., J. FUCHS & I. SCHUBERT, 1998 A comparison of sequence resolution on plant chromosomes: PRINS versus FISH. *Theor Appl Genet* **97**: 1314-1320.
- MESTRIES, E., L. GENTZBITTEL, D. TOURVIELLE DE LABROUHE, P. NICOLAS & F. VEAR, 1998 Analyses of quantitative trait loci associated with resistance to *Sclerotinia*

- sclerotiorum in sunflower (*Helianthus annuus* L.) using molecular markers. *Mol Breed* **4**: 215-226.
- MEYERS, B. C., K. A. SHEN, P. ROHANI, B. S. GAUT & R. W. MICHELMORE, 1998 Receptor-like genes in the major resistance locus of lettuce are subject to divergent selection. *Plant Cell*: 1833–1846.
- MICIC, Z., V. HAHN, E. BAUER, A. E. MELCHINGER, S. J. KNAPP et al., 2005a Identification and validation of QTL for *Sclerotinia* midstalk rot resistance in sunflower by selective genotyping. *Theor Appl Genet* **111**: 233-242.
- MICIC, Z., V. HAHN, E. BAUER, C. C. SCHÖN & A. E. MELCHINGER, 2005b QTL mapping of resistance to *Sclerotinia* midstalk rot in RIL of sunflower population. *Theor Appl Genet* **110**: 1490-1498.
- MICIC, Z., V. HAHN, E. BAUER, C. C. SCHON, S. J. KNAPP et al., 2004 QTL mapping of *Sclerotinia* midstalk-rot resistance in sunflower. *Theor Appl Genet* **109**: 1474-1484.
- MILLER, J. F., & T. J. GULYA, 1991 Inheritance of resistance to race 4 of downy mildew derived from interspecific crosses in sunflower. *Crop Sci* **31**: 40-43.
- MISSIAGGIA, A., & D. GRATTAPAGLIA, 2006 Plant microsatellite genotyping with 4-color fluorescent detection using multiple-tailed primers. *Genetics and Molecular Research* **1**: 72-78.
- MOKRANI, L., L. GENTZBITTEL, F. AZANZA, L. FITAMANT, G. R. AL-CHAARANI et al., 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.
- MORGANTE, M., M. HANAFEY & W. POWELL, 2002 Microsatellites are preferentially associated with nonrepetitive DNA in plant genomes. *Nature Genetics* **30**: 194-200.
- MOSCONE, E. A., M. A. MATZKE & A. J. M. MATZKE, 1996 The use of combined FISH/GISH in conjunction with DAPI counterstaining to identify chromosomes containing transgene inserts in amphidiploid tobacco. *Chromosoma* **105**: 231-236.
- MOURAS, A., & NEGRUTIU, I., 1989 Localization of the T-DNA on marker chromosomes in transfomed tobacco cells by in situ hybridization. *Theor Appl Genet* **78**: 715–720.
- MOURAS, A., M. W. SAUL, S. ESSAD & I. POTRYKUS, 1987 Localization by in situ hybridization of a low copy chimaeric resistance gene introduced into plants by direct gene transfer. *Mol Gen Genet* **207**: 204-209.
- MUTEGNA, C. S., P. B. TONGOONA & ITHOLE-NIANG., 2005 Genetic studies and a search for molecular markers that are linked to *Striga asiatica* resistance in sorghum. *African Journal of Biotechnology* **4**: 1355-1361.
- NASU, S., J. SUZUKI, R. OHTA, K. HASEGAWA, R. YUI et al., 2002 Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Research* **9**: 163–171.
- NATALI, L., S. SANTINI, T. GIORDANI, S. MINELLI, P. MAESTRINI et al., 2006 Distribution of Ty3-gypsy- and Ty1-copia-like DNA sequences in the genus *Helianthus* and other Asteraceae. *Genome* **49**: 64-72.

- NICOT, N., V. CHIQUET, B. GANDON, L. AMILHAT, F. LEGEAI et al., 2004 Study of simple sequence repeat (SSR) markers from wheat expressed sequence tags (ESTs). *Theor Appl Genet* **109**: 800-805.
- NUEZ, F., & J. M. CARRILLO, 2002 Panorama general de la genómica en relación con la mejora vegetal., pp. 9-35 in *Genómica y mejora vegetal*, edited by F. NUEZ, J. M. CARRILLO and R. LOZANO. Junta de Andalucía. Conserjería de Agricultura y Pesca, Sevilla, España.
- ÖZDEMİR, N., R. HORN & W. FRIEDT, 2004 Construction and characterization of a BAC library for sunflower (*Helianthus annuus* L.). *Euphytica* **138**: 177-183.
- PAETKAU, D., & C. STROBECK, 1995 The molecular basis and evolutionary history of a microsatellite null allele in bears. *Mol. Ecol.* **4**: 519-520
- PAGLIA, P., A. M. OLIVER & M. MORGANTE, 1998 Towards second-generation STS (sequence-tagged site) linkage maps in conifers: a genetic map of Norway spruce. *Mol Gen Genet* **258**: 466-478.
- PANIEGO, N., M. ECHAIDE, M. MUNOZ, L. FERNANDEZ, S. TORALES et al., 2002 Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). *Genome* **45**: 34-43.
- PANIEGO, N., R. HEINZ, P. FERNANDEZ, P. TALIA, V. NISHINAKAMASU *et al.*, 2007 Sunflower, pp. 155-179 in *Genome Mapping and Molecular Breeding in Plants*, edited by C. KOLE. Springer-Verlag, Berlin Heidelberg.
- PARIDA, S. K., K. ANAND RAJ KUMAR, V. DALAL, N. K. SINGH & T. MOHAPATRA, 2006 Unigene derived microsatellite markers for the cereal genomes. *TAG Theoretical and applied genetics* **112**: 808-817.
- PARNISKE, M., K. E. HAMMOND-KOSACK, C. GOLSTEIN, C. M. THOMAS, D. A. JONES *et al.*, 1997 Novel disease resistance specificities result from sequence exchange between tandemly repeated genes at the Cf-4/9 locus of tomato. *Cell* **91**: 821-832.
- PASHLEY, C. H., J. R. ELLIS, D. E. MCCAULEY & J. M. BURKE, 2006 EST databases as a source for molecular markers: lessons from *Helianthus*. *J Hered* **97**: 381-388.
- PEDERSEN, C., & I. LINDE-LAURSEN, 1994 Chromosomal locations of four minor rDNA loci and a marker microsatellite sequence in barley. *Chrom Res* **2**: 65-71.
- PEDERSEN, C., J. ZIMNY, D. BECKER, JAHNE-GARTNER & H. LORZ, 1997 Localization of introduced genes on the chromosomes of transgenic barley, wheat and triticale by fluorescence in situ hybridization. *Theor Appl Genet* **94**: 749-757.
- PEERBOLTE, R. P., & J. PELEMAN, 1996 The Cartisol sunflower RFLP map (146 loci) extended with 291 AFLP markers. In *Proc. 18th Sunflower Res. Workshop*, Fargo, ND. 11-12 Jan. 1996.
- PEREZ-VICH, B., B. AKHTOUCH, S. J. KNAPP, A. J. LEON, L. VELASCO et al., 2004 Quantitative trait loci for broomrape (*Orobanche cumana* Wallr.) resistance in sunflower. *Theor. Appl. Genet.* **109**: 92-102.

- PHILLIPS, R. L., & I. K. VASIL, 2001 DNA-based markers in plants, advances in cellular and molecular biology of plants. Kluwer, Dordrecht 6.
- PLOCIK, A., J. LAYDEN & R. KESSELI, 2004 Comparative análisis of NBS domain sequences of NBS-LRRR disease resistance genes from sunflower, lettuce, and chicory. *Mol Phylogen evol* **31**: 153-163.
- POGGIO, L., V. CONFALONIERI, C. COMAS, GONZALEZ & C. NARANJO, 1999 Genomics affinities of *Zea luxurians*, *Z. diploperennis* and *Z. perennies*: meiotic behavior of their hybrids and genomic "in situ" hybridization (GISH). *Genome* **42**: 993-1000.
- POGGIO & NARANJO, 2005. Citogenética. Biotecnología y Mejoramiento vegetal. Ediciones INTA. Parte II. Capítulo 5.
- POMPANON, F., A. BONIN, E. BELLEMAIN & P. TABERLET, 2005 Genotyping errors: causes, consequences and solutions. Nature Publishing Group: 1-13.
- RADWAN, O., M. F. BOUZIDI, P. NICOLAS & S. MOUZEYAR, 2004 Development of PCR markers for the P15/P18 locus for resistance to *Plasmopara halstedii* in sunflower, *Helianthus annuus* L. from complete CC-NBS-LRR sequences. *Theor Appl Genet* **109**: 176-185.
- RADWAN, O., M. F. BOUZIDI, F. VEAR, J. PHILIPPON, D. T. DE LABROUHE et al., 2003 Identification of non-TIR-NBS-LRR markers linked to the P15/ P18 locus for resistance to downy mildew in sunflower. *Theor Appl Genet* **106**: 1438-1446.
- RADWAN, O., S. GANDHI, A. HEESACKER, B. WHITAKER, C. TAYLOR *et al.*, 2008 Genetic diversity and genomic distribution of homologs encoding NBS-LRR disease resistance proteins in sunflower. *Mol Genet Genomics* **280**: 111–125.
- RAFALSKI, A., 2002 Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* **502**: 94-100.
- RAFALSKI, A., & M. MORGANTE, 2004 Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet.* **20**: 103–111.
- RAHIM, M., C. C. JAN & T. J. GULYA, 2002 Inheritance of resistance to sunflower downy mildew races 1, 2 and 3 in cultivated sunflower. *Plant Breed.* **121**: 57-60.
- RAICU, P., VRANCEANU, V., MIHAILESCU, A., POPESCU, C. & KIRILLOVA MOTZ, M. 1976 Research of the chromosome complement in *Helianthus* L. genus. *Caryologia* **29**: 307-316.
- RAMSAY, L., M. MACAULAY, L. CARDLE, M. MORANGTE, D. S. IVANISSEVICH *et al.*, 1999 Intimate association of microsatellite repeats with retrotransposons and other dispersed repetitive elements in barley. *Plant J* **17**: 415-425.
- RAMSAY, L., M. MACAULAY, S. D. IVANISSEVICH, K. MACLEAN, L. CARDLE et al., 2000 A simple sequence repeat-based linkage map of barley. *Genetics* **156**: 1997-2005.
- RAVEN, P. H., & D. I. AXELROD, 1974 angiosperm biogeography and past continental movements *Annals Missouri Bot. Garden* **61**: 539-673.

- RIESEBERG, L. H., 1998 Genetic mapping as a tool for studying speciation. In D.E. Soltis et al. (ed.) *Molecular systematics of plants*. 2nd edition. Chapman and Hall, New York.
- RIESEBERG, L. H., H. CHOI, R. CHAN & C. SPORE, 1993 Genomic map of a diploid hybrid species. *Heredity* **70**: 285-293.
- ROATH, W. W., J. F. MILLER & T. J. GULYA, 1981 Registration of RHA 801 sunflower germplasm. *Crop Sci* **21**: 479.
- ROCCO, M., 2002 *Citogenética clásica y molecular en girasol*. Universidad de Buenos Aires, Buenos Aires.
- RUNGIS, D., Y. BERUBE, J. ZHANG, S. RALPH, C. E. RITLAND et al., 2004 Robust simple sequence repeat markers for spruce (*Picea* spp.) from expressed sequence tags. *Theor Appl Genet* **109**: 1283-1294.
- SADDER, M. T., N. PONELIES, U. BORN & G. WEBER, 2000 Physical localization of single-copy sequences on pachytene chromosomes in maize (*Zea mays* L.) by chromosome in situ suppression hybridization. *Genome* **43**: 1081–1083.
- SAGHAI-MAROOF, M. A., K. M. SOLIMAN, R. A. JORGENSEN & R. W. ALLARD, 1984 Ribosomal DNA spacer-length polymorphisms in barley: mendelian inheritance, chromosomal location, and population dynamics. *Proc Natl Acad Sci U S A* **81**: 8014-801
- SAHA, M., M. ROUF MIAN, I. EUJAYL, J. ZWONITZER, L. WANG *et al.*, 2004 Tall fescue EST-SSR markers with transferability across several grass species. *Theor. Appl. Genet.* **109**: 783–791.
- SALVO-GARRIDO, H., S. TRAVELL, T. SCHWARZACHER, W. A. HARWOOD & J. W. SNAPE, 2001 An efficient method for physical mapping of transgenes in barley using in situ hybridization. *Genome* **44**: 104–110.
- SAMBROOK, J., E. F. FRITSCH & T. MANIATIS, 1989 *Molecular Cloning: A Laboratory Manual*, 2nd edn. Cold Spring Harbor: Cold Spring Harbor Laboratory Press.
- SAN MIGUEL, P., & J. L. BENNETZEN, 1998 Evidence that a recent increase in maize genome size was caused by the massive amplification of intergenic retrotransposon. *Ann Bot* **82**: 37-44.
- SANTINI, S., A. CAVALLINI, L. NATALI, S. MINELLI, F. MAGGINI et al., 2002 Ty1 /copia- and Ty3 /gypsy-like DNA sequences in *Helianthus* species. *Chromosoma* **111**: 192-200.
- SCHIEX, T., & C. GASPIN, 1997 CARTHAGENE, construction and joining maximum likelihood genetic maps. in: *Proceeding of ISMB, Porto Carras, Halkidiki, Greece* **5**: 258-267.
- SCHLÖTTERER, C., 2004 The evolution of molecular markers - just a matter of fashion? *Genetics* **5**: 63-69.
- SCHLÖTTERER, C., 2000 Evolutionary dynamics of microsatellite DNA. *Chromosoma* **109**: 365-371.

- SCHNEIDER, K., 2005 Mapping Population and Principles of Genetic Mapping, pp. 4-20 in The Handbook of Plant Genome Mapping, edited by K. MEKSEM and K. GÜNTER.
- SCHNEIDER, K., B. WEISSHAAR, D. C. BORCHARDT & F. SALAMINI, 2001 SNP frequency and allelic haplotype structure of Beta vulgaris expressed genes. Mol. Breed. **8**: 63-74.
- SCHRADER, O., AHNE, R., FUCHS, J. & SCHUBERT, I. 1997 Karyotype analysis of Helianthus annuus using Giemsa banding and fluorescence in situ hybridization. Chromosome Res. **5**: 192-200.
- SCHUG, M. D., K. A. WETTERSTRAND, M. S. GAUDETTE, R. H. LIM, C. M. HUTTER *et al.*, 1998 The distribution and frequency of microsatellite loci in Drosophila melanogaster Molecular Ecology. Molecular Ecology **7**: 57-70.
- SCHUSTER, I. & CRUZ, C.D., 2004 Estatística genômica aplicada a populações derivadas de cruzamentos controlados. UFV Viçosa, Brasil.
- SCHWARZACHER, T., 2003 DNA, chromosomes and *in situ* hybridization. BIOS Scientific Publishers: 203.
- SEILER, G. J., & L. H. RIESBERG, 1997 Systematics, origin and germplasm resources of the wild and domesticated sunflower. In Sunflower technology and production, Schneiter, A.A. (ed.) Amer. Soc. Agronomy, Madison, WI: 21-66.
- SHIGYO, M., T. WAKO, A. KOJIMA, N. YAMAUCHI and Y. TASHIRO, 2003 Transmission of alien chromosomes from selfed progenies of a complete set of Allium monosomic additions: the development of a reliable method for the maintenance of a monosomic addition set. Genome **46**: 1098-1103.
- SMULDERS, M. J. M., G. BREDEMEIJER, W. RUS-KORETEKAAS, P. ARENS & VOSMAN B., 1997 Use of short microsatellites from database sequences to generate polymorphism among Lycopersicon species. Theor. Appl. Genet. **97**: 264-272.
- SOMERS, D., P. ISAAC & E. K., 2004 A high-density microsatellite consensus map for bread wheat (Triticum aestivum L.) Theor. Appl. Genet. **109**: 1105-1114.
- SUMNER, A. T., 2003 Chromosome organization and function North Berwick, United Kingdom.
- SUN, Q., N. C. COLLINS, M. AYLIFEE, S. M. SMITH, J. DRAKE *et al.*, 2001 Recombination between paralogues at the Rpl rust resistance locus in maize. Genetics **158**: 423-438.
- SWED, F. S., & C. ELISENHART, 1943 Tables for Testing Randomness of Grouping in a Sequence of Alternatives. Ann. Math. Statist **14**: 66-87.
- TABERLET, P., S. GRIFFIN, B. GOOSSENS, S. QUESTIAU, V. MANCEAU *et al.*, 1996 Reliable genotyping of samples with very low DNA quantities using PCR. Nucleic Acids Res. **24**: 3189-3194
- TANG, S., C. G. HASS & S. J. KNAPP, 2006 Ty3/gypsy-like retrotransposon knockout of a 2-methyl-6-phytyl-1,4-benzoquinone methyltransferase is non-lethal, uncovers a cryptic paralogous mutation, and produces novel tocopherol (vitamin E) profiles in sunflower. Theor Appl Genet **113**: 783-799.

- TANG, S., V. K. KISHORE & S. J. KNAPP, 2003 PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower. *Theor Appl Genet* **107**: 6-19.
- TANG, S., J. K. YU, B. SLABAUGH, K. SHINTANI & J. KNAPP, 2002 Simple sequence repeat map of the sunflower genome. *Theor Appl Genet* **105**: 1124-1136.
- TAO, Q., Y. CHANG, J. WANG, H. CHEN, M. ISLAM-FARIDI et al., 2001 Bacterial Artificial Chromosome-Based Physical Map of the Rice Genome Constructed by Restriction Fingerprint Analysis. *Genetics* **158**: 1711-1724.
- TAYLOR, B. A., 1978 Origins of Inbred mice In Morse. Academic Press: 423-438.
- TEMNYKH, S., G. DECLERCK, A. LUKASHOVA, L. LIPOVICH, S. CARTINHOOR et al., 2001 Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. *Genome Res* **11**: 1441-1452.
- TEMNYKH, S., W. D. PARK, N. AYERS, S. CARTINHOOR, N. HAUCK et al., 2000 Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.) *Theor. Appl. Genet.* **100**: 697-712.
- THIEL, T., W. MICHALEK, R. K. VARSHNEY & A. GRANER, 2003 Exploiting EST databases for the development and characterization of gene-derived SSR-markers in barley (*Hordeum vulgare* L.). *Theor Appl Genet* **106**: 411-422.
- THIEL, T., R. KOTA, I. GROSSE, N. STEIN & A. GRANER, 2004 SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. *Nucleic Acids Research* **32**: 1-5.
- TÓTH, G., Z. GÁSÁRI & J. JURKA, 2000 Microsatellite in different eukaryotic genomes: survey and analysis. *Genome Res* **10**: 967-981.
- VALLUNEN, J. A., 2006 Protocols for Working with M13-Tailed Primers (v. 2.3.1) 1-4.
- VAN DEN BERG, J. H., S. D. CHASALOW & R. WAUGH, 1997 RFLP Mapping Of Plant Nuclear Genomes: Planning Of Experiments, Linkage Map Construction And Qtl Mapping. *Plant Molecular Biology, A Laboratory Manual*: 335-396.
- VANZELA, A. L., C. F. RUAS, M. F. OLIVEIRA & P. M. RUAS, 2002 Characterization of diploid, tetraploid and hexaploid *Helianthus* species by chromosome banding and FISH with 45S rDNA probe. *Genetica* **114**: 105-111.
- VARSHNEY, R. K., A. GRANER & M. E. SORRELLS, 2005a Genomics- assisted breeding for crop improvement. *Trends Plant Sci* **10**: 621-630.
- VARSHNEY, R. K., A. GRANER & M. E. SORRELLS, 2005b Genic microsatellite markers in plants: features and applications. *Trends in Biotechnology* **23**: 48-55.
- VARSHNEY, R. K., T. THIEL, N. STEIN, P. LANGRIDGE & A. GRANER, 2002 *In silico* analysis on frequency and distribution of microsatellites in ESTs of some cereal species. *Cell Mol Biol Lett* **7**: 537-546
- VEAR, F., L. GENTZBITTEL, J. PHILIPPON, S. MOUZEYAR, E. MESTRIES et al., 1997 The genetics of resistance to five races of downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). *Theor Appl Genet* **95**: 584-589.

- VOGL, C., & S. XU, 2000 Multipoint mapping of viability and segregation distorting loci using molecular markers. *Genetics* **155**: 1439-1447.
- VOORRIPS, R. E., 2002 MapChart, Software for the graphical presentation of linkage maps and QTLs. *J Heredity* **93**: 77-78.
- VRANCEANU, V., & F. STOENESCU, 1970 Immunity to sunflower downy mildew due to a single dominant gene. *Probleme Agric* **22**: 34-40.
- WANG, C. J., & C. C. CHEN, 2005 Cytogenetic mapping in maize. *Cytogenet Genome Reshen*.
- WANG, Y., TANG X., CHENG Z., MUELLER, L. GIAVANNONU J., & S. TANKSLEY 2006 Eucromatic and Pericentromeric Heterochromatin: Comparative Composition in the Tomato Genome. *Genetics* **172**: 2529-2540
- WEBER, J. L., 1990 Informativeness of human (cC-dA)n (dGdT)n polymorphisms. *Genomics* **7**: 524-530.
- WEI, W. H., W. P. ZHAO, Y. C. SONG, L. H. LIU, L. Q. GUO et al., 2003 Genomic in situ hybridization analysis for identification of introgressed segments in alloplasmic lines from *Zea mays* × *Zea diploperennis*. *Hereditas* **138**: 21-26.
- WINTER, P., A. M. BENKO-ISEPPON, B. HÜTTEL, M. RATNAPARKHE, A. TULLU et al., 2000 A linkage map of the chickpea (*Cicer arietinum* L.) genome based on recombinant inbred lines from a *C. arietinum* × *C. reticulatum* cross: localization of resistance genes for fusarium wilt races 4 and 5. *Theor. Appl. Genet.* **101**: 1155–1163.
- YI, G., J. M. LEE, S. LEE, D. CHOI & B. D. KIM, 2006 Exploitation of pepper EST-SSRs and an SSR-based linkage map. *Theor Appl Genet* **114**: 113-130.
- YOUNG, N.D. 1994 Constructing a plant genetic linkage map with DNA markers, En: DNA-based markers in plant. Edited by Ronald & L. Phillips, DNA-based markers in plants p. 39-57.
- YOUNG, N.D. 2000 Constructing a plant genetic linkage map with DNA markers, En: DNA-based markers in plant. Edited by Phillips, R:L. & J.K. Vasil, DNA-based markers in plants p. 31-45.
- YU, J. K., T. DAKE, S. SINGH, D. BENSCHER, L. WANLONG *et al.*, 2004 Development and mapping of EST-derived simplesequence repeat (SSR) markers for hexaploid wheat. *Genome* **47**: 805–818.
- YU, J. K., S. TANG, M. B. SLABAUGH, A. HEESACKER, G. COLE et al., 2003 Towards a saturated molecular genetic linkage map for cultivated sunflower. *Crop Sci* **43**: 367–387.
- YÜKSEL, B., & A. H. PATERSON, 2005 Construction and characterization of a peanut HindIII BAC library. *Theor Appl Genet* **111**: 630-639.
- ZHANG, L. S., V. BECQUET, L. SHAO-HUA & D. ZHANG, 2003 Optimization of Multiplex PCR and Multiplex Gel Electrophoresis in Sunflower SSR Analysis Using Infrared Fluorescence and Tailed Primers. *Acta Botanica Sinica* **45**: 1312-1318.

- ZHANG, P., W. LI, J. FELLERS, B. FRIEBE & B. S. GILL, 2004a BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. *Chromosoma* **112**: 288-299
- ZHANG, P., W. LI, B. FRIEBE & B. S. GILL, 2004b Simultaneous painting of three genomes in hexaploid wheat by BAC-FISH. *Genome* **47**: 979-987.
- ZHANG, L. S., V. LE CLERO & D. ZHANG, 2005 Establishment of an effective set of simple sequence repeat markers for sunflower variety identification and diversity assessment. *Can. J. Bot.* **83**: 66-72.
- ZHANG, D. Q., Z. Y. ZHANG, K. YANG & L. TIAN, 2002 Segregation of AFLP markers in a (*Populus tomentosa* x *P. bolleana*) x *P. Tomentosa* Carr. BC1 family. *For Stud China* **4**: 21-26.
- ZHOU, Y., T. BUI, L. D. AUCKLAND & C. G. WILLIAMS, 2002 Direct Fluorescent Primers Are Superior to M13-Tailed Primers for Pinus taeda Microsatellites. *Biotechniques* **32**: 46-52.
- ZHU, Y. L., Q. J. SONG, D. L. HYTEN, C. P. VAN TASSELL, L. K. MATUKUMALLI et al., 2003 Single nucleotide polymorphisms in soybean. *Genetics* **163**: 1123-1134.

TABLA ANEXA

Información completa de los marcadores SSR, EST-SSR e EST-InDeL incluidos en la matriz de datos. Se presenta el nombre del marcador, el número de acceso en el GenBank, el tipo de repeticiones, el motivo y el número de repetición, el fluoróforo, el tamaño del alelo presente en el parental RHA266 en pb, el tamaño del alelo presente en el parental PAC2 en pb, el tamaño de alelo esperado en pb, la secuencia del oligonucleótido sentido, la secuencia del oligonucleótido antisentido y la localización en los grupos de ligamiento en el mapa genético.

Nombre del marcador	Nº de acceso en el Genbank	tipo de motivo SSR	Motivo	Fluoróforo	alelo RHA266	alelo PAC2	Tamaño de alelo esperado	Oligonucleótido sentido	Oligonucleótido antisentido	GL
HA102	BV727861	dinucleótido	GA	FAM	159	149	160	CCTCAAAACACATCTCCCTCTC	GAAGAGGGGTGAGGGTGAAGAGG	-
HA1108	BV727970	trinucleótido	ATT	HEX	186	144	185	GGACCTTCTATTTAGGAGGG	GGGAACATGGAATAAGGG	10
HA1155	BV727896	dinucleótido	GA-GT	HEX	106	90	110	CAATCAAAACATGTGTATGTG	GGATTCTGCTGAAAGGAGACAAC	12
HA1258	BV727912	dinucleótido	GT	HEX	153	151	151	TCCTCCCTTCTCATCAGTG	CGGTTGTGATGATTGGGTTT	4
HA140	G67517	dinucleótido	GA	FAM	158	143	147	CTAGCAACCAACCTCATTG	GTCTCCTTCTCTTCTCGGC	5
HA1402	BV727899	trinucleótido	ATT	FAM	168	153	204	GTGTGAATGTGAGTGTGAATCC	GTCACAGTTGCATATCCATCC	17
HA1590	BV727994	dinucleótido	GT	FAM	68	66	71	GTGGGAGAGAAGATAGGAG	GGGGATTATGCAACCTTGCC	-
HA1626	G67417	trinucleótido	ATT	HEX	148	170	170	GATGTTACACGTTAGCAACG	GAACTCAGCCTAAAAGTC	13
HA1752	BV728001	trinucleótido	ATC	FAM	139	143	147	CCTTCTCTCAAATCCCG	CGTCATTGTCATTGAGAAGATCG	-
HA1770	BV727917	dinucleótido	GA	HEX	179	181	186	GCCGGTCACCCATTCCCCCTTAC	GGCGGAAGCCAGGCGTTGTAGCG	3
HA1837	BV728003	trinucleótido	ATT	FAM	196	187	154	GTTATAAGCGCAAATGGG	CACACACGTGCACTTTAAC	15
HA1848	BV728005	dinucleótido	GT	HEX	233	245	196	CATCCCCCTTCTGAATAGAAA	TAGGCTCGTTAAACTTACGG	7
HA196	G67518	di-mononucleótido	GA G	HEX	180	176	176	GGTGTTAGAGTAAAGTATGCC	GAGGGGTTTCTTTTCAG	-
HA2053	BV728111	trinucleótido	ATT	FAM	130	133	186	CACCATCACCCACCTGCACAAG	CATCAAAGCCATTCTCAGCCC	9
HA2063	BV728113	dinucleótido	GA-GT	FAM	181	172	173	CACCGCAGATGGCCAATTCAAC	GGTTCGTAACTGCAATGCG	9
HA2077	BV727907	dinucleótido	GA	HEX	116	132	130	GATAAGGTTTTCTCTCTCCC	GAGAAAAATGAGCTGATACCG	14
HA2147	BV728115	dinucleótido	GA	FAM	84	86	90	CCCATAAACATTTACACCC	CCTGAGAAATCGGTAGATAGTG	-
HA2190	BV728130	dinucleótido	GA	HEX	188	192	189	CAAACCCTAATCGCCCAATTGC	GGAAGCTTGTGAATCTGGAGGG	3
HA2191	BV728131	dinucleótido	GA	FAM	204	200	211	GGCAACTCAATATCCACACCCC	GGGTTCTCAAGGAGAATGTGAC	16
HA2193	BV727902	dinucleótido	GA	FAM	127	137	132	CAATACACGCAAACCCCGAAC	GCACTGCCAAAACAGTTGAC	16
HA2564	BV728053	dinucleótido	GT	FAM	180	178	162	CTTGGGTTCTTCATAACTC	CATGTAATCACCTTTCAAG	9
HA2579	BV728051	dinucleótido	GT	FAM	187	189	190	CGGATTTATGATTTGATTCACCCG	GGATAACACTACCTTTGATGTTG	10
HA2598	BV728185	dinucleótido	GA-GT	HEX	137	141	242	ATGTCTCCAAGGAATATGC	CAAAACAGGTGTTGCCTGAATC	13
HA2600	BV728055	dinucleótido	GA	FAM	162	160	171	GCCGAATTTGATGATTAGTGG	GAAAGTGAAGTTGGATGTGGCGG	10
HA2604	BV728057	dinucleótido	GT	γ-P ³²	160-180	160-180	162	CTTGGGTTCTTCATAACTC	CATGTAATCACCTTTCAAG	-
HA2605	BV728347	dinucleótido	GA	HEX	76	89	151	GCAGATTACAATATGCACAC	GTGTTGAAAGTAGTTGGTTGC	8
HA2714	BV728147	di-mononucleótido	GA-G	FAM	227	224	223	CCCTAGTGTATAGCAACCTTTC	GTAGTGTGTATGAGGGAGATGG	14
HA2920	BV728085	trinucleótido	ATG	FAM	170	175	178	AACGTGTTCAATTCGGATGC	GAAGATTGGTATGATGGGGC	3
HA293	G67519	dinucleótido	GA	HEX	113	122	122	GGGGACATCTCCCGTCCACC	CCTCATCCATCTCCATCCAATC	14
HA2940	BV728088	dinucleótido	GT	NED	115	98	116	CAGTGAAGCGTTGCATACTCG	CCAATCCTTGAGCAAATCAG	-
HA2989	BV728090	hexa-trinucleótido	ATCCTC-ATC	HEX	127	142	145	GCCTGTGTCTCTCAGAGTC	GTTATGTGGGCTGGCTGCACC	14
HA3039	BV728158	dinucleótido	GA	FAM	184	174	186	GGTTGATTAAGTGTGGGTC	CCTGTGTATGACTCGTTATC	10
HA3059	BV728163	dinucleótido	GA	NED	237	255	256	GAGTGAGAGTAATCAATG	CTCCCTGTCATCACAATCAC	12
HA3073	BV728166	dinucleótido	GA	NED	252	255	261	GAGTCAATTGCTGTATG	GCACTGAAACATTAGCATC	12
HA3102	BV728366	dinucleótido	GA	NED	226	229	225	CATGATGAAGTACCACCTCCA TC	CCTCTCTCATCATGTAAGCC	15
HA3103	BV728174	dinucleótido	GA	NED	184	174	186	CCT TAAAAGGCAAACGAT TCTCC	GTTACAGTACGGCGAGAATGTAGG	7
HA3204	BV728181	trinucleótido	ATC	NED	116	119	121	GCCCTTCAATCCTACCATTAA C	GGATATTGAGTTGGTTGTTGGG	10

Nombre del marcador	Nº de acceso en el Genebank	tipo de motivo SSR	Motivo	Fluoróforo	alelo RHA266	alelo PAC2	Tamaño de alelo esperado	Oligonucleótido sentido	Oligonucleótido antisentido	GL
HA3208	BV728183	trinucleótido	ATG	NED	157	160	157	CCCTATCTCTAATCTTCCTC	GTTCTTGAAGATCTGATGTG	2
HA3239	BV728208	trinucleótido	ATG	NED	133	143	149	CTTAATTCGACATAAGCAGG	CTTTCATTATTCACAACCCCC	2
HA3272	BV728357	dinucleótido	TA GA	FAM	124	128	123	TGTTCTCACACTTTTCACCA	TAGGCCCTTAAATTCAGATG	7
HA3278	BV728212	dinucleótido	GA TA	FAM	160	158	163	TAGCCAACCTTCTCTTTCCA	CTTCCCCTGCACCCATA	8
HA3288	BV728101	dinucleótido	GA	FAM	194	208	176	GCATTAGTGAAGGTGGAAAG	AGGTCTTCAAACGTCAAAAC	-
HA3312	BV728075	dinucleótido	GA	FAM	110	100	115	TAACGTCAACAGACATGGAA	ATAACGATTGCACAACACAA	6
HA3319	BV728105	dinucleótido	GA	NED	170	168	170	GTGCGAAATGCTAGATTGCTCC	CTTCATCACCAGATACCAGCC	-
HA3325	BV728106	dinucleótido	GA	NED y FAM	247	242/250	222	GTCATTGCTTGAGTGTGAGA	CGACGCTGAACATATACTTG	2
HA3330	BV728245	dinucleótido	GA	NED	148	119	182	GGCTGAGTAAATGCCAAATACGG	GGTTGTTGATTACAAGCTCTCC	13
HA3373	BV728080	dinucleótido	GT	NED	182	180	184	GCGGCGAAGAAGAAATTGAGG	GGAAAGGGTTCTGTTCTCC	-
HA3396	BV728228	di-trinucleótido	GT ATT	FAM	176	189	253	CACTGGAACACCTGAAAGAT	ATTCTTTGTTGCCTTTGAG	12
HA3417	BV728230	dinucleótido	GA	FAM	126	124	122	TAATTGATTGGGGTAAATG	TATGATTTGGTGTGCTCAGA	13
HA3448	BV728225	dinucleótido	TA GT	FAM	262	266	266	TCAAGATAGAACCCTGATGTCG	TTTATTAAGTTTCCGGGTTT	11
HA3513	BV728249	trinucleótido	ATT/CAT	γ-P32	120-170	120-170	179	CCACATGACCCATTCAACTTC	TCATGGTTCTGATGAG	8
HA3555	BV728375	trinucleótido	CAT	FAM	102	104	110	GATATCTCTCATAAGTGCCG	GGTCTTGTGATGACGAAGC	12
HA3581	BV728253	trinucleótido	ATG	NED	140	144	147	GTAGAGGAGTATCGCCAATAGC	GTTCCAATGGCCAACGATTGTG	8
HA3582	BV728254	trinucleótido	ATG	NED	121	118	121	TGACCTTAATCCTTGGAGAG	CCTCTAATTCGAAAACCTTAAC	16
HA360	G67406	dinucleótido	(GA)15	FAM	225	236	194	CTCACCCCTTCATCTCCTTC	CAACAAGGAACCGATAACTG	16
HA3621	BV728258	trinucleótido	ATG	NED	138	145	159	CTGGTGCAACTTTCTCAGTG	GTAACCTACTCGCTCCACCTC	-
HA3627	BV728237	dinucleótido	GT	HEX	182	200	179	GATGAAATGCCTCACAGACT	CATTGAAGATTGATGGATGA	5
HA3640	BV728259	dinucleótido	GA	NED	186	180	184	GCCATATCAATAACCAATCCC	CAAAGCATGTGAGGACCAGAATG	11
HA3652	BV728274	dinucleótido	TA GA	HEX	213	215	151	TCAGGCTCCATCACTACTCT	CACGATAAGGTGGTCTTGTA	-
HA3683	BV728278	dinucleótido	GA GT	FAM	186	182	189	TGCCCTTATCTCCTAATTG	GTGGATCTTGGACAGTGTTC	16
HA3700	BV728302	dinucleótido	GA	FAM	178	176	168	GAAAAGATTGGGAGAGACCT	TTCTTCTCCACCCACCTAT	5
HA3703	-	dinucleótido	GA	HEX	215	213	217	CAAATGCTGATTCCACACTA	ATGGTTTCTGTTTGAATTG	4
HA3847	BV728286	dinucleótido	GT	NED	148	143	145	CATCACTTCAACATGCCTCC	CCTAGCTCCTTTATGTAACCTC	10
HA3878	BV728314	trinucleótido	ATC ATT	FAM	199	196	262	TTTGTTTAGCATCATCATCATC	GAGACCCTAACCATACATGA	7
HA3886	BV728315	trinucleótido	ATT	NED	186	188	188	GCTTATGTGGAAGCATGTGCG	CAGGTTTCTACTTACGGGTCG	14
HA3938	BV728186	trinucleótido	ATT	NED	241	232	225	GTCAGTAAAAAGTTACCATTGTTGC	GCTTGAAGATATAAAGTTGTTGC	3
HA3950	BV728296	dinucleótido	C GA	NED	129	137	145	CATTCTCACCAACATAACA	CCCTTGAGTATGCTTCAAAC	14
HA3971	BV728318	dinucleótido	GA	HEX	183	178	179	GGATCTCTCCTGGTCTAGT	GATCAATATCTTCCATGTTAT	17
HA4011	BV728360	dinucleótido	GA	FAM	115	125	126	ACTTCTACCCTCCCCTTCTT	CTGTACACGTGCTGCTTTAG	13
HA4021	BV728320	dinucleótido	GA	HEX	194	185	188	GTCTGTCTGTACAGAATTCTC	GAAGCAGAAGGTTAGGGCTATG	-
HA4023	BV728189	dinucleótido	GA	NED	203	197	254	CGTCGTCTATCTCCAGTAACC	GAATCTTGGACCTGAGTCACC	9
HA4031	BV728329	di-trinucleótido	ATC-GA	FAM	186	167	245	TCAAAAACAACAACACCA	AGCTGTGAATGTGATGATGA	-
HA4057	BV728333	dinucleótido	GT	HEX	206	201	172	AAACCCTTCCGACTTATCTC	TAAAGAGAGACCCAACAAG	3
HA4058	BV728334	dinucleótido	AT	NED	163	161	156	GATTATGATTCCATGAGCCAAG	CATCCTGTAATCCTGTCAATGTC	14
HA4074	BV728323	dinucleótido	GT	NED	178	174	179	CCTTCACACCATACTTATCACC	CAACTTCTACAAGCTCCTAGC	-
HA4090	BV728336	dinucleótido	GT	NED	181	174	184	GCCATTAATGGCTAAGGTTTCG	TGCGTTACCAGACAACAAGG	-
HA4094	BV728192	dinucleótido	GA-GT	HEX	172	174	178	GGGACAAAATGGGAAATAC	ATTTGAGACCCGAACACTCC	-
HA4103	BV728338	dinucleótido	GA	FAM	141	150	192	TCACTTATCACCAATCTCATCCA	CTGTCTCAAATCGGGTGGTT	6
HA4105	BV728354	dinucleótido	GA	NED	113	119	118	GGGACTAGTTTGAATATCTCTC	GTTTGAGAGTGGTATAGGTTATG	3
HA4112	BV728340	dinucleótido	GA	HEX	213	215	183	TTGACTCTCCTTCTCGTCTC	GCTCTCAAGAAATCGGTTAG	2
HA4129	-	dinucleótido	GA	NED	83	85	79	CCCCTCGATTGAGCAGAAATG	CCTTCCGCTTCTGTCTTCAAC	-

Nombre del marcador	Nº de acceso en el Genebank	tipo de motivo SSR	Motivo	Fluoróforo	alelo RHA266	alelo PAC2	Tamaño de alelo esperado	Oligonucleótido sentido	Oligonucleótido antisentido	GL
HA4149	BV728355	dinucleótido	GA	HEX	188	198	111	TCAATTCATCGTGCATATCG	CCAAAGTCCACCAATCTTCC	17
HA4208	BV728198	trinucleótido	ATT	NED	306	303	290	CCCACAATTGAATACGGACATC	CATCTCGTTGCCGTTAACTATC	13
HA4222	BV728199	trinucleótido	ATT	HEX	254	272	219	TCAATAATATGAGGCACCAA	CCCCATAAAAAACATCACA	16
HA4239	BV728202	trinucleótido	ATT	FAM	113	109	116	GAATGATAGTGAATTGAGACAGG	CTGGCATCTATATCCATGGATAG	15
HA4264	BV728346	dinucleótido	GT	HEX	181	179	169	TCCCAAACCTACACCTCTAA	AAACAAACTATGGATTAACAAGG	14
HA432	G67407	dinucleótido	GT	HEX	166	170	168	GGGTTTAGTGGCCAGTAGTTGTC	CTTTATCCCCACCCCCTCC	4
HA477	BV727933	dinucleótido	GT	AgNO3	220-240	220-240	233	CGAGCGAGTGCTAGTTGTGAAG	GGATCCACACATATTACTGAC	9
HA503	BV727935	dinucleótido	GA-TA	FAM	116	114	100	GTAGATCTTCCCTGCACCC	GATGCGTTAGAGATAACCTG	8
HA541	BV727938	dinucleótido	GT	HEX	213	229	220	CAACACACGCTTTGTTGGGTG	CGATAAGGAGGGAGGGGAAGATAG	3
HA77	BV727857	dinucleótido	GA	HEX	247	259	249	TGTAATCTGTATCACTCCACC	GTTGTTCTGTTAGGTCGTTCCG	1
HA806	G67410	dinucleótido	GA-GT	HEX	192	186	191	GATGTTCTTCTGCAC	GGTTGGATAATGGGGCAGC	-
HA817	BV727885	dinucleótido	GA	FAM	110	100	114	GTCTCCTAATACACACCC	GTGTAGAGAGAGAAAGCTTG	-
HA876	BV727953	trinucleótido	TGG	HEX	140	131	146	CGTTATTATAAGACCCACCC	CAACTTGGATGTGGCCGGAG	15
HA911	BV727888	dinucleótido	GA	FAM	186	179	179	CAAAGTTCACCTCGTTTTTC	AAGTGGGAAGGTCTACGAGT	8
HA920	BV727955	trinucleótido	ATC	FAM	177	185	171	CCCCATGTAATCGCTTAG	GTGGGGATAGTGTTCGGC	10
HA928	BV728013	di-trinucleótido	GT-ATT	HEX	217	221	222	CCTTTGTAGTCCCTTACTGG	GGTCGATCATGTATGCGTGTG	10
HA95	G67405	dinucleótido	GA	HEX	145	159	162-166	ACGCTTGATAGACAAATGCT	TAGGCAGCAGGTTTTACTCT	5
HA969	BV727890	trinucleótido	TGG	FAM	160	162	168	CAGTTCTCACTTGTA	CATCAATAAGTGAAGACGGG	-
HA991	G67520	dinucleótido	GA	FAM	137	139	148	GCCCCCTTGATGCCCTTTTC	GAATCGCCATTTGAATCGCCAG	4
ORS1009	BV006395	dinucleótido	CT	AgNO3	340-360	340-360	344	CCATTTACCCTTTACACGGATCT	TATGTATGGTCGCCAATTTACCA	9
ORS1024_1	BV006408	dinucleótido	AG	FAM	214	231	224	GGGAAGTGGGCTTGCTATGTAT	AACACACCGAAATCACCTATGAA	5
ORS1035	BV006417	dinucleótido	CT	NED	326	328	321	CAACCCAACCTTCTCCTCATAACC	AGGGCTGATATTCACTTACACA	2
ORS1040	BV006422	dinucleótido	AG	AgNO3	165-185	165-185	176	CTGCTGATCGTTTCTGGATAGA	TGCTAATCCTTCTAATCAACTCCAC	17
ORS1041	BV006423	dinucleótido	CT	FAM	281	183	279	AAACAAACCTTAATGGGGTCGTA	ATATTGGCTGGTTGATGCTGAT	7
ORS1043	BV006425	dinucleótido	CT	γ-P32	200-210	200-210	204	CCAAACCGCATGTTCTATGTTT	AGTGTATTGCGAATTGTAGTGC	8
ORS1045	BV006427	dinucleótido	AG	AgNO3	145-165	145-165	155	GGTGACAAGAGGGGTTCAG	GGCAAAATGAGTGTGGGATTT	-
ORS1065	BV006445	dinucleótido	CT	NED	297	276	295	ACCGCTGTCAACACCTTAAACTC	GGCTGGGAATCAACTGCTACTAC	2
ORS1068	BV006448	dinucleótido	AG	HEX	346	348	348	AATTTGTGCGACGGTGACGATAG	TTTTGTCATTTTATTACCCAAGG	4
ORS1085_1	BV006463	dinucleótido	CT	HEX	281	277	280	GACCTCAAGGCATGCTAACACTC	ACTAAGTGTGTGGACGGGGAAA	12
ORS1086_1	BV006464	dinucleótido	CT	FAM	145	127	140	TTGTTTGTGCGCACTCAAGATT	ATTATCGGCACATCTTTGGATTT	14
ORS1097	BV006475	dinucleótido	CT	FAM	164	160	161	GACCAAGTGGACTGACACGAG	CGGTGGTGGCTTAGATTGTATAGT	17
ORS1120	BV006490	dinucleótido	GT-GA	γ-P32	320-340	320-340	321	TACCGTGTGCAGGTCTGTCTAAA	TGCAAATAATAAGATAATGACCGATT	-
ORS1127	BV006496	dinucleótido	CT	AgNO3	400	420	424	AAACTGAAACCGTACGAAAGAC	CCTTTTATTTGAGTGTATAATCAG	9
ORS1128	BV006497	dinucleótido	CT	NED	257	259	260	TCTGTGCCATACCACTTTATTCTG	GGCGATAATAGATGCGCACTC	14
ORS1144	BV006509	dinucleótido	CT-CA	AgNO3	125-145	125-145	130	TTAATGTTGTTGCAGCCGTCTC	GCCCGTAAGGTGATCTGTTAGTAGT	10
ORS1146	BV006511	dinucleótido	AG	FAM	344	378	379	GGCTACACTTGCATCTATTGT	TGAAGACACCATCTCCAATGC	11
ORS1152_1	BV006515	dinucleótido	AG	HEX	269	264	275	CCCAAGCTCCCTTTCATCTTA	TGTTTCATGCCTCATTATGTTG	14
ORS1159	BV006522	dinucleótido	CT	NED	212	210	261	TTTCGTGATGGTGATTGATGATT	CAGCAACTTGACCGTTTCATTA	5
ORS1194	BV006553	dinucleótido	CT	FAM	102	108	107	CCTGATCCTGATTCTGATTGTTT	GGAGTAGGAGAGAAGATGGGTGA	2
ORS121	-	dinucleótido	CA	AgNO3	240-260	240-260	241	CAATCTCTGATTTCCCGGAA	CAAAAATAGCCGGTGAAGGA	15
ORS1219	BV006576	dinucleótido	CT	AgNO3	420-440	420-440	442	CAAACCTTGCATGTTATGTCCTTT	GATCTATCTCAACCACCACTGATCT	12
ORS123_1	BV005815	dinucleótido	CT	NED	147	145	147	TATGACCACGCCACTGACAC	ACATCCATCACAGTCCATTTTG	14
ORS1233	BV006589	trinucleótido	AAG	FAM	439	437	448	ACAGAAAGCTGTTAGCTCAGAACA	CAGTCTACATCACCATTGGCTTT	6
ORS1242	BV006598	dinucleótido	CT	HEX	265	255	269	GCAATCGTTTCACTCTTCCATTC	TGGTCGTAGAATTGTCGGTCAT	15

Nombre del marcador	Nº de acceso en el Genebank	tipo de motivo SSR	Motivo	Fluoróforo	alelo RHA266	alelo PAC2	Tamaño de alelo esperado	Oligonucleótido sentido	Oligonucleótido antisentido	GL
ORS1247	BV006602	dinucleótido	CT	FAM	339	337	339	GTCTCGACTTGGGATCTGCTG	AGTGTGAGGGTGTGTAGGATTGA	17
ORS1259	BV006612	dinucleótido	CT	AgNO ₃	210-230	210-230	227	ACCCATGTTGATAGCCAACCTT	TAGATTCCGAGGTGTGAGGGTAT	-
ORS126_1	-	-	-	-	-	-	-	CACTGTCCCTTCTGGTAGTTCC	TTCCACGCAAACCTCAATTCC	16
ORS1262	BV006615	dinucleótido	AG	HEX	336	38	336	CTCCACCATCAACATTAGACCAC	GCTTAAACCATTACTTGCCACAT	-
ORS1265	BV006617	dinucleótido	CT	HEX	229	227	222	GGGTTTAGCAAATAATAGGCACA	ACCCTTGGAGTTTAGGGATCA	9
ORS128	BV005818	dinucleótido	CA	AgNO ₃	365	385	378	GACCGTCCACGTGTCAGC	CAAAATAGCGTTGACGAGCA	16
ORS1287_1	-	dinucleótido	CT	-	-	-	139			6
ORS166	BV012434	dinucleótido	CA	FAM	339	327	344	CAGCCACATGCCCTCTGAC	TGTTAAGAACC GCGACA ACTGC	-
ORS169	-	dinucleótido	CA	-	-	-	-	TGGAAGTGAATGGACCCAAG	GCACTGCACCATTATGAGAAG	17
ORS176	BV005824	dinucleótido	GT	AgNO ₃	430-450	430-450	462	CCCTAACTGGTTTTCTGACCC	AACTTTTGTGTTGTGTCAGG	-
ORS203	BV012457	dinucleótido	CA	AgNO ₃	240-260	240-260	264	GCCCAAGATGTGAAGCGAATG	GTCAGAACAGGACCGAACCCT	2
ORS229	BV012471	trinucleótido	AAC	NED	172	175	176	TCCGACCCGAATCTTATGAACC	GACCCGAATGAGACCCAAACTG	2
ORS243	BV012478	dinucleótido	GT	NED	173	169	170	GGGATGACGTGCGTTTGG	ACCACCATTCTACCGTTTCTC	8
ORS297	BV006634	dinucleótido	GT	HEX	226	222	225	GTGTCTGCACGAACTGTGGT	TGCAAAGCTCACACTAACCTG	17
ORS301	BV005905	dinucleótido	GT	NED	235	237	235	CGTGACCTGTGAAACACCAA	CGATAACCGTGTGAATCGTG	14
ORS303_1	BV005906	dinucleótido	CT	FAM	283	290	289	CGAGAGGACAAGAAATCTGTGA	AGAACGACGGATGAAGATCG	16
ORS31_1	BV006879	trinucleótido	GAA	HEX	280	292	286	AATTCATGCCCAAGAGATG	CACAATTCATGCATTTCTCTGG	5
ORS316	BV005917	dinucleótido	AC	FAM	182	179	189	TGGCGTCTTCATAGCATCAG	GAGATTTGAGCTTCGTGTTGC	13
ORS323	BV006639	dinucleótido	GT	AgNO ₃	380-400	380-400	405	GCCGGAGGATTAGAGGAGTT	CGGGAAACTAGGATCAGAGG	-
ORS329	BV006840	dinucleótido	GT	AgNO ₃	220-240	220-240	236	CATCCTCTCACCAACCAGA	GGGAAATCTTCTAAACGGTATGG	8
ORS331_1	BV005927	dinucleótido	CT-AC	FAM	283	290	177	TGAAGAAGGGTTGTTGATTACAAG	GCATTGGGTTACCATTCT	7
ORS333	BV006841	dinucleótido	GT	HEX	451	453	467	CGGTTAAGATGGTTCAGTTGG	ATATTAAGTTTTGGTTTTAGCCAGAA	16
ORS337	BV006640	dinucleótido	AC	AgNO ₃	180-200	180-200	184	GGGTTGGTGGTTAATTCGTC	TTGGTTCATTATCCTTGTC	-
ORS342	BV005933	dinucleótido	GT	HEX	340	330	342	TGTTTCATCAGTTTGTCTCCA	CACCAGCATAGCCATTCAA	2
ORS354	BV006644	dinucleótido	AT-GT	HEX	162	173	161	TTATCAACCATGACCGCATC	AAATTTTCTTCTCAAATGAACCTC	11
ORS364	BV005947	dinucleótido	AC	NED	223	364	228	ACGGAAGCTCTTGAAGCA	GCGGGCATTCCAAGTAGTAA	-
ORS365	BV005948	dinucleótido	CT-CA	NED	246	242	509	CGAGGCAAAGGGTGTCTAAA	GAAGCGAAGGAATGGTGGTT	1
ORS418	BV005975	dinucleótido	GT	-	-	-	397	TCGTTCTCAACATCACCTTCC	CACGGTATCACTTCCATTCTAACA	16
ORS455	BV005995	dinucleótido	GT	FAM	167	165	167	CTTGATCCACAACGCACAAC	AATCTTCTCTCTCATCAGA	16
ORS457	BV005997	dinucleótido	AC-CT	FAM	226	230	230	TGCATACCCAATCTACCAGCTA	AAGACGAAGGTGCAACCAGT	-
ORS460	BV005998	dinucleótido	GA	FAM	302	300	302	ACTCGGCTACCACCTCACAC	GCCCTTGACCTAACCAAG	7
ORS483	BV006013	dinucleótido	GT	FAM	267	269	271	CCGAACAACAATCTCCACAA	GGTTTAGGTGTCGCATCACA	6
ORS485	BV006015	dinucleótido	CA	AgNO ₃	300-320	300-320	297	ATAAGAGCCGCCCTGAATA	GCTAAAAGTGAACATGCTGGAA	-
ORS491	BV006678	trinucleótido	CAT	AgNO ₃	375-395	375-395	388	ACCCACACACCATAAGTTCCG	CGAAGTGTGAGGATGAACG	-
ORS495_1	BV006020	dinucleótido	GT	AgNO ₃	140-160	140-160	152	CCAGGATTAGGTAGCTTAGTTCCG	GCGATCTGAGGTTGACTCGT	16
ORS499	BV006022	dinucleótido	AC-AT	NED	221	213	226	CACGGTTGGTCAATCACATC	TTTAGGCTGCACATTATGGACT	15
ORS5_1	BV006871	trinucleótido	GGA	AgNO ₃	290-310	290-310	311	ATGTGGAGCAGCAAATTCAG	CTGCTGCCACCATACTG	11
ORS509	BV006029	dinucleótido	GT	AgNO ₃	200-220	200-220	198	CAACGAAAAGACAGAATCGAAA	CGGGAATTTTACAAGGTGA	1
ORS510	BV006030	dinucleótido	GT	HEX	258	249	255	CATCGCGTCCCTCTCTTAA	CCAACCATCACAGCAATCAG	9
ORS511	BV006031	dinucleótido	AT-GT	AgNO ₃	140-160	140-160	156	TGGCTCAGATTAAGTTCACACAG	CGGTTGCGAGTAACAGGTA	13
ORS523_1	BV006683	dinucleótido	AC	HEX	166	164	167	TCGAGAAAGAAAGAGAGTGTGA	GGATGAAAGATATGACCTGGATG	5
ORS525_1	BV006038	dinucleótido	AC	AgNO ₃	275-295	275-295	288	TGCAAGGAGAAGAGGGTTTG	GGACCAAGGTGAAAGAAGGA	2

Nombre del marcador	Nº de acceso en el Genebank	tipo de motivo SSR	Motivo	Fluoróforo	alelo RHA266	alelo PAC2	Tamaño de alelo esperado	Oligonucleótido sentido	Oligonucleótido antisentido	GL
ORS53	-	mononucleótido	T	-	-	-	-	GCTGGCAATTTCTGATACACGAT	CATCTAGACAACGACAGAAGATG	1
ORS533	BV006044	dinucleótido	CT	FAM	162	160	157	TGGTGGAGGTCACTATTGGA	AGGAAAGAAGGAAGCCGAGA	5
ORS578	BV006071	dinucleótido	CT	NED	242	246	238	CTCTCAATCCCTAAAGTCCCCT	TGGTGGATGTGGTTGTTGAT	14
ORS591	BV006699	dinucleótido	CT	NO3Ag	115-135	115-135	127	GTTGCAGGAAAGGTCGAAAG	TTTTATTCTACGCAGACAAGACG	10
ORS613	BV006091	dinucleótido	AG	HEX	226	230	228	GTAACCCTAGGTCAATTTGCAG	ATCTCCGGAAAAATTCTCG	-
ORS620	BV006854	dinucleótido	AG	FAM	244	246	246	TCATGGCGGTTAGAGACGAT	CACTTCTTTCATCAACCCTCT	4
ORS621	BV006097	dinucleótido	CT	γ-P32	230-250	230-250	249	CGCCTTATGCTGAGAGGAAA	CCTGAAGCGAAGAAGAATCG	-
ORS624	BV006855	dinucleótido	CT	AgNO3	100-120	100-120	116	CACGACCTGACACAAAGTTAATG	ATCAGAAGCTCCACCACGAC	8
ORS630	BV006710	dinucleótido	CT	AgNO3	330-350	330-350	347	GCACGACCCGGATATGTAAC	TGTGCTGAGGATGATATGCAG	13
ORS653	BV006115	dinucleótido	CT	HEX	288	317	312	CACCCACCAAGAACCCTAGA	CCGATACATACCATAGCCGATT	2
ORS656	BV006715	dinucleótido	CT	NED	202	206	196	TCGTGGTAAGGGAAGACAACA	ACGGACGTAGAGTGGTGGAG	16
ORS657	BV006118	dinucleótido	CT	NED	140	156	254	AGCCGTCGTCCTCTTTGAT	ACACATTATGGTAGACGCCAGA	3
ORS662	BV006121	dinucleótido	AG	NED	313	317	314	CGGGTTGGATATGGAGTCAA	CCTTTACAACGAAGCACAAATTC	1
ORS665	BV006123	dinucleótido	AG	HEX	290	288	286	GCACATGGATATGGATCTCCT	TGCAAATACAACCTCGGGAAA	16
ORS671_1	BV006127	dinucleótido	AG	NED	182	184	182	CGAGTGGTAGGAATGCTTGG	GGGTGTCAGATTTCTCATCTCA	4
ORS677	BV006132	dinucleótido	CT	AgNO3	380-400	380-400	383	TCCTTCTTCTCATTCTGCATT	CTCCATACCGTTGCCATCAT	17
ORS679	BV006133	dinucleótido	GA	HEX	225	223	240	CCCTCCTCCCTCTTCACTTT	CTCATCGGACAACCAGAACCC	12
ORS682	BV006134	dinucleótido	AG	AgNO3	250-270	250-270	268	CCACAACCTCAAACCGAAACC	CACCGCCAATCTTCAGTCTT	14
ORS687	BV006138	dinucleótido	CT	HEX	166	164	168	ACCGTTACACTTATTGGTTATTTTATT	GGGGTTTGTGTTCTGTTTTG	15
ORS691	BV006140	dinucleótido	CT	FAM	465	461	447	GCATCTGAGCAACTGCGTTA	ACCGTCCTTAGCTCTTGTGAG	10
ORS718	BV006158	dinucleótido	CT	HEX	272	274	278	CACCTTACGCACACCAAACC	ATGCAACACCCGAATCAAAG	3
ORS727	BV006163	dinucleótido	AG	HEX	154	196	192	GGTGGCAAGTGGTGGTTAGA	AAAGCTGGTCATCTCAAGGGTA	17
ORS733	BV006733	dinucleótido	AG	AgNO3	190-210	190-210	194	GGACTCCAACGAGAGAATCAGT	TATGAGTTGGCAAGGGCTTC	11
ORS774	BV006192	dinucleótido	AG	HEX	307	303	309	CCGATTTCTGCAACATTGTAAG	GAACCTTCCAATGGTTTGGGA	-
ORS78	BV006898	trinucleótido	GAA	AgNO3	180-200	180-200	161	GTTTCGTCGAGTACATGTTCTGC	TTTCCCTCTGAAAGTTGTCA	10
ORS788	BV006202	dinucleótido	GA	FAM	264	262	263	CTGGATAAAGATGGATAAAGAGAG	GGACCCACCAAGATTTGTTTT	16
ORS8	BV012512	dinucleótido	CT	AgNO3	180-200	180-200	195	TTGGATCGATTGATGATTGTTG	GAATCCGTCATGTATAAAACGA	15
ORS803	BV006214	dinucleótido	CT	HEX	167	165	165	TTCCACCCATCTTCATCTC	ACCGCCGCATATAAAGGAGT	1
ORS805	BV006216	dinucleótido	AG	NED	271	279	276	CATGGATTATAAGAACGGGTGTT	AATCCCAGGGGTAAAATTGC	9
ORS806	-	dinucleótido	CT	FAM	232	236	239	CGGTTGTTGTTGATGTGGTG	ACGCCACGTTATTTAGGTT	-
ORS807	BV006217	dinucleótido	AG	FAM	256	267	226	CCGATATTTTGACCGATATTTTGC	TCTCACCTTCATCTCCTTCC	10
ORS844	BV006252	dinucleótido	AC	FAM	304	302	307	ACGATGCAAAGAATATACTGCAC	CATGTTTAATAGGTTTTAATTCTAGGG	-
ORS878	BV006281	dinucleótido	AC	FAM	191	200	203	TGCAAGGTATCCATATCCACAA	TATACGCACCGGAAAGAAAGTC	10
ORS887	BV006290	dinucleótido	AC	FAM	247	250	252	TCGAAAACGACTAATCCAATTTTC	GAGCATGAACAAGAATTGACACA	9
ORS889	BV006292	dinucleótido	GT	NED	111	140	138	ATCAACTACGTCACGATACTCC	GTTCTCATGGATTCTCACAACCTC	10
ORS894_1	BV006297	dinucleótido	GT	NED	247	250	252	TTTCTCATGATCCCATTCTAT	TGCATTACCTAATTTCTAGTGGGTTT	8
ORS899	BV006302	dinucleótido	AC	HEX	310	338	323	GCCACGTATAACTGACTATGACCA	CGAATACAGACTCGATAAACGACA	16
ORS925	BV006327	dinucleótido	AC	γ-P32	200-220	200-220	201	ATGATTCTAAGTTGCGGTAGTGC	GTTGGGTTTAAAGTTGTTGCTTCC	1
ORS934	BV006334	dinucleótido	AG	NED	255	253	252	CGTTCGATCTACAAGGTAAAGG	AACATACACAAATCCACGCAGAG	6
ORS959	BV006356	dinucleótido	GT	HEX	247	236	240	CCGCTAAGTATAAACCGCTATT	CGTCTCTTCGCATCAATCTTAT	1
ORS993	BV006383	dinucleótido	AG	FAM	334	330	327	GTGTTACAATCCTTGCCTCGATA	CAAACCACAGGGACCAAAATG	16

Nombre del marcador	Nº de acceso en TIGR	tipo de motivo SSR	Motivo	Fluoróforo	alelo RHA266	alelo PAC2	Tamaño de alelo esperado	Oligonucleótido sentido	Oligonucleótido antisentido	GL
TC12357	TC12357	trinucleótido	ATG	HEX	300	302	279	AGTATTCAGTGAAGCGATG	CCAACCAAACCTACAGCAACT	-
TC15366	TC15366	InDel	GGTTA	FAM	284	290	282	ATGAATTGAAGCATGCAGTAG	ATGAGTTCTTTATCAGCCAGTAT	10
TC18146	TC18146	InDel	CCAAATCT	HEX	504	513	447	ATGAATTTTGGCTTGAAGAAGG	TGCCGGGTATGTATATGGGAAT	10
TC18194	TC18194	dinucleótido	TG	HEX	306	308	305	AGAGGGAAGACACAAACTGAC	AATCAACTCATCATCAATCCA	3
TC20625	TC20625	trinucleótido	GAG	FAM	209	206	209	GGTGGTTCACAAACATAAGG	CTTCTTCTCCATTTCCAATCT	4
TC21542	TC21542	trinucleótido	ATA	FAM	240	242	237	AGTGAGGAGCACAAGCAA	TGAACATCAACATCAAATCAA	-
TC21689	TC21689	trinucleótido	GTG	HEX	362	356	357	GTTACGGAGGTGGTGGTC	CCAATAAACACAGAATCCAAA	4
TC21990	TC21990	dinucleótido	AT	FAM	126	120	136	CCATAAGTGTTGCTCCCTTT	TGATTCTCAGTTCCATTCCA	-
TC22997	TC22997	trinucleótido	ATG	HEX	240	246	232	GGTGTAAGGATGATTGGAAG	AACATACGAAGAAAGCAGTAAAG	-
TC23078	TC23078	trinucleótido	AGG	FAM	194	200	198	GTCACGGAGGAAATACTG	GCACAATAATAAGCACACCC	-
TC23602	TC23602	tetranucleótido	GTTT	FAM	198	201	200	TTCCAACACAAGGCATCA	AAACAACAGTCACAATACATCAA	8
TC23762	TC23762	InDel	GTAGAAATTCCTTTCTGAGGTA	HEX	787	744	779	AACTTCACGCAATGAGACT	CAAGCTTCCCTACAAAGAAA	14
TC24343	TC24343	tetranucleótido	TTGT	FAM	218	210	220	TAAGGCAAAGAACCAGGGAC	ATGAAGCACAACCTAAAGGCA	16
TC24992	TC24992	trinucleótido	AAT	FAM	181	184	180	ACATTCACCTTCTCTCCATCA	AGCCATACTGTCCTCCATACT	8
TC25054	TC25054	pentanucleótido	TTTAC	HEX	313	307	318	TACCTTGACGATGACCAGA	ACCTTACATTCCACCAAACA	16
TC26323	TC26323	dinucleótido	AT	HEX	303	301	303	ATCACAACCTCCTGACAAC	CAACTCTTGGGTGGCTAA	17
TC26869	TC26869	trinucleótido	GAT	HEX	277	280	288	AGAGGAGGAGTTTGTGATGT	CCAATAAACACAGAATCCAAA	16