

Tesis Doctoral

Diversidad genómica y mapeo por asociación para la resistencia a la podredumbre húmeda del capítulo causada por *Sclerotinia sclerotiorum* en girasol

Filippi, Carla Valeria

2015

Este documento forma parte de la colección de tesis doctorales y de maestría de la Biblioteca Central Dr. Luis Federico Leloir, disponible en digital.bl.fcen.uba.ar. Su utilización debe ser acompañada por la cita bibliográfica con reconocimiento de la fuente.

This document is part of the doctoral theses collection of the Central Library Dr. Luis Federico Leloir, available in digital.bl.fcen.uba.ar. It should be used accompanied by the corresponding citation acknowledging the source.

Cita tipo APA:

Filippi, Carla Valeria. (2015). Diversidad genómica y mapeo por asociación para la resistencia a la podredumbre húmeda del capítulo causada por *Sclerotinia sclerotiorum* en girasol. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

http://hdl.handle.net/20.500.12110/tesis_n5695_Filippi

Cita tipo Chicago:

Filippi, Carla Valeria. "Diversidad genómica y mapeo por asociación para la resistencia a la podredumbre húmeda del capítulo causada por *Sclerotinia sclerotiorum* en girasol". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2015.

http://hdl.handle.net/20.500.12110/tesis_n5695_Filippi

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales

**Diversidad genómica y mapeo por asociación
para la resistencia a la podredumbre
húmeda del capítulo causada por
Sclerotinia sclerotiorum en girasol**

Tesis presentada para optar al título de Doctor de la Universidad de Buenos Aires
en el área **CIENCIAS BIOLÓGICAS**

Lic. Carla Valeria Filippi

Directora de Tesis: Dra. Verónica V. Lia

Directora Asistente: Dra. Norma B. Paniego

Consejero de Estudios: Dr. H. Esteban Hopp

Instituto de Biotecnología, CICVyA, INTA Castelar
Buenos Aires, 2015

Fecha de defensa: 16 de abril de 2015.

DIVERSIDAD GENÓMICA Y MAPEO POR ASOCIACIÓN PARA LA RESISTENCIA A LA PODREDUMBRE HÚMEDA DEL CAPÍTULO CAUSADA POR *SCLEROTINIA SCLEROTIORUM* EN GIRASOL

Resumen

Argentina tiene una larga tradición en el mejoramiento de girasol, siendo su germoplasma un recurso genético invaluable a nivel mundial. Nuestro país es el tercer exportador mundial de aceite crudo y segundo exportador de harinas proteicas y pellet. Actualmente, la producción del cultivo presenta una importante brecha entre el rendimiento real y el rendimiento potencial debido principalmente a las limitantes generadas por los estreses bióticos y abióticos.

El mapeo por asociación (MA) es un método de mapeo de QTL (del inglés, *Quantitative trait loci*) que tiene el potencial de identificar las bases genéticas de características cuantitativas complejas, alcanzando resolución a nivel de genes individuales. Los objetivos de este trabajo comprendieron: 1) el estudio de la diversidad genómica en colecciones de girasol cultivado conservadas en el Banco Activo de Germoplasma de INTA Manfredi (BAG-IM); y 2) la identificación de genes involucrados en la defensa a la Podredumbre Húmeda del Capítulo (PHC) causada por *Sclerotinia sclerotiorum* utilizando la estrategia de MA.

La población de mapeo por asociación (PMA) para caracteres complejos de INTA, constituida por 137 líneas endocriadas pertenecientes al programa de mejoramiento de girasol, fue genotipificada con un panel de 384 polimorfismos de nucleótido simple (SNPs) basado en la tecnología Veracode (Illumina), 28 genes candidato y 42 marcadores microsatélite (SSRs). Los estudios de diversidad realizados mostraron que tanto los SSR como los SNPs resultaron informativos para la caracterización del germoplasma. En general, las estimas de variabilidad genética fueron moderadas, al tiempo que se obtuvieron evidencias de la existencia de tres grupos genéticos diferentes en la PMA. La respuesta a PHC en las líneas de la PMA fue evaluada a campo durante cinco campañas sucesivas con inoculación asistida con esporas del hongo, registrando como medidas fenotípicas la incidencia, severidad y período de incubación de la enfermedad.

El análisis estadístico de las asociaciones fenotipo-genotipo llevado a cabo, ya sea mediante el uso de modelos lineales mixtos que contemplan la existencia de estructuración y relaciones de parentesco entre los individuos de la población de mapeo, o modelos

bayesianos de interrogación simultánea de *loci*, permitió la identificación de polimorfismos asociados con reducción de la enfermedad ($p < 0,05$).

Los resultados obtenidos demuestran que los estudios de asociación son herramientas útiles a la hora de identificar genes implicados en caracteres complejos para ser usados como insumos para el mejoramiento genético del cultivo de girasol.

Palabras clave: Girasol, *Sclerotinia sclerotiorum*, mapeo por asociación, gen candidato, SNPs, diversidad genómica.

GENOMIC DIVERSITY AND ASSOCIATION MAPPING FOR SCLEROTINIA HEAD ROT RESISTANCE IN SUNFLOWER

Argentina has a long tradition of sunflower breeding, and its germplasm is a valuable genetic resource worldwide. Our country is the third global oil exporter and the second protein meal pellet exporter. Nowadays, there is a significant gap between actual and potential yield, mainly due to biotic and abiotic stresses.

Association mapping (AM) is a method for QTL mapping that has a potential resolution to the level of individual genes. The aims of this thesis were to: a) study the genomic diversity in a set of sunflower accessions preserved at the Active Germplasm bank of INTA Manfredi (AGB-IM); and b) identify gene loci involved in resistance to *Sclerotinia* Head Rot disease in sunflower by an association mapping approach.

The association mapping population (AMP) for complex characters of INTA, composed of 137 inbred lines from the sunflower breeding program, was genotyped with an Illumina 384-plex oligo pool assay, 28 candidate genes and 42 microsatellite markers (SSR). The analysis of polymorphism in the set of sunflower accessions studied here showed that both the microsatellites and Single Nucleotide Polymorphism (SNP) markers were informative for germplasm characterization. In general, the estimates of genetic variability were moderate, while evidence for the existence of three different genetic groups was found in the AMP. The AMP was evaluated for disease incidence, severity and incubation period (PI) after assisted inoculation with the pathogen across five campaigns in replicated field trials.

Mixed linear models accounting for population structure and kinship relatedness, as well as a Bayesian approach for multilocus association mapping, were used for the statistical analysis of phenotype-genotype associations, allowing the identification of polymorphisms associated with disease reduction ($p < 0.05$).

These results demonstrate the potential of candidate gene association mapping for complex trait dissection in sunflower, contributing to the generation of useful tools for sunflower molecular breeding.

Keywords: Sunflower, *Sclerotinia sclerotiorum*, association mapping, candidate genes, SNPs, genomic diversity.

Agradecimientos

Quisiera expresar mi agradecimiento a todas las personas e instituciones que han hecho su aporte para que este trabajo de Tesis pudiera realizarse:

A la Agencia Nacional de Promoción Científica y Técnica (ANPCyT) y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por las becas otorgadas para la realización de este trabajo de Tesis.

Al Instituto de Biotecnología del Instituto Nacional de Tecnología Agropecuaria (INTA) y a sus directivos por abrirme las puertas de esta prestigiosa institución para realizar este trabajo con libertad y confianza. A toda la gente que allí trabaja, por generar un grato ambiente laboral. Muchas gracias!

A mis directoras Norma y Verónica, por la confianza que han depositado en mí hace casi cinco años, por sus enseñanzas que me han hecho crecer tanto profesional como personalmente. No podría haber tenido mejores directoras!

A Norma Paniego quisiera agradecerle por abrirme las puertas y recibirme con tanta calidez en su grupo de investigación, por compartir su amplia experiencia conmigo, por motivar y apoyar mi formación académica y por su serenidad... que más de una vez me ha ayudado a “bajar un cambio”. Muchas gracias Norma!!

A Verónica Lía quisiera agradecerle por la paciencia y la buena predisposición que ha tenido en la enseñanza. Por estar presente en todo momento, celebrar mis logros y acompañarme en los momentos de flaqueza... Por escuchar mis ideas y confiar en mi criterio. Por todas las correcciones, principalmente las de “último minuto”. No sabés cuánto he aprendido de vos...! Gracias Vero, de corazón!!

A Ruth Heinz y Esteban Hopp por su activa participación en este trabajo, por las enseñanzas, discusiones, ideas y sugerencias brindadas, y por tomarse un tiempo para la lectura y corrección de la Tesis. Muchas gracias!

A Corina Fusari, por ser mi guía en los primeros pasos en el laboratorio, en el Instituto y en el mapeo por asociación. Gracias, Cori!

Al MSc. Julio Di Rienzo por la asistencia con el análisis estadístico de datos fenotípicos y de asociación... por la CANTIDAD de corridas estadísticas que ha hecho (que duraban días), muchas de

ellas derivadas de mis “es posible...?” ... Por su ayuda con la sección MyM, los scripts “*for dummies*” y principalmente por la paciencia (tras la avalancha de mails)! Muchas gracias, Julio!!!

Al Dr. Eduardo Cappa, con quien he hablado personalmente pocas veces, pero que ha respondido a los cientos de mails que le he enviado bajo el *subject* “consulta” y que tenían una carilla de extensión, con cuestiones relacionadas principalmente con el uso de BLR. Gracias Eduardo, aprendí mucho!!!

Si voy a hablar de BLR y de Tassel, Pamela Villalba también merece su apartado, por las horas que hemos pasado intentando hacer andar (bien) ambos programas. Y por prestarme su súper-computadora para los análisis estadísticos utilizando R!! Gracias Pame!!

A Andrea Puebla, Vero Nishinakamasu, Marianne Muñoz, Pablo Vera y Natalia Aguirre de UGB, por todas las corridas en el ABI, secuenciador automático y BeadXpress que han hecho durante este trabajo de Tesis, que por lo general necesitaba “a la brevedad”. Son unos genios!!

A Valeria Moreno, por su ayuda con la multiplicación de las líneas de la PMA para el ensayo, por el ADN de las líneas “B” y “R”, por las correcciones del paper... por todo! Gracias Vale

A Jeremías Zubrzycki por la ayuda con la recolección de muestras en el campo y con la construcción de los mapas de ligamiento...Gracias!!!

A “The mate’s girls” (Lau R., Flavia D., Pame V., Noe U., Moni F. y Naty A.) por las charlas matutinas (¡¡oh-la-la!!), los almuerzos del aquelarre, las salidas gastronómicas y culturales... las re quiero chicas! Ustedes hicieron de este un tiempo muy feliz!!

A toda la gente del “locutorio” por todos los momentos compartidos, mates y galletas de por medio, que han hecho más ameno el trabajo en computadora. Muchas gracias!!

Al grupo de fitopatología de girasol de Balcarce (Facundo Quiroz, Alberto Escande, Carla Maringolo, Silvio Giuliani, Choco, Nereo), por todo lo que me han enseñado sobre “el campo” y la ayuda y soporte que me han dado para que año a año los ensayos fenotípicos pudieran realizarse... Porque cuando he ido con las semillas preparadas para sembrar el suelo ya estaba removido y los surcos marcados, porque cuando he ido a muestrear el ensayo ya había sido raleado, porque a la hora de inocular nunca me faltaron bolsas de papel manila, porque cuando decidimos hacer el ensayo bajo riego tuve cañerías, canillas etc, porque nunca estuve sola a la hora de inocular ni de destapar... todas esas cosas, que son MUCHAS, no se hicieron mágicamente, ustedes siempre estuvieron atrás ocupándose... Gracias de verdad!

A los becarios de INTA Balcarce, a la gente que se hospeda/ó en el Casino, y a mi gran amiga Jesi C. (que Monterrey nos presentó y Balcarce nos hizo amigas) porque han hecho más amenos los meses que he pasado cada verano allí, lejos de casa y en el medio del campo!!

A mis compañeros de grupo desde que comencé Jere Z., Fede E., Sebas M., y a los que se incorporaron después, Moni F., Juan M y Mariana L... aprendí mucho junto a ellos!! Y nos hemos divertido también!! Gracias chicos!!

A Juan Gabriel Rivas, por su ayuda con la genotipificación de SSR para el paper (y por los geles!!)... qué cantidad de PCR que has hecho!!

A Alberto Maligne por la ayuda con el HPLC, por estar ahí cuando algo no funcionaba, y por poner el equipo “a punto” cada vez que decía “este año sí voy a correr los GC que me quedan”... Gracias!

A todos los jefes y compañeros del Grupo Plantas, por las charlas y consejos, porque siempre es bueno tener la visión de otros! Y especialmente a toda la gente de los laboratorios Plantas I y II por los consejos y la ayuda en los experimentos, protocolos, PCRs...Muchas gracias!

A los ingenieros Daniel Álvarez y Diego Cordes del Banco Activo de Germoplasma de INTA Manfredi, por la selección de los materiales para constituir la PMA, por encargarse de la multiplicación de los mismos y por cederme año a año semillas para que pudiera realizar los ensayos a campo. Muchas gracias!

A todos los que me han llevado a Capital (o a alguna estación de trenes) cada vez que perdía el chárter de regreso!!

Y a quienes quizá olvidé mencionar, pero que han puesto su “granito de arena” para que este trabajo pudiera realizarse en tiempo y forma... ¡GRACIAS!

Dedicatorias

*A mis padres y hermanas,
por alentarme a recorrer este camino*

*A mi Nonna,
por iluminarlo*

*A Juan,
por recorrerlo conmigo*

Parte de los resultados presentados en este trabajo de Tesis fueron publicados:

Filippi CV, Aguirre N, Rivas JG, Zubryzcki JE, Cordes D, Moreno MV, Fusari CM, Alvarez D, Heinz RA, Hopp HE, Paniego NB, Lia VV. **Population Structure and Genetic Diversity Characterization of a Sunflower Association Mapping Population using SSR and SNP Markers.** *BMC Plant Biology* 2015, **15**:52.

Carla Filippi, Jeremias Zubrzycki, Verónica Lia, Ruth Heinz, Norma Paniego, Esteban Hopp. **Genetics and Genomics applied to sunflower breeding.** En “Sunflowers: growth and development, environmental influences and pest/diseases”, Nova Science Pubs. Nueva York, USA, 2014. pp. 61–94.

ÍNDICE

ABREVIATURAS	1
INTRODUCCION GENERAL	3
Taxonomía e historia del cultivo	4
Importancia económica	6
El girasol en Argentina	6
Enfermedades “mayores” que afectan el cultivo de girasol	8
Estado del conocimiento del genoma de girasol	10
Estrategias de mejoramiento para resistencia a enfermedades	12
Mapeo por asociación y desequilibrio de ligamiento en girasol	17
Bancos de germoplasma y mejoramiento	19
HIPÓTESIS DE TRABAJO	22
OBJETIVOS	22
Objetivo general	22
Objetivos específicos	22
C A P Í T U L O 1	
CARACTERIZACIÓN FENOTÍPICA DE LA POBLACIÓN DE MAPEO POR ASOCIACIÓN PARA EL CARÁCTER RESISTENCIA A LA PODREDÚMBRE HÚMEDA DEL CAPÍTULO	24
INTRODUCCIÓN	25
MATERIALES Y MÉTODOS	29
Material vegetal	29
Caracterización de las líneas que componen la PMA	29
Preparación del inóculo	30
Tratamiento pre-siembra	30
Siembra	30
Inoculación y registro de enfermedad	31
Análisis estadístico	33
RESULTADOS	36
Constitución de la población de mapeo por asociación	36
Características generales de los ensayos de evaluación de PHC a campo	37
Incidencia de PHC	41
AUDPC Incidencia	45
Severidad de PHC	48
Período de incubación	52
Correlación entre medidas fenotípicas	57
DISCUSIÓN	59

C A P Í T U L O 2

ESTUDIO DE LOS NIVELES DE DIVERSIDAD GENÉTICA Y ESTRUCTURA POBLACIONAL EN LA POBLACIÓN DE MAPEO POR ASOCIACIÓN DE GIRASOL DE INTA _____ 69

INTRODUCCIÓN _____ 70

MATERIALES Y MÉTODOS _____ 74

Material vegetal y marcadores moleculares _____ 74

Genotipificación con marcadores SSR _____ 75

Desarrollo del panel de 384 SNPs para el sistema VeraCode BeadXpress de Illumina _____ 79

Caracterización de la PMA utilizando marcadores tipo SNP _____ 81

Análisis de la diversidad genética _____ 84

Análisis de la estructura poblacional y distancia genética _____ 84

RESULTADOS _____ 86

Estudio de la diversidad genética usando marcadores tipo SSR _____ 86

Diversidad en la PMA utilizando marcadores SNP _____ 87

Estructura poblacional _____ 88

DISCUSIÓN _____ 98

C A P Í T U L O 3

BÚSQUEDA, SELECCIÓN Y GENOTIPIFICACIÓN DE GENES CANDIDATO DE RESPUESTA A ESTRÉS BIÓTICO _____ 105

INTRODUCCIÓN _____ 106

MATERIALES Y MÉTODOS _____ 110

Líneas de la PMA _____ 110

Selección de genes candidatos derivados de un análisis de expresión diferencial en respuesta a *S. sclerotiorum* en girasol _____ 110

Selección de genes candidato derivados de estudios de expresión diferencial en *B. napus* sobre un microarreglo de *Arabidopsis* (Zhao *et al.* 2007b) _____ 111

Selección de genes candidato descritos en literatura como partícipes en los procesos de defensa frente a patógenos fúngicos _____ 113

Selección de marcadores identificados mediante análisis de QTL implicados en procesos de resistencia a *S. sclerotiorum* en poblaciones biparentales de girasol _____ 114

Construcción de Colecciones Núcleo (CN) _____ 115

Diseño de oligonucleótidos _____ 115

Reacciones de PCR _____ 116

Purificación de los productos de PCR, clonado y secuenciación automática _____ 117

Confirmación de la identidad de los genes candidato derivados de *Arabidopsis* _____ 117

Identificación de SNPs y análisis de variabilidad _____ 117

Selección del método de genotipificación de los genes candidato en la PMA _____ 118

Genotipificación por métodos de detección de moléculas heterodúplex _____ 120

a. Corte del Heterodúplex con CEL1 _____ 121

b. Método de dHPLC _____ 122

c. Asignación de genotipos _____ 122

Secuenciación _____ 123

Construcción del mapa genético _____ 123

RESULTADOS	125
Amplificación de genes candidato derivados de un microarreglo de EST de girasol	125
Identificación de ortólogos y amplificación de genes candidato derivados de estudios de expresión diferencial en <i>B. napus</i> (Zhao <i>et al.</i> 2007)	125
Amplificación de GC descriptos como partícipes en procesos de defensa y de marcadores que colocalizaron con QTL de resistencia a <i>Sclerotinia</i>	130
Construcción de colecciones núcleo	130
Caracterización de las regiones candidatas en la Colección Núcleo de líneas de girasol	131
Genotipificación de genes candidato para análisis de asociación	134
Mapeo de SNPs	139
DISCUSIÓN	140
C A P Í T U L O 4	
MAPEO POR ASOCIACIÓN PARA LA RESISTENCIA A LA PODREDUMBRE HÚMEDA DEL CAPÍTULO EN GIRASOL	147
INTRODUCCIÓN	148
MATERIALES Y MÉTODOS	152
Estudio de la homogeneidad genética de las líneas que constituyen la PMA	152
Evaluación fenotípica	153
Marcadores utilizados para el establecimiento de asociaciones	153
Análisis estadístico	154
1. Análisis estadístico mediante el uso de modelos lineales mixtos	154
2. Análisis bayesiano de interrogación simultánea de loci	155
Establecimiento de la estructura poblacional y relaciones de parentesco	157
Caracterización de los GC asociados	158
RESULTADOS	159
Modelo lineal mixto: análisis estadístico de un <i>locus</i> por vez	161
Incidencia de PHC	161
AUDPC incidencia de PHC	163
Severidad	165
Período de incubación	166
Método bayesiano: análisis estadístico con interrogación simultánea de <i>loci</i>	167
Resumen conjunto de los análisis de asociación	169
Análisis de los perfiles de expresión por tejido	173
DISCUSIÓN	176
C O N C L U S I O N E S y P E R S P E C T I V A S	190
B I B L I O G R A F Í A	194
A P É N D I C E	222

ÍNDICE DE TABLAS Y FIGURAS

INTRODUCCION GENERAL

Figura 1.1 Mapas de las variaciones en superficie sembrada, producción y rendimiento del cultivo de girasol en la República Argentina _____	7
Figura 1.2 Enfermedades “mayores” del cultivo de girasol _____	8
Figura 1.3 Comparaciones de los niveles resolutivos alcanzados entre estudios de ligamiento tradicionales y MA _____	13
Figura 1.4 Comparación esquemática de distintos métodos para la identificación de polimorfismos asociados a características complejas _____	14
Figura 1.5 Principios del desequilibrio de ligamiento y mapeo por asociación _____	17
Figura 1.6 Cuellos de botella generados en las especies cultivadas durante la domesticación y las prácticas modernas de mejoramiento vegetal _____	19

CAPÍTULO 1: CARACTERIZACIÓN FENOTÍPICA DE LA POBLACIÓN DE MAPEO POR ASOCIACIÓN PARA EL CARÁCTER RESISTENCIA A PODREDUMBRE HÚMEDA DEL CAPÍTULO

Figura 1.1 Ensayo fenotípico para el carácter resistencia a PHC _____	32
Figura 1.2 Características morfológicas de las líneas de la PMA _____	37
Tabla 1.1 Medias de humedad relativa promedio, temperatura máxima y mínima promedio para las fechas de inoculación, por campaña _____	38
Figura 1.3 Niveles de enfermedad observados para cinco campañas de evaluación de PHC a campo _____	39
Figura 1.4 Distribución del número de plantas evaluadas por surco para las 5 campañas _____	39
Figura 1.5 Otras enfermedades fúngicas detectadas por infección natural en el ensayo de MA _____	40
Figura 1.6 Curvas de progreso de la enfermedad _____	41
Figura 1.7 Relación respecto a la media por línea y por campaña para la variable Incidencia de PHC _____	44
Figura 1.8 Histograma de medias ajustadas para IEmax a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente _____	45
Figura 1.9 Área bajo la curva de progreso de la enfermedad (incidencia) _____	46
Figura 1.10 Histograma de medias ajustadas para AUDPCinc a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente _____	48
Figura 1.11 Distintos grados de avance de la enfermedad _____	48
Figura 1.12 Relación respecto a la media por línea y por campaña para la variable severidad de PHC _____	51
Figura 1.13 Histograma de medias ajustadas para severidad a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente _____	52
Figura 1.14 Relación respecto a la media por línea y por campaña para la variable período de incubación _____	55
Figura 1.15 Histograma de medias ajustadas para Pinc a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente _____	57
Tabla 1.2 Coeficientes de correlación de Spearman para las distintas variables fenotípicas _____	57
Figura 1.16 Gráfico de dispersión para las cuatro variables fenotípicas evaluadas _____	58

CAPÍTULO 2: ESTUDIO DE LOS NIVELES DE DIVERSIDAD GENÉTICA Y ESTRUCTURA POBLACIONAL EN LA POBLACIÓN DE MAPEO POR ASOCIACIÓN DE GIRASOL DE INTA

Figura 2.1 Efecto de la estructura poblacional en el DL y en las asociaciones fenotípicas _____	72
Tabla 2.1 Información de los marcadores SSR utilizados para genotipificar la PMA _____	76
Tabla 2.2 Conjuntos de iniciadores usados para PCR Multiplex _____	79
Figura 2.2 Distribución de los SNPs que integran el SOPA de acuerdo al número de EST considerados para la detección _____	81
Figura 2.3 Esquema de trabajo para la genotipificación de una placa de 96 pocillos utilizando la tecnología GoldenGate BeadXpress de Illumina _____	83
Figura 2.4 Salidas del <i>software</i> GenomeStudio _____	84

Tabla 2.3 Resumen de las estimas de diversidad genética obtenidas para las entradas de girasol incluidas en este estudio _____	86
Tabla 2.4 Tasa de éxito del ensayo de genotipificación _____	87
Figura 2.5 Distribuciones de las frecuencias alélicas observadas _____	88
Tabla 2.5 Resumen de las estimas genéticas obtenidas para el panel de Mapeo por Asociación usando marcadores tipo SNP _____	88
Tabla 2.6 Análisis de la varianza molecular (AMOVA) _____	89
Figura 2.6 Estimaciones de estructura poblacional para el conjunto total de entradas de girasol (PMA, CP y OP) _____	90
Tabla 2.7 Porcentaje de individuos asignados a poblaciones (coeficiente de pertenencia >0,70) utilizando diferentes sets de marcadores moleculares _____	91
Figura 2.7 Resultados de STRUCTURE para K = 3 _____	92
Figura 2.8 Gráficos de dispersión de DAPC mostrando las dos primeras componentes principales _____	93
Tabla 2.8 Porcentaje de individuos asignados al mismo grupo usando STRUCTURE y DAPC _____	93
Figura 2.9 Contribución de los alelos de SNPs a la diferenciación entre grupos dentro de la PMA _____	94
Figura 2.10 Filogramas de Neighbor-Joining para las 137 líneas endocriadas que conforman la PMA _____	97

CAPÍTULO 3: BÚSQUEDA, SELECCIÓN Y GENOTIPIFICACIÓN DE GENES CANDIDATO DE RESPUESTA A ESTRÉS BIÓTICO

Tabla 3.1 Genes candidato seleccionados a partir de un microarreglo de girasol _____	110
Tabla 3.2 Genes candidato seleccionados a partir del análisis de expresión a las 24 horas <i>post</i> inoculación en plantas de <i>B. napus</i> resistentes a <i>S. sclerotiorum</i> _____	113
Tabla 3.3 Genes candidato seleccionados a partir de literatura por su participación en procesos de defensa _____	114
Tabla 3.4 Marcadores que colocalizaron con QTL de resistencia a <i>S. sclerotiorum</i> en poblaciones biparentales de girasol _____	115
Figura 3.1 Diagrama de trabajo para la selección de la metodología de genotipificación de los GC _____	119
Tabla 3.5 Identificación de ortólogos mediante reconstrucción filogenética _____	125
Figura 3.2 Consenso de los Árboles de Máxima Parsimonia (MP) obtenidos para los genes candidatos tomados de Zhao <i>et al.</i> (2007) _____	128
Tabla 3.6 Resultados de las búsquedas BLASTx de las secuencias de girasol correspondientes a los genes candidatos tomados de Zhao <i>et al.</i> (2007) _____	129
Tabla 3.7 Posibles conformaciones de la Colección Núcleo de 8 líneas _____	130
Figura 3.3 SNPs sinónimos (π_{sin}) y no-sinónimos (π_{no-sin}) en las 27 regiones candidatas analizadas _____	132
Tabla 3.8 Características de los 27 genes candidatos seleccionados para realizar mapeo por asociación en girasol _____	133
Figura 3.4 Número de haplotipos encontrados y diversidad haplotípica en los 27 genes de girasol analizados _____	134
Figura 3.5 Confirmación de los fragmentos de digestión observados en gel de agarosa 3% mediante corrida electroforética automática en Fragment Analyzer _____	135
Tabla 3.9 Características de los genes candidato genotipificados por la metodología de detección de moléculas heterodúplex _____	135
Figura 3.6 Nuevos haplotipos detectados con dHPLC y confirmados por secuenciación directa _____	137
Tabla 3.10 Genotipificación de GC en la PMA _____	138
Figura 3.7 Incorporación de marcadores funcionales al mapa saturado de girasol RHA266 X PAC2 _____	139

CAPÍTULO 4: MAPEO POR ASOCIACIÓN PARA LA RESISTENCIA A LA PODREDUMBRE HÚMEDA DEL CAPÍTULO EN GIRASOL

Tabla 4.1 Características de los SSRs utilizados para la evaluación de la estructura poblacional _____	152
Figura 4.1 Metodología de trabajo para la conducción de los análisis de asociación en girasol para la resistencia a la PHC _____	154

Figura 4.2 Esquema de validación cruzada <i>dejando uno afuera</i> (LOO)	156
Figura 4.3 Medias ajustadas de las cuatro medidas fenotípicas descriptoras de la enfermedad PHC en girasol en las 135 líneas analizadas de la PMA	160
Figura 4.4 Resultado de la inferencia de estructura poblacional utilizando el programa STRUCTURE	160
Figura 4.5 Distribución acumulada de los p-valores obtenidos para las cuatro características fenotípicas	161
Tabla 4.2 Análisis estadístico - Marcadores asociados con menor incidencia de PHC	162
Tabla 4.3 Líneas de la PMA portadoras de al menos seis de los nueve alelos de los marcadores asociados con menor incidencia de la enfermedad	163
Tabla 4.4 Análisis estadístico - Marcadores asociados con menor AUDPCinc de PHC	164
Tabla 4.5 Líneas de la PMA portadoras de al menos cuatro de los seis alelos de los marcadores asociados con menor AUDPCinc de la enfermedad	164
Tabla 4.6 Análisis estadístico - Marcadores asociados con menor severidad de PHC	165
Tabla 4.7 Líneas de la PMA portadoras de al menos cuatro de los cinco alelos de los marcadores asociados con menor severidad de la enfermedad	166
Tabla 4.8 Análisis estadístico - Marcadores asociados con mayor período de incubación	166
Tabla 4.9 Líneas de la PMA portadoras de al menos dos de los tres alelos de los marcadores asociados con mayor período de incubación de la enfermedad	167
Tabla 4.10 Estimación de las hipervarianzas para las distribuciones <i>a priori</i> de las varianzas genética y residual	167
Figura 4.6 Gráficos diagnóstico de convergencia de MCMC	168
Tabla 4.11 Análisis Bayesiano	169
Figura 4.7 Relación entre los resultados obtenidos mediante la implementación de los métodos de análisis unilocus y multilocus	170
Tabla 4.12 Marcadores asociados con resistencia a PHC en girasol	172
Figura 4.8 Niveles de expresión por tejido de los distintos genes candidato	174
Figura 4.8 Niveles de expresión por tejido de los distintos genes candidato (continuación)	175

APÉNDICE

Tabla A1.1 Características generales de las líneas endodriadas de girasol incluidas en este estudio	223
Tabla A1.2 Parámetros de los efectos aleatorios para el modelo [1.1] – Incidencia	227
Tabla A1.3 Parámetros de los efectos aleatorios para el modelo [1.2] – Incidencia	227
Tabla A1.4 Parámetros de los efectos aleatorios para el modelo [1.3] – Incidencia	227
Tabla A1.5 Parámetros de los efectos aleatorios para el modelo [1.1] – AUDPCinc	227
Tabla A1.6 Parámetros de los efectos aleatorios para el modelo [1.2] – AUDPCinc	227
Tabla A1.7 Parámetros de los efectos aleatorios para el modelo [1.3] – AUDPCinc	227
Tabla A1.8 Parámetros de los efectos aleatorios para el modelo [1.1] – Severidad	228
Tabla A1.9 Parámetros de los efectos aleatorios para el modelo [1.2] – Severidad	228
Tabla A1.10 Parámetros de los efectos aleatorios para el modelo [1.3] – Severidad	228
Tabla A1.11 Parámetros de los efectos aleatorios para el modelo [1.1] - Período de incubación	228
Tabla A1.12 Parámetros de los efectos aleatorios para el modelo [1.2] - Período de incubación	228
Tabla A1.13 Parámetros de los efectos aleatorios para el modelo [1.3] - Período de incubación	228
Tabla A1.14 Medias ajustadas por línea para cada carácter	229
Tabla A2.1 Características generales de las poblaciones de polinización abierta y compuestas de girasol incluidas en este estudio	231
Tabla A2.2 Lista y origen de los SNPs incluidos en el ensayo de 384 SNPs de Illumina que resultaron polimórficos en las líneas de la PMA	233
Tablas A2.3 Análisis de la varianza molecular (AMOVA) completos con datos de tipo SSR	237
Tablas A2.4 Análisis de la varianza molecular (AMOVA) completos con datos de tipo SNP	238
Tablas A2.5 Análisis de la varianza molecular (AMOVA) completos con datos de tipo SSR+SNP	238

Figura A2.1 Gráficas de los valores de deltaK de STRUCTURE para el conjunto total de materiales (PMA+OP+CP)	239
Figura A2.2 Gráficas de los valores de deltaK de STRUCTURE para las líneas que constituyen la PMA	239
Tabla A2.6 Líneas endocriadas asignadas a los tres grupos obtenidos por el análisis de DAPC	240
Tabla A3.1 Identificación utilizada (ID) para reconocer el género y la especie de las secuencias de GenBank utilizadas en la construcción de los árboles MP	243
Tabla A3.2 Iniciadores utilizados para amplificar, secuenciar y genotipificar los genes candidato en la PMA	245
Tabla A3.3 Genotipos de los genes candidato	249
Tabla A3.3 Genotipos de los genes candidato (continuación)	253
Tabla A4.1 Medias ajustadas de los cuatro caracteres fenotípicos	257
Tabla A4.2 Resultados de los análisis de asociación	259

ABREVIATURAS

ADN	Ácido desoxirribonucleico
ADNc	ADN copia
AIC	Criterio de información de Akaike, del inglés <i>Akaike information criterion</i>
AMOVA	Análisis molecular de la Varianza
ARN	Ácido ribonucleico
ARNm	Ácido ribonucleico mensajero
AUDPC	Área bajo la curva de progreso de la enfermedad, del inglés <i>area under the disease progress curve</i>
BAG-IM	Banco Activo de Germoplasma INTA Manfredi
BIC	Contenido de información bayesiana, del inglés <i>Bayesian Information content</i>
°C	Grados centígrados
cAMP	AMP cíclico, del inglés <i>cyclic AMP</i> , ó <i>3'-5'-cyclic adenosine monophosphate</i>
CEL1	Endonucleasa extraída de apio, del inglés <i>celery 1</i>
cM	Centimorgans
CN	Colección Núcleo de líneas de girasol
CP	Poblaciones Compuestas
DAPC	Análisis discriminante de componentes principales, del inglés <i>Discriminant Analysis of Principal Components</i>
DBCA	Diseño de Bloques completos aleatorizados
dHPLC	Cromatografía líquida de alta resolución en condiciones desnaturalizantes, del inglés <i>denaturing High Performance Liquid Chromatography</i>
DL	Desequilibrio de Ligamiento
dpi	Días post inoculación
ECF	Electroforesis Capilar con detector de fluorescencia inducida por láser
EST	Secuencias que se expresan, del inglés <i>expressed sequence tags</i>
<i>et al.</i>	Y otros, del latín <i>Et alii</i>
GC	Gen/nes candidato
GL	Grupo de ligamiento
GLP	Proteína Tipo Germina, del inglés <i>Germin Like Protein</i>
GO	<i>Gene Ontology</i>
GWAS	Mapeo por asociación a través del relevamiento del genoma completo, del inglés <i>Genome-Wide Association</i>
GxA	Genotipo por ambiente (interacción)
h	Horas
He	Heterocigosis esperada
Ho	Heterocigosis observada
Inc	Incidencia de PHC
IE _{max}	Incidencia máxima de la enfermedad
InDels	Inserciones/deleciones
JA	Ácido jasmónico, del inglés <i>jasmonic acid</i>
Kb	Kilobases
MA	Mapeo por Asociación

MAF	Frecuencia del alelo minoritario, del inglés <i>Minor Allele Frequency</i>
MCMC	Cadenas de Markov Monte Carlo, del inglés <i>Markov Chain Monte Carlo</i>
min	Minutos
ml	Mililitro
mM	Milimolar
MP	Máxima Parsimonia
MR	Moderadamente resistente a PHC
Nº	Número
OP	Poblaciones de polinización abiertas, del inglés <i>Open Pollinated Populations</i>
OPA	Ensayo de genotipificación de SNPs de Illumina, del inglés <i>Oligo pool Assay</i>
ORF	Marco abierto de lectura, del inglés <i>Open Reading Frame</i>
PHC	Podredumbre húmeda del Capítulo
PMA	Población de Mapeo por Asociación
pb	Pares de bases
PCR	Reacción en cadena de la Polimerasa, del inglés <i>Polymerase Chain Reaction</i>
PI	Probabilidad de identidad, del inglés <i>Probability of identity</i>
PIsibs	Probabilidad de identidad considerando que los individuos están emparentados, del inglés <i>Probability of identity siblings</i>
PIC	Contenido de información polimórfica, del inglés <i>Polymorphic information content</i>
Pinc	Período de incubación
ppm	Partes por millón
QTL	<i>Loci</i> de caracteres cuantitativos, del inglés <i>Quantitative Trait loci</i>
RILs	Líneas endocriadas recombinantes, del inglés <i>Recombinant Inbred Lines</i>
rrm	Relación Respecto de la Media
s	Segundos
S	Susceptible a PHC
Sev	Severidad de PHC
SNP	Polimorfismo de Simple Nucleótido, del inglés <i>Single Nucleotide Polymorphism</i>
SOPA	Ensayo de genotipificación de SNPs de Illumina para girasol, del inglés <i>Sunflower Oligo pool Assay</i>
SSRs	Microsatélites, del inglés <i>Simple Sequence Repeats</i>
µg	micro gramo
µL	Microlitro
UV	Luz ultravioleta
Vs.	Versus

INTRODUCCIÓN GENERAL

INTRODUCCION GENERAL

Taxonomía e historia del cultivo

Dentro del universo de las plantas con flor, pocas resultan tan variadas y atractivas como aquellas pertenecientes a la familia de las Compuestas. Entre éstas, se destaca el girasol (*Helianthus annuus* L.) por ser una especie alimenticia, oleaginosa y ornamental.

Según su clasificación taxonómica, *H. annuus* pertenece a la familia *Asteraceae* (*Compositae*), subfamilia *Asteroideae*, tribu *Helianthae*, subtribu *Helianthinae*, género *Helianthus* (Seiler & Rieseberg 1997). La familia de las Compuestas, con más de 1000 géneros y aproximadamente 20000 especies, es la más grande y diversificada dentro de las Angiospermas (Heywood 1993), conteniendo el 10 % de todas las especies de plantas con flor (Stevens 2010). Representantes de esta familia cosmopolita pueden ser hallados tanto en semidesiertos, como en los trópicos o en los países árticos, aunque se localizan mayoritariamente en regiones templadas o subtropicales. Las Compuestas incluyen cultivos económicamente importantes, flores salvajes raras y hermosas, hierbas invasivas y diversas especies que contienen alérgenos comunes y valiosas moléculas de utilidad en medicina (Dempewolf *et al.* 2008). Se las conoce como Compuestas porque aquello que luce como una única gran flor es en realidad un conjunto de muchas, quizá miles, pequeñas flores, inflorescencia compuesta que recibe el nombre de capítulo.

El género *Helianthus* engloba cerca de 50 especies de hierbas anuales y perennes nativas de América (Heiser & Schilling 1981), e incluye especies diploides, tetraploides y hexaploides, todas con el número cromosómico básico de 17 (Rieseberg 1991). *H. annuus* está compuesto por tres subespecies: *H. annuus* ssp. *macrocarpus* (el girasol cultivado), *H. annuus* ssp. *lenticularis* y *H. annuus* ssp. *annuus*, ambos parientes silvestres del cultivado. El girasol es una especie diploide ($2n = 2x = 34$), con un genoma haploide medio de 3500 Mpb (Baack *et al.* 2005). Es una planta típicamente alógama y auto-incompatible, aunque su domesticación no solo permitió la generación de auto-compatibilidad, sino también el aumento del tamaño y del vigor de la semilla y un mayor número de semillas por capítulo. Su tiempo generacional es relativamente corto, de 70 a 100 días y el rendimiento es de aproximadamente 1000 semillas por capítulo.

El amplio potencial de *H. annuus* fue descubierto hace más de 4000 años por civilizaciones pre-colombinas que habitaban la región centro-oeste de América del Norte (Crites 1993; Harter *et al.* 2004; Smith & Yarnell 2009; Bowers *et al.* 2012) quienes lo domesticaron no solo con fines alimenticios, sino también ornamentales, dado que la belleza de su flor y la amplia gama de pigmentos que podían extraerse de ella lo convertían en una especie valiosa para la obtención de tintes naturales y como ornamento en ceremonias. Asimismo, las culturas aztecas yotomíes de México, y los Incas de Perú, lo consideraban un símbolo que representaba a la deidad del sol, quizá por su forma similar al gran astro.

En cuanto al origen del girasol cultivado, la hipótesis más fuerte fue postulada primeramente por Heiser (Heiser *et al.* 1969; Heiser & Schilling 1981) en base a registros morfológicos, geográficos y arqueológicos, y sostenida posteriormente por estudios moleculares (Rieseberg & Seiler 1990). La misma postula que el girasol silvestre *H. annuus* subespecie *annuus*, de capítulos pequeños y ramificados, utilizado como fuente de alimento por civilizaciones que habitaban la región oeste de Estados Unidos, se convirtió en hierba invasiva, llegando de este modo a la parte central de América del Norte, región donde fue domesticado. Parte de este proceso de domesticación incluyó la selección de plantas de un solo capítulo, pero de mayor tamaño y cantidad de semillas, dando lugar al girasol monocéfalo *H. annuus* subespecie *macrocarpus*. En apoyo de la teoría de domesticación en la región central de América del Norte, se han encontrado restos de girasol domesticado en diversos sitios arqueológicos de la región centro, mientras que en la región sureste sólo se ha reportado el hallazgo de piezas fosilizadas de girasol silvestre.

Luego de su introducción en Europa, probablemente por las expediciones españolas (Carter & Putt 1978), el girasol se esparció rápidamente por el continente, donde fue cultivado principalmente con fines ornamentales. Su potencial como cultivo oleaginoso no fue reconocido hasta el siglo XVIII (Carter & Putt 1978), momento en el que se comenzó la selección activa para alto contenido de ácido oleico en Rusia (Heiser & Schilling 1981). Desde entonces, el girasol ha sido foco de esfuerzos para su mejoramiento. La introducción de la heterosis, descrita por primera vez en 1966 (Leclercq 1966), la incorporación de la esterilidad citoplasmática masculina luego del cruzamiento interespecífico con *H. petiolaris* Nutt (Leclercq 1969), y el desarrollo de líneas restauradoras de la fertilidad por Kinman en 1970 (Miller & Fick 1997) permitieron el desarrollo de los híbridos de girasol, portadores de

mayor rendimiento y calidad, alta homogeneidad, sincronización en el tiempo de maduración y mejor adaptación a las condiciones de campo.

Importancia económica

Desde su domesticación, el girasol ha sido cultivado con diversos propósitos, entre los que se destacan la extracción de aceite comestible, su uso principal; el uso como especie ornamental y el consumo directo de sus semillas (girasol confitero). Asimismo, en los últimos años ha surgido un creciente interés en el uso de proteínas de girasol para nutrición humana. Esto se debe a que el contenido proteico remanente en el capítulo, posterior a la extracción de semilla, es de aproximadamente 30-50 % (Dorrell & Vick 1997) y a que las propiedades de las proteínas de girasol son similares a las de la soja y otras leguminosas (González-Pérez *et al.* 2005; Zilic *et al.* 2010).

El girasol es actualmente la cuarta fuente de aceite vegetal comestible a nivel mundial (<http://www.fas.usda.gov/>). El aceite de girasol es considerado *premium* debido a su alto contenido de ácidos grasos insaturados (oleico y linoleico) y bajo contenido de ácidos saturados (linolénico). Hoy en día, se destinan más de 23 millones de hectáreas a nivel mundial al cultivo de girasol (www.fao.org), con una producción anual que asciende a 32 millones de toneladas métricas, principalmente concentradas en la Federación Rusa, Ucrania, India y Argentina (*National Sunflower Association*, 2013).

El girasol en Argentina

La historia de introducción y adaptación del girasol en Argentina está íntimamente relacionada con la de los flujos migratorios humanos. El cultivo llegó al país por primera vez de la mano de inmigrantes provenientes del sur de Rusia, que trajeron consigo pequeñas cantidades de semillas. Luego, la introducción de materiales tempranos de Rusia, Canadá y Rumania, así como la introgresión con especies silvestres del género *Helianthus* permitieron la emergencia del germoplasma argentino, el cual tiene una constitución genética distintiva y está bien adaptado a las condiciones de crecimiento locales (Álvarez 2002; Bertero de Romano 2003; Moreno *et al.* 2013).

Actualmente, Argentina se posiciona como el cuarto productor mundial y el tercer exportador de aceite (Bowers *et al.* 2012; *National Sunflower Association*, 2013). En el país,

el área potencialmente cultivable se extiende entre los 24 y 38 grados de latitud sur, abarcando una amplia gama de ambientes. Esta gran adaptación del cultivo a diversas condiciones agroecológicas lo convierte en el segundo cultivo oleaginoso en importancia, después de la soja, cultivo con el cual compite marcadamente por el uso del suelo.

Como se puede observar en los mapas (La Argentina en mapas: Evolución de la agricultura – CONICET, <http://www.imhicihu-conicet.gov.ar/ARGENTINAenMAPAS/index.html>), la provincia de Buenos Aires ha ocupado históricamente el primer lugar en superficie sembrada. El segundo lugar, sin embargo, ha ido oscilando entre Santa Fe y Córdoba, y últimamente lo ocupa la provincia de La Pampa, demostrando la marginalización del cultivo hacia áreas menos productivas (Figura I.1).

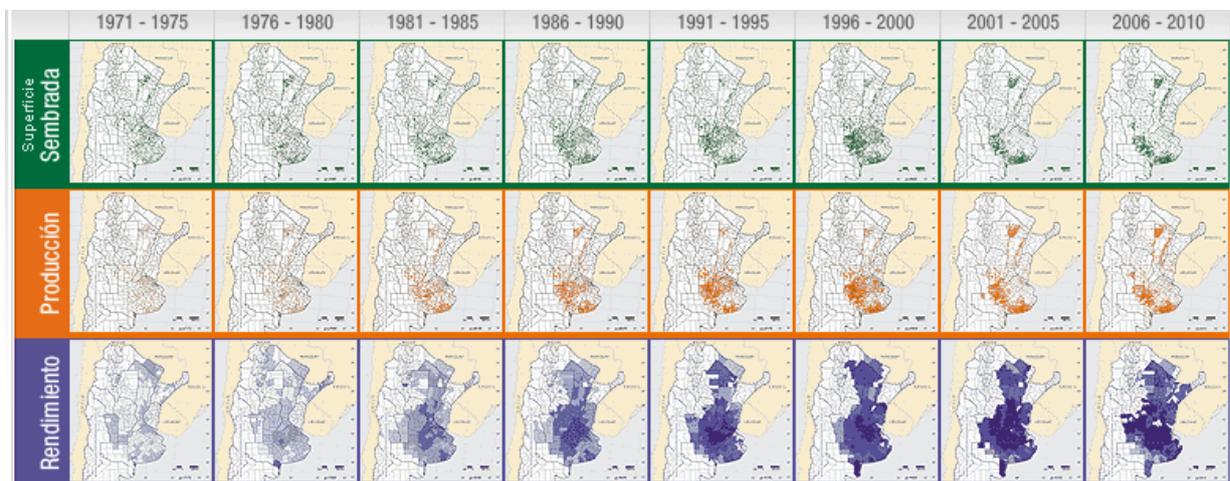


Figura I.1 Mapas de las variaciones en superficie sembrada, producción y rendimiento del cultivo de girasol en la República Argentina Datos divididos por quinquenios desde 1971 hasta 2010. Fuente: La Argentina en Mapas (CONICET).

En los mapas también puede observarse que el rendimiento promedio nacional (medido en kg/ha) ha aumentado considerablemente desde 1975 hasta 1995 aproximadamente, producto de la aplicación de fuertes cambios tecnológicos y de la adopción masiva de híbridos. A partir de entonces, los rendimientos se han estabilizado, lo cual va en concordancia con la relocalización del cultivo en ambientes de menor calidad agroecológica. En este aspecto, la Asociación Argentina de Girasol (ASAGIR) está llevando a cabo un proyecto con el cual se busca estimar la diferencia, o brecha, existente entre el rendimiento potencial del cultivo (esperable al utilizar híbridos adaptados y ensayos conducidos en condiciones exentas de estrés, entre otros) y el rendimiento promedio alcanzado en el país.

Se estima que esta brecha es de aproximadamente 2-2,5 veces, y se atribuye principalmente a la presencia de estreses, tanto bióticos (malezas, enfermedades y plagas), como abióticos (reserva inicial de agua, precipitación, tipo de suelo), y a condiciones de manejo (fecha de siembra, uso de insumos, labranza) (Hall *et al.* 2010).

En este contexto, los programas de mejoramiento de girasol cobran importancia. Si bien mediante mejoramiento tradicional ha sido posible aumentar el rendimiento potencial del cultivo de girasol, al tiempo que ha sido posible controlar la resistencia a ciertas enfermedades fúngicas, pestes y malezas (Sala *et al.* 2012), el advenimiento y desarrollo de marcadores moleculares y mapas genéticos posicionan al mejoramiento asistido por marcadores moleculares como la herramienta más promisoría para el entendimiento de las bases genéticas de diferentes caracteres agronómicos.

Enfermedades “mayores” que afectan el cultivo de girasol

Como todo cultivo, el girasol es susceptible a enfermedades causadas por múltiples agentes, entre ellos virus, bacterias y hongos. Las enfermedades mayores de girasol, denominadas así por su alto potencial de daño, capacidad de producir mermas en el rendimiento, y por la ocurrencia reiterada de epifitias, o manifestación de la enfermedad en extensas áreas del cultivo, son cuatro, a saber: roya negra causada por *Puccinia helianthi*, mildiu o enanismo causado por *Plasmopara halstedii*, marchitez causada por *Verticillium dahliae* y podredumbre húmeda del capítulo causada por *Sclerotinia sclerotiorum* (Figura 1.2, A-D).

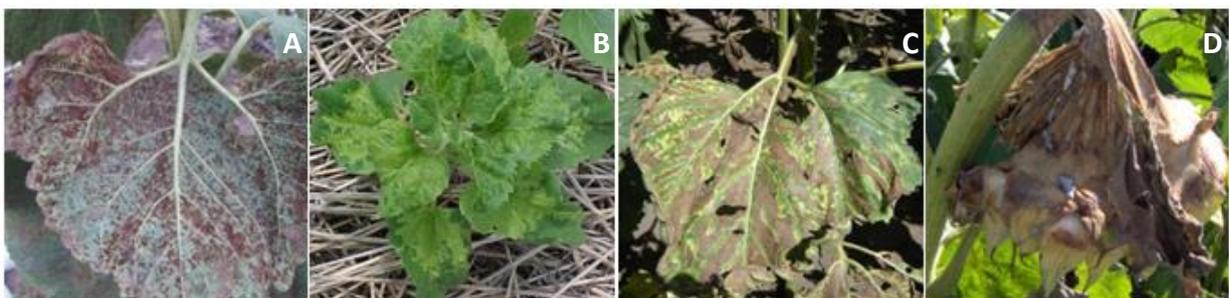


Figura 1.2 Enfermedades “mayores” del cultivo de girasol (A) Roya negra; (B) Mildiu o enanismo; (C) marchitez por *Verticillium dahliae*; (D) Podredumbre húmeda del capítulo.

El hongo ascomicete *Sclerotinia sclerotiorum* (Lib.) de Bary es uno de los patógenos de plantas más devastadores y cosmopolitas. Más de 60 nombres se han utilizado para referirse a las enfermedades causadas por este patógeno fúngico (Purdy 1979), incluyendo

la pudrición algodonosa, la pudrición blanda acuosa, podredumbre del tallo, pudrición de la corona, el tizón de flores y podredumbre húmeda del capítulo. Este patógeno necrotrófico exhibe poca especificidad de huésped e infecta a más de 400 especies de plantas en todo el mundo, incluyendo cultivos importantes y numerosas malezas (Boland & Hall 1994). Como su nombre lo indica, el hongo produce esclerocios, estructuras de resistencia. Los esclerocios pueden germinar en dos formas, carpogénicamente para formar apotecios a partir de la cual se liberan las ascosporas o miceliogénicamente, para producir hifas. En girasol, la germinación carpogénica de esclerocios, y la consecuente liberación de ascosporas al aire son los causantes de la podredumbre húmeda del capítulo (PHC). Por su parte, la germinación miceliogénica puede iniciar directamente el proceso de infección de las raíces y de la base del tallo que resulta en la marchitez y pudrición (Hegedus & Rimmer 2005).

Los síntomas de PHC se manifiestan durante el llenado del grano como lesiones en el receptáculo de consistencia blanda y color té con leche, llegando en su condición más extrema a la pudrición y pérdida total del capítulo. La enfermedad produce mermas en el rendimiento y reducciones en el peso del grano, al tiempo que se aplican penalizaciones en el pago de la cosecha, como consecuencia del aumento en la proporción de corpúsculos extraños y la acidificación del aceite que se extrae de sus semillas (Aguero *et al.* 2001).

El mecanismo de invasión de *S. sclerotiorum* conlleva dos etapas. La primera, denominada etapa de patogenicidad temprana, implica la producción de ácido oxálico y la expresión de enzimas de degradación de pared celular, como poligalacturonasas (SSPG1) y proteasas (ASP), en los extremos de la lesión. Esto conlleva la liberación de pequeñas moléculas (oligo-galacturónidos y péptidos), que inducen la expresión de una segunda oleada de enzimas degradativas que finalizan con la disolución total de tejido de la planta. El ácido oxálico generado y otros metabolitos y enzimas suprimen las defensas del huésped durante la fase de patogenicidad, al tiempo que inician la respuesta de muerte celular, conduciendo a la formación de tejido necrótico. La fase patogénica es seguida por una segunda fase saprofita, transición mediada por la disminución de los niveles de AMP cíclico (cAMP) y aumento en la disponibilidad de glucosa. Los bajos niveles cAMP y la acidificación del entorno generada por la secreción de ácido oxálico promueven el desarrollo de esclerocios y la finalización del ciclo de vida (Hegedus & Rimmer 2005).

A partir de la epifitía de PHC ocurrida en 1987/88 en el Sudeste bonaerense, diversos estudios epidemiológicos e investigaciones se han centrado en la búsqueda de herramientas de manejo de la enfermedad. Las herramientas de manejo actualmente utilizadas se basan en una buena elección del cultivar, una fecha de siembra que permita escapar a los momentos más peligrosos y en la aplicación de desecantes para anticipar la cosecha de cultivos infectados (Escande *et al.* 2002).

Debido a que el control químico es poco práctico y sumamente costoso, el control genético a través del desarrollo de líneas resistentes emerge como la mejor estrategia protectora para el cultivo. Hasta el momento, no ha podido identificarse inmunidad total al ataque del hongo, aunque un amplio rango de respuestas fenotípicas (desde susceptible a medianamente resistente) ha sido descrito para diversas líneas de girasol, variedades e híbridos (Bazzalo *et al.* 1991; Gulya *et al.* 1997). No se conoce resistencia de tipo vertical contra *S. sclerotiorum*, siendo toda la resistencia heredable de tipo horizontal o cuantitativa (Hahn 2002; Ronicke *et al.* 2005; Talukder *et al.* 2014), con acción génica predominantemente aditiva (Bert *et al.* 2002; Castaño *et al.* 1993; Gentzbittel *et al.* 1998; Micic *et al.* 2005), de modo que la estrategia a aplicar en programas de mejoramiento supone la combinación de genes favorables seleccionados a partir de diferentes genotipos de girasol.

Estado del conocimiento del genoma de girasol

Las primeras descripciones del mapa de ligamiento de girasol fueron reportadas en 1995 por Gentzbittel *et al.* (1995) y Berry *et al.* (1995), ambas basadas en la segregación de marcadores de tipo RFLP (*Restriction Fragment Length Polymorphism*). Desde entonces, diversos mapas de ligamiento de diferente longitud, densidad y tipo de marcador molecular han sido desarrollados para girasol, ya sea basados en RFLP (Berry *et al.* 1996, 1997; Jan *et al.* 1998; Gentzbittel *et al.* 1999), RAPD (Rieseberg *et al.* 1993; Rieseberg 1998), AFLPs (Peerbolte & Paleman 1998; Flores Berrios *et al.* 2000; Gedil *et al.* 2001; Al-Chaarani *et al.* 2002), DALPs (*Direct Amplification of Length Polymorphisms*) (Langar *et al.* 2003), o TRAPs (*Target Region Amplification Polymorphisms*), que usan información contenida en las bases de datos de EST (Secuencias que se expresan, *expressed sequence tags*) para generar marcadores polimórficos alrededor de secuencias génicas blanco (Hu & Vick 2003). Del

mismo modo, el desarrollo de marcadores tipo microsatélite (SSR) para el cultivo posibilitó la generación de mapas públicos de referencia, capaces de unificar los mapas desarrollados previamente (Flores Berrios *et al.* 2000; Paniego *et al.* 2002, 2007; Tang *et al.* 2002; Yu *et al.* 2003; Al-Chaarani *et al.* 2004; Kiani *et al.* 2007). Todos los mapas genéticos desarrollados hasta el momento se encuentran accesibles a través de la base de datos cMap de girasol (<http://www.sunflower.uga.edu/cmap/>).

Estos mapas basados en SSR han sido usados para identificar *loci* responsables de la determinación de caracteres cuantitativos (*quantitative trait loci*, QTL) involucrados en numerosos rasgos de importancia agronómica (Tang *et al.* 2006; Kiani *et al.* 2007; Wills & Burke 2007; Ebrahimi *et al.* 2008; Poormohammad Kiani *et al.* 2009; Haddadi *et al.* 2010, 2011), así como también para investigar variaciones entre la estructura genómica de girasol y la de otras especies del género *Helianthus* (Burke *et al.* 2002; Heesacker *et al.* 2008).

En el último tiempo, las nuevas tecnologías de secuenciación y plataformas de genotipificación de alta procesividad han permitido la interrogación simultánea de miles de SNPs distribuidos a lo largo del genoma (Gupta *et al.* 2008), haciendo posible de este modo el desarrollo de mapas genéticos altamente saturados. Un panel de genotipificación de 10K SNPs de tecnología Infinium de Illumina fue desarrollado recientemente (Bachlava *et al.* 2012) y utilizado en diversas poblaciones, con el fin de constituir un mapa de ligamiento consenso del genoma de girasol (Bowers *et al.* 2012). Este mapa genético de alta densidad tiene el potencial de facilitar el ensamblado de secuencias del genoma de girasol (Kane *et al.* 2011), al tiempo que permite la caracterización genotípica eficiente de colecciones de germoplasma, y provee una herramienta importante para estudios genómicos en el cultivo.

A pesar de la importancia económica de diversas especies de la familia *Asteraceae*, no existe hasta el momento un genoma de referencia, lo que dificulta los procesos de investigación y mejoramiento. Debido a que el genoma de girasol es muy grande (3500 millones de bases de longitud) (Baack *et al.* 2005), y complejo (78,5% de secuencias repetitivas), durante varios años se ha considerado demasiado costosa, y prácticamente inviable, su secuenciación. Sin embargo, las nuevas tecnologías de secuenciación masiva han vuelto cada vez más factible y menos costoso el proceso de secuenciación y ensamblado de genomas *de novo* (Mardis 2008; Schatz *et al.* 2010; Kane *et al.* 2011). Kane *et al.* (2011) reportaron el desarrollo de una estrategia para encarar la secuenciación del genoma completo de girasol mediante una combinación de secuenciación de tipo *shot-gun*,

utilizando las plataformas Solexa (Illumina) y 454 (Life Sciences), y la generación de mapas genéticos y físicos de alta densidad que sirven como anclaje para el conjunto de secuencias generadas. Este enfoque híbrido combina las ventajas de las dos tecnologías: la gran profundidad de lectura de Solexa y la capacidad de 454 para generar largas secuencias, aun en regiones de baja complejidad (Dalloul *et al.* 2010; Schatz *et al.* 2010). El desempeño de este enfoque se ve reforzado por la construcción de un mapa físico que proporciona marcadores posicionados cada 5-6 kb en todo el genoma.

Análisis preliminares de la secuencia genómica generada indican que casi el 78,5 % del genoma de girasol está constituido por secuencias repetitivas, lo que dificulta aún más el proceso de ensamblado de las lecturas (Kane *et al.* 2011). No obstante, los autores reportaron que el 76 % de los *contigs* mayores a 5 kb pudieron ser asignados a una posición única en base al mapa físico o genético, sugiriendo que esta aproximación tiene un gran potencial para la generación de un genoma de referencia para girasol (Kane *et al.* 2011). El consorcio que se encuentra actualmente a cargo de la secuenciación del genoma prevé la liberación del primer borrador para el año 2015 (www.sunflowergenome.org).

Estrategias de mejoramiento para resistencia a enfermedades

Gran parte de los caracteres de importancia agrícola o evolutiva están controlados por múltiples *loci* que determinan caracteres cuantitativos. El mapeo genético y la caracterización molecular de los *loci* funcionales facilitan el desarrollo de programas de mejoramiento basados en marcadores moleculares, para caracteres como aumento en el rendimiento, resistencia a enfermedades o tolerancia a sequía, entre otros. Las herramientas más utilizadas para la disección de caracteres complejos son el análisis de ligamiento y el mapeo de asociación (MA) (Risch & Merikangas 1996; Mackay 2001). Los análisis de ligamiento exploran la ocurrencia de herencia compartida entre polimorfismos funcionales y marcadores adyacentes, dentro de familias o poblaciones de *pedigree* conocido. En plantas, este tipo de estudio ha sido típicamente conducido utilizando poblaciones experimentales derivadas de cruzamientos biparentales. Si bien se basan en el mismo principio fundamental de recombinación genética, los estudios de mapeo por asociación, también conocido como mapeo por desequilibrio de ligamiento, implican el

estudio de colecciones de individuos, muchas veces de ascendencia desconocida, la cual puede remontarse a muchas generaciones en el pasado (Yu & Buckler 2006, Figura I.3).

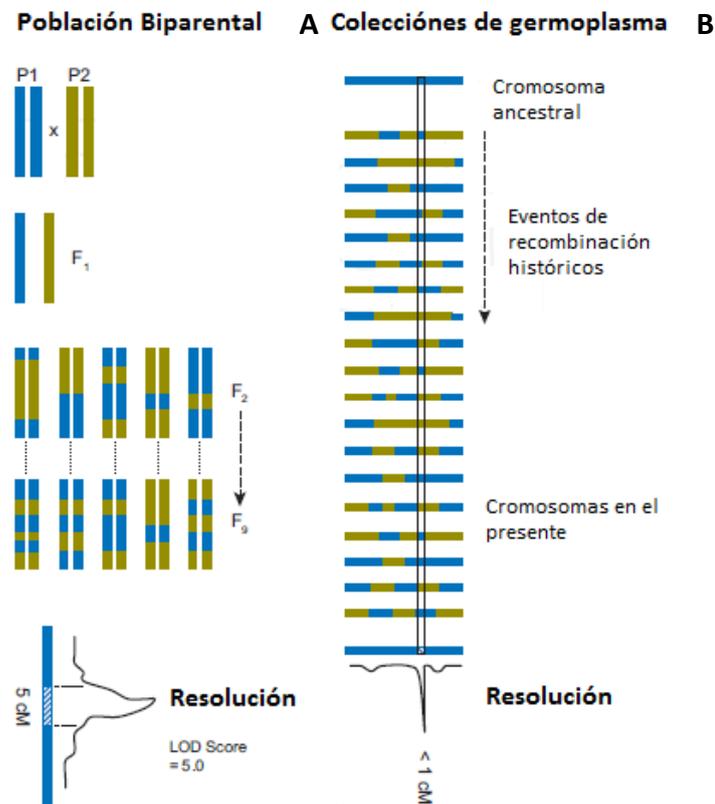


Figura I.3 Comparaciones de los niveles resolutivos alcanzados entre estudios de ligamiento tradicionales y MA (A) Poblaciones de mapeo biparental (RILS). El mapeo en poblaciones biparentales presenta baja resolución en la detección de QTL, como consecuencia de la acumulación de pocos eventos de recombinación; (B) Colección de germoplasma de diferentes orígenes. El MA presenta mayor resolución, a nivel de genes individuales, debido al gran número de eventos de recombinación acumulados durante la historia de la población. Figura adaptada de Soto-Cerda *et al.* (2012).

Esencialmente, los estudios de MA se basan en la exploración de la diversidad genética natural y los eventos de recombinación ancestrales ocurridos a nivel poblacional (Risch & Merikangas 1996; Nordborg & Tavaré 2002). Mediante la exploración profunda de esta genealogía poblacional, en lugar del *pedigree* familiar, los estudios de asociación ofrecen tres ventajas sobre el mapeo de ligamiento clásico en poblaciones biparentales: mayor resolución de mapeo, mayor número de alelos, y menor tiempo para establecer asociaciones, debido a que no es necesario el desarrollo de poblaciones de mapeo (Buckler & Thornsberry 2002; Flint-Garcia *et al.* 2003) (Figura I.4). Para una revisión más detallada de las distintas aproximaciones para la identificación de polimorfismos asociados a

características complejas, como clonado posicional, asociación por *pedigree*, o líneas isogénicas, entre otros, ver Yu *et al.* (2006).

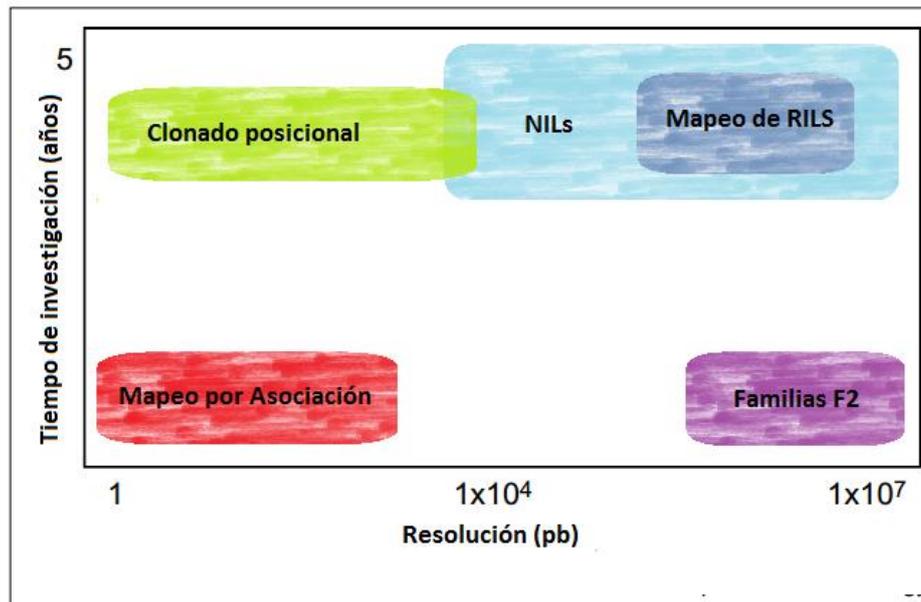


Figura I.4 Comparación esquemática de distintos métodos para la identificación de polimorfismos asociados a características complejas en términos de resolución, tiempo de investigación y número alélico. Figura adaptada de Yu *et al.* (2006).

Desde su primera utilización en plantas (Thornsberry *et al.* 2001), el MA ha cobrado importancia en la investigación genética, apoyado principalmente por el desarrollo de tecnologías de secuenciación y genotipificación de alto rendimiento, y las mejoras en los métodos estadísticos (Rafalski 2002). De acuerdo a la escala y objetivo de cada estudio en particular, el mapeo por asociación puede ser abordado mediante dos estrategias alternativas: a través de la caracterización de genes candidato o a través del relevamiento del genoma completo (*genome-wide*). En la aproximación a través de genes candidato, se pone a prueba la hipótesis de la existencia de correlación entre polimorfismos a nivel de ADN de unos pocos genes de interés y un determinado carácter fenotípico. Esta aproximación supone tener conocimiento de la bioquímica y la genética del carácter (Rafalski 2010). Las primeras aproximaciones de MA en una determinada especie suelen hacerse a través de esta metodología: se realiza la genotipificación con marcadores genéticos en un *locus* que se cree está involucrado con el fenotipo, y luego se prueba la existencia de asociación entre dichos marcadores y el genotipo. La elección de genes

candidato y de marcadores asociados involucra una tarea ardua de búsqueda, selección y genotipificación.

Por su parte, el mapeo de genoma completo (GWAS, *Genome-Wide Association*) supone evaluar para asociación la totalidad de segmentos del genoma, mediante la genotipificación de marcadores moleculares ampliamente distribuidos en todos los cromosomas, de modo de hacer un estudio más exhaustivo del carácter de interés. Los estudios de GWAS han revolucionado el mapeo genético en humanos (Altshuler *et al.* 2008; Donnelly 2008) y esta aproximación ha comenzado a ser adoptada masivamente en estudios en plantas (Nordborg & Weigel 2008). Sin embargo, dado que el primer paso de este proceso involucra la búsqueda de grandes colecciones de marcadores moleculares, principalmente de tipo SNP, este tipo de aproximaciones sólo suele ser llevada a cabo en especies modelo, cuyos genomas están completamente secuenciados, o por grandes consorcios que pueden afrontar la búsqueda *de novo* de polimorfismos en especies que no son modelo.

El desarrollo de un estudio de MA involucra la genotipificación de un conjunto apropiado de marcadores moleculares y la evaluación fenotípica del carácter de interés sobre un conjunto de líneas altamente representativo de la especie para la cual se realizará la inferencia de asociación (Ersoz *et al.* 2007).

El primer paso clave en todo estudio de asociación para caracteres complejos consiste en la selección del germoplasma que conformará la población de mapeo (PMA) (Zhu *et al.* 2008). La diversidad genética, la extensión del desequilibrio de ligamiento (DL) y la presencia de estructura y/o relaciones de parentesco entre miembros de la población determinarán la resolución del mapeo, la densidad de marcadores necesarios y los métodos estadísticos a utilizar. Adicionalmente, debe considerarse la capacidad de adaptación del germoplasma a las condiciones de cultivo locales, ya que esto garantizará el éxito de la fenotipificación (Myles *et al.* 2009). Generalmente, las poblaciones para estudios de asociación pueden ser clasificadas en cinco grupos: (1) población ideal, con estructura poblacional y parentesco inexistente; (2) muestras multi-familia, (3) población estructurada, (4) población estructurada con presencia de materiales relacionados, y (5) población severamente estructurada con presencia de materiales relacionados. Debido a procesos de adaptación local, selección e historia en el mejoramiento, la mayoría de las poblaciones utilizadas en ensayos de asociación en plantas caen en la cuarta categoría (Zhu & Salmeron 2007).

Alternativamente, las PMA pueden clasificarse de acuerdo a la fuente de obtención del germoplasma, en: (1) colecciones núcleo de los bancos de germoplasma, (2) poblaciones sintéticas o (3) líneas elite representativas del germoplasma utilizado en los programas de mejoramiento (Breseghello & Sorrells 2006). De acuerdo a la constitución de la PMA, la aplicación de los estudios de asociación puede diferir en varios aspectos.

Las colecciones núcleo de bancos de germoplasma resultan de interés para estudios de asociación debido a la gran diversidad que albergan en relativamente pocos materiales. Desafortunadamente, la presencia de materiales heterogéneos, como especies nativas o poblaciones abiertas, obstaculizan los pasos posteriores del mapeo (obtención de ADN, genotipificación, etc), como así también la obtención de datos fenotípicos robustos, debido a la dificultad de adaptar estos materiales a condiciones de cultivo normales (Breseghello & Sorrells 2006).

Las poblaciones sintéticas parecieran ser la mejor aproximación a una PMA ideal, dado que se diseñan y mantienen para minimizar los efectos de endocría. Si bien la diversidad alélica de estas poblaciones depende del número y divergencia de los fundadores utilizados para crear la población, la resolución de los análisis de asociación se maximiza debido a la existencia de panmixia y se minimizan los riesgos de encontrar falsos positivos (Breseghello & Sorrells 2006). Sin embargo, si la población se encuentra sometida a selección de forma continua, los análisis de asociación deben contemplar la genotipificación de los individuos en las sucesivas generaciones para poder reflejar correctamente la constitución genética de la población en estudio, siendo esta su principal desventaja de utilización.

Las líneas elite, debido a su estabilidad genética y gran adaptación a condiciones de cultivo, son sumamente atractivas para estudios de asociación de caracteres de baja heredabilidad. Como contrapartida, dado que en los programas de mejoramiento las líneas emparentadas tienden a ser admitidas simultáneamente en las sucesivas etapas de selección de materiales, la estructura poblacional se convierte inevitablemente en un factor a corregir, ya sea mediante información de *pedigree*, o utilizando marcadores moleculares. Es por ello que este tipo de poblaciones suelen utilizarse en estudios de asociación de tipo *genome-wide* (Breseghello & Sorrells 2006).

Mapeo por asociación y desequilibrio de ligamiento en girasol

Los conceptos de desequilibrio de ligamiento (DL) y MA suelen ser utilizados indistintamente en literatura. Sin embargo, presentan sutiles diferencias. De acuerdo con Gupta *et al.* (2005), MA hace referencia a la asociación de un marcador en un determinado *locus* con un carácter fenotípico, mientras que DL se refiere a asociaciones no aleatorias entre dos marcadores o dos genes/QTL (Figura 1.5).

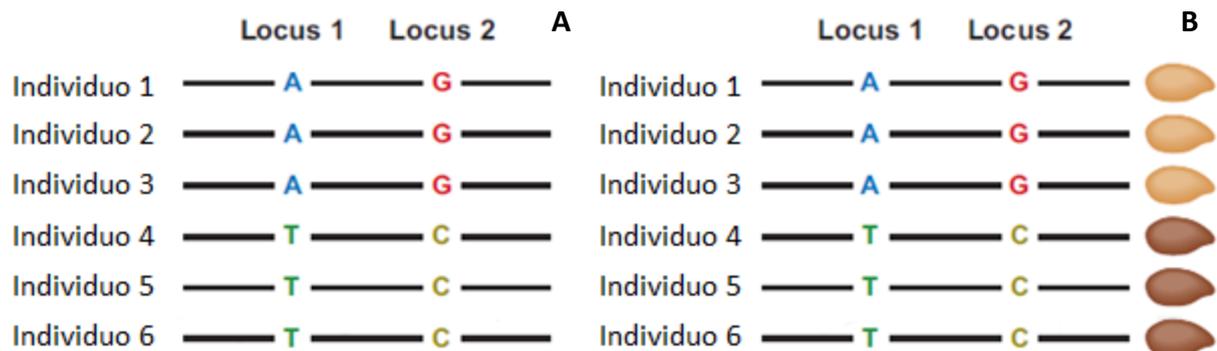


Figura 1.5 Principios del desequilibrio de ligamiento y mapeo por asociación (A) Desequilibrio de ligamiento. Los *loci* 1 y 2 presentan un patrón inusual de asociación entre los alelos AG y TC, presentando una desviación de lo esperado por azar, pero sin correlación significativa con un fenotipo; (B) Mapeo por asociación. Los *loci* 1 y 2, que están en desequilibrio de ligamiento, presentan una covarianza significativa con la medida fenotípica “color de semilla”, lo cual se considera evidencia de asociación. Figura adaptada de Soto-Cerda *et al.* (2012).

Entonces, el MA es en realidad una aplicación del DL. En otras palabras, dos marcadores en DL representan una asociación no aleatoria entre alelos, pero no tienen que estar necesariamente asociadas/correlacionadas con un fenotipo en particular, mientras que asociación implica una significancia estadística en la covarianza de un marcador y un fenotipo de interés. Cabe destacar que se puede detectar DL significativo no solo entre *loci* cercanos, sino también entre *loci* distantes, aun localizados en diferentes cromosomas. Esto es el resultado de fuerzas como selección, sistemas de apareamiento, estructura poblacional, entre otros (Soto-Cerda & Cloutier 2012).

El análisis del DL en distintos cultivares y la demanda de conocimiento sobre los factores que afectan los distintos caracteres han posibilitado estudios de genética de asociación en numerosas especies de interés agronómico como maíz, trigo, sorgo, papa, cebada y raigrás perenne (Thornsberry *et al.* 2001; Palaisa *et al.* 2004; Kraakman *et al.* 2006; Breseghello &

Sorrells 2006; Wei & Croft 2006; Malosetti *et al.* 2007; Agrama *et al.* 2007; Skøt *et al.* 2007; Casa *et al.* 2008).

El alcance del DL en girasol fue estudiado por diversos autores (Liu & Burke 2006; Kolkman *et al.* 2007; Fusari *et al.* 2008; Mandel *et al.* 2013), en base a distintos tipos de materiales (silvestres y cultivados) y mediante diferentes aproximaciones (sobre un panel de regiones cromosómicas seleccionadas o en todo el genoma). A partir de estos trabajos, se observó que la caída del DL es claramente diferente en materiales silvestres comparado con materiales cultivados. En el primer grupo, el DL es fuerte solo a distancias menores a 200 pb (Liu & Burke 2006), mientras que en las líneas endocriadas el DL persiste al menos hasta las 700 pb (Fusari *et al.* 2008). Estos mismos patrones se observaron para el grupo de líneas endocriadas estudiado por Kolkman *et al.* (2007) al ser comparados con los accesos cultivados primitivos y mejorados estudiados por Liu y Burke (2006).

En un estudio publicado recientemente por Mandel *et al.* (2013), en el cual un conjunto de líneas endocriadas de girasol fue genotipificado con un panel de 10k SNPs (Bachlava *et al.* 2012), los autores describieron variaciones sustanciales en los niveles de DL observados a lo largo del genoma, donde las islas de alto DL generalmente correspondían a regiones genómicas relacionadas con caracteres que habían sido históricamente blanco de selección, y grandes regiones en las cuales el decaimiento del DL era relativamente rápido.

Al momento, cuatro estudios de MA han sido reportados para girasol cultivado. El primero de ellos se enfocó en el estudio de la resistencia a PHC causada por *S. sclerotiorum* mediante la genotipificación de 16 genes candidato seleccionados por su participación en procesos de defensa, sobre una población de 94 líneas endocriadas pertenecientes al programa de mejoramiento de girasol de INTA (Fusari *et al.* 2012). En este estudio, se reportó la ocurrencia de asociación significativa entre un alelo del gen candidato HaRIC_B y una reducción del 20 % de incidencia de la enfermedad.

Un estudio combinado de mapeo de asociación y de ligamiento fue llevado a cabo por Cadic *et al.* (2013) para identificar QTL y genes involucrados en el control del tiempo de floración. Un panel de 384 líneas endocriadas y una población de 273 líneas recombinantes endocriadas (RILs) fueron genotipificadas con un set de 5923 marcadores tipo SNP y fenotipificadas. En este trabajo, parte de los QTL identificados como asociados con los caracteres de interés mediante mapeo biparental fueron confirmados utilizando la aproximación de MA.

La aproximación de MA reportada por Mandel *et al.* (2013) se basó en un total de 271 líneas que fueron genotificadas con el panel de 10k SNPs desarrollado por Bachlava *et al.* (2012), y fenotificadas para dos caracteres: días a floración y número de ramificaciones. Se encontraron asociaciones positivas para ambos caracteres, al tiempo que se detectó que existe una correlación negativa entre ambos.

El estudio de MA más reciente fue publicado por Talukder *et al.* (2014), quienes aplicaron la aproximación de MA sobre un panel de 260 entradas de girasol cultivado para encontrar variantes alélicas relacionadas con menor incidencia de Podredumbre de Base de Tallo causada por *S. sclerotiorum*, utilizando un total de 8 genes candidato, seleccionados por ser partícipes en procesos de defensa en *Arabidopsis thaliana*.

Bancos de germoplasma y mejoramiento

El valor de las colecciones de germoplasma para el mejoramiento de cultivos fue reconocido por primera vez por Nikolai Vavilov, quien sostenía que éstas podrían obrar como “ladrillos” a la hora de “construir” especies mejoradas (Vavilov 1940). Fue esta premisa la que motivó el establecimiento de bancos de germoplasma, es decir, colecciones de semillas enfocadas principalmente en razas y especies que estuvieran fuertemente relacionadas con los cultivos comerciales. Esta necesidad de preservar la diversidad genética de cultivares, especialmente aquella presente en las poblaciones de ancestros silvestres (Tanksley 1997), está fundamentada principalmente en el estrechamiento que sufre la diversidad genética de los cultivos, producto de su domesticación temprana y de la selección para características favorables, tales como la pérdida de la inhibición de la germinación, crecimiento compacto, y mejor preservación de semillas, entre otros. Luego de las etapas iniciales de domesticación y selección humana, la variación genética de los cultivos se ha visto reducida por otra fuerza: el mejoramiento moderno, que ha dado lugar al surgimiento de las variedades comerciales de alto rendimiento en las que hoy en día la agricultura se basa (Figura I.6).

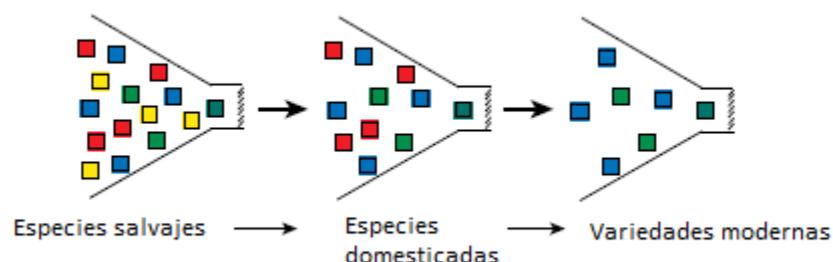


Figura I.6 Cuellos de botella generados en las especies cultivadas durante la domesticación y las prácticas modernas de mejoramiento vegetal Los cuadrados representan variaciones alélicas de los genes presentes en las especies salvajes, pero que se van perdiendo gradualmente a través de las etapas de domesticación y mejoramiento. Figura adaptada de Tanksley, 1997.

La disminución de la diversidad genética de los cultivos los vuelve más vulnerables a enfermedades y epidemias de insectos y pone en peligro el mejoramiento genético sostenido en el largo plazo (Harlan 1987). Un ejemplo de esto lo constituye la epidemia de tizón de la hoja de maíz ocurrida en 1970. Esta enfermedad redujo drásticamente los rendimientos de maíz en los Estados Unidos y se atribuyó a la amplia utilización de un único factor de esterilidad masculina genética que, por desgracia, estaba ligado a la susceptibilidad a la enfermedad (Ullstrop 1978).

El mismo riesgo se encuentra latente en los materiales de girasol sembrados globalmente, debido a que el mejoramiento de híbridos está basado en una única fuente de sistema de esterilidad citoplasmática (CMPS), conocido como PET1 o citoplasma francés, derivado del cruzamiento interespecífico entre *Helianthus petiolaris* y *Helianthus annuus* (Leclercq 1969). En este contexto, resulta de importancia el incremento en los conocimientos en materia de diversidad genética de girasol, a fin de reducir el riesgo asociado a vulnerabilidad genética.

Hoy en día, existen a nivel mundial más de 1750 colecciones de semillas documentadas, que contienen aproximadamente 7 millones de entradas (McCouch *et al.* 2012). A pesar de la importancia que tiene el almacenamiento y la preservación de la variación genética en los bancos de semillas, estas actividades no resultan suficientes para asegurar la productividad futura de la agricultura: el establecimiento de los bancos de semillas debe estar acoplado a la habilidad de utilizar activamente estos recursos. La gran mayoría de estas entradas almacenadas en bancos no hacen contribuciones a las variedades modernas, de modo que el mejoramiento de cultivos todavía se hace sobre una base genética estrecha.

Los estudios de diversidad genética basados en el uso de marcadores moleculares, así como la aplicación de metodologías que permitan vincular variantes alélicas con respuestas fenotípicas se presentan como herramientas útiles, que tienen la capacidad de descubrir el verdadero potencial de los materiales almacenados en bancos de semillas para ser utilizados en mejoramiento de cultivos.

El presente trabajo se enmarca dentro de un grupo de investigaciones llevadas a cabo por el grupo genómica de girasol de INTA tendientes a estudiar el o los mecanismos de respuesta

de girasol frente a distintos tipos de estreses (bióticos y abióticos). Al respecto, en trabajos recientes del grupo, se ha avanzado en la caracterización fenotípica del germoplasma de girasol cultivado en relación con la resistencia a PHC (Trogia 2003) y en el análisis de los perfiles de expresión de transcritos y de metabolitos en genotipos de respuesta contrastante frente al patógeno (Peluffo 2010; Fernandez *et al.* 2012b). Asimismo, se ha abordado el estudio del carácter tanto en poblaciones biparentales (Maringolo 2007; Zubrzycki *et al.* 2012b) como en poblaciones no estructuradas (Fusari *et al.* 2012), al tiempo que se están estudiando genes candidatos identificados a partir de bibliografía y trabajos previos del grupo (Ehrenbolger *et al.* 2012). Este trabajo de Tesis se propone continuar la línea de trabajo en poblaciones no estructuradas de girasol iniciada por Fusari (2010). La constitución de una nueva y mayor población de mapeo, el desarrollo y aplicación de nuevas tecnologías de tipificación de distintos niveles de procesividad, y la utilización de nuevos métodos estadísticos de análisis constituyen las bases de este trabajo.

La meta principal es contribuir al desarrollo de sistemas eficientes para la exploración de los recursos genéticos con vistas al mejoramiento asistido del cultivo de girasol a través de la asociación de marcadores moleculares y genes con parámetros de resistencia a la enfermedad Podredumbre Húmeda del Capítulo. Este estudio involucra la evaluación fenotípica en ensayos a campo con inoculación asistida con esporas del hongo (Capítulo 1); la caracterización de la Población de Mapeo por asociación mediante el estudio de los niveles de diversidad y estructura poblacional presentes en las líneas que la conforman (Capítulo 2); y la conducción de un estudio de mapeo por asociación en girasol para el carácter resistencia a PHC, utilizando paneles de genotipificación de alta *performance* (Illumina) y un conjunto de genes seleccionados por ser partícipes en procesos de defensa frente a patógenos fúngicos (Capítulo 3 y Capítulo 4).

HIPÓTESIS DE TRABAJO

Las hipótesis que sustentan este trabajo de Tesis son:

- Existen regiones cromosómicas (QTL) no descritas asociadas a la resistencia a la PHC que aportan diferentes mecanismos de defensa al patógeno.
- Es posible identificar diferentes fuentes de resistencia a la PHC en genotipos de comportamiento diferencial al ataque del hongo.
- La integración de los datos moleculares (SNPs) y funcionales (fenotipos) obtenidos del análisis de poblaciones no estructuradas correctamente caracterizadas, permite asociar regiones genómicas, variantes alélicas, y genes candidato con la resistencia a PHC.

OBJETIVOS

Objetivo general

En este trabajo se propone contribuir al estudio de las regiones genómicas asociadas a la defensa del girasol frente al ataque del patógeno *S. sclerotiorum*, a través del análisis integral de la diversidad nucleotídica presente en genotipos de girasol de comportamiento diferencial ante la enfermedad, e implementar estrategias de mapeo por asociación para identificar alelos útiles presentes en colecciones de germoplasma disponibles como insumos para el mejoramiento genético del cultivo.

Objetivos específicos

- Caracterizar fenotípicamente la Población de Mapeo por Asociación (PMA) de girasol de INTA en campañas sucesivas en el campo utilizando infecciones asistidas con suspensiones de esporas del hongo;
- Diseñar y evaluar paneles de genotipificación masiva utilizando plataformas de alta procesividad;

- Estudiar los niveles de diversidad, estructura poblacional y relaciones de parentesco entre las líneas que conforman la PMA;
- Desarrollar un estudio de MA para la resistencia a la PHC en girasol.

CAPÍTULO 1

**CARACTERIZACIÓN FENOTÍPICA DE LA POBLACIÓN DE
MAPEO POR ASOCIACIÓN PARA EL CARÁCTER
RESISTENCIA A LA PODREDUMBRE HÚMEDA DEL
CAPÍTULO**

INTRODUCCIÓN

Los estudios de asociación para el mejoramiento vegetal suelen ser proyectos a largo plazo, ya que se necesita contar con ensayos fenotípicos conducidos a lo largo de muchos años, en múltiples ambientes (Flint-Garcia *et al.* 2005). Contar con datos fenotípicos de alta calidad, tomados a lo largo del tiempo, garantiza no solo la potencia de la detección de QTL (Arbelbide & Bernardo 2006; Yu & Buckler 2006) sino que también allana el camino para una futura transición de una estrategia de MA por genes candidato a una de tipo *genome-wide* (Zhu *et al.* 2008).

El Banco Activo de Germoplasma de Girasol de INTA Manfredi (BAG-IM) funciona desde hace más de 30 años en asociación al programa de mejoramiento, y alberga cerca de 1200 accesiones provenientes de diversos orígenes geográficos (Moreno *et al.* 2013). Las líneas endocriadas generadas y/o mantenidas en el BAG-IM constituyen una herramienta de utilidad para estudios de asociación debido a sus altos niveles de diversidad genética. Dado que estas líneas forman parte de programas de mejoramiento llevados a cabo en INTA, cuentan con información fenotípica preexistente sobre varios caracteres de interés, entre ellos el comportamiento ante enfermedades fúngicas, como la PHC. En este contexto, el desarrollo de una PMA a partir de líneas elegidas de programas de mejoramiento bien adaptadas a las condiciones de crecimiento locales, tiene la ventaja de aportar un balance entre diversidad genética y adaptación de germoplasma.

En los últimos años, el desarrollo de nuevas tecnologías de secuenciación y genotipificación ha permitido la obtención de grandes volúmenes de datos genotípicos. Como contrapartida, la obtención de datos fenotípicos robustos continúa siendo un desafío, haciendo que en este momento, la fenotipificación haya tomado el lugar que otrora ocupaba la genotipificación como el principal cuello de botella y de restricción operativa en los programas de mejoramiento asistido por marcadores moleculares (Furbank & Tester 2011).

La fenotipificación para resistencia a patógenos en plantas merece especial atención, dado los múltiples factores que deben considerarse para el desarrollo de las mismas, incluyendo presencia del patógeno, condiciones climáticas favorables, y huésped susceptible. En este contexto, garantizar las condiciones de ambiente favorable y umbral mínimo de inóculo resulta necesario para la correcta identificación de materiales con diferentes grados de susceptibilidad, que aseguren el éxito del mapeo.

La caracterización del comportamiento de los cultivares de girasol frente a la podredumbre en condiciones naturales es dificultosa debido a la alta influencia que tienen las condiciones ambientales sobre la capacidad de este hongo para instalarse y producir la enfermedad (Vear & Tourvieille de Labrouche 1984).

Dada la importancia de este carácter, y que las frecuencias y magnitudes de ataque natural por el hongo muestran variaciones de acuerdo al año, se han desarrollado distintos protocolos de infección asistida buscando emular las condiciones de infección natural para favorecer la detección de genotipos de comportamiento diferencial. Entre las distintas estrategias descritas pueden mencionarse la evaluación de la capacidad infectiva mediante inoculación directa con micelio del hongo, filtrado de cultivo de micelio, soluciones de ácido oxálico e inoculación con esporas del hongo, resultando esta última estrategia la más efectiva, por mostrar mayor similitud con la infección natural (Vear & Tourvieille de Labrouche 1988; Baldini *et al.* 2002). Al respecto, en la Estación Experimental INTA Balcarce se utiliza exitosamente un protocolo de inoculación asistida con esporas del hongo, que consiste en la modificación del método descrito por Vear *et al.* (1984) para su adaptación a las condiciones locales. Los resultados obtenidos bajo inoculación asistida han sido confrontados con condiciones naturales de infección, observándose una alta correlación entre ambas evaluaciones (Quiroz *et al.* 1998).

Es muy común en los ensayos a campo que las condiciones ambientales no sean homogéneas, de modo que otro punto a considerar para la optimización de los datos fenotípicos obtenidos es el uso de diseños experimentales que resulten eficientes en el campo, como por ejemplo la aplicación de bloques para una mejor estimación de los efectos de variabilidad introducidos por cualquier factor que afecte la variable respuesta (pendiente del terreno, distribución de equipos de riego, presencia de diferentes calidades de suelo, etc). La existencia de condiciones que den lugar a la delimitación de bloques, puede generar correlaciones entre observaciones del mismo ensayo. De este modo, el análisis estadístico de los datos se convierte también en un importante factor a considerar. Por lo general, resulta difícil utilizar los modelos lineales clásicos de análisis de varianza y regresión para este tipo de análisis, dado que no se cumplen los supuestos de independencia, normalidad, igualdad de varianzas o, incluso, linealidad. La modelación en el marco de los modelos mixtos brinda la posibilidad de analizar datos con estructuras de dependencia, desbalances y falta de normalidad. Es así posible relajar los supuestos tradicionales del modelo lineal

general y modelar, de manera flexible, estructuras complicadas de datos utilizando diversas estrategias que, bajo el mismo marco teórico, permiten modelar la variabilidad sobre, y más allá, de la componente usual de varianza asociada a los términos de error (Balzarini *et al.* 2010).

Finalmente, dado que los principales caracteres de importancia agronómica-forestal y económica (como el rendimiento, o la resistencia a enfermedades) están altamente influenciados por el ambiente, es necesario poner a prueba la presencia de interacciones Genotipo×Ambiente (G×A), *i.e.* la respuesta diferencial de diferentes genotipos a través de un rango de ambientes (Kang 1998).

La estimación de los niveles de enfermedad es un parámetro requerido en todo estudio epidemiológico. Como bien postuló Kranz (1988), "sin la cuantificación de la enfermedad no sería posible llevar a cabo estudios en epidemiología, la evaluación de pérdidas de cosechas y el desarrollo de aplicaciones para contrarrestar las enfermedades de las plantas". Las mediciones también son necesarias para caracterizar la dinámica temporal y espacial de la enfermedad y evaluar diversas tácticas de manejo, tales como el uso de fungicidas y cultivares resistentes. Mucho ha sido escrito acerca de los procedimientos para estimar los niveles de enfermedad en plantas (Madden & Campbell 1990), determinar su medida más precisa y certera (James & Teng 1979; Nutter *et al.* 1993) y analizar e interpretar los datos generados (Kranz 1990). Sin embargo, la identificación de las medidas fenotípicas que mejor describen a la enfermedad PHC en girasol aún continúa siendo materia de debate. Al momento, las determinaciones más utilizadas han sido: 1) incidencia de la enfermedad, que es el porcentaje (o proporción) de plantas enfermas en una muestra; 2) severidad, que refiere al porcentaje de la planta o tejido afectado por la enfermedad; y 3) período de incubación, que representa el tiempo hasta la aparición del primer síntoma. Estas medidas han sido utilizadas con éxito para la identificación de QTL asociados a la resistencia a PHC en girasol. En base a la severidad de la enfermedad, Mestries *et al.* (1998) y Rönicke *et al.* (2005) identificaron ocho QTL que explicaban entre el 10,6 y el 17,5 % de la varianza fenotípica. Combinando medidas de severidad y período de incubación, Bert *et al.* (2002; 2004) y Yue *et al.* (2008) lograron identificar 36 QTL relacionados con disminución de la enfermedad. Por su parte, mediante la exploración de los niveles de incidencia de la enfermedad Fusari *et al.* (2012), encontraron asociación significativa entre un alelo del gen HaRIC_B y menor enfermedad.

Dado el carácter cuantitativo de la resistencia, resulta de interés el estudio del incremento temporal de la enfermedad. Existen modelos que describen el aumento de la enfermedad en el tiempo (Berger & Luke 1979; Pennypacker 1980; Luke & Berger 1982; Thal & Campbell 1986; Subbarao *et al.* 1996) o en tiempo y espacio (Berger & Luke 1979; Jeger *et al.* 1983; Headrick & Pataky 1988; Damicone *et al.* 1990). El uso de estos datos de progreso de la enfermedad ha sido implementado mediante el cálculo del área bajo la curva de progreso de la enfermedad (AUDPC). Esta medida se ha utilizado para promediar la variación a menudo observada en las curvas de progreso de la enfermedad (Royle 1994), y también para integrar todos los aspectos del progreso de la enfermedad en relación con el desarrollo y crecimiento del hospedador. Si bien muchos estudios han sido reportados a la fecha utilizando el cálculo de AUDPC para la evaluación de resistencia de tipo cuantitativa entre plantas y patógenos fúngicos (Wilcoxson *et al.* 1975; Shaner & Finney 1977; Parker & Navia 1991; Das *et al.* 1992; Hartman & Wang 1993; Singh 1993; Ma *et al.* 1995; Birhman & Singh 1995; Broers *et al.* 1996; Conles *et al.* 2011), al momento no ha sido utilizado para describir el avance de la enfermedad PHC en girasol en estudios de mapeo de QTL.

En este capítulo se presenta la conformación de la Población de Mapeo por Asociación de INTA para caracteres complejos en girasol, y el análisis fenotípico del comportamiento de la misma, determinado como incidencia, severidad, período de incubación y AUDPC, tras campañas sucesivas en el campo utilizando infecciones asistidas con suspensiones de esporas de *S. sclerotiorum*, conjuntamente con la aplicación de métodos estadísticos robustos, que contemplan todos los factores involucrados en la evaluación experimental de la enfermedad.

MATERIALES Y MÉTODOS

Material vegetal

Del total de entradas albergadas en el BAG-IM, se seleccionó un conjunto de 137 líneas endocriadas, parte de las cuales ya contaban con datos sobre incidencia de infección frente a *S. sclerotiorum* en distintas campañas (2002-2003 hasta 2007-2008), y habían demostrado amplitud de respuesta fenotípica. Dado el interés en que esta población de mapeo por asociación se convirtiera en una herramienta para diversos estudios en girasol, no sólo resistencia a PHC, también se decidió la inclusión de materiales contrastantes para otros caracteres, destacándose un panel de líneas endocriadas provenientes del programa de resistencia a estrés hídrico que INTA lleva a cabo. Si bien los orígenes geográficos de las entradas coleccionadas en el Banco son diversos, entre ellos, Alemania Oriental, Argentina, Armenia, Australia, Bolivia, Brasil, Bulgaria, Canadá, Chile, China, España, Estados Unidos, Francia, Grecia, Hungría, Israel, Italia, Marruecos, Moldavia, Polonia, Rumania, Rusia, Siria, Turquía, Uruguay y Yugoslavia, se priorizó la inclusión de materiales argentinos, aunque materiales de otros países fueron también incluidos para aumentar la variabilidad. La información de *pedigree* de cada línea, su origen, característica principal e historia en programas de mejoramiento se detallan en el Apéndice (Tabla A1.1).

Semillas de las 137 entradas que conforman la PMA para caracteres complejos de INTA, fueron suministradas año a año por el BAG-IM para la realización de los ensayos fenotípicos a campo.

Caracterización de las líneas que componen la PMA

Con el fin de caracterizar en profundidad la PMA, se realizaron observaciones a campo de las características morfológicas de las líneas que la componen. Se consideraron las siguientes características: número de capítulos (línea multiflora o uniflora), forma de hojas (aserrada, acorazonada), color de nervaduras (verdes, violáceas), color de las flores liguladas (amarillo pálido, rojizo, amarillo fuerte), color del receptáculo (marrón claro u oscuro).

Preparación del inóculo

Esclerocios derivados de una población natural de *S. sclerotiorum* fueron colectados en la localidad de Balcarce (Buenos Aires, Argentina) y almacenados en bolsas de papel a 13 °C por tres meses. La producción de apotecios a partir de los mismos se realizó según el procedimiento descrito por Prioletta *et al.* (1995). Brevemente, los esclerocios previamente expuestos a -18 °C ± 2 °C durante siete días fueron enterrados a 1 cm de profundidad en suelo pasteurizado, en condiciones de oscuridad y riego diario, hasta su emergencia. Posteriormente, se procedió a la incubación de los mismos a 16 °C y a 2500 lux de luz continua. Los apotecios maduros se colectaron hacia arriba en una caja de Petri y se incubaron durante 4 h para favorecer la generación de ascosporas. Finalmente, las ascosporas se reservaron a -18 °C hasta su uso. Por cada fecha de inoculación, se preparó inóculo fresco a una concentración de 2500 ascosporas/mL en agua bidestilada estéril. El conteo de esporas se realizó utilizando una cámara de Neubauer.

Tratamiento pre-siembra

Para asegurar un alto porcentaje de germinación, las semillas de las 137 líneas que conforman la PMA fueron incubadas durante 60 minutos en una solución 100 ppm de ácido giberélico, seguido por almacenamiento a 10 °C, siguiendo las recomendaciones de Pereyra *et al.* (2002) para ruptura de dormición en semillas.

Dada la alta incidencia de *Plasmopara halstedii*, agente causal del mildiu o enanismo de girasol, en el suroeste bonaerense, las semillas fueron tratadas con el fungicida Metalaxil M® (METALAXIL-M APRON GOLD, SYNGENTA AGRO S.A., Argentina), siguiendo la dosis recomendada por la empresa (600 cm³/100 kg de semillas).

Siembra

Los ensayos a campo se realizaron en la EEA INTA Balcarce (37° 50' 0" S, 58° 15' 33" W, Buenos Aires, Argentina) en cinco campañas sucesivas: 2009-2010, 2010-2011, 2011-2012, 2012-2013 y 2013-2014. La siembra se realizó en el campo experimental durante la primera semana de diciembre, de acuerdo a un diseño de dos bloques completos aleatorizados (DBCA), en dos ambientes: seco (campañas 2009-2010, 2010-2011, 2011-2012) y con

riego por aspersión (campañas 2012-2013 y 2013-2014), en un suelo Arguidol típico conteniendo 5 % de materia orgánica a pH 6,2. Los surcos (8 m de longitud) conteniendo 32 plantas por línea de girasol se espaciaron entre sí 70 cm, con una distancia entre bloques de 1 m. El bloqueo se realizó tomando en consideración la pendiente del terreno, y en el caso de tratamiento bajo riego, la disposición del sistema de riego. Las aleatorizaciones se realizaron utilizando el programa SAS (SAS Institute, 2001).

Debido a que la enfermedad es altamente dependiente de las condiciones ambientales, y dado que las inoculaciones con esporas del hongo se realizaron en distintas etapas dentro de la misma campaña, es decir, de acuerdo a la fecha en la cual el 50 % del surco a inocular se encontraba en estadio R5.2 de la escala de Schneiter y Miller (1981), fue necesario realizar controles de la fecha de inoculación. Así, con el fin de evaluar la variabilidad inherente al lote y, a modo de control de infección, en cada campaña se sembró, además de las líneas que componen la PMA, una dupla de híbridos testigo de comportamiento conocido ante la enfermedad: DK3820 (moderadamente resistente) y ACA885 (susceptible). Estos híbridos comerciales fueron elegidos por su comportamiento contrastante y constante frente a la enfermedad, luego de repetidas mediciones realizadas a campo con inoculación asistida con el hongo por la Red Nacional de Ensayos de Cultivares Comerciales de Girasol que conduce el INTA, publicadas en la página de la Asociación Argentina de Girasol (ASAGIR) (www.asagir.org.ar). Para evitar la competencia con las líneas, estos híbridos fueron sembrados en un lote contiguo al ensayo de asociación, en fechas de siembra escalonadas (4 fechas anteriores a la siembra de la PMA, y 4 fechas posteriores), para asegurar la coincidencia de floración de los materiales control con la de las distintas líneas que conforman la PMA.

Adicionalmente, y para corroborar la homogeneidad genotípica de las líneas sembradas, se tomaron muestras de hoja en plantas de tres semanas de edad en todas las campañas (estos datos se analizarán en el Capítulo 4 de esta Tesis).

Inoculación y registro de enfermedad

Como se mencionó previamente, la inoculación con la suspensión de ascosporas del hongo se llevó a cabo cuando al menos el 50% del surco se encontraba en estadio R5.2 de la escala de Schneider y Miller (1981), el cual corresponde al momento en que las dos líneas de flores

externas se encuentran en antesis. Estudios previos han demostrado que éste es el estadio de mayor susceptibilidad de la planta al ataque por el hongo (Troglia 2003). Acto seguido, se cubrió cada capítulo inoculado con bolsa de papel manila (EMBAIRES S.A., Buenos Aires, Argentina), hasta los 10 días post inoculación (dpi). En aquellos ensayos realizados bajo riego, el mismo fue realizado desde el momento de la inoculación hasta el final de las observaciones, 15 minutos por día, en el momento de máxima temperatura. Las evaluaciones fenotípicas fueron realizadas los días 10, 14, 17, 21, 24 y 28 post inoculación, registrándose tres medidas fenotípicas relacionadas con la infección por *S. sclerotiorum*: 1) **Incidencia de PHC** que se obtiene como el número de plantas enfermas sobre el número de plantas totales inoculadas de cada línea; 2) **Severidad de la enfermedad**, o porcentaje del capítulo enfermo, promediado por el número de plantas enfermas del surco; y 3) **Período de incubación**, correspondiente al número de días desde la inoculación hasta la aparición del primer síntoma, determinado por planta, promediado por el número total de plantas enfermas. En la Figura 1.1 se presenta un diagrama de la conducción del ensayo, y una vista panorámica del mismo. La lista completa de datos fenotípicos se encuentra en http://bioinformatica.inta.gov.ar/download_files/tesis_filippi/



Figura 1.1 Ensayo fenotípico para el carácter resistencia a PHC (A) Esquema del desarrollo del ensayo fenotípico para la enfermedad PHC en girasol; (B) Vista panorámica del frente del ensayo (campaña 2013/14). Pueden observarse algunas líneas inoculadas y tapadas con bolsas de papel manila.

Análisis estadístico

El análisis estadístico de los datos se realizó utilizando modelos lineales mixtos (MLM) y modelos lineales generalizados (GLM) según se detalla a continuación. En ambos casos se emplearon los módulos incorporados en el programa Infostat (Di Rienzo *et al.* 2008).

Se evaluó incidencia máxima (IEmax) como número de plantas enfermas a tiempo final sobre número de plantas totales inoculadas por surco. Dado que estos datos son de tipo binomial (presencia/ausencia de enfermedad), el análisis se realizó utilizando modelos lineales generalizados mixtos, seleccionando familia “Binomial” con función de enlace “logit”, utilizando como covariable el número de plantas inoculadas por línea. En el primer caso, se consideró un modelo en el que las “líneas” fueron contempladas como efecto fijo, y “campaña”, “bloque dentro de campaña” y “fecha de inoculación” como efectos aleatorios, de acuerdo al modelo [1.1]:

$$Y_{ijkm} = \mu + L_i + C_j + B(C)_{k(j)} + F_m + \varepsilon_{ijkm} \quad [1.1]$$

donde y_{ijkm} es la incidencia (proporción) de la línea (genotipo) i , en la campaña j ; μ es la media general; L_i es el efecto fijo de la línea i con $i = 1-137$; C_j es el efecto aleatorio de la campaña j con $j = 1-5$; $B(C)_{k(j)}$ es el efecto aleatorio del bloque k dentro de la campaña j ; F_m es el efecto aleatorio de la fecha de inoculación con $m = 1-48$; y ε_{ijkm} es el término de error aleatorio asociado a la observación y_{ijkm} .

A fin de evaluar la existencia de interacción GxA, se consideró un segundo modelo [1.2] en el cual el ambiente estuvo representado por las distintas fechas de inoculación, de acuerdo a:

$$Y_{ijkm} = \mu + L_i + C_j + B(C)_{k(j)} + F_m + LF_{(im)} + \varepsilon_{ijkm} \quad [1.2]$$

donde y_{ijkm} es la incidencia (proporción) de la línea (genotipo) i , en la campaña j ; μ es la media general; L_i es el efecto fijo de la línea i con $i = 1-137$; C_j es el efecto aleatorio de la campaña j con $j = 1-5$; $B(C)_{k(j)}$ es el efecto aleatorio del bloque k dentro de la campaña j ; F_m es el efecto aleatorio de la fecha de inoculación con $m = 1-48$; $LF_{(im)}$ es el efecto aleatorio de la línea i con la fecha de inoculación j ; y ε_{ijkm} es el término de error aleatorio asociado a la observación y_{ijkm} .

Finalmente, con el fin ulterior de cuantificar la contribución del efecto “genotipo” a la variabilidad, se consideró un tercer modelo [1.3] en el cual todos los componentes fueron considerados aleatorios, de acuerdo a:

$$y_{ijkm} = \mu + L_i + C_j + B(C)_{k(j)} + F_m + LF_{(im)} + \epsilon_{ijkm} \quad [1.3]$$

donde y_{ijkm} es la incidencia (proporción) de la línea (genotipo) i , en la campaña j ; μ es la media general; L_i es el efecto aleatorio de la línea i con $i = 1-137$; C_j es el efecto aleatorio de la campaña j con $j = 1-5$; $B(C)_{k(j)}$ es el efecto aleatorio del bloque k dentro de la campaña j ; F_m es el efecto aleatorio de la fecha de inoculación con $m = 1-48$; $LF_{(im)}$ es el efecto aleatorio de la línea i con la fecha de inoculación j ; y ϵ_{ijkm} es el término de error aleatorio asociado a la observación y_{ijkm} .

Al ajustar distintos modelos a un mismo conjunto de datos, es necesario utilizar criterios para la comparación de los ajustes y, por tanto, para la selección de un modelo. Los indicadores utilizados fueron el criterio de información de Akaike (AIC) y el criterio de información bayesiano (BIC). Estos se definen como:

$$AIC = 2k - 2\ln(L)$$

$$BIC = -2 \ln L + k \ln(n)$$

donde k es el número de parámetros en el modelo estadístico, y L es el máximo valor de la función de verosimilitud para el modelo estimado (Akaike 1974). Bajo estas expresiones de AIC y BIC, el mejor modelo resulta ser aquel con menor valor para el indicador.

Con la totalidad de los datos registrados de incidencia, se realizó el cálculo de AUDPC (*Area Under the Disease Progress Curve* o área debajo de la curva de progreso de la enfermedad), mediante la fórmula [1.4] descrita por Shaner y Finney (1977), para obtener un solo dato por línea que describa la enfermedad:

$$AUDPC = \sum((X_{i+1} + X_i)/2)(T_{i+1} - T_i) \quad [1.4]$$

Donde X_i es la proporción de plantas enfermas en la observación i , $(T_{i+1} - T_i)$ corresponde al tiempo (días) entre dos observaciones.

Con esta medida fenotípica, se analizaron los mismos tres modelos descritos anteriormente (1.1, 1.2 y 1.3). Dado que los datos de AUDPC son áreas, una medida continua, los datos se analizaron como normales, previa observación de que la distribución de los residuos se ajustara al modelo.

Tal como se describió en la sección anterior, la severidad de la enfermedad hace referencia al porcentaje del capítulo enfermo, promediado por el número de plantas enfermas del surco. De esto se desprende que la severidad igual a cero no está definida, de modo que para el análisis de este carácter se contó con menos datos que para el carácter “incidencia”, ya que debieron ser descartadas del análisis aquellas líneas que no presentaron enfermedad, es decir, aquellas con incidencia igual a cero. Dado que los datos reflejan proporciones, se realizó previamente una transformación angular de los mismos, usando la función arcoseno de la raíz cuadrada de “x”, para poder analizarlos como normales. Nuevamente, se probaron para este carácter los mismos tres modelos: 1.1, 1.2 y 1.3.

Para el caso de período de incubación (Pinc), por la naturaleza de la variable (días hasta que aparece el primer síntoma en cada planta enferma del surco), para la obtención de las medias ajustadas de cada línea que conforma la PMA se recurrió a un modelo de distribución gamma y función de enlace “inversa”. Este modelo es una generalización del modelo Exponencial, utilizado para modelar variables que describen *el tiempo hasta que se produce p veces un determinado suceso*. Nuevamente, se probaron para este carácter los mismos tres modelos: 1.1, 1.2 y 1.3.

En todos los casos, se consideraron los datos fenotípicos obtenidos en las cinco campañas realizadas a campo desde el año 2009 hasta 2014.

Se testeó la existencia y magnitud de correlación entre las variables fenotípicas registradas, utilizando el método de Spearman implementado en el *software* Infogen (Balzarini & Di Rienzo 2011).

Las gráficas de relación respecto a la media (rrm) y efecto del ambiente fueron realizadas con el *software* Infogen (Balzarini & Di Rienzo 2011).

RESULTADOS

Constitución de la población de mapeo por asociación

Para la constitución de la PMA para caracteres complejos de INTA, se tomó como punto de partida la población de mapeo utilizada por Fusari *et al.* (2012) constituida por 94 entradas de girasol, 84 pertenecientes al programa de mejoramiento de INTA y 10 líneas públicas de interés por ser parentales de diversas poblaciones de mapeo. De estos 94 materiales, se descartaron 34 por no ser líneas de endocría avanzada. Los 60 materiales que se conservaron de la población inicial de Fusari *et al.* (2012) estaban constituidos por 50 líneas pertenecientes al programa de mejoramiento de INTA y 10 líneas públicas. Dada la necesidad de generar una PMA que sea una herramienta de utilidad para múltiples aproximaciones, se incluyeron 54 líneas provenientes del programa de mejoramiento para tolerancia a déficit hídrico en girasol que INTA lleva a cabo, 7 líneas públicas de orígenes diversos, de importancia en programas de mejoramiento internacionales, y un panel de 16 líneas endocriadas de valioso aporte a la diversidad. Cabe destacarse la constitución gradual de la PMA, que fue conformada en respuesta a las necesidades de los programas de mejoramiento locales. De este modo, en las campañas 2009-10 y 2010-11 se evaluaron 69 materiales, en la campaña 2011-12 el número de materiales evaluados se incrementó a 112 y recién en las campañas 2012-13 y 2013-14 se evaluó la totalidad de las líneas que componen la PMA definitiva de INTA, es decir, 137 líneas endocriadas.

La inspección visual de características morfológicas como número de capítulos, forma de hojas, color de nervaduras, de flores liguladas y de receptáculo reveló la existencia de variación fenotípica entre las líneas. Del conjunto total, 77 fueron multiflora, mientras que las 60 restantes uniflora.

Se observaron diferencias en la coloración de las flores liguladas, desde amarillo pálido, a rojo, y en el receptáculo, de marrón claro a oscuro. La forma de las hojas también fue variable, entre acorazonadas, borde aserrado, etc, al igual que las nervaduras (claras, violáceas) (Figura 1.2, A-F).



Figura 1.2 Características morfológicas de las líneas de la PMA Algunas variaciones detectadas en forma y color en capítulo y hojas. (A) Lígulas amarillo pálido, receptáculo oscuro; (B) Lígulas rojizas, receptáculo oscuro; (C) Lígulas amarillas, receptáculo claro; (D) Hoja aserrada, nervaduras claras; (E) Hoja acorazonada, nervaduras violáceas; (F) Línea uniflora (izquierda) vs. línea multiflora (derecha).

Características generales de los ensayos de evaluación de PHC a campo

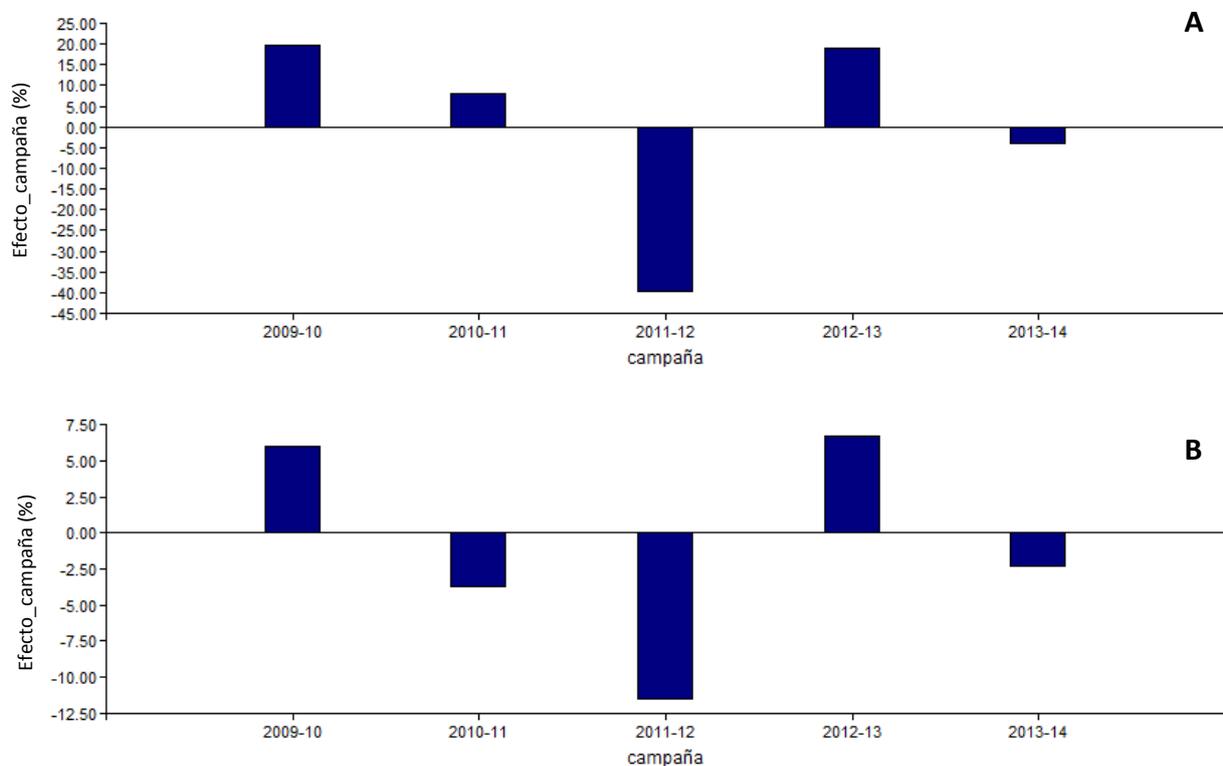
Del total de cinco ensayos a campo realizados durante los meses de diciembre-marzo de 2009/10, 2010/11, 2011/12, 2012/13 y 2013/14, los tres primeros se hicieron en condición de secano y los dos últimos bajo riego, a fin de aumentar los ambientes para la evaluación fenotípica. Las condiciones de temperatura y humedad difirieron año a año, siendo la campaña 2011/12 la de mayor temperatura máxima y menor humedad, y la de 2012/13 la de mayor humedad, de acuerdo a los registros de la estación meteorológica de la EEA INTA Balcarce (Tabla 1.1).

Tabla 1.1 Medias de humedad relativa promedio, temperatura máxima y mínima promedio para las fechas de inoculación, por campaña

Campaña	Humedad (%)	T° máxima (°C)	T° mínima (°C)
2009-10	73,67	24,00	13,74
2010-11	76,17	25,08	14,59
2011-12	60,88	31,45	15,63
2012-13	78,11	26,73	15,27
2013-14	62,75	28,59	13,84

Estas diferencias climáticas se relacionaron directamente con los niveles de enfermedad obtenidos, siendo las campañas 2011/12 y 2013/14 las de menor incidencia y severidad debido al arresto de la enfermedad en etapas tempranas, y las campañas 2009/10 y 2012/13 las de mayor incidencia y severidad (Figura 1.3, A y B).

Del mismo modo, al analizar los valores de período de incubación por línea, se observa que en los años más desfavorables para el desarrollo de la enfermedad (2011/12 y 2013/14) las líneas mostraron períodos de incubación mayores, mientras que en los años más favorables para el desarrollo de la enfermedad, los períodos de incubación se acortaron. Estas observaciones sugieren una gran dependencia de estas medidas fenotípicas con el ambiente (Figura 1.3, C).



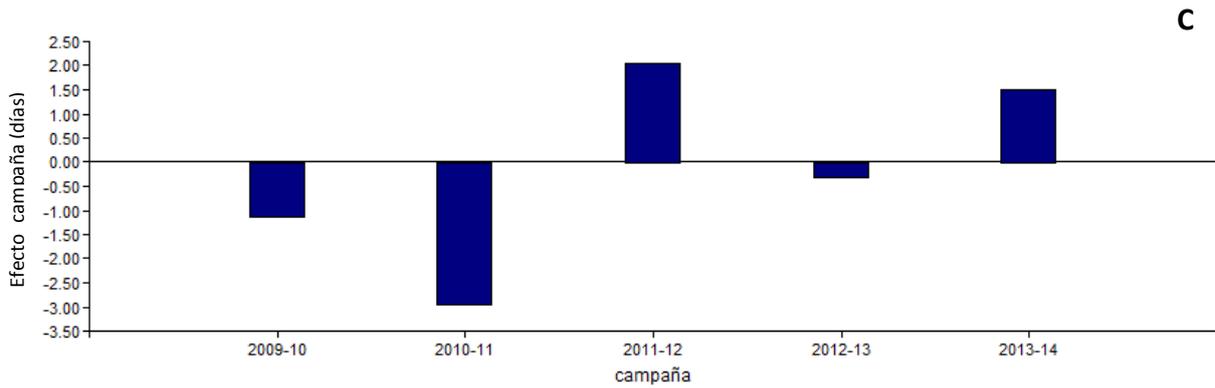


Figura 1.3 Niveles de enfermedad observados para cinco campañas de evaluación de PHC a campo (A) Incidencia; (B) Severidad; (C) Período de incubación. La línea central representa la media global (considerando las incidencias/severidades/períodos de incubación de todos los genotipos en todas las campañas). Las barras de valor positivo indican aquellos años en los que la media anual fue superior a la media global, mientras que las barras de valor negativo corresponden a aquellos años en los que la media anual fue menor a la global.

El número de plantas inoculadas y evaluadas por línea fue variable, desde 2 plantas hasta más de 36, con un promedio general para todos los ensayos de 15 plantas por surco (Figura 1.4). Las dos primeras campañas fueron las más críticas en cuanto al número de plantas obtenidas, la primera debido a un bajo índice de germinación, que fue resuelto en campañas sucesivas realizando a las semillas un tratamiento previo con ácido giberélico para romper dormición, y la segunda porque las condiciones del terreno (suelo compactado, exceso de malezas) no permitieron la emergencia del total de líneas. La ventana de fechas de inoculación fue de aproximadamente tres semanas, determinada por la fecha de floración de las líneas tempranas y tardías en la evaluación.

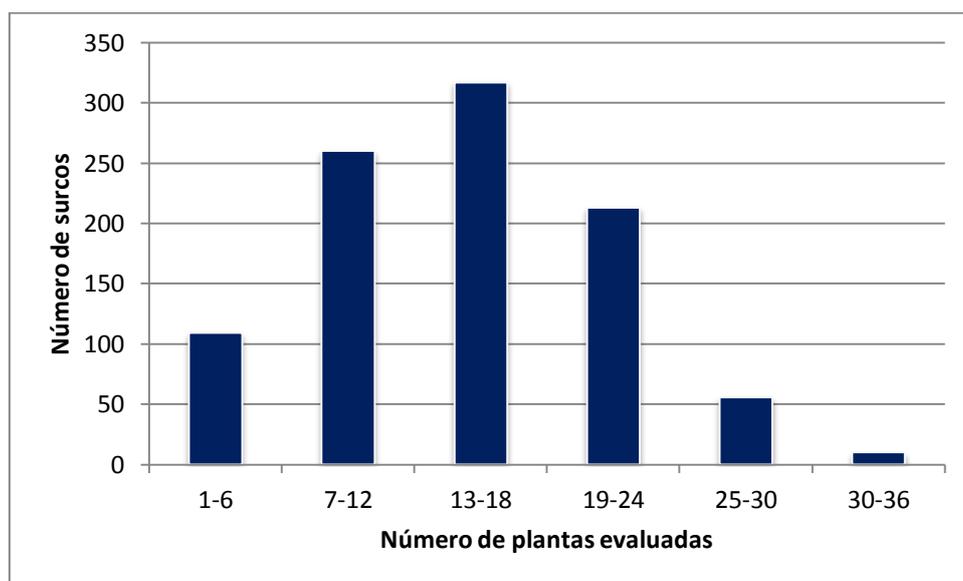


Figura 1.4 Distribución del número de plantas evaluadas por surco para las 5 campañas.

Dado que las condiciones del ensayo no pueden ser rigurosamente controladas, debido a que se realiza a campo, se registraron otras enfermedades fúngicas en girasol, como marchitez anticipada causada por *Verticillium dahliae* (con niveles de daño que alcanzaron en algunos casos el 100 % de enfermedad), roya blanca, alternaria, entre otras. Dependiendo del grado de daño alcanzado, no se pudo continuar con la evaluación fenotípica de la PHC en algunas líneas (Figura 1.5, A-C).



Figura 1.5 Otras enfermedades fúngicas detectadas por infección natural en el ensayo de MA (A) Roya blanca; (B) Mildiu o enanismo; (C) Marchitez producida por *Verticillium*.

El tratamiento realizado con el curasemilla Metalaxil M® para la prevención de mildiu fue efectivo en la gran mayoría de los casos, aunque se detectaron escapes, observándose plantas enanas, con el capítulo orientado hacia arriba en algunos surcos.

Dada la amplia campana de fechas de inoculación, producto de la diferencia de las líneas endocriadas en tiempos a floración, se considera útil contar con controles de comportamiento contrastante ante la enfermedad (susceptibles y moderadamente resistentes), a fin de corroborar la efectividad de la fecha de inoculación. Los híbridos comerciales ACA885 (susceptible) y DK3820 (medianamente resistente), sembrados como controles de ensayo, se comportaron en todas las campañas de acuerdo a lo esperado. Sin embargo, pese a realizar una siembra escalonada en el tiempo de los distintos materiales control para garantizar la presencia de híbridos susceptibles y resistentes en cada fecha de inoculación de las líneas de la PMA, esto no siempre pudo ser posible.

Incidencia de PHC

Como se describió en el apartado “Materiales y Métodos”, la medida “incidencia” hace referencia al número de plantas enfermas respecto del total de plantas inoculadas. Si bien se realizaron distintas medidas de incidencia, desde el día 10 post inoculación (que coincidía con el destape de las plantas) hasta el día 28 post inoculación, una de las medidas fenotípicas de mayor importancia a nivel agronómico es la Incidencia Máxima de la enfermedad (IEmax), ya que es la que da idea del comportamiento de la línea frente a la enfermedad en el momento de la cosecha de girasol. En la Figura 1.6 se presentan las curvas de progreso de la enfermedad para cuatro líneas de comportamiento contrastante frente a la enfermedad (dos susceptibles y dos moderadamente resistentes). En ellas, es posible distinguir una etapa previa de latencia, coincidente con los primeros puntos de la evaluación, seguida por una etapa de crecimiento exponencial de la enfermedad, y finalmente una etapa final estable, en donde ya no se observaba más aumento en la incidencia, punto denominado IEmax.

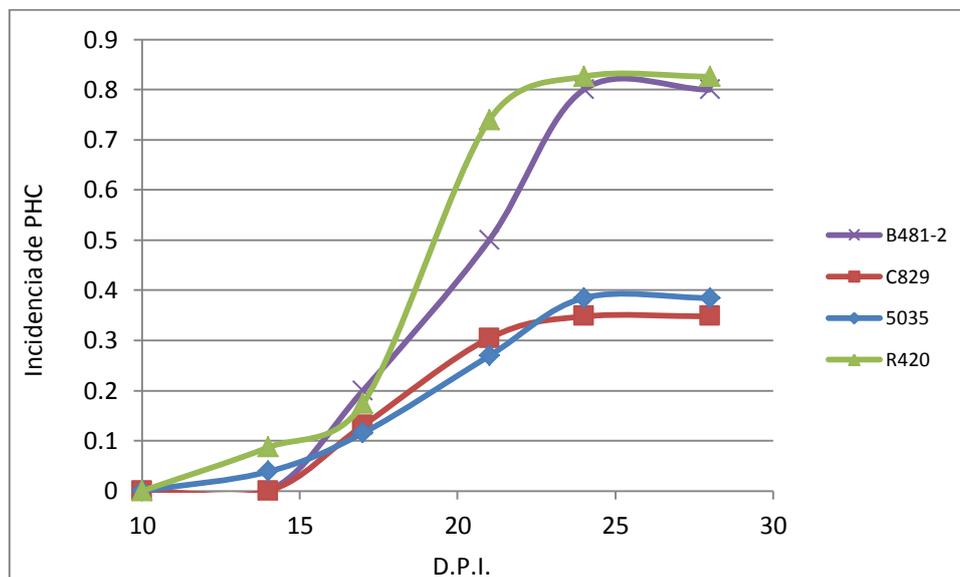


Figura 1.6 Curvas de progreso de la enfermedad para dos líneas moderadamente resistentes (5035, azul; C829, rojo), y dos líneas susceptibles (B481-2, violeta; R420, verde). Datos tomados de la campaña 2013/14.

Los valores de IEmax observados para el total de las líneas evaluadas en todos los años variaron entre 0 (ninguna planta enferma en el surco) y 100 % (todas las plantas del surco con síntomas de enfermedad). A fin de estudiar el comportamiento por línea y por año respecto a la enfermedad, se construyeron gráficos de relación respecto a la media (rrm),

los cuales se presentan en la Figura 1.7 (A-E). En estos gráficos, si la media de incidencia de una línea es igual a uno, quiere decir que se comportó como el promedio de la población. Si es mayor, se trata de una línea susceptible. Si por el contrario es menor, se trata de una línea moderadamente resistente. Se observa que, si bien los valores absolutos de incidencia por línea mostraron fluctuaciones año a año, sus rrm tienden a mantenerse constante en una alta proporción de las líneas.

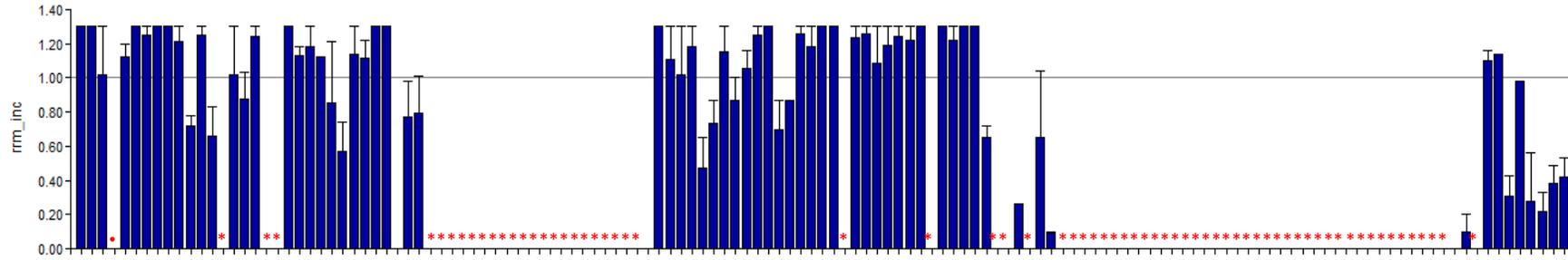
Para poder analizar en conjunto todos los datos de incidencia obtenidos a lo largo de las cinco campañas fenotípicas, se testeó la *performance* de distintos modelos. En todos ellos, se tuvo en consideración el número total de plantas inoculadas, el número de plantas enfermas, la campaña (o ambiente), la fecha de inoculación y la repetición o bloque (dos repeticiones por año, anidadas en la campaña).

El primer modelo testeado fue aquel en el que las líneas (o genotipos) fueron considerados como efecto “fijo”, y no se consideró el efecto de la interacción GxA (modelo 1.1). En este caso, el modelo tuvo un AIC = 4780,91 y un valor de BIC = 5463,00. El efecto de las “líneas” fue altamente significativo ($p < 0,0001$). Para los efectos “aleatorios”, la “campaña” y la “fecha de inoculación” fueron los parámetros que mayor parte de la varianza explicaban, y la “repetición” el que menos (Apéndice, Tabla A1.2). Al realizar la prueba de comparaciones múltiples DGC (Di Rienzo *et al.* 2008), utilizando un nivel de significación del 0,05, pudieron diferenciarse tres grandes grupos.

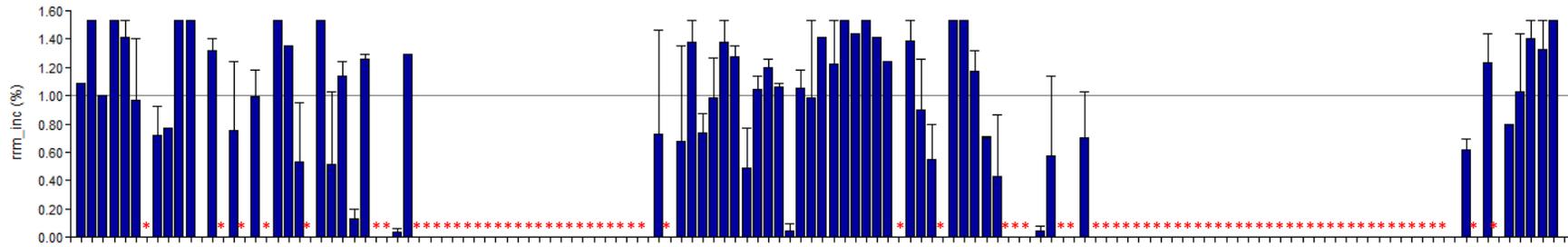
El segundo modelo evaluado, también consideró a las líneas (genotipos) como efecto “fijo”, pero esta vez sí se tuvo en cuenta el efecto de la interacción aleatoria GxA (modelo 1.2). Este modelo tuvo un valor de AIC = 4303,88 y BIC = 4990,85, un poco por debajo que los obtenidos para el modelo anterior, donde no se consideró la interacción.

Al igual que en el caso anterior, la mayor parte de la varianza estaba explicada por la “campaña”, seguida de la “fecha de inoculación”, y la menor medida por la “repetición” (Apéndice, Tabla A1.3). La interacción ($p < 0,05$) se llevó gran parte de la varianza, estimada en 18,83 %. Al realizar la prueba de comparaciones múltiples DGC (Di Rienzo *et al.* 2008), utilizando un nivel de significación del 0,05, pudieron diferenciarse dos grandes grupos.

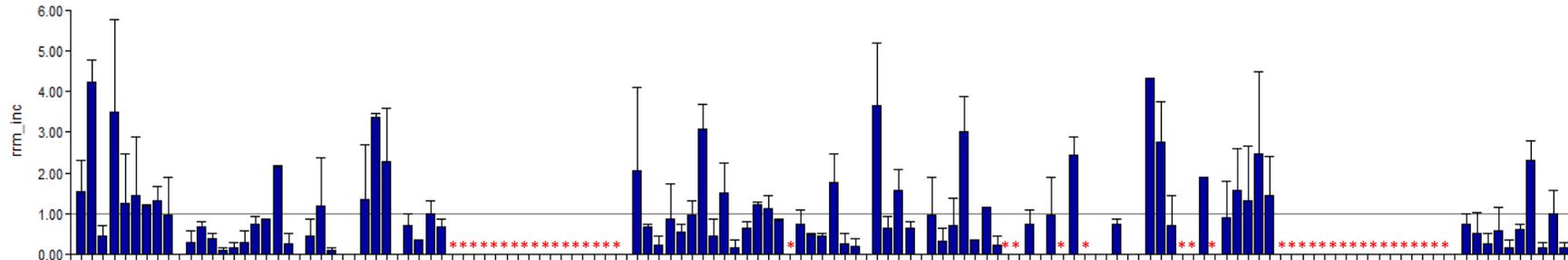
A



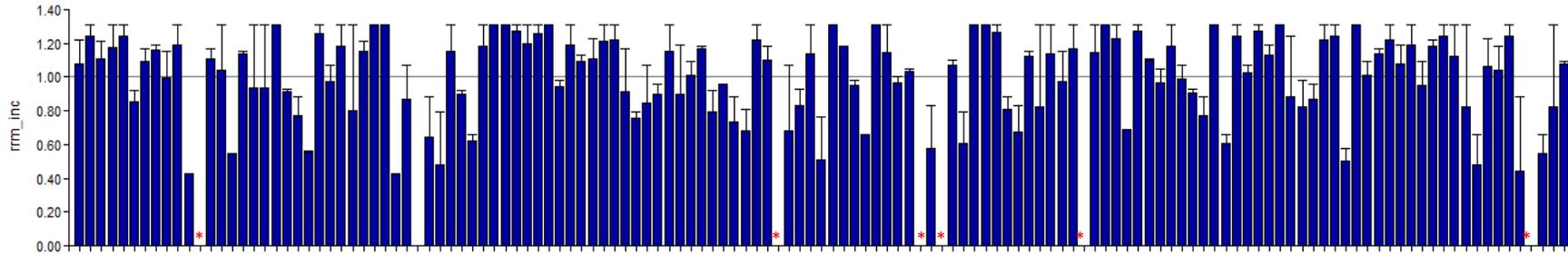
B



C



D



E

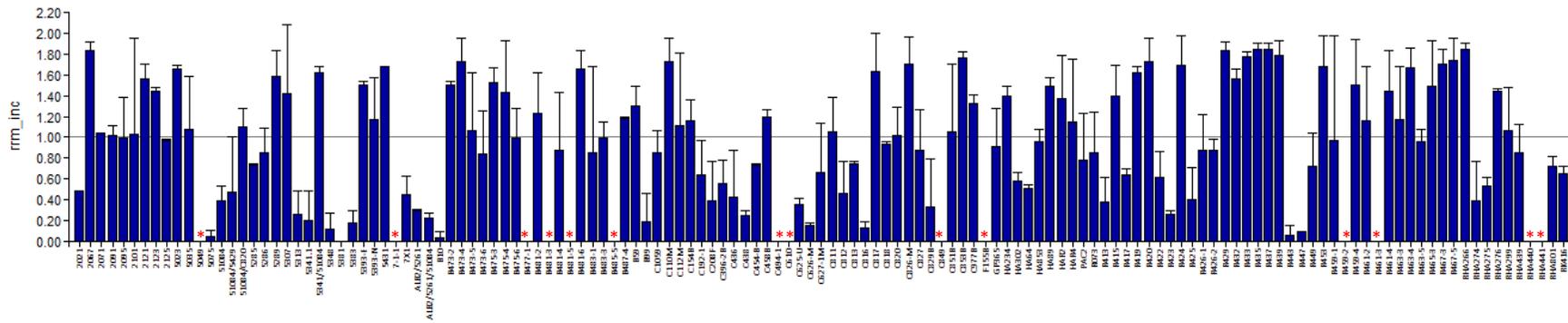


Figura 1.7 Relación respecto a la media por línea y por campaña para la variable Incidencia de PHC (A) Campaña 2009-10; (B) Campaña 2010-11; (C) Campaña 2011-12; (D) Campaña 2012-13; (E) Campaña 2013-14. Las barras representan el error estándar. Los puntos rojos indican ausencia de dato. Se graficó el promedio de los dos bloques por línea por campaña.

Finalmente, y con el objetivo de cuantificar la contribución del efecto de las “líneas” a la varianza total, se evaluó un tercer modelo en el que las líneas se consideraron efectos “aleatorios”, contemplando también la interacción GxA (modelo 1.3). Este modelo tuvo un AIC = 4330,74 y BIC = 4359,97, siendo el valor de AIC mayor al obtenido para el modelo anterior, mientras que el BIC fue un poco menor. La contribución de las “líneas” a la varianza total, fue estimada en 13 % (Apéndice, Tabla A1. 4).

Las medias ajustadas por línea obtenidas según el segundo modelo testeado se muestran en el Apéndice, Tabla A1.14. Con estos valores, se construyó un histograma de distribución de medias (Figura 1.8).

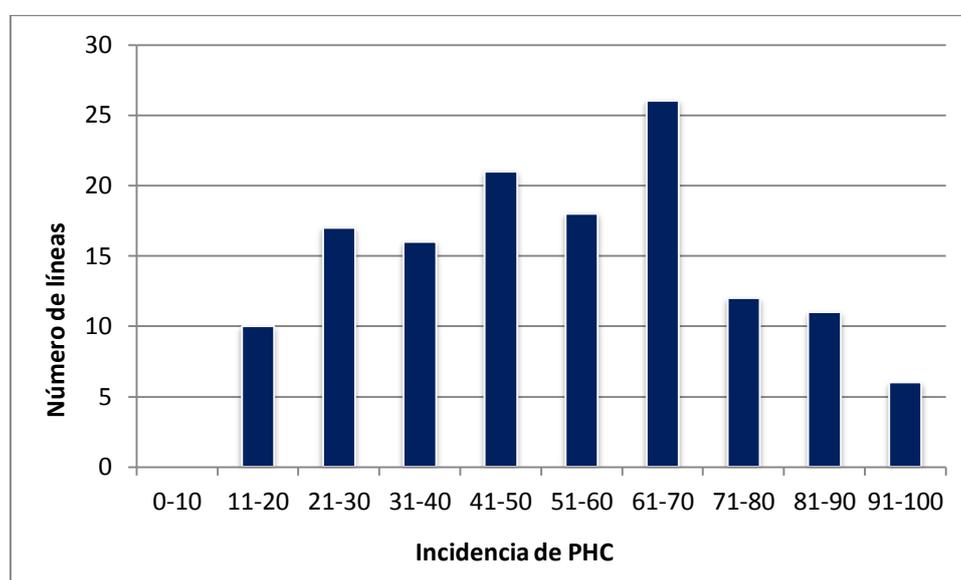


Figura 1.8 Histograma de medias ajustadas para IEmax a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente

AUDPC Incidencia

Las estimas de incidencia a día 14, 17, 21, 24 y 28 post inoculación fueron utilizadas para calcular el área bajo la curva de progreso de la enfermedad (AUDPCinc), utilizando la regla trapezoidal, de modo de tener una medida numérica que describa más ampliamente el desarrollo de la enfermedad.

A modo de ejemplo, se graficaron las AUDPC acumuladas para las cuatro líneas endocriadas de comportamiento contrastante frente a la enfermedad PHC presentadas en la Figura 1.6. Se observa que aquellas líneas que mostraron menor IEmax (rojo y azul) también muestran

menor AUDPC, demostrando una correspondencia entre ambas medidas fenotípicas (Figura 1.9).

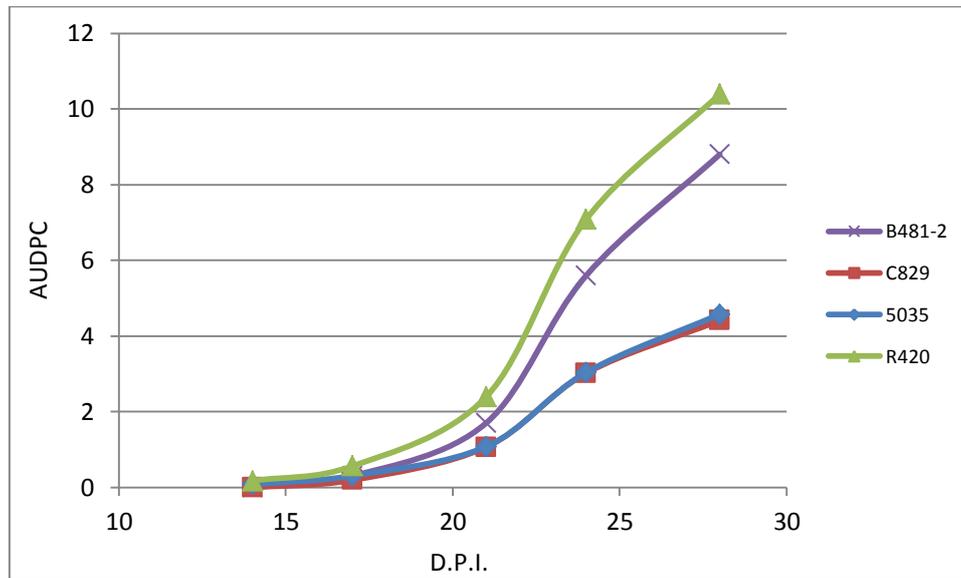


Figura 1.9 Área bajo la curva de progreso de la enfermedad (incidencia) para dos líneas moderadamente resistentes (5035, azul; C829, rojo), y dos líneas susceptibles (B481-2, violeta; R420, verde). Datos tomados de la campaña 2013/14.

Del total de datos obtenidos, se descartaron del análisis aquellas líneas que no tuvieron el total de las observaciones fenotípicas, correspondientes a los días 14, 17, 21, 24 y 28 dpi (a fin de hacer una estima de AUDPC sobre la misma cantidad de puntos), quedando un total de 899 datos que pudieron ser analizados para todas las líneas.

Al analizar el conjunto global de datos fenotípicos, se observa que los valores de AUDPC estimados estuvieron en el rango de 0 (correspondiente a aquellos casos en los que no se observó enfermedad en el surco) a 14 (correspondiente a aquellos casos en los que ya desde el día 14 post inoculación todas las plantas del surco presentaron síntomas de enfermedad).

Al igual que para la variable IEmax, para poder analizar en conjunto todos los datos de AUDPC obtenidos a lo largo de las cinco campañas fenotípicas, se evaluó la *performance* de distintos modelos. En todos ellos, se tuvo en consideración el efecto de la línea (o genotipo), la campaña (o ambiente), la fecha de inoculación y la repetición/bloque (dos repeticiones por año, anidadas en la campaña).

El primer modelo evaluado fue aquel en el que las líneas (o genotipos) fueron considerados como efecto “fijo”, y no se consideró el efecto de la interacción GxA (modelo 1.1). En este

caso, el modelo tuvo un AIC = 4189,82 y un valor de BIC = 4866,80. El efecto de las “líneas” fue altamente significativo ($p < 0,0001$). Para los efectos “aleatorios”, la “campaña” y la “fecha de inoculación” fueron los parámetros que mayor parte de la varianza explicaban, y la “repetición” el que menos (Apéndice, Tabla A1.5). Al realizar la prueba de comparaciones múltiples DGC (Di Rienzo, Guzmán y Casanoves, 2001), utilizando un nivel de significación del 0,05, pudieron diferenciarse dos grandes grupos.

El segundo modelo evaluado, también consideró a las líneas (genotipos) como efecto “fijo”, pero esta vez sí se tuvo en cuenta el efecto de la interacción aleatoria GxA (modelo 1.2). Este modelo tuvo un valor de AIC = 4115,05 y BIC = 4796,84, un poco por debajo de los obtenidos para el modelo anterior, donde no se consideró la interacción.

Al igual que en el caso anterior, la mayor parte de la varianza estuvo explicada por la “campaña”, seguida de la “fecha de inoculación”, y en menor medida por la “repetición” (Apéndice, Tabla A1.6). La interacción ($p < 0,05$) explicó un 23,17 % de la varianza total.

Finalmente, y con el objetivo de cuantificar la contribución del efecto de las “líneas” a la varianza total, se evaluó un tercer modelo en el que las líneas se consideraron efectos “aleatorios”, contemplando también la interacción GxA (modelo 1.3). Este modelo tuvo un AIC = 4438,39 y BIC = 4472,00, ambos valores mayores a los obtenidos para los modelos anteriores. La contribución de las “líneas” a la varianza total, fue estimada en 14,43% (Apéndice, Tabla A1.7).

Las medias ajustadas por línea obtenidas según el segundo modelo testeado se presentan en el Apéndice, Tabla A1.14. Un histograma de distribución de medias se presenta en la Figura 1.10.

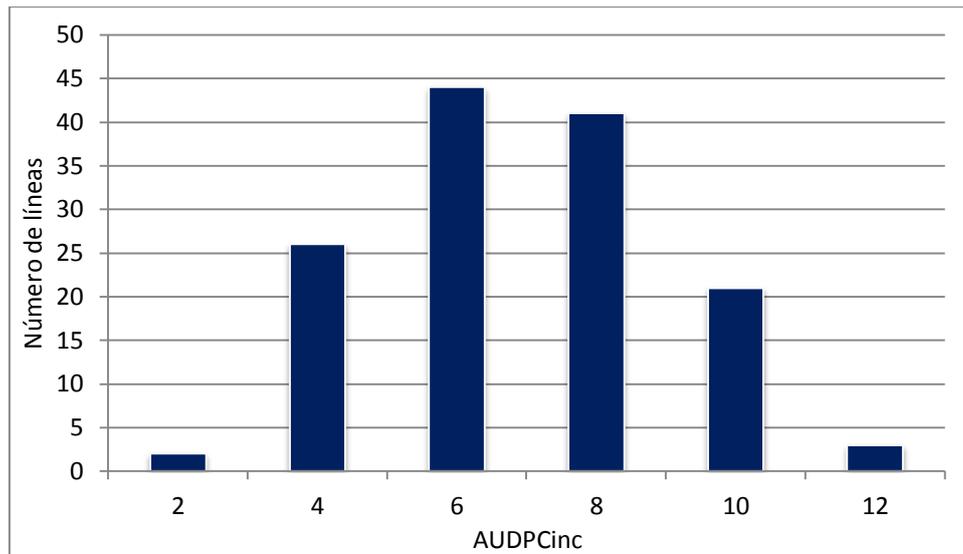


Figura 1.10 Histograma de medias ajustadas para AUDPCinc a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente

Severidad de PHC

Las medidas de severidad se estimaron mediante observación visual, detectándose líneas con severidades entre 1% (mancha inicial que no progresó con el paso del tiempo) y 100 % (capítulo caído) (Figura 1.11, A-C) en todas las campañas fenotípicas. En la Figura 1.12 se presentan los valores respecto a la media de severidad por línea por campaña (Figura 1.12, A-E).



Figura 1.11 Distintos grados de avance de la enfermedad (A) 10 % de severidad; (B) 80 % de severidad; (C) 100 % de severidad, capítulo caído.

La transformación angular de los datos de severidad obtenidos utilizando la fórmula $\arcsen(\sqrt{x})$ permitió su análisis como si se tratara de datos de distribución normal. Cabe destacar que el volumen de datos analizado para este carácter es menor al analizado para el carácter IEmax debido a que, como la severidad "0" no está definida, se pierden del

análisis aquellas líneas que tuvieron incidencia “0”. De este modo, del total de datos obtenidos, 863 datos pudieron ser utilizados para las estimas de severidad.

Al igual que en los dos casos anteriores, se testearon dos modelos, uno en el que no se consideró la interacción GxA y otro en el que sí, y se estimó también la contribución del efecto “líneas” a la varianza total al aplicar un tercer modelo en el que el efecto de las mismas fue considerado “aleatorio”.

El primer modelo evaluado fue el 1.1, el cual considera a las líneas como efecto “fijo” y no tiene en cuenta la existencia de interacción GxA. El mismo tuvo un valor de AIC = 737,25 y un valor de BIC = 1422,46. El efecto de las líneas fue altamente significativo ($p < 0,0001$).

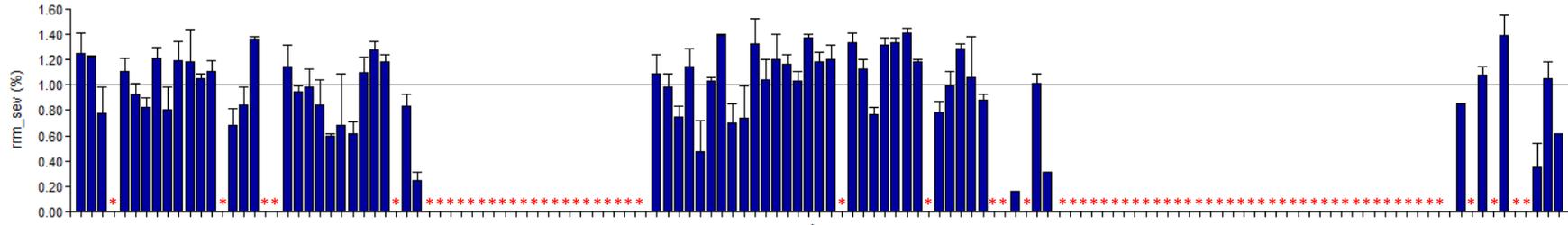
Según este modelo, la mayor parte de la varianza estaba explicada por la “fecha de inoculación”, seguido por la “campaña”, siendo el efecto de la “repetición” casi despreciable (Apéndice, Tabla A1.8). La comparación mediante DGC permitió discriminar dos grandes grupos.

El segundo modelo evaluado, en el cual se consideró a las líneas (genotipos) como efecto “fijo”, al tiempo que se tuvo en cuenta el efecto de la interacción aleatoria GxA (modelo 1.2), obtuvo un valor de AIC = 730,07 y BIC = 1420,14, levemente por debajo que los obtenidos para el modelo anterior, donde no se consideró la interacción.

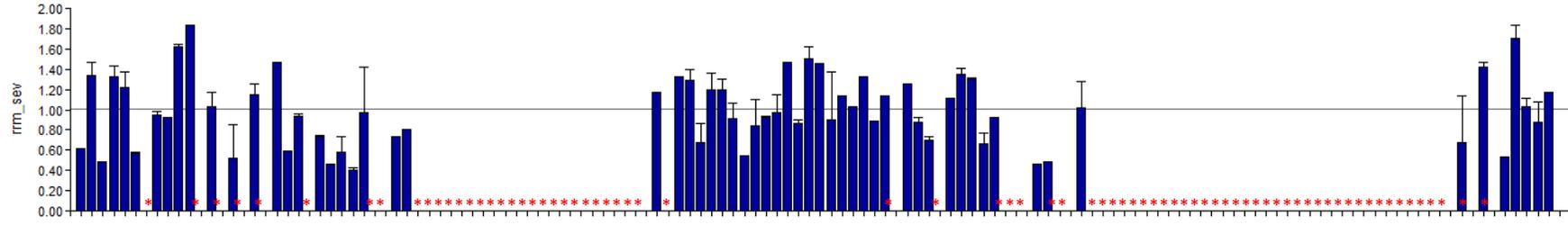
En este caso, puede observarse que la mayor parte de la varianza es explicada por la “fecha de inoculación”, seguida por la “campaña”, y la interacción GxA (Apéndice, Tabla A1.9). Como era esperable, la “repetición” no se llevó gran parte de la varianza.

Al analizar un modelo en el que todos los efectos fueron considerados como “aleatorios”, pudo identificarse la contribución del efecto de las líneas (genotipos) a la varianza total (20 %) (Apéndice, Tabla A1.10).

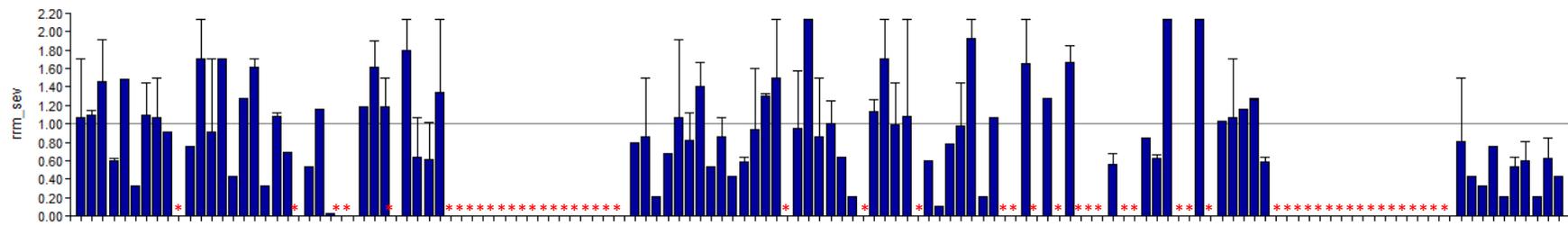
A



B



C



Las medias ajustadas por línea obtenidas según el segundo modelo testeado se presentan en el Apéndice, Tabla A1.14. Un histograma de distribución de medias se muestra en la Figura 1.13.

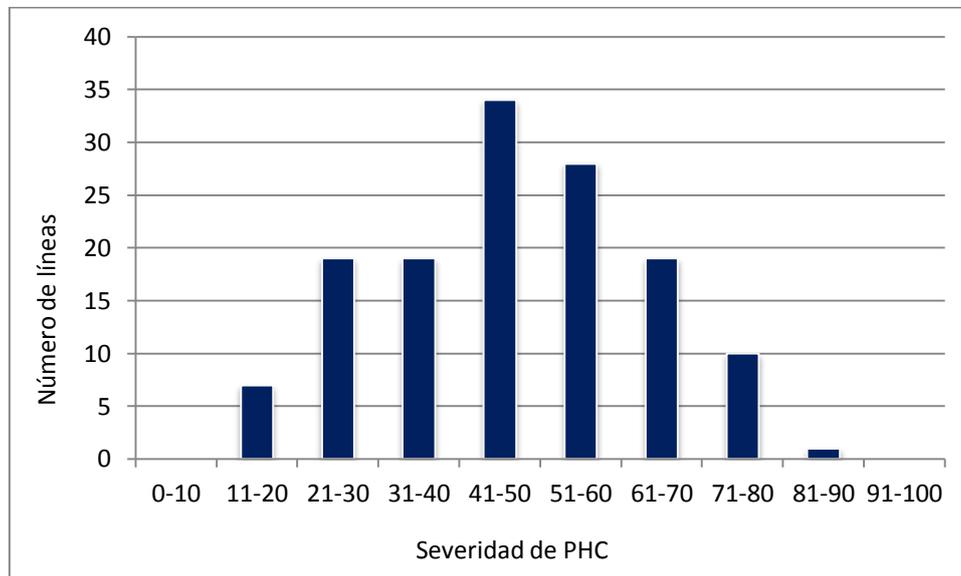


Figura 1.13 Histograma de medias ajustadas para severidad a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente

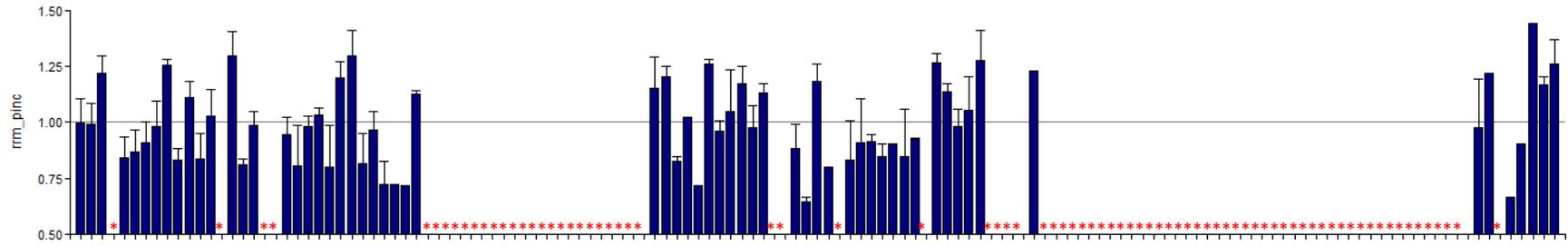
Período de incubación

Dado que esta medida fenotípica se relaciona directamente con las plantas enfermas, sólo se consideraron como medidas informativas aquellas obtenidas para surcos con al menos tres plantas enfermas, ya que si se estimara esta variable a partir de sólo una o dos plantas podría incurrirse en error.

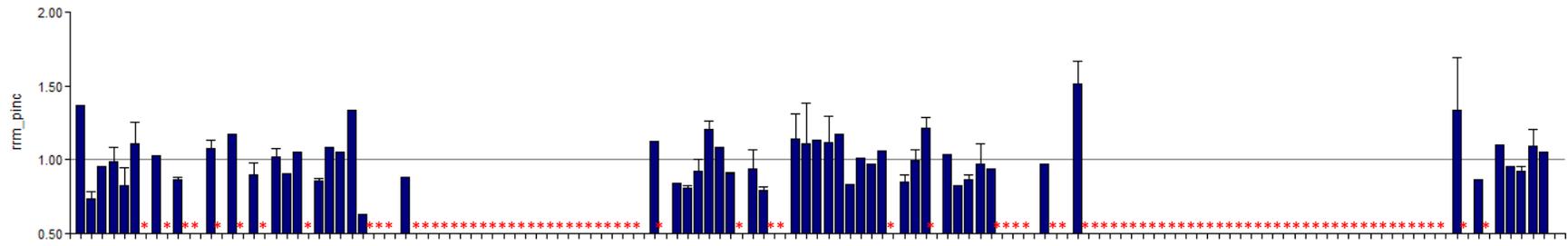
De este modo, luego de descartar aquellos casos en los que se observaron entre 0-2 plantas enfermas, se pudo contar con un total de 723 datos para las estimas de período de incubación. A fin de estudiar el comportamiento por línea y por año respecto a la enfermedad, se construyeron gráficos de relación respecto a la media (rrm), los cuales se presentan en la Figura 1.14 (A-E). En estos gráficos, si la media de período de incubación de una línea es igual a uno, quiere decir que se comportó como el promedio de la población. Si es mayor, se trata de una línea moderadamente resistente. Si por el contrario es menor, se trata de una línea susceptible. Se observa que, si bien los valores absolutos de período de

incubación por línea mostraron fluctuaciones año a año, sus rrm tienden a mantenerse constante en una alta proporción de las líneas.

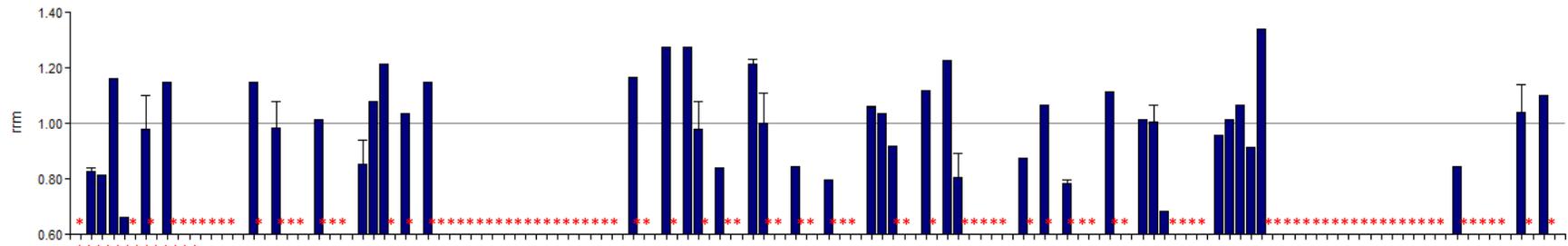
A



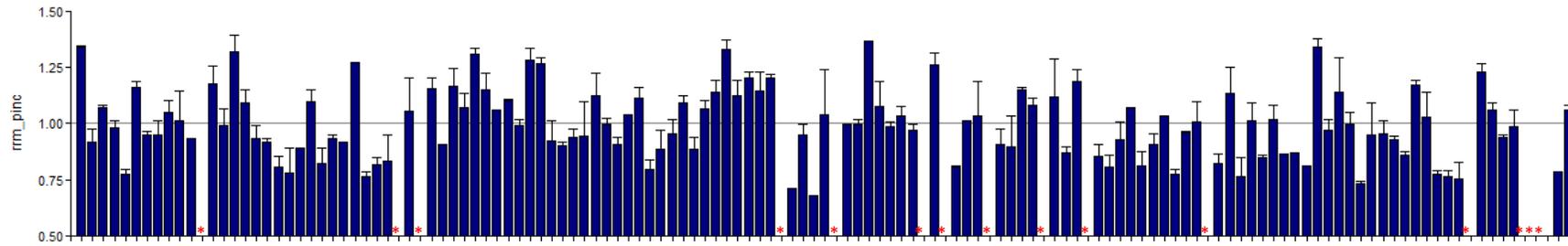
B



C



D



E

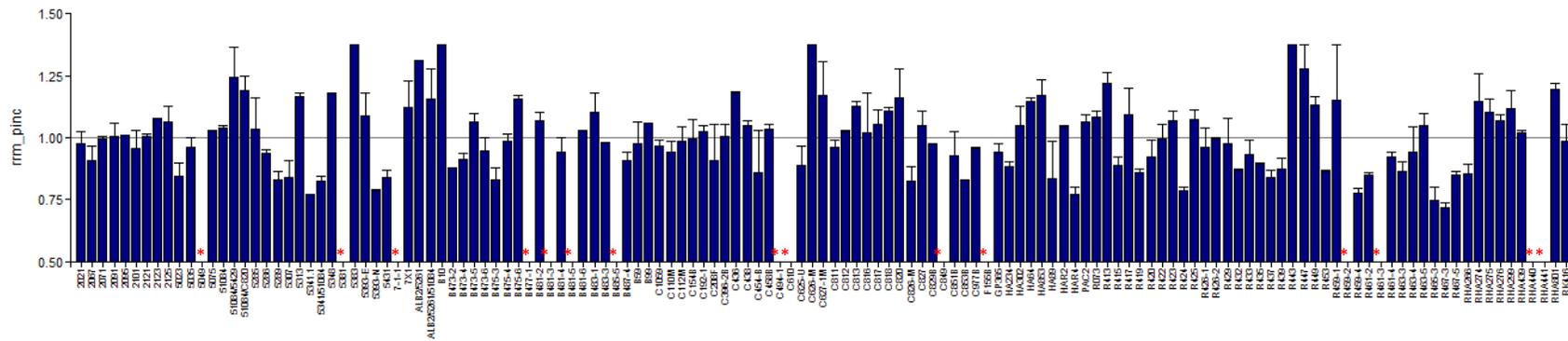


Figura 1.14 Relación respecto a la media por línea y por campaña para la variable período de incubación (A) Campaña 2009-10; (B) Campaña 2010-11; (C) Campaña 2011-12; (D) Campaña 2012-13; (E) Campaña 2013-14. Las barras representan el error estándar. Los puntos rojos indican ausencia de datos. Se graficó el promedio de los dos bloques por línea por campaña.

Al analizar este conjunto de datos con el modelo 1.1, en el cual no se consideró la interacción GxA, se obtuvo un valor de AIC = 3339,71 y un valor de BIC = 3981,38. El efecto de las líneas fue altamente significativo ($p < 0,0001$).

Según este modelo, la mayor parte de la varianza estaba explicada por la “fecha de inoculación”, seguido por “campaña”, siendo el efecto de la “repetición” casi despreciable (Apéndice, Tabla A1.11). La comparación mediante DGC permitió discriminar dos grandes grupos.

El modelo 1.2 obtuvo un valor de AIC = 2941,02 y un valor de BIC = 3587,29 para este set de datos, siendo el efecto de la interacción GxA significativa ($p < 0,05$). En este caso, la interacción explicó la mayor parte de la varianza, seguida por la “fecha de inoculación” y de la “campaña” en tercer lugar. Nuevamente, la “repetición” sólo explicó una pequeña parte de la varianza total (Apéndice, Tabla A1.12).

Finalmente, siguiendo los ejemplos anteriores, se aplicó el modelo 1.3, según el cual todos los componentes fueron considerados “aleatorios”, a fin de estimar la contribución de las líneas a la varianza total (10,57 %). Este modelo obtuvo valores de AIC y BIC iguales a 2830,13 y 2862,24, respectivamente. Al igual que en el caso anterior, la interacción explicó la mayor parte de la varianza, seguida por la “fecha de inoculación”, mientras que el efecto de la “línea” quedó en tercer lugar (Apéndice, Tabla A1.13).

Las medias ajustadas por línea obtenidas según el segundo modelo testeado se presentan en Apéndice A1.14. Un histograma de distribución se muestra en la Figura 1.15. En la misma se observa que el valor de Pinc más frecuente se encuentra alrededor de 15 dpi.

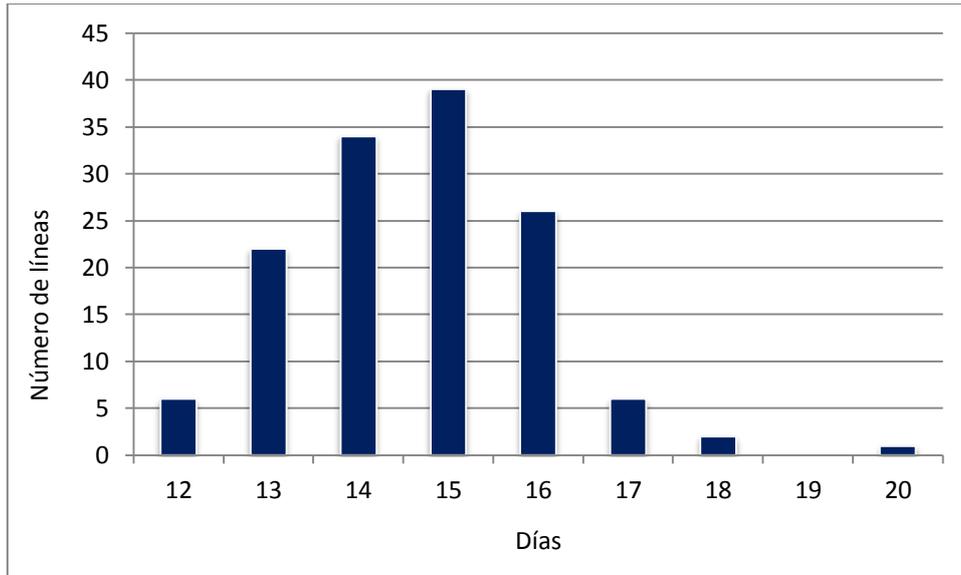


Figura 1.15 Histograma de medias ajustadas para Pinc a partir del análisis de cinco campañas fenotípicas bajo un modelo estadístico que contempla la presencia de interacción genotipo por ambiente

Correlación entre medidas fenotípicas

El estudio de la existencia de correlación entre las medidas fenotípicas registradas permitió determinar que las medidas IEmax, AUDPCinc y severidad se encuentran positivamente correlacionadas entre sí, mientras que las mismas presentan correlación negativa con la variable período de incubación. De acuerdo a lo esperado, sólo las medidas IEmax y AUDPCinc presentaron correlación fuerte (0,90), ya que ambas derivan de una misma determinación fenotípica. En Tabla 1.2 se presentan los valores de correlación, y en Figura 1.16 los gráficos de dispersión de las distintas variables.

Tabla 1.2 Coeficientes de correlación de Spearman para las distintas variables fenotípicas

Variables	r	p-valor
IEmax vs. Sev	0,57	<0,0001
IEmax vs. AUDPCinc	0,90	<0,0001
Sev vs. AUDPCinc	0,72	<0,0001
IEmax vs. Pinc	-0,50	<0,0001
AUDPCinc vs. Pinc	-0,70	<0,0001
Sev vs. Pinc	-0,70	<0,0001

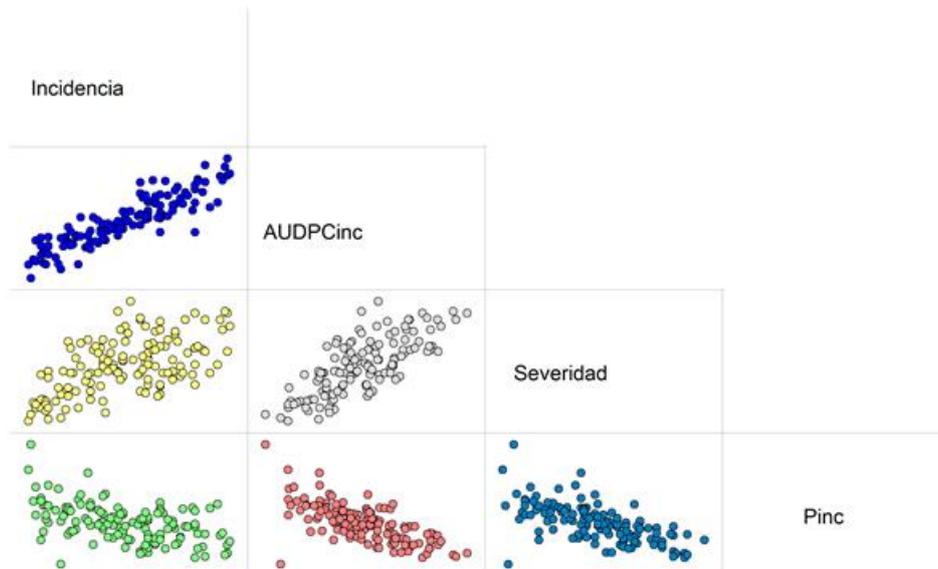


Figura 1.16 Gráfico de dispersión para las cuatro variables fenotípicas evaluadas Incidencia, AUDPCinc, Severidad y Período de incubación.

DISCUSIÓN

La caracterización fenotípica es un factor crucial de los estudios de mapeo de asociación. Del mismo modo, contar con una población de mapeo altamente diversa, de modo que pueda ser representativa de la especie para la cual se realizará la inferencia de asociación es un paso clave para el éxito del mapeo (Ersoz *et al.* 2007).

El panel de 137 líneas que conforman la PMA para caracteres cuantitativos complejos de INTA cumple con este requisito: observaciones *a priori* del material sembrado permitieron detectar diferencias morfológicas inherentes a la constitución de cada línea. Se observaron diferencias en los patrones de coloración de las flores liguladas, del receptáculo, la forma, tamaño número de capítulos y hojas, en la coloración de nervaduras, la presencia de múltiples capítulos y en los días a floración. Toda esta variabilidad detectada se traduce en variabilidad genotípica, lo cual es un indicador de la diversidad de las líneas. Las características morfológicas constituyen importantes determinantes de valor agronómico, ya que además de ser altamente informativos, son evaluaciones directas, económicas y fáciles de tomar (Cholastova & Knotova 2012).

Tal como se describió anteriormente, el germoplasma que conforma una PMA puede provenir de distintas fuentes, y de acuerdo a cuál sea su constitución, la aplicación de los estudios de asociación puede diferir (Breseghello & Sorrells 2006). Para la conformación de esta PMA, se seleccionaron materiales provenientes de programas de mejoramiento que el INTA está llevando a cabo en girasol para distintos caracteres, entre ellos resistencia a estreses bióticos y abióticos (principalmente, resistencia a Podredumbre húmeda del capítulo y a déficit hídrico). Entre los criterios de selección de las líneas a incluir, tuvieron gran peso: (a) que las líneas contaran con al menos un dato de comportamiento frente al ataque del patógeno *Sclerotinia sclerotiorum*, a fin de trabajar con materiales de gran variabilidad de respuesta frente a la enfermedad y (b) que la mayoría de las entradas fueran de origen nacional, a fin de poder contribuir a la caracterización del germoplasma local. Asimismo, se decidió la inclusión en la PMA de líneas públicas relevantes para los programas de mejoramiento locales e internacionales. El uso de este tipo de materiales aporta dos grandes ventajas al estudio de asociación que se está llevando a cabo, a saber: (a) la posibilidad de obtener datos fenotípicos robustos en ensayos a campo, ya que los materiales están adaptados a las condiciones de crecimiento locales y (b) dado que los

materiales pertenecen a programas de mejoramiento, de encontrarse alelos asociados al carácter de interés, se verá allanado el camino para su introgresión con vistas al mejoramiento genético del cultivo (Myles *et al.* 2009).

Actualmente, se están llevando a cabo en el país diversos programas de mejoramiento para girasol, conducidos tanto por instituciones oficiales como privadas, cuyo objetivo ulterior es la generación de nuevas combinaciones para el posterior desarrollo de líneas y la creación y evaluación de nuevos híbridos. Si bien es cierto que el objetivo principal de cualquier programa es lograr mayor rendimiento, que para el caso de girasol, se expresa en kilos de aceite por hectárea, dado que existen muchos factores que condicionan el rendimiento, se suman diversos aspectos como objetivos de los programas. Dentro de éstos, se encuentran el control eficiente de malezas, el manejo de enfermedades, la nutrición mineral y la fisiología del cultivo (Bruniard 2002). El INTA está llevando a cabo diferentes programas de mejoramiento, siendo el programa de resistencia a enfermedades uno de los más importantes que involucra el trabajo articulado de tres unidades, el BAG-IM, el grupo de Patología de Girasol de la EEA Balcarce y el grupo de Genómica de Girasol del Instituto de Biotecnología de INTA Castelar. Este programa apunta a la búsqueda de resistencia para dos de las enfermedades más relevantes para el cultivo que son la PHC y la marchitez por *Verticillium*. En este trabajo se aborda la búsqueda y caracterización de nuevas fuentes de resistencia a partir de una aproximación de mapeo de asociación como complemento de la caracterización de QTL a partir del estudio de poblaciones segregantes.

Los ensayos a campo llevados a cabo en la Estación Experimental INTA Balcarce fueron efectivos para la correcta generación de datos fenotípicos. No obstante, para que esto sucediera, fue necesario atender determinados asuntos agronómicos. La falta de vigor, propia de las semillas de líneas endocriadas, fue compensada con el tratamiento con ácido giberélico para romper dormición, el cual resultó clave para garantizar el éxito de los distintos ensayos llevados a cabo. Esto fue evidente al comparar la tasa de germinación del primer ensayo (2009/10), donde no se realizó este tratamiento previo, con los últimos cuatro, donde el tratamiento sí fue aplicado resultando en un incremento del 50 % en el número promedio de plantas inoculadas (de 10 plantas en 2009/10 a aproximadamente 15 por línea en el resto de los ensayos).

La concentración utilizada del curasemilla Metalaxil M[®] fue efectiva para prevenir mildiu en las líneas de la PMA, aunque algunos escapes fueron detectados, principalmente en las

últimas campañas. Ya ha sido reportado por varios grupos de trabajo la aparición de ciertas razas de *Plasmopara halstedii* con resistencia al Metalaxil M® (Molinero-Ruiz *et al.* 2008), de modo que para ensayos futuros podrían necesitarse nuevas combinaciones de agentes protectores. Como contrapartida, en las campañas 2010/11 y 2012/13 se detectó alta incidencia de *Verticillium dahliae* por infección natural, lo cual provocó marchitamiento en algunas de las líneas de la PMA, llegando en los casos más extremos, a impedir las evaluaciones fenotípicas de PHC. Esta alta presencia del hongo en el suelo se atribuye a que la parcela donde se sembró el ensayo esas dos campañas sólo fue sembrada con girasol en los últimos 10 años, generándose de esta forma un infectario natural de enfermedades fúngicas del cultivo. Por este motivo, se procedió al cambio de lote para el ensayo 2013/14 y para los futuros ensayos.

El método de inoculación con ascosporas del hongo implementado en este trabajo deriva del método propuesto por Vear *et al.* (1984), que consiste en la aspersión de 5 mL de suspensión de 5000 ascosporas/mL por capítulo en estadio R5.2 de la escala de Schneidher y Miller (1981). Sin embargo, esta alta concentración de esporas genera una alta presión de inóculo en la planta, haciendo que las incidencias por línea sean cercanas a 1 en todos los casos. De este modo, sólo es posible discriminar las líneas por variantes como severidad y período de incubación, o índice de latencia. El método de inoculación puesto a punto en la EEA INTA Balcarce consiste en la aspersión de 2500 ascosporas/mL por capítulo demostrando ser más informativo, ya que permitió una buena discriminación de las líneas en cuanto a su respuesta fenotípica para las cuatro variables seleccionadas: incidencia, severidad, AUDPCinc y período de incubación.

Dada la gran diversidad de las líneas que conforman la PMA existen diferencias en tiempo a floración entre ellas. Debido a ello, no fue posible la inoculación de todas las líneas en una única fecha, lo cual hubiera sido deseable para disminuir el término de error asociado al ambiente. Las diferencias generadas debido a la amplia ventana de fechas de inoculación (aproximadamente 17 días de diferencia entre el primer grupo de materiales inoculados y el último), suele ser homogeneizada mediante la siembra escalonada de materiales control (S y MR), con los cuales relativizar los niveles de enfermedad obtenidos (Vear & Tourvieille de Labrousche 1984). Dado que en este caso no pudieron obtenerse datos de este tipo para todos los ensayos, se salvó este inconveniente mediante la consideración en todos los modelos estadísticos evaluados de la “fecha de inoculación” como factor ambiental.

El hecho de trabajar con medidas fenotípicas que no responden a una distribución normal, como es el caso de la incidencia o del período de incubación, el uso de tablas de datos incompletas o desbalanceadas y la aplicación de medidas repetidas, como en el caso de AUDPC, plantea el problema de la selección de métodos estadísticos robustos para el modelado de los datos. Dado que no se cumple ninguno de los supuestos, incursionar en análisis de la varianza (ANOVA) clásicos hubiera derivado en resultados erróneos o poco confiables (Bolker 2008). La modelación de datos experimentales obtenidos para todas las medidas fenotípicas en el marco teórico de los modelos lineales generalizados mixtos permitió el análisis de datos con estructuras de dependencia, desbalances y falta de normalidad. De este modo, fue posible relajar los supuestos tradicionales del modelo lineal general y modelar, de manera flexible, la estructura compleja de este tipo de datos (Balzarini *et al.* 2010).

La interacción Genotipo×Ambiente ($G \times A$), *i.e.* la respuesta diferencial de diferentes genotipos a través de un rango de ambientes, es una característica universal relacionada a los seres vivos desde bacterias a plantas y humanos (Kang 1998). El tema es de relevancia en agricultura, ya que los principales caracteres de importancia agronómica (como el rendimiento, o la resistencia a estreses) están altamente influenciados por el ambiente mostrando variación continua. Debido a la omnipresencia de $G \times A$ en caracteres cuantitativos, las evaluaciones de genotipos se llevan a cabo en experimentos multiambientales (Balzarini *et al.* 2010). En este caso, los ambientes estuvieron dados por los distintos años de campaña, y la presencia/ausencia de riego. Al analizar distintos modelos estadísticos (*i.e.*, contemplando o no la presencia de interacción) y utilizando el criterio de menores AIC y BIC, se observó interacción $G \times A$ significativa en todos los casos, lo cual demuestra la alta dependencia de la enfermedad con el ambiente. Esta dependencia de la enfermedad con el ambiente también se hizo evidente al observar el comportamiento (determinado ya sea como incidencia, severidad o período de incubación) de las líneas año a año. Al predecir los efectos de interacción para inferir, no sólo en sentido amplio sino también en sentido estricto (inferencia específica de cada ambiente, es decir, anual), sobre el desempeño de genotipos fue posible cuantificar el efecto del ambiente sobre los niveles de enfermedad. En este aspecto, es notorio el cambio de comportamiento de las líneas en la campaña 2011-12, año de extrema sequía y altas temperaturas, donde se observa que las medias anuales de enfermedad se situaron ampliamente por debajo de la media global,

mientras que en años más predisponentes como 2009/10 y 2012/13 las medias anuales la superaron. Diversos autores han dado cuenta de la existencia de la interacción GxA al analizar niveles de PHC en girasol (Bert *et al.* 2002, 2004; Ronicke *et al.* 2005).

Si bien es cierto que la severidad de la enfermedad suele ser una medida más apropiada para caracterizar enfermedades fúngicas que no afectan la planta en su totalidad (Kranz 1988), ya que describe mejor la dinámica que gobierna el incremento de la enfermedad y su expansión en tiempo y espacio (Campbell & Madden 1990), la incidencia de la enfermedad es una medida mucho más objetiva y fácil de tomar, de modo que resulta conveniente en ensayos donde se realizan observaciones extensas, o cuando las observaciones son realizadas por personas con poca experiencia (Madden & Hughes 1995). De este modo, utilizar un método de inoculación asistida que permita la discriminación con ambas medidas fenotípicas resulta estratégico a la hora de caracterizar una enfermedad compleja como la PHC.

Campbell y Madden (1990) definieron incidencia de la enfermedad como “el número de plantas que están visiblemente enfermas, usualmente relativo al número total de unidades evaluadas”. De este modo, se desprende que los datos de incidencia son de tipo binario, es decir, cada observación (planta) puede tomar uno de dos valores posibles: está enferma o no. Dado que cada observación es independiente, estos datos siguen una distribución binomial, tal cual lo describió Bernoulli hace casi 300 años (Collett 1991). La aplicación de un MLGM con un modelo de distribución binomial permitió el modelado de los datos para la obtención de las medias ajustadas. La media global de incidencia registrada fue de 52 %. Este valor fue similar al reportado por Fusari *et al.* (2012) (50,4 %) para incidencia de PHC en el conjunto de 94 líneas endocriadas evaluado durante dos campañas, también en la EEA INTA Balcarce. Se observa que si bien los valores absolutos de incidencia por línea mostraron fluctuaciones año a año, exceptuando casos aislados, el comportamiento de cada línea frente a la enfermedad respecto a la media anual se mantuvo invariable.

El AUDPC es una medida más abarcativa que la IEmax, ya que da cuenta de cómo fue el proceso de desarrollo en el tiempo de la enfermedad. Contar con esta medida, derivada de las determinaciones en el tiempo de incidencia, ayuda a describir mejor el comportamiento de cada genotipo frente a la PHC. En el panel de datos obtenidos, se observa una gran variabilidad entre las curvas de progreso de la enfermedad correspondientes a cada año estudiado: hubo curvas que progresaron rápidamente, alcanzando en poco tiempo

incidencias de 80-100 %, particularmente en las campañas 2009/10, 2010/11 y 2012/13, mientras que otras fueron progresando más lentamente (datos no mostrados). Exceptuando observaciones puntuales, no se observa gran diferencia en incidencia entre las observaciones efectuadas ente los 24 y 28 dpi, ni tampoco inicio de la enfermedad a días posteriores a 21 dpi, sugiriendo que la determinación efectuada a 28 dpi puede ser considerada Incidencia a tiempo final. Es notorio que las curvas epidémicas que comenzaron tarde alcanzaron siempre incidencia menor al 50 %.

La selección de un modelo matemático apropiado para describir estas curvas de progreso de la enfermedad permite comprender, cuantificar y comparar el fenómeno biológico que significa la epidemia en el cultivo (Campbell & Madden 1990; Conles *et al.* 2011). Debido a la complejidad de los datos, dada principalmente por el uso de matrices incompletas y la falta de independencia (ya que las medidas en el tiempo se registraron sobre las mismas plantas), la aplicación de modelos lineales generalizados mixtos emerge como la mejor propuesta para el análisis de datos de tipo AUDPC.

Pese a ser ampliamente informativa, el uso de esta variable para describir el comportamiento de la PHC en girasol no ha sido reportado en la bibliografía, aunque sí para otras enfermedades fúngicas en cultivos comerciales como trigo (Wilcoxson *et al.* 1975; Shaner & Finney 1977; Das *et al.* 1992; Singh 1993; Ma *et al.* 1995; Broers *et al.* 1996), papa (Parker & Navia 1991; Birhman & Singh 1995), y tomate (Hartman & Wang 1993), entre otros.

La severidad es la cantidad de enfermedad que afecta a cada planta en una unidad de muestreo. Esta medida fenotípica cobra especial importancia en aquellas enfermedades fúngicas que afectan visiblemente sólo cierta parte de la planta (*i.e.*, capítulo), ya que se facilita la obtención de los datos fenotípicos (Seem 1984). Particularmente en girasol, ésta ha sido la medida más ampliamente utilizada para cuantificar los niveles de PHC, junto con período de incubación o índice de latencia (Vear & Tourvieille de Labrousche 1984; Bert *et al.* 2004; Vear *et al.* 2007; Yue *et al.* 2008). Las evaluaciones suelen realizarse por observación visual, una práctica que lleva más de 100 años (Cobb 1892; Chester 1950; Bock *et al.* 2010), y se puede estimar como proporción (0 a 1) o como porcentaje (0 a 100). La transformación de estos datos porcentuales utilizando la función angular permitió el análisis de los mismos bajo un modelo lineal mixto de distribución normal. La severidad promedio para todos los años fue cercana al 50 %. Este valor se encuentra en el rango de medias reportadas por Bert

et al. (2002), para dos ensayos independientes de estudio de la resistencia a *S. sclerotiorum* en familias F3 derivadas del cruzamiento de las líneas XRQ y PSC8 (41-65 %) y levemente inferiores si se los compara con los valores reportados por Yue *et al.* (2008) para familias F2:3 y F2:4 derivadas del cruzamiento de las líneas públicas HA441xRHA439, en los que se obtuvieron valores de severidad promedio cercanos al 80 %. Al igual que para IEmax, las relaciones respecto a la media de cada línea se mantuvieron constantes con el paso de los años, aunque los valores absolutos de severidad variaron campaña a campaña.

La variable “período de incubación”, también conocida como “periodo de latencia” determina el tiempo ocurrido entre la infección y la primera aparición de síntomas (Shearer & Zadoks 1972). Estudios previos han sugerido que esta medida fenotípica podría representar el componente más importante de resistencia parcial en ciertas relaciones planta-patógeno (Parlevliet & Van Ommeren 1975; Parlevliet 1976; Carson 1995).

Para el estudio de la PHC en girasol, esta variable ha sido utilizada ampliamente como “índice de latencia (IL)” (Vear & Tourvieille de Labrouche 1984; Vear & Tourvieille de Labrouche 1988; Castaño *et al.* 1993), y computándose como la relación entre el período de incubación de cada línea sobre el período de incubación de un control susceptible. Dado que en el presente trabajo no se contaba con información de control susceptible para todos los casos, se analizaron los datos de periodo de incubación como días hasta la aparición del primer síntoma. Dada la complejidad de este tipo de datos, la modelización de la aparición de síntomas utilizando la distribución gamma proporciona un método para reducir la inexactitud (error) causada, ya sea como consecuencia de la falta de evaluaciones diarias o de la heterogeneidad de los períodos de tiempo transcurridos entre una y otra (Lovell *et al.* 2004), ya que las mismas se hacían cada tres o cuatro días. Por su parte, contemplar la presencia de interacción GxA ayuda a disminuir el error asociado a variaciones ambientales, ya que se ha demostrado que la misma puede variar ampliamente dependiendo de condiciones ambientales, principalmente cambios de temperatura (Parlevliet & Van Ommeren 1975; Lovell *et al.* 2004).

La media global de esta variable fue de 15,15 días. Si se considera que las observaciones fenotípicas se extendieron entre los 10 dpi (momento del destape del capítulo) hasta los 28 dpi, se aprecia que el tiempo de observación fue adecuado para lograr describir las curvas de enfermedad en su totalidad.

Al evaluar las heredabilidades o contribución del genotipo a la varianza total, se observa que éstas son bajas para las cuatro variables fenotípicas tomadas. La variable “severidad” fue la que presentó mayor heredabilidad (20 %), en segundo lugar AUDPCinc (14,43 %), seguido inmediatamente por la variable de la cual deriva, incidencia (13,21 %), y finalmente por período de incubación (10,57 %), lo cual contribuye a demostrar nuevamente la notoriedad de las variaciones ambientales sobre estas medidas fenotípicas. *A priori* podría decirse que los valores obtenidos son bajos si se los compara con los reportados por otros autores para poblaciones biparentales (Castaño *et al.* 1993; Mestries *et al.* 1998; Bert *et al.* 2002; Collard *et al.* 2008) pero debe tenerse en consideración que las heredabilidades son válidas para la población en estudio, el ambiente o localidad en donde se evalúan los materiales y el método estadístico para estimarla (Falconer & Mackay 1996).

Al evaluar la existencia de correlación entre las medidas fenotípicas registradas, se observa que la mayor correlación ($r = 0,89$) se da entre IEmax y AUDPCinc. Esto es de esperarse, ya que las dos medidas derivan de una misma. Si bien moderada, se detectó también correlación positiva significativa entre IEmax y severidad, y severidad y AUDPCinc. Pese a existir correlación entre incidencia y severidad, cabe destacar las discrepancias entre ambas variables, ya que la ocurrencia de una en altas proporciones no tiene que corresponderse con la ocurrencia de la otra en altas proporciones: esto se hace evidente en la campaña 2010/11 en la que se observaron altos valores de incidencia pero bajos valores de severidad, al compararlos con la media global. Los valores de correlación detectados para incidencia y severidad ($r = 0,6$) son similares a los reportados por Yue *et al.* (2008) para estudios similares en girasol (0,57- 0,64). Estudios de correlación entre incidencia y severidad en otras especies vegetales también arrojaron valores positivos (*i.e.*, $r = 0,6$ en trigo infectado con *Fusarium graminearum*, Groth *et al.* 1999), llegando a postularse métodos para predecir severidad en función de datos de incidencia, especialmente para aquellos casos de baja severidad (Seem 1984; Groth *et al.* 1999).

La variable periodo de incubación se correlacionó en forma negativa con IEmax, severidad y AUDPCinc. Esta correlación negativa implica que en años más predisponentes para el desarrollo de la enfermedad, los tiempos de latencia se acortan, mientras que en aquellos años en los que las condiciones no son tan favorables para el establecimiento de la enfermedad, los tiempos de latencia se alargan. Si bien la correlación más fuerte se dio con severidad, es preciso destacar que el comportamiento de la variable período de incubación

fue exactamente el opuesto al de incidencia en todas las campañas, si se utiliza como parámetro el comportamiento respecto a la media. En las campañas 2009/10, 2010/11 y 2012/13 se observaron niveles de incidencia por encima de la media y periodos de incubación por debajo; por el contrario, en las campañas 2011/12 y 2013/14 se observaron niveles de incidencia por debajo de la media, y períodos de incubación por encima. Esta relación entre incidencia y período de incubación se corresponde con estudios previos, que postularon que el porcentaje de severidad constituye una estima de la resistencia a la extensión del hongo en los tejidos de la planta, mientras que el período de incubación y la incidencia indicarían medidas de resistencia a la penetración del hongo (Vear & Tourvieille de Labrouche 1984; 1988), de modo que ambas variables estarían estimando fuentes de resistencia relacionadas.

La presencia de correlaciones negativas entre severidad y período de incubación también fue reportada por Bert *et al.* (2002), aunque en menor medida (entre -0,293 y -0,477), Bert *et al.* (2004) (-0,55). Sorprendentemente, Gentzbittel *et al.* (1998) reportaron la existencia de correlación positiva entre estas dos variables ($r = 0,76$).

En cuanto al comportamiento frente a la enfermedad observado en las líneas de la PMA durante las cinco campañas fenotípicas llevadas a cabo desde 2009 a 2014 y su relación con los registros de enfermedad con los que se contaba para algunas de estas líneas (Maringolo, 2007, Fusari *et al.* 2012), se observó que si bien muchas líneas se comportaron según lo esperado, algunas no mantuvieron el comportamiento de resistencia o susceptibilidad reportado con antelación. Claros ejemplos de esto son las líneas HA89 y HA302 (reportadas como susceptibles), que en este trabajo se comportaron como moderadamente resistentes. Esto corrobora una vez más la importancia de realizar múltiples ensayos fenotípicos para este carácter.

De este modo, puede concluirse que la selección de las variables fenotípicas incidencia, severidad y período de incubación como descriptoras de la enfermedad demostró abarcar de forma complementaria distintos aspectos de la resistencia. Adicionalmente, el uso del carácter “incidencia de la enfermedad” no solo como Incidencia máxima, sino también como AUDPC permitió caracterizar la enfermedad en su totalidad.

La PMA de girasol de INTA, actualmente constituida por 137 materiales de gran diversidad fenotípica, constituye en sí una gran herramienta para estudios de asociación para caracteres cuantitativos en este cultivo. Además de ser utilizada para el carácter resistencia

a PHC, actualmente se está probando su variabilidad de respuesta para el carácter tolerancia a déficit hídrico (Valeria Moreno, *comunicación personal*). Por su parte, estudios que se están llevando a cabo en INTA relacionados con senescencia foliar y resistencia a distintas razas de *Verticillium dahliae* se basan en genotipos seleccionados a partir del conjunto de líneas que componen esta PMA por presentar variabilidad de respuesta a los caracteres de interés, variabilidad que fue observada en los distintos ensayos efectuados desde la campaña 2009/10 a la campaña 2013/14 (Sebastián Moschen, Juan Montecchia, *comunicación personal*).

Por último, la amplia caracterización de la PMA en relación con la respuesta a la enfermedad PHC efectuada en múltiples ambientes con repetición, ha generado un conjunto de datos robustos para su uso no solo en estudios de mapeo por asociación para la resistencia a PHC, sino también en futuros estudios de predicción genómica para este carácter.

CAPÍTULO 2

**ESTUDIO DE LOS NIVELES DE DIVERSIDAD GENÉTICA Y
ESTRUCTURA POBLACIONAL EN LA POBLACIÓN DE
MAPEO POR ASOCIACIÓN DE GIRASOL DE INTA**

INTRODUCCIÓN

Los altos niveles de diversidad fenotípica y molecular que caracterizan al girasol convierten a este cultivo en un candidato ideal para estudios de mapeo de asociación (MA). A la fecha, se han utilizado con éxito en estudios de MA diversas colecciones de germoplasma provenientes del Banco Activo de Germoplasma de INTA Manfredi, el USDA *North Central Regional Plant Introduction Station* (NCRPIS), el *USDA-ARS North Central Regional Plant Introduction Station* y el *Institut National de la Recherche Agronomique*, INRA (Fusari *et al.* 2012; Mandel *et al.* 2013; Cadic *et al.* 2013; Talukder *et al.* 2014).

Dada la compleja historia de domesticación y dispersión del girasol cultivado, resulta esperable que los niveles de diversidad varíen de una región a otra. De modo que, para poder completar la caracterización global de la diversidad del cultivo, resulta necesario evaluar en profundidad los *pooles* o acervos génicos provenientes de las distintas regiones girasoleras del mundo. La estructura poblacional y los niveles de diversidad genética de los recursos provenientes de América del Norte y Europa ya han sido evaluados exhaustivamente por Coque *et al.* (2008) y Mandel *et al.* (2011). Sin embargo, el conocimiento de la constitución genética y los niveles de variabilidad de las colecciones de germoplasma argentino es aun escaso. Tal como se detalló en el capítulo anterior, la PMA de girasol para caracteres complejos de INTA está conformada principalmente por entradas seleccionadas por representar la diversidad genética presente en los materiales de mejoramiento de INTA, al tiempo que posee contribuciones de germoplasma Ruso, Canadiense, Rumano y Estadounidense y de especies silvestres como *H. petiolaris*. Es por ello que el estudio acabado de los niveles y distribución de la variación genética contenida en esta PMA resulta fundamental a la hora de completar el mapa mundial de la diversidad del girasol cultivado, y de asistir a los programas de mejoramiento y conservación locales.

Se encuentran disponibles diferentes tipos de marcadores para estudios moleculares en girasol, siendo los microsatélites (SSR, *single sequence repeats*) y los Polimorfismos de Nucleótido Simple (SNP, *single nucleotide polymorphisms*) los más utilizados en la actualidad. Un SNP se define como una diferencia a nivel de base nucleotídica entre dos secuencias de ADN. En teoría, las variantes de tipo SNP pueden involucrar las cuatro posibilidades determinadas por las bases nitrogenadas A, T, G ó C, pero en general, solo se observan dos alelos para un determinado *locus*. De este modo, la información que aportan

estos marcadores bialélicos puede ser algo limitada, comparada con la multiplicidad de alelos de los marcadores microsatélites (Kruglyak 2008). Esto, que en principio es una desventaja, se ve contrarrestado por la gran abundancia y estabilidad de los *loci* SNPs respecto de los *loci* SSRs (Oraguzie *et al.* 2007). Al momento, más de 2000 marcadores tipo SSR han sido desarrollados sobre regiones codificantes y no codificantes del genoma de girasol (Dehmer & Friedt 1998; Paniego *et al.* 2002; Tang *et al.* 2002; Yu *et al.* 2002), mientras que el uso de paneles de genotipificación de SNPs de mediana y alta procesividad ha comenzado a ser reportado más recientemente (Bachlava *et al.* 2012; Zubrzycki *et al.* 2012a; Mandel *et al.* 2013; Pegadaraju *et al.* 2013; Cadic *et al.* 2013; Filippi *et al.* 2015).

Diversos estudios han descrito la ocurrencia de altos niveles de diversidad de SNPs en girasol (Kolkman *et al.* 2007; Fusari *et al.* 2008; Bachlava *et al.* 2012). Estos altos niveles de polimorfismo, en conjunción con su abundancia y estabilidad en el genoma, sugieren que dichos marcadores serían ideales para aproximaciones genómicas en la especie, tales como análisis de diversidad genética, estudios de asociación, de selección genómica, entre otros (McCouch *et al.* 2010; Tung *et al.* 2010; Thomson *et al.* 2011). Hoy en día, existen plataformas de genotipificación de mediana a alta procesividad que permiten la interrogación simultánea de múltiples SNPs en todo el genoma, las cuales varían en términos de sensibilidad, reproducibilidad, capacidad de multiplexado, entre otros (Syvänen 2001; Sobrino *et al.* 2005). La tecnología GoldenGate de Illumina contempla la realización de ensayos de genotipificación de SNPs basados en procesos de hibridación alelo-específica seguidos de extensión, que permiten genotipificar grandes colecciones de muestras (hasta 480 individuos), con niveles variables de SNPs (de 48 a 3072). Esta tecnología ha sido aplicada con éxito en una variedad de especies vegetales conteniendo genomas de diferente tamaño y complejidad, demostrando altos niveles de reproducibilidad (Rostoks *et al.* 2006; Pavy *et al.* 2008; Hyten *et al.* 2008; Akhunov *et al.* 2009; Close *et al.* 2009; McMullen *et al.* 2009; Zhao *et al.* 2010; Thomson *et al.* 2011). A pesar de estos avances, la búsqueda y genotipificación de SNPs continúa siendo un desafío en especies como el girasol, para las cuales aún no se cuenta con un genoma completamente secuenciado y ensamblado (Seeb *et al.* 2011).

Uno de los pasos críticos en todo estudio de mapeo por asociación es la determinación de los niveles de estructuración y relaciones de parentesco dentro de la PMA. La omisión de estos factores puede llevar al establecimiento de asociaciones espurias debido a que,

independientemente de la proximidad física entre dos *loci*, la magnitud del desequilibrio de ligamiento (DL) aumenta en poblaciones estructuradas cuyas subpoblaciones difieren en las correspondientes frecuencias alélicas (Zhu *et al.* 2008; Soto-Cerda & Cloutier 2012). Un caso extremo de estructuración es el de una población creada a partir de dos poblaciones independientes, provocando la mezcla de cromosomas de diferentes linajes y con diferentes frecuencias alélicas (Figura 2.1). Frecuentemente esto provoca que el DL se extienda hasta sitios no ligados, incluso en cromosomas distintos, pero en general este DL se rompe a lo largo de generaciones de apareamiento aleatorio (Pritchard & Przeworski 2001).

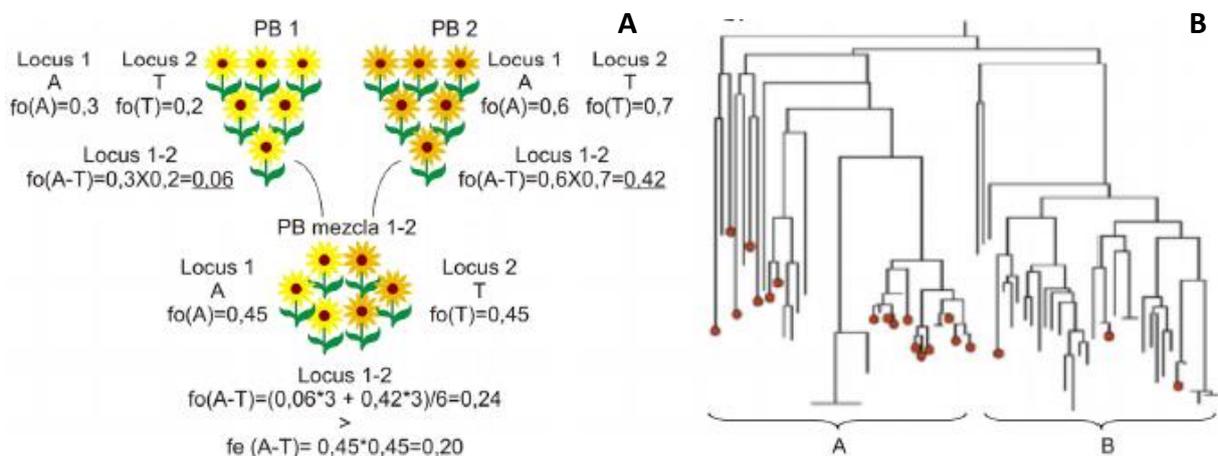


Figura 2.1 Efecto de la estructura poblacional en el DL y en las asociaciones fenotípicas (A) Cuando la población analizada (PB1-2) está formada por mezcla de dos poblaciones (PB1 y PB2), la frecuencia observada (*fo*) de co-segregación del alelo A del *locus 1* con el alelo T del *locus 2* es mayor que la frecuencia esperada (*fe*), induciendo a inferir que ambos *loci* están ligados; (B) Cuando el carácter estudiado (puntos rojos) es muy común en el subgrupo A pero muy raro en el subgrupo B, cualquier marcador con distribución alélica desigual entre los dos subgrupos dará asociación positiva, independientemente de su ubicación en el genoma. Figura adaptada de Rafalski (2010).

En los estudios de MA, la estructura poblacional suele ser comúnmente estimada utilizando información derivada de marcadores tipo SSR, dado su demostrada utilidad para este tipo de inferencias (Fusari *et al.* 2012; Xiao *et al.* 2012; Zhao *et al.* 2014), y su alto contenido informativo respecto de los marcadores bialélicos (Dreisigacker *et al.* 2005; Roussel *et al.* 2005; Reif *et al.* 2013; Würschum *et al.* 2013). No obstante, en los últimos tiempos, la creciente disponibilidad de marcadores tipo SNP y de tecnologías cada vez más rápidas y automatizadas de genotipificación, ha motivado su uso en estudios de diversidad y evaluación de estructura poblacional en diversas especies vegetales.

Al estudiar la repetibilidad de los resultados genotípicos y la proporción de datos faltantes para marcadores tipo SSR y SNP, Jones *et al.* (2007) encontraron una clara superioridad para los SNPs. Por su parte, Hamblin *et al.* (2007) estudiaron la utilidad de 89 SSRs versus 847 SNPs para la estimación de relaciones de parentesco y la evaluación de los niveles de diversidad obtenidos en un conjunto de líneas públicas de maíz, demostrando que los SSRs se comportan mejor a la hora de asignar las líneas a subpoblaciones. Estos autores sugieren que es necesario un número considerablemente mayor de SNPs para obtener un poder discriminante comparable al de los SSRs. Resultados similares a los de Hamblin *et al.* (2007) fueron encontrados en uva, soja y trigo (Hamblin *et al.* 2007; Li *et al.* 2010; Emanuelli *et al.* 2013). Sin embargo, hasta el momento no se ha reportado ningún estudio que compare la *performance* de ambos tipos de marcadores para estimas de diversidad y estructura poblacional en girasol.

En este capítulo se presenta la caracterización genética de las 137 líneas endocriadas que conforman la PMA de girasol de INTA y de un conjunto de poblaciones de polinización abierta y compuestos relacionados pertenecientes al BAG-IM, teniendo como objetivos: (a) el diseño y la caracterización de un panel de genotipificación masiva de SNPs bajo la tecnología Illumina, (b) la evaluación de los niveles de diversidad molecular y estructura poblacional de las entradas antes mencionadas usando marcadores SNP y SSR, y (c) la comparación de la utilidad de los diferentes tipos de marcadores para la obtención de estimas de diversidad y estructura poblacional en girasol.

MATERIALES Y MÉTODOS

Material vegetal y marcadores moleculares

Los materiales analizados en este capítulo fueron: el panel de 137 líneas endocriadas de girasol que conforman la Población de Mapeo por Asociación de INTA (PMA), 13 poblaciones de polinización abierta (OP) y 20 poblaciones compuestas (CP) pertenecientes al Banco Activo de Germoplasma de INTA Manfredi (BAG-IM). Información de *pedigree*, ID y origen se detallan en las Tablas A1.1 y A2.1 del Apéndice.

Para proceder con la genotipificación, se colectaron hojas de 3, 6 y 9 individuos de tres semanas de edad, dependiendo de si se trataba de materiales pertenecientes a la PMA, CP u OP, respectivamente. La recolección de las mismas se realizó utilizando nitrógeno líquido para garantizar la integridad del ADN, conservándose a -80°C hasta su liofilización. Una vez liofilizadas, la disrupción del tejido se realizó mecánicamente con el sistema TissueLyser (QIAGEN, Países Bajos).

El ADN genómico se aisló a partir de 20 mg de material liofilizado utilizando el kit NucleoSpin Plant II (Macherey-Nagel, Alemania), siguiendo las indicaciones del fabricante. La verificación de la calidad del mismo se realizó mediante corridas electroforéticas en gel de agarosa 0,8 % (TAE 1X: Tris-Acetato 20 mM, pH 8,5, EDTA 1mM), de acuerdo a lo descrito por Sambrook *et al.* (1989).

La cuantificación de la concentración se realizó con Picogreen® *dsDNA Assay Kit* (Intercalante de ADN, Invitrogen, EUA). Para la cuantificación se construyó una curva de calibración a partir de ADN stock de soja en concentración 200 ng/ μL . Éste fue diluido mediante diluciones secuenciales con buffer TE 1X (Tris-EDTA, pH 8), para generar una curva de siete puntos. Las cuantificaciones de cada punto se hicieron por quintuplicado en equipo NanoDrop 3300 (Thermo Scientific, EUA). Se preparó una solución de trabajo 1:200 de Picogreen® por dilución del *stock* en buffer TE 1X. Cada ADN de concentración desconocida fue mezclado con un volumen igual de dicha solución de trabajo; la cuantificación final se realizó a partir de 2 μL de la mezcla. Finalmente, la concentración de ADN fue normalizada a 25 ng/ μL previo a su uso.

Genotipificación con marcadores SSR

Todas las muestras de ADN fueron genotipificadas utilizando 22 SSR genómicos (gSSR) seleccionados de trabajos previos del grupo (Paniego *et al.* 2002; Talia *et al.* 2010) y 20 EST-SSR (SSR derivados de secuencias que se expresan o *expressed sequence tag*) seleccionados a partir de los trabajos publicados por Chapman *et al.* (2008) y Mandel *et al.* (2011), teniendo en consideración que los 17 grupos de ligamiento (GL) de girasol estén representados por al menos dos marcadores. Los gSSR fueron seleccionados en base a su neutralidad, alto nivel de polimorfismo, facilidad de lectura de los cromatogramas y localización en los mapas genéticos, mientras que los EST-SSR fueron seleccionados por su distribución en los 17 GL, y con el fin de poder realizar comparaciones entre las estimas de diversidad y estructura obtenidas para nuestra población de mapeo y aquellas reportadas por Mandel *et al.* (2011). A continuación se presenta la lista de los marcadores tipo SSR incluidos en el estudio (Tabla 2.1).

Tabla 2.1 Información de los marcadores SSR utilizados para genotipificar la PMA

GL	Alias	Tipo	Motivo	Min	Max	Iniciador <i>Forward</i>	Iniciador <i>Reverse</i>
1	HT0974	EST-SSR	GT	173	195	HEX-AAAGGCGATTGAGTTTTTCGAT	GCAAGAATGAAAGGCCAACTTTA
1	HT0595	EST-SSR	GA	291	312	NED-TTGTGTGGTGACTTGATTTAT	TCTCTTAAATACACTTCCTCTCT
1	HA77	gSSR	GA	247	262	HEX-TGTAATCTGTATCACTCCACC	GTTGTTCTGTTAGGTCGTTCCG
2	HT0761	EST-SSR	TG	126	128	NED-GGTGGGTGTTTGATTTTGTGTG	TCTCCTCTCCCTTTGTCTGC
2	HT0572	EST-SSR	TA	300	321	FAM-CACGCCATATGAAAGCAACTCT	TCATGGATGTTTCATCACCAGAC
2	HA3239	gSSR	ATG	134	144	NED-CTTAATCCGACATAAGCAGG	CTTTCATTATTCACAACCCCC
3	HT0292	EST-SSR	TC	311	334	HEX-ACTGTCAACACCTCCTTCGACT	GGCTTACACTTCTCTCCATCTCAA
3	HT0734	EST-SSR	TA	294	300	NED-GTTTCGCAGCTCTAGCATTGA	CCCCCTACAAGGCAACAAAATA
3	HA2920	gSSR	ATG	163	187	FAM-AACGTGTTTCATTCGGATGC	GAAGATTGGTATGATGGGGC
4	HT1045	EST-SSR	NA	312	325	NED-GGATGAGAGACCGGAGAAATA	AAAGGAAGTATTGATTTATGTATGGA
4	HT0375	EST-SSR	TA	398	418	NED-GCTGAAACGGGTCAATAAAGTC	TGAAGAAATGCTTCCAATCTGA
4	HA991	gSSR	GA	137	148	FAM-GCCCCCTTGATGCCCTTTC	GAATCGCCATTGAAATCGCCAG
5	HT0440	EST-SSR	AGG	236	248	NED-GAAGTTGGGAGGGTTGTTCAAG	CCTCCTGTTGGAACACCAAAT
5	HT0591	EST-SSR	TTC	310	333	HEX-AGAGGAATGAGATCGGGTTGAT	GTGGGACAACCTCAGCAACGTC
6	HA4103	gSSR	GA	141	152	FAM-TCACTTATCACCAATCTCATCCA	CTGTCTCAAATCGGGTGGTT
6	HA3312	gSSR	GA	98	110	FAM-TAACTGCAACAGACATGGAA	ATAACGATTGCACAACACAA
7	HA1848	gSSR	ATT	230	268	HEX-CATCCCCTTCTGAATAGAAA	TAGGCTCGTTAAACTTACGG
7	ORS1041	gSSR	CT	274	282	FAM-AAACAAACCTTAATGGGGTCGTA	ATATTGGCTGGTTGATGCTGAT
8	HT0519	EST-SSR	AT	252	274	HEX-TTTCGTGACCGAGAAAGGTAT	TGTGGCATACATAGAAATGATCTAA
8	HA3581	gSSR	ATG	128	147	NED-GTAGAGGAGTATCGCCAATAGC	GTTCCAATGGCCAACGATTGTG
9	HA2063	gSSR	GA	167	180	FAM-CACCGCAGATGGCCAATTCAAC	GGTCTGCTAACCTGCAATGCG
9	ORS510	gSSR	GT	247	257	HEX-CATCGCGTCCCTCTCTCTAA	CCAACCATCACAGCAATCAG
10	HA928	gSSR	GTATT	217	227	HEX-CCTTTGTAGTCCCTTACTGG	GGTCGATCATGTATGCGTGTG
10	HA3204	gSSR	ATC	113	118	NED-GCCCTTCAATCCTACCATTAAC	GGATATTGAGTTGGTTGTTGGG
11	c5783	EST-SSR	CTA	204	207	FAM-GCCATCACAACATCAAATGG	AAGCTTGACGTCATACAC
11	HT0426	EST-SSR	TA	312	324	HEX-AACCGTAAATGAAATCGGTGTG	GAGGGCAAAGTTGGGATACTCT
12	HA3555	gSSR	CAT	101	103	FAM-GATATCTCTCATAAGTGCCG	GGTCTTGTGATGACGAAGC

GL	Alias	Tipo	Motivo	Min	Max	Iniciador <i>Forward</i>	Iniciador <i>Reverse</i>
12	ORS767	gSSR	NA	348	378	FAM- ACAGGAGCTTTCGGTCTTCA	GGCACAGTCTCGTTCATATCC
13	HT0382	EST-SSR	AGG	359	390	NED-GAAGTCTGCTTTGGTGTGGTC	ATGGCTCTCTCATTTCCTACTG
13	HT0271	EST-SSR	TA	242	257	HEX-CCCATTGTTGCTAATTCAAGAT	TTGGAGTGCATAACACGTATCAG
14	HA293	gSSR	GA	113	123	HEX-GGGGACATCTCCCGTCCACC	CCTCATCCATCTCCATCCAATC
14	HA2077	gSSR	GA	108	146	HEX-GATAAGGTTTTCTCTCTCCC	GAGAAAAATGAGCTGATACCG
15	HT1024	EST-SSR	NA	254	257	FAM-AAGCAGGTTGCATGAAGAGAAG	GTCGAAACGGGTCAGGTTGTAT
15	HT0586	EST-SSR	TTAA	303	330	FAM-TGAACTCTTGTGTTGGCATCT	AAATGTGGATTTATGTATCTCAGTAA
15	HA4239	gSSR	ATT	86	112	FAM-GAATGATAGTGAATTGAGACAGG	CTGGCATCTATATCCATGGATAG
16	c2070	EST-SSR	ATGT	235	261	FAM-GGGTAATGCAAAGTACTAAGATGTG	GCATCATCCAACAACTAGAAGG
16	HT0723	EST-SSR	CAA/CCG	279	303	FAM-TCAAGCAATCAGACACCACATC	TGAACACAACCAAGAAATCCAA
16	HA3582	gSSR	ATG	119	139	NED-TGACCTTAATCCTTGGAGAG	CCTCTAATCCGAAAACCTAAC
17	HT0998	EST-SSR	CAA	374	383	FAM-CATGTCCCGATCAAAGAGTTGT	CCATATCCTGGTTGTTGTGGAG
17	HA1402	gSSR	GT	147	167	FAM-GTGTGAATGTGAGTGTGAATCC	GTCACAGTTGCATATCCATCC
NA	HA1752	gSSR	ATC	139	142	FAM-CCTTCTTCTCAAATCCCG	CGTCATTGTCATTGAGAAGATCG
NA	HA2147	gSSR	GA	74	104	FAM-CCCATAAACATTTACACCC	CCTGAGAAATCGGTAGATAGTG

GL: grupo de ligamiento. NA: no hay información.

Para la genotipificación de los SSR se utilizaron iniciadores portadores de marca fluorescente (HEX, NED y FAM). Mientras que los gSSR fueron amplificados en forma independiente, para los EST-SSR las PCRs (reacción en cadena de la polimerasa, *Polymerase Chain Reaction*) se realizaron en formato *multiplex*. Las PCR de los gSSR se realizaron en volumen final de 12 μ L, con la siguiente composición: 2 mM MgCl₂, 0,2 mM dNTPs, 1 U Polimerasa Taq Platinum (Invitrogen, EUA) y 0,25 mM de cada uno de los oligonucleótidos iniciadores específicos. Con el objetivo final de evitar amplificaciones espurias, la reacción de amplificación fue llevada a cabo por enriquecimiento del templado mediante disminución progresiva de la temperatura de hibridación (*touchdown*) (Don *et al.* 1991; Paniego *et al.* 2002) con el siguiente ciclado: una desnaturalización inicial de 4 min a 94 °C, 24 ciclos de (a) 30 s a 94 °C, (b) 30 s a 64 °C (-0,5°C/ciclo, *touchdown* 64-52) y (c) 30 s a 72 °C, seguido por 35 ciclos de (a) 30 s a 94 °C, (b) 30 s a 52 °C y (c) 30 s a 72 °C, seguido por una extensión final de 10 min a 72 °C.

Los productos de PCR obtenidos fueron visualizados mediante cromatografía en geles de agarosa 1,5%, utilizando como método de detección un escáner de fluorescencia (Typhoon Trio, GE Healthcare Bio-Sciences, Reino Unido), a fin de corroborar la correcta incorporación de la marca fluorescente en los mismos. La genotipificación se realizó utilizando la técnica de electroforesis capilar con detector de fluorescencia inducida por láser (ECF, ABI 3130 XL, Applied Biosystem, EUA). A fin de economizar recursos, los fragmentos que poseían diferentes tamaños y/o llevaban distintas marcas fluorescentes, fueron combinados para realizar una co-electroforesis capilar.

La selección de los iniciadores para la PCR *multiplex* de los EST-SSR, se realizó utilizando el *software* libre Multiplex Manager (disponible en www.multiplexmanager.com). Los parámetros seleccionados incluyeron: un máximo de hasta tres fluoróforos por reacción (FAM, HEX y NED), un máximo de 7 *loci* por reacción, y una diferencia mínima de longitud entre alelos de *loci* con iniciadores portadores de marca del mismo color fijada en 20 pb. Las PCR se realizaron utilizando el *kit* Type-it Multiplex PCR Master Mix (QIAGEN, Alemania), siguiendo las instrucciones del fabricante, pero ajustando el volumen final a 10 μ L, a fin de optimizar recursos. El programa de ciclado utilizado fue el siguiente: desnaturalización inicial de 5 minutos a 95 °C, 28 ciclos compuestos por (a) 30 s a 95 °C, (b) 90 s a 60 °C y (c) 30 s a 72 °C, seguido de una extensión final de 30 minutos a 60 °C. Los conjuntos de marcadores utilizados para cada mezcla de reacción se muestran en Tabla 2.2.

Tabla 2.2 Conjuntos de iniciadores usados para PCR Multiplex

Pool	Marcador	Mínimo esperado	Máximo esperado	Fluoróforo
1	HT0761	126	156	NED
1	HT0292	311	367	HEX
1	HT0734	298	323	NED
1	HT0375	369	441	NED
1	C5783	140	230	FAM
1	HT0271	257	286	HEX
1	HT0723	265	345	FAM
2	HT0595	189	331	NED
2	HT0591	326	384	HEX
2	HT0382	377	407	NED
2	HT1024	257	282	FAM
2	HT0586	312	348	FAM
2	HT0998	383	406	FAM
3	HT0974	173	209	HEX
3	HT0572	301	344	FAM
3	HT1045	311	357	NED
3	HT0440	247	268	NED
3	HT0519	261	294	HEX
3	HT0426	328	356	HEX
3	C2070	235	271	FAM

Los tamaños esperados fueron obtenidos del trabajo de Mandel *et al.* (2011) y Chapman *et al.* (2008).

El análisis de los fragmentos obtenidos se realizó utilizando el *software* GeneMapper® 4.0 (Applied Biosystems®, EUA), usando un estándar comercial para la asignación del tamaño de los alelos (GeneScan ROX 500, Applied Biosystems®, EUA). La designación automática de los alelos fue confirmada luego de revisar en forma manual todos los electroferogramas.

Desarrollo del panel de 384 SNPs para el sistema VeraCode BeadXpress de Illumina

El desarrollo del panel 384 SNPs utilizado en este trabajo para la genotipificación con la tecnología BeadXpress (Illumina, EUA) fue llevado a cabo por el grupo Genómica de Girasol de INTA (Zubrzycki *et al.* 2012a; Filippi *et al.* 2013). Dicho panel se generó a partir de un análisis *in silico* de variantes alélicas usando una rutina informática desarrollada localmente que tomó como insumo la salida del programa CAP3 “*output.ace*”. Esta última fue obtenida como resultado del ensamblado de un total aproximado de 130000 secuencias de EST provenientes de diferentes genotipos de girasol cultivado disponibles en la base de datos dbEST de NCBI (www.ncbi.nlm.nih.gov), usado para construir el índice de genes “*Sunflower Unigene*” disponible en la base de datos ATGC-SUR (atgc-sur.inta.gob.ar, Fernandez *et al.* 2012a).

Los criterios que se tuvieron en cuenta para la selección del panel de SNPs a ser interrogados por el sistema VeraCode fueron los siguientes: que las posiciones variables se encontraran a un mínimo de 100 pb de los extremos de la secuencia, que los SNP se encontraran espaciados entre sí por un mínimo de 60 pb, y que la secuencia consenso entre SNPs tuviera un contenido de GC de entre 40 y 60 %. Además de estos criterios, se tuvieron en cuenta únicamente los *contigs* (genes tentativos) que resultaron del ensamblado de 5 o más secuencias de EST, y dentro de cada *contig*, únicamente las secuencias que presentaban una sola posición bialélica con frecuencia del alelo minoritario no inferior al 5%. Bajo estos criterios se encontraron 1283 SNPs presentes en 695 *contigs*. Todos ellos, sumados a un conjunto de 119 SNP caracterizados por re-secuenciación en líneas endocriadas de girasol (Kolkman *et al.* 2007; Fusari *et al.* 2008, 2012) fueron ordenados según el formato establecido por Illumina, y analizados con el programa *Assay Design Tool* (Illumina, EUA). Para el diseño final del panel de genotipificación de girasol (SOPA, *Sunflower Oligo Pool Assay*) se consideraron aquellos SNPs cuyo *design score*, parámetro utilizado por el sistema para predecir la eficiencia de genotipificación para un *locus* dado, fuera $\geq 0,6$, ya que Illumina establece que las regiones que poseen un valor superior a dicho umbral tienen una probabilidad alta de ser exitosas para la tecnología GoldenGate. Finalmente, para la selección definitiva de los 384 SNP se consideró la anotación funcional establecida para los *contigs* disponible en el repositorio ATGC-SUR (atgc-sur.inta.gob.ar) (Fernandez *et al.* 2012b), eligiéndose aquellos asociados a funciones de resistencia a estrés biótico. En total, el panel de 384 SNPs estuvo constituido por 97 SNPs definidos por resecuenciación, y 287 SNPs obtenidos por búsquedas *in silico*. Cabe destacar que estos 287 SNPs fueron obtenidos a partir del alineamiento de un número variable de EST, desde 5 hasta 65, con una media de 8 EST alineados por SNP. En la Figura 2.2 se presenta el número de EST considerados para la detección *in silico* de los SNPs que integran el SOPA.

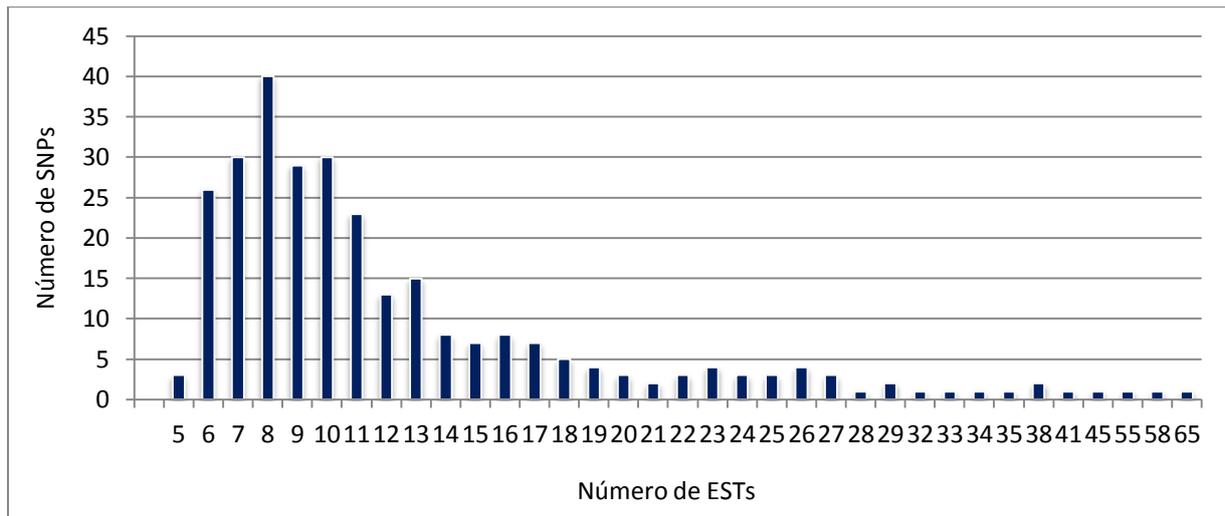


Figura 2.2 Distribución de los SNPs que integran el SOPA de acuerdo al número de EST considerados para la detección

Caracterización de la PMA utilizando marcadores tipo SNP

Para ahondar en la caracterización genética de las 137 líneas de la PMA, las mismas fueron también evaluadas utilizando el panel de 384 SNPs (SOPA), detallado en la sección previa “Desarrollo del panel de 384 SNPs para el sistema VeraCode BeadXpress de Illumina”.

La genotipificación de los SNPs fue llevada a cabo utilizando el equipo Illumina GoldenGate, BeadXpress (Illumina, San Diego, CA) perteneciente al Instituto de Biotecnología de INTA Castelar, siguiendo el protocolo provisto por el fabricante (Fan *et al.* 2006), utilizando como punto de partida 250 ng de ADN genómico por línea.

En la Figura 2.3 se presenta un esquema del ensayo de GoldenGate BeadXpress. El primer paso del ensayo consiste en la activación del ADN, mediante la cual las moléculas de ADN genómico se unen a partículas paramagnéticas (paso 1). Este ADN activado se incubó posteriormente con un *buffer* de hibridización y el conjunto de oligonucleótidos diseñados específicamente para cada SNP para la hibridización (paso 2). Por cada *locus* SNP se requiere de tres oligonucleótidos: dos alelo-específicos (ASO, *allele-specific oligo*) que difieren en su base 3' y un tercer oligonucleótido locus-específico (LSO, *locus-specific oligo*), que hibridiza río abajo del sitio SNP. De este modo, 1152 oligonucleótidos debieron ser diseñados para conducir un ensayo de 384 SNPs. Además de ser complementarios a la secuencia a amplificar, los tres oligonucleótidos/SNP poseen secuencias para el pegado de iniciadores universales; el LSO además contiene una secuencia única, denominada *address*, que permite la identificación del SNP. Durante la etapa de hibridización, los ASOs y LSOs se unen a las

moléculas de ADN genómico, que se encuentran previamente unidas a partículas paramagnéticas. Acto seguido, se procede al lavado de las mismas para la remoción del exceso de óligos no hibridados. La extensión de cada ASO y su posterior ligación a su LSO permite la unión de la información genotípica del SNP con su secuencia *address* contenida en el LSO (paso 3).

Este producto de ligación generado/SNP sirve como templado para realizar una PCR con iniciadores universales P1, P2 y P3 (paso 4). Los iniciadores P1 y P2 poseen las marcas fluorescentes Cy3 y Cy5, respectivamente. Luego del paso de PCR, se separan las dos hebras generadas, se elimina la hebra complementaria (paso 6) y aquellas hebras portadoras de marca se unen a su *Veracode bead* complementaria por la secuencia *address* (paso 7). Esta hibridación de los productos de PCR generados a *beads* específicas permite la lectura de cada genotipo de SNP en forma individual utilizando el lector BeadXpress, el cual es capaz de identificar el código de cada *bead* y de detectar de la señal fluorescente generada (paso 8).

Resulta interesante destacar que todo este proceso puede ser realizado en dos días de trabajo, y lleva a la generación de hasta 3684 datos genotípicos por ensayo de 384 SNPs por placa de 96 pocillos, *i.e.* 96 individuos diferentes.

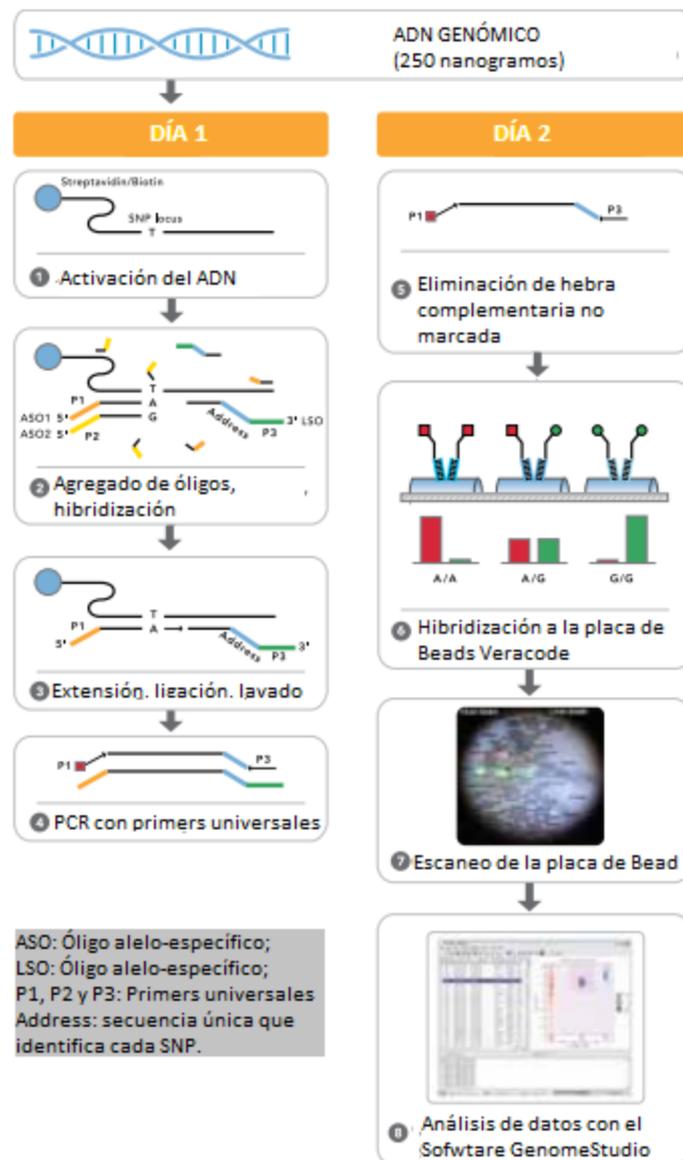


Figura 2.3 Esquema de trabajo para la genotipificación de una placa de 96 pocillos utilizando la tecnología GoldenGate BeadXpress de Illumina

Para evaluar la reproducibilidad del ensayo de genotipificación, se utilizó como control el ADN de cinco individuos. Los datos generados fueron analizados utilizando el *software* GenomeStudio (Illumina, EUA), determinándose un nivel de calidad para cada genotipo. Se utilizó un valor de corte de *GenCall* igual a 0,25 para establecer los genotipos válidos para cada SNP, mientras que sólo se consideraron como válidos aquellos SNPs que tuvieron un valor de *GenTrain* superior a 0,25, lo que representa un criterio estricto utilizado en la mayoría de los estudios genéticos reportados (Lepoittevin *et al.* 2010). Los valores de *GenCall* y *GenTrain* miden la confiabilidad en la asignación del SNP, basándose en la distribución de las clases genotípicas (AA, AB y BB). Las agrupaciones se inspeccionaron

visualmente para asegurar la obtención de datos de alta calidad. En la Figura 2.4 se presentan las tres posibles salidas del análisis de SNPs: SNP polimórfico (A), SNP monomórfico (B) o falla en el agrupamiento (C).

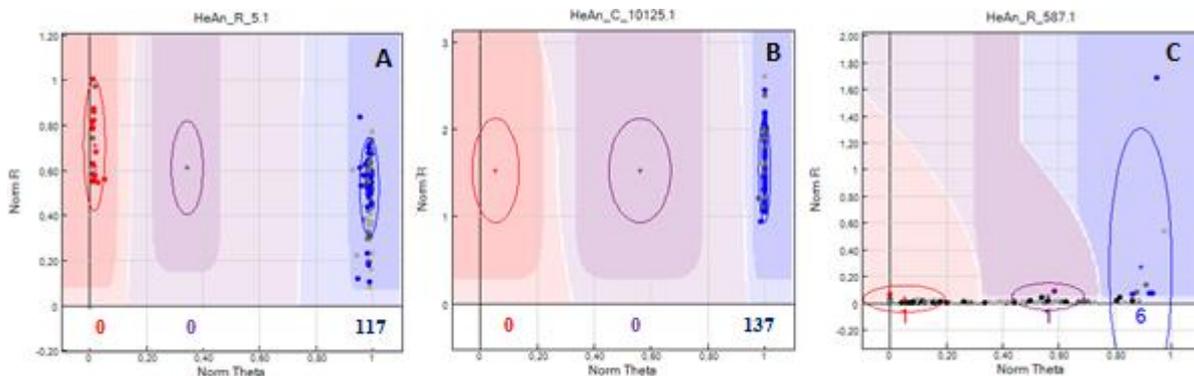


Figura 2.4 Salidas del software GenomeStudio (A) SNP polimórfico; (B) SNP monomórfico; (C) Falla en el agrupamiento.

Análisis de la diversidad genética

Las medidas de diversidad genética, incluyendo número total de alelos, número promedio de alelos por locus (A), heterocigosis esperada (H_e) (Nei 1978), heterocigosis observada (H_o), y contenido de Información Polimórfica (PIC) fueron estimados a partir de los *sets* de datos derivados de los marcadores SSR y SNP, utilizando el *software* PowerMarker v. 3.51 (Liu & Muse 2005).

La probabilidad de identidad (PI), probabilidad de identidad considerando similitud entre hermanos (PI_{sibs}) (Waits *et al.* 2001) y la frecuencia del alelo minoritario (MAF) fueron calculados en toda la matriz de datos utilizando GenAlEx (Peakall & Smouse 2006).

Análisis de la estructura poblacional y distancia genética

La presencia de estructura poblacional en los materiales en estudio se evaluó mediante un análisis molecular de la varianza (AMOVA; Excoffier *et al.* 1992), utilizando el *software* GenAlEx (Peakall & Smouse 2006). La significación estadística se evaluó haciendo 999 permutaciones.

La existencia de estructura poblacional se analizó, además, utilizando el método bayesiano implementado en el *software* STRUCTURE (Pritchard *et al.* 2000a). Para cada análisis, el número de grupos genéticos (K) evaluados fue de 1 a 20, con 5 corridas por valor de K. Para

cada corrida, el período de *burn-in* seleccionado fue de 500000 iteraciones, con 500000 iteraciones posteriores de la cadena de Markov (MCMC, *Markov Chain Monte Carlo*). Se utilizó el modelo de *admixture* (mezcla), considerando frecuencias alélicas independientes, y sin incorporar información previa acerca del origen de los individuos (Falush *et al.* 2003). Para determinar el valor de K más probable se utilizó el método de deltaK descrito por Evanno *et al.* (2005), implementado en Structure Harvester (Earl & VonHoldt 2011). Los resultados de STRUCTURE fueron plasmados utilizando el *software* Distruct (Rosenberg 2003). Se realizaron análisis de correlación de Spearman entre los diferentes grupos inferidos por STRUCTURE, utilizando el *software* Infostat (Di Rienzo *et al.* 2008).

Las relaciones genéticas entre la líneas que conforman la PMA fueron también examinadas aplicando el Análisis Discriminante de Componentes Principales (*Discriminant Analysis of Principal Components*, DAPC; Jombart *et al.* 2010), usando el paquete Adegenet (Jombart 2008) para el *software* R 3.0.2 (R *development Core Team*, 2013). Se ejecutó la función DAPC usando los grupos (o *clusters*) identificados por *k-means* (Legendre & Legendre 1998). Se determinó el número de grupos utilizando la función '*find.clusters*', evaluándose un rango de valores de 1 a 40. El número óptimo de grupos fue elegido en base al criterio de información bayesiano (BIC). La contribución de los alelos individuales a la estructura poblacional fue estimada utilizando la función '*loadingplot*' del paquete Adegenet (Jombart 2008).

Las medidas de distancia genética entre las líneas que conforman la PMA fueron estimadas utilizando la proporción de alelos compartidos. Las correlaciones entre matrices de distancia fueron analizadas utilizando el *test* de Mantel implementado en el *software* GenAlEx (Peakall & Smouse 2006).

Finalmente, se construyeron árboles de Neighbor-Joining (NJ) en base a las distancias genéticas calculadas entre pares de entradas. Los análisis de agrupamiento y re-muestréos de *bootstrap* (1000 pseudo-réplicas) fueron realizados utilizando el *software* libre PowerMarker 3.25 (Liu & Muse 2005). El soporte de cada rama se visualizó usando el programa Consense incluido en el *software* PHYLIP v. 3.68 (Felsenstein 2005). Los árboles resultantes fueron visualizados y editados utilizando el programa FigTree versión 1.3.1 (Morariu *et al.* 2008).

En el caso de la PMA, todos los análisis precedentes se realizaron considerando la información obtenida utilizando los siguientes *sets* de marcadores: SSR, SNP y SSR+SNP.

RESULTADOS

Estudio de la diversidad genética usando marcadores tipo SSR

Un total de 170 entradas de girasol correspondientes a la PMA (137 entradas, 3 individuos por cada una), CP y OP (33 entradas, 6 y 9 individuos por categoría, respectivamente) fueron analizadas usando 42 marcadores tipo SSR. El porcentaje de datos faltantes correspondió al 4,57 % del total de la matriz de datos. Para el panel completo de entradas, la probabilidad de identidad (PI) fue igual a $3,5 \times 10^{-27}$, la PI_{sibs} fue estimada en $3,3 \times 10^{-12}$, y el contenido promedio de información polimórfica (PIC) fue de 0,50. Asimismo, el número total de alelos identificado fue de 208, con un rango que varió entre 2-14 por *locus*, y un promedio de alelos por *locus* de 4,95. La heterocigosis esperada promedio (He) fue de $0,51 \pm 0,16$ (media \pm error estándar). Del total de 208 alelos identificados, 10 fueron privados o únicos de la PMA. En contraste, se identificaron 36 alelos privados para el grupo formado por OP+CP. En total, la PMA y las OP+CP comparten 162 alelos.

Para el análisis de diversidad dentro de la PMA, la misma fue subdividida en cuatro categorías diferentes, basadas en su origen geográfico (Argentina u otro origen) e historia en programas de mejoramiento (HA: Mantenedoras; RHA: Restauradoras). Se identificaron 25 alelos privados para las líneas mantenedoras (HA) y 16 para las restauradoras (RHA), mientras que se identificaron 36 alelos privados para el germoplasma argentino al compararlo con el germoplasma de otro origen.

Las estimas de diversidad genética calculadas para todas las entradas y categorías analizadas se detallan en Tabla 2.3.

Tabla 2.3 Resumen de las estimas de diversidad genética obtenidas para las entradas de girasol incluidas en este estudio

Marcador	Muestra	N ent.	N ind.	A	a	He	Ho
SSR (todos)	Total	169	646	208	4,95 \pm 2,60	0,51 \pm 0,17	0,06 \pm 0,04
	OP + CP	33	235	198	4,71 \pm 2,50	0,52 \pm 0,16	0,09 \pm 0,06
	PMA	137	411	172	4,09 \pm 2,16	0,48 \pm 0,17	0,01 \pm 0,03
	HA	59	177	156	3,71 \pm 1,91	0,44 \pm 0,19	0,01 \pm 0,03
	RHA	78	234	147	3,50 \pm 1,76	0,46 \pm 0,16	0,01 \pm 0,03
	Argentina	104	312	167	3,97 \pm 2,16	0,47 \pm 0,17	0,01 \pm 0,03
	Otro origen	33	99	136	3,24 \pm 1,69	0,47 \pm 0,19	0,01 \pm 0,03
gSSR	Total	169	646	121	5,50 \pm 3,20	0,56 \pm 0,14	0,07 \pm 0,03
	OP + CP	33	235	113	5,14 \pm 3,06	0,56 \pm 0,13	0,10 \pm 0,05

Marcador	Muestra	N ent.	N ind.	A	a	He	Ho
	PMA	137	411	102	4,63±2,68	0,53±0,01	0,00±0,00
	HA	59	177	89	4,05±2,36	0,48±0,19	0,00±0,01
	RHA	78	234	85	3,86±2,16	0,51±0,13	0,00±0,00
	Argentina	104	312	99	4,50±2,68	0,52±0,14	0,00±0,00
	Otro origen	33	99	78	3,55±2,11	0,54±0,14	0,00±0,01
EST-SSR	Total	169	646	87	4,35±1,60	0,46±0,19	0,05±0,05
	OP + CP	33	235	85	4,25±1,65	0,47±0,18	0,08±0,06
	PMA	137	411	70	3,50±1,19	0,42±0,19	0,02±0,04
	HA	59	177	67	3,35±1,22	0,39±0,19	0,02±0,03
	RHA	78	234	62	3,10±1,07	0,41±0,18	0,03±0,04
	Argentina	104	312	68	3,40±1,19	0,42±0,18	0,02±0,04
	Otro origen	33	99	58	2,90±1,02	0,40±0,20	0,03±0,04

N ent.: Número de entradas de girasol; N Ind.: Número de individuos analizados; A: Número total de alelos; a: Número promedio de alelos por *locus*; He: Heterocigosis esperada, Ho: Heterocigosis observada.

La lista completa de datos genotípicos generados a partir de los marcadores SSR se encuentra en http://bioinformatica.inta.gob.ar/download_files/tesis_filippi/

Diversidad en la PMA utilizando marcadores SNP

La PMA fue caracterizada con mayor profundidad utilizando el panel de 384 SNPs desarrollado para el sistema BeadXpress de Illumina. Del total de *loci* analizados, fueron eliminados aquellos marcadores que resultaron monomórficos (80/384), y los que mostraron más del 10 % de datos perdidos, o generaron datos ambiguos (122/384). Así, el *set* de datos resultante quedó conformado por 182 marcadores altamente informativos que fueron reportados en el repositorio público ATGC-SUR (Tabla 2.4). La proporción media de datos faltantes fue de 0,91 %. La PI para los SNPs resultantes fue de $1,0 \times 10^{-46}$ y el PI_{sibs} fue $3,3 \times 10^{-24}$. Se estimó el PIC promedio en 0,232. En la Tabla A2.2 del Apéndice se presenta la lista y origen de los SNPs incluidos en el ensayo de 384 SNPs que resultaron polimórficos en las líneas de la PMA.

La lista completa de datos genotípicos generados a partir de los marcadores SNP se encuentra en http://bioinformatica.inta.gob.ar/download_files/tesis_filippi/

Tabla 2.4 Tasa de éxito del ensayo de genotipificación

CATEGORIA	# de SNPs	% de SNPs
Polimórfico	182	47,4
Monomórfico	80	20,83
Falla	122	31,77
Total	384	100

La Figura 2.5 muestra una comparación gráfica de la distribución de frecuencias alélicas obtenidas para SSR y SNP.

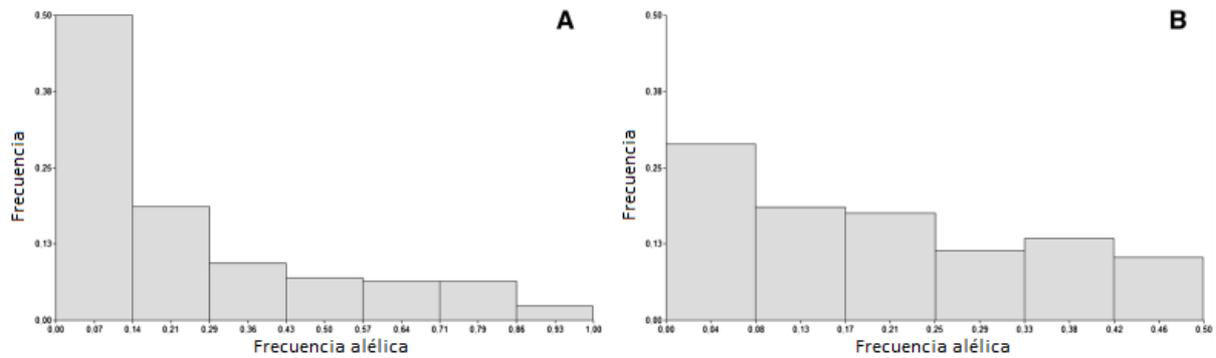


Figura 2.5 Distribuciones de las frecuencias alélicas observadas (A) Para los 42 SSRs; (B) Para los 182 SNPs.

La frecuencia del alelo minoritario (MAF) fue mayor a 0,01 para el 91,76 % de los 182 SNP polimórficos. La heterocigosis esperada ($He = 0,29$) fue menor para los datos moleculares derivados del conjunto de SNPs que para los SSRs. Como se esperaba, la heterocigosis observada obtenida fue baja para ambos conjuntos de marcadores (0,01 y 0,03 para SSR y SNP, respectivamente). Estas estimas se resumen en la Tabla 2.5.

Tabla 2.5 Resumen de las estimas genéticas obtenidas para el panel de Mapeo por Asociación usando marcadores tipo SNP

Muestra	N ent.	A	a	He	Ho
PMA	137	364	2,00±0,00	0,29±0,17	0,03±0,03
HA	59	354	1,95±0,23	0,29±0,17	0,02±0,03
RHA	78	347	1,90±0,29	0,27±0,17	0,03±0,03
Argentina	104	359	1,97±0,16	0,28±0,17	0,03±0,03
Otro origen	33	349	1,92±0,27	0,28±0,17	0,03±0,03

N ent.: Número de entradas de girasol; A: Número total de alelos; a: Número medio de alelos por *locus*; He: Heterocigosis esperada, Ho: Heterocigosis observada. Las entradas de girasol fueron agrupadas de acuerdo a las categorías descritas en “Materiales y Métodos”.

Estructura poblacional

Se condujo un análisis molecular de la varianza (AMOVA) para poner a prueba la existencia de posibles diferencias entre los grupos predefinidos en este trabajo: PMA vs. OP+CP, y entre los subgrupos en los cuales la PMA fue subdividida *a priori* (“HA” vs. “RHA”, “Argentina” vs. “Otro origen”). Se encontraron diferencias significativas entre AMP y OP+CP

($F_{ST} = 0,025$, $p < 0,001$, Tabla A2.3). Dentro de la PMA, los análisis se realizaron utilizando los conjuntos de datos: SSR, SNP y SSR+SNP. En cada caso, el AMOVA reveló diferencias significativas entre los grupos en los que la PMA estaba delimitada, aunque esas diferencias solo explicaban el 2-3% de la varianza total (Tabla 2.6), quedando la varianza restante dividida entre los componentes dentro de grupos y entre individuos. Los resultados completos se presentan en las Tablas A2.3 (SSR), A2.4 (SNP) y A2.5 (SSR+SNP) del Apéndice.

Tabla 2.6 Análisis de la varianza molecular (AMOVA)

Marcador	Comparación	Fst	P-valor
SSR	PMA vs. OP+CP	0,025	0,001
	Argentina vs. "Otro origen"	0,028	0,001
	HA vs. RHA	0,016	0,001
SNP	Argentina vs. "Otro origen"	0,018	0,001
	HA vs. RHA	0,003	0,032
SSR+SNP	Argentina vs. "Otro origen"	0,021	0,001
	HA vs. RHA	0,007	0,002

Como primera aproximación al análisis bayesiano de la estructura poblacional, se evaluaron las 137 líneas que componen la PMA, conjuntamente con las OP y CP, utilizando únicamente los marcadores SSR, ya que las OP y CP no fueron genotipificadas con el panel de SNP. Dado que el logaritmo de la verosimilitud incrementaba progresivamente con el aumento de los valores de K, se utilizó el método de Evanno *et al.* (2005) como criterio para inferir el valor de K más probable. El valor máximo de ΔK fue detectado a $K = 2$, con un segundo máximo a $K = 5$. A pesar de que se distingue una clara señal de estructura poblacional, es difícil determinar el valor de K más probable, ya que no es posible identificar ninguna característica morfológica, agronómica o de origen, que permita explicar la constitución de los grupos inferidos con $K = 2$ o $K = 5$ (Figuras 2.6, A, B, y A2.1 del Apéndice). El análisis de la gráfica de DAPC obtenida también revela la presencia de estructura genética entre estas entradas. En este caso, el algoritmo secuencial *k-means* identificó 14 grupos, mientras que los *eigenvalues* mostraron que la estructura genética era capturada por las tres primeras Componentes Principales (CPs, Figura 2.6, C). Nuevamente, no pudo identificarse evidencia morfológica o agronómica que esté asociada a los grupos hallados.

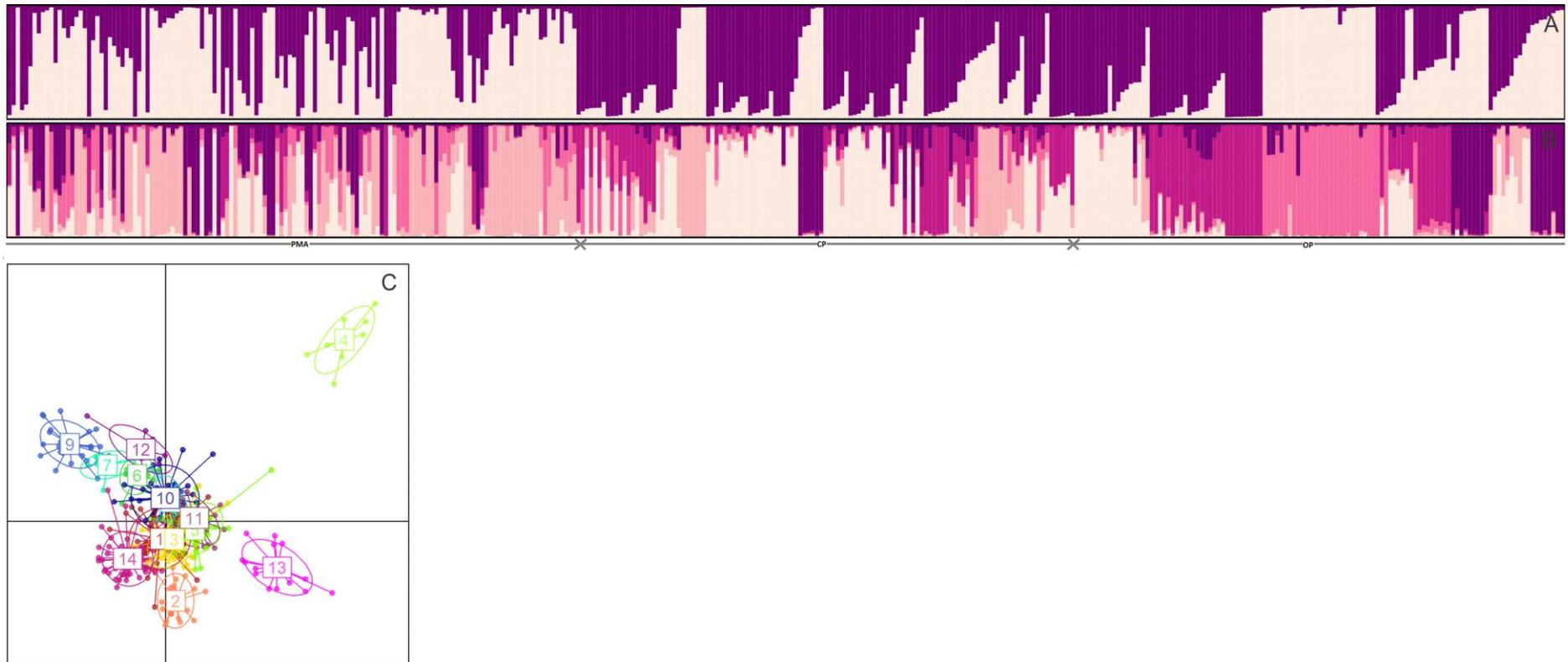


Figura 2.6 Estimaciones de estructura poblacional para el conjunto total de entradas de girasol (PMA, CP y OP) (A) Resultados de STRUCTURE para K = 2 ordenado por tipo de entrada; (B) Resultados de STRUCTURE para K = 5 ordenado por tipo de entrada; (C) Gráficos de dispersión de DAPC mostrando las dos primeras componentes principales.

La disponibilidad de los genotipos SNP para la PMA permitió evaluar el desempeño de los diferentes *sets* de marcadores (SSRs, SNPs y SSRs+SNPs) en la predicción de estructura poblacional. El método de Evanno *et al.* (2005) logró identificar tres picos a $K = 2$, $K = 3$, $K = 5$, para los datos derivados de SSR y SNP; y $K = 2$, $K = 3$, $K = 4$ para el conjunto de SSR+SNP, siendo el pico de $K = 2$ el máximo para SNP y SSR+SNP, y el de $K = 3$ el mayor para SSR (Apéndice, Figura A2.2). Dado que los picos de ΔK en $K = 2$ suelen considerarse un artefacto de la técnica (Vigouroux *et al.* 2008), y que los tres *sets* de datos mostraron picos a $K = 3$, la representación gráfica de la estructura poblacional se realizó utilizando la separación de poblaciones en $K = 3$ (Figura 2.7, A-C). El porcentaje de individuos asignados a cada población, con coeficiente de pertenencia determinado arbitrariamente como $>0,70$, fue menor para los SNPs que para los otros dos conjuntos de marcadores, independientemente del valor de K considerado (Tabla 2.7).

Tabla 2.7 Porcentaje de individuos asignados a poblaciones (coeficiente de pertenencia $>0,70$) utilizando diferentes *sets* de marcadores moleculares

Marcadores	k = 2	k = 3	k = 4	k = 5
SSRs	83,94	67,88	-	69,34
SNPs	67,88	52,55	-	48,91
SSR+SNP	77,37	57,66	59,12	-

Se observa que los grupos 1 (violeta) y 3 (verde) están compuestos principalmente por líneas mantenedoras, mientras que las líneas restauradoras fueron agrupadas principalmente en el grupo 2 (celeste). Para evaluar la consistencia de las asignaciones de individuos a cada grupo utilizando los distintos conjuntos de marcadores, se realizaron correlaciones de Spearman entre los coeficientes de pertenencia obtenidos por STRUCTURE para cada conjunto de marcadores. Las correlaciones entre las salidas generadas utilizando los datos derivados de SSR y SNPs fueron significativas para los tres grupos ($R_{G1} = 0,6$; $R_{G2} = 0,51$ y $R_{G3} = 0,49$; $p < 0,0001$). También se obtuvo correlación significativa al comparar SSR con SSR+SNP ($R_{G1} = 0,7$; $R_{G2} = 0,65$ y $R_{G3} = 0,68$; $p < 0,0001$) y entre SNP vs. SSR+SNP ($R_{G1} = 0,93$; $R_{G2} = 0,83$ y $R_{G3} = 0,82$; $p < 0,0001$). Las medidas de diferenciación entre los grupos genéticos identificados por STRUCTURE, estimadas a partir del índice “distancias nucleotídicas”, estuvieron en el rango de 0,13 a 0,16 para los grupos 1 y 2, de 0,12 a 0,14 para los grupos 1 y 3, y entre 0,06 a 0,09 para los grupos 2 y 3, dependiendo del conjunto de datos considerados.

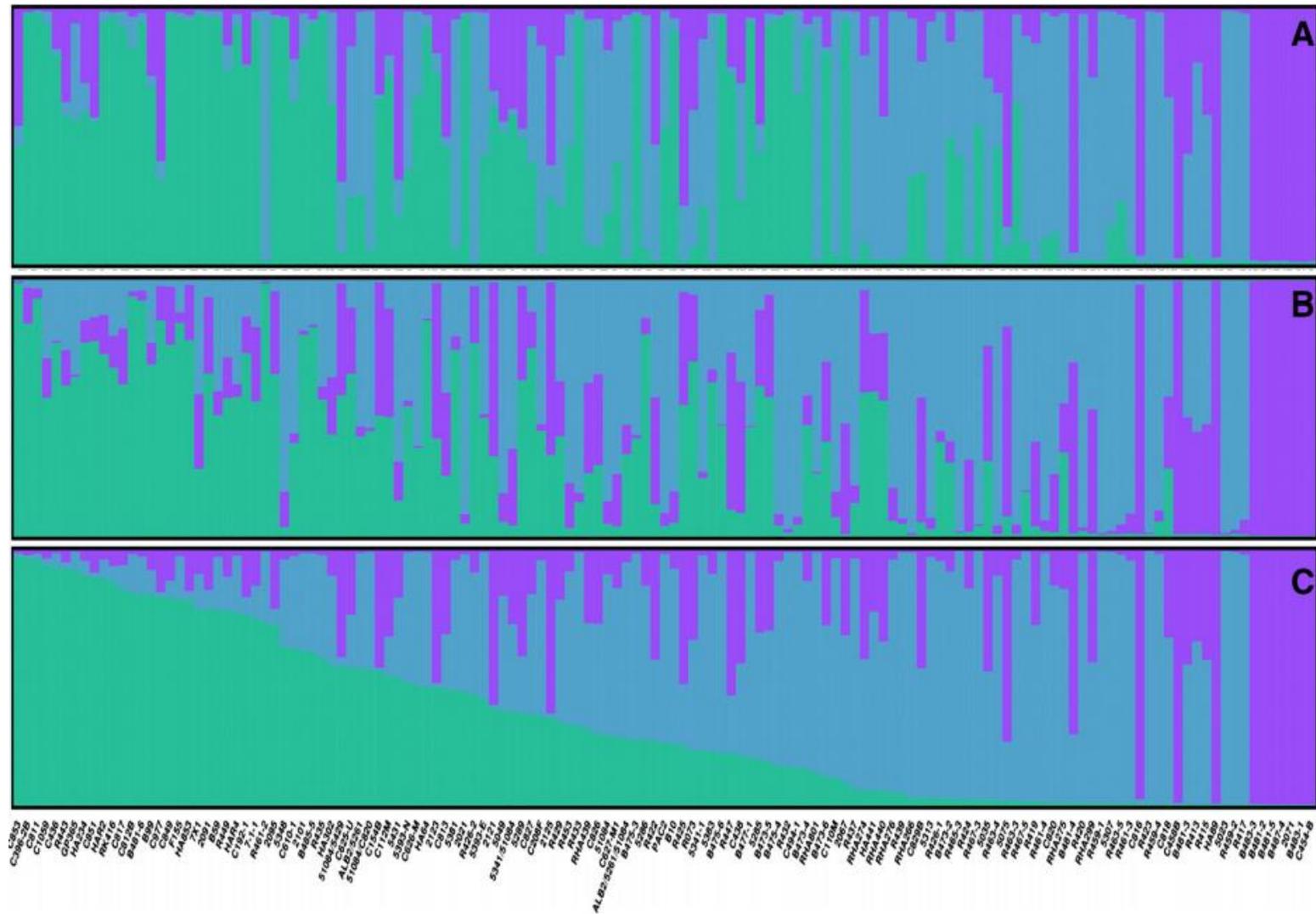


Figura 2.7 Resultados de STRUCTURE para K = 3 (A) Estructura poblacional en la PMA estimada con SSR; (B) Estructura de la PMA estimada con SNP; (C) Estructura poblacional en la PMA estimada con SSR+SNP.

El método de DAPC también reveló la presencia de estructura genética entre los materiales que conforman la PMA. En concordancia con el análisis de STRUCTURE, el algoritmo *k-means* logró identificar tres grupos, independientemente del *set* de datos utilizado (Figura 2.8, A-C, y Tabla A2.6, Apéndice).

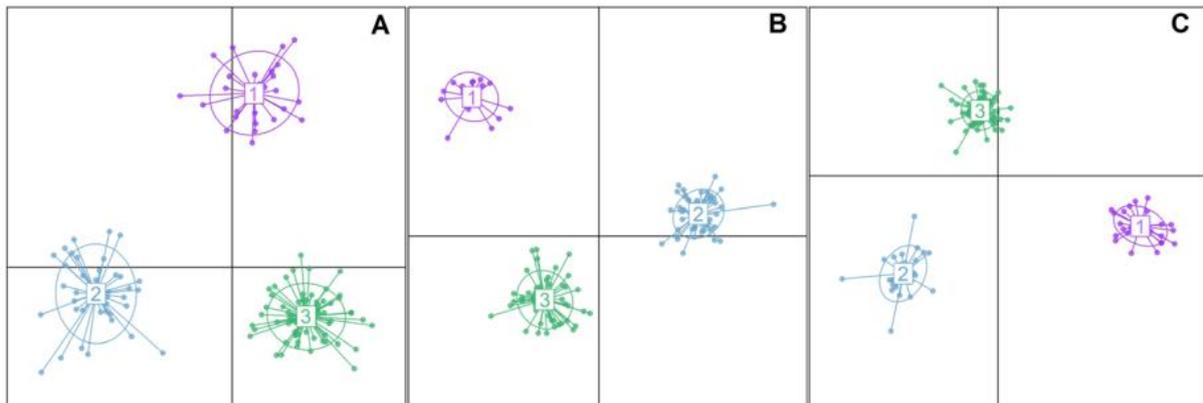


Figura 2.8 Gráficos de dispersión de DAPC mostrando las dos primeras componentes principales Se presentan los tres *sets* de datos. (A) Estructura poblacional en la PMA estimada con SSR; (B) Estructura de la PMA estimada con SNP; (C) Estructura poblacional en la PMA estimada con SSR+SNP.

Para evaluar la correspondencia entre los grupos obtenidos utilizando la aproximación bayesiana y el DAPC, se computó el porcentaje de individuos asignados a un grupo dado por STRUCTURE que fueron asignados al mismo grupo por DAPC (Tabla 2.8).

Tabla 2.8 Porcentaje de individuos asignados al mismo grupo usando STRUCTURE y DAPC

	SSR			SNP			SSR+SNP		
	Grupo 1 STR	Grupo 2 STR	Grupo 3 STR	Grupo 1 STR	Grupo 2 STR	Grupo 3 STR	Grupo 1 STR	Grupo 2 STR	Grupo 3 STR
DAPC	100	79,48	97,56	100	53,84	86,36	100	63,16	88,88

STR: Structure

Un aspecto interesante del método DAPC es que permite la identificación de aquellos alelos que contribuyen más a la delimitación entre grupos. Para obtener mayor conocimiento de las causas que generan la separación entre los grupos detectados para la PMA, se realizó una inspección de los alelos asociados para los datos de SNPs, ya que son los únicos para los cuales se cuenta con anotación funcional. Se identificaron 13 SNPs como los de mayor contribución: SNP 30, 34, 44, 69, 72, 105, 116, 168, 178, 192, 193 (ambos alelos) y SNP 139 y 147 (1 alelo). La representación gráfica de los mismos se presenta en la Figura 2.9. Al analizar las anotaciones de *gene ontology* (GO) asociadas a cada marcador, puede observarse que siete de ellas estaban relacionados con la categoría “procesos metabólicos”.

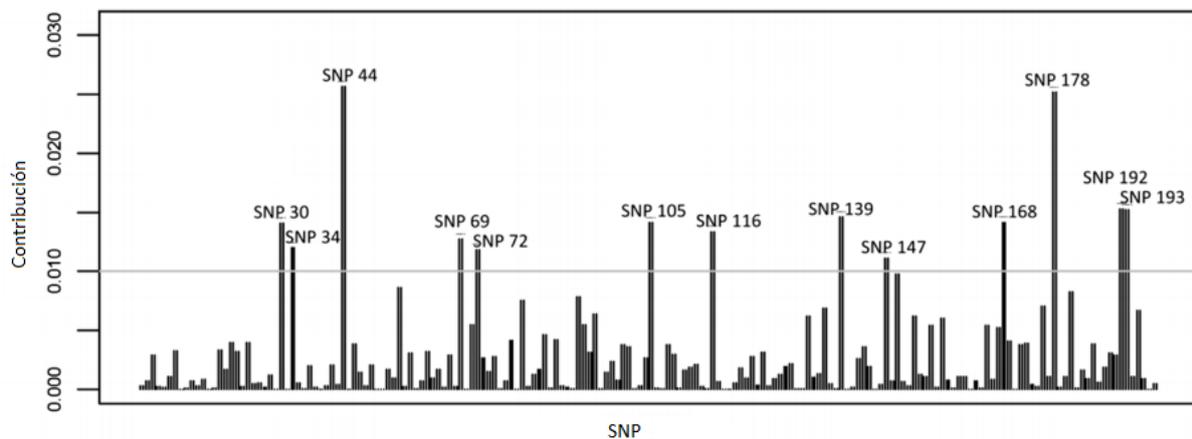
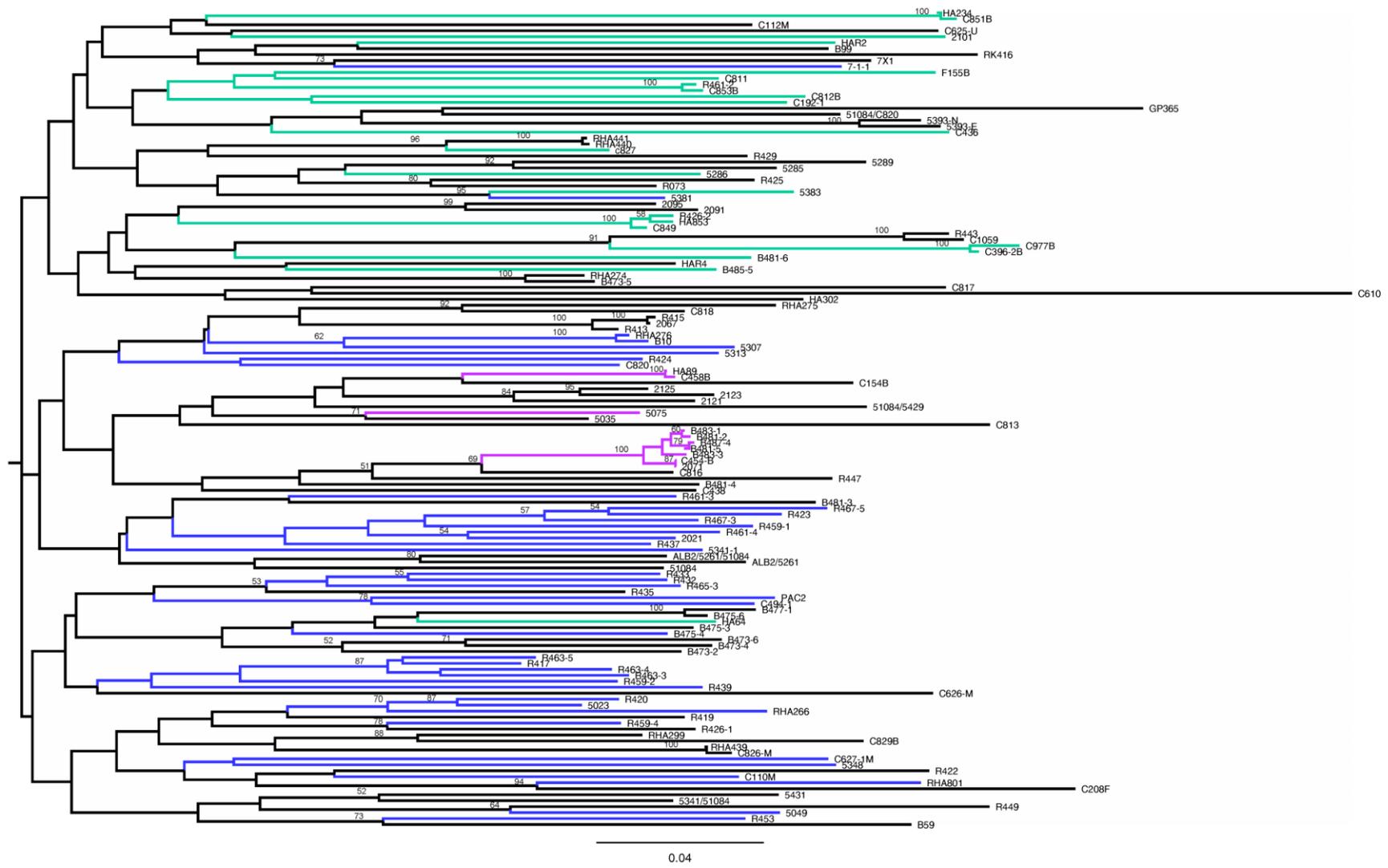


Figura 2.9 Contribución de los alelos de SNPs a la diferenciación entre grupos dentro de la PMA La altura de cada barra es proporcional a la contribución del correspondiente alelo. Solo se indican aquellos alelos cuya contribución supera el límite arbitrario de 0,010 (línea gris horizontal). Se muestran los datos para la primera función discriminante.

Para el análisis de agrupamiento, se construyeron matrices de distancia basadas en el porcentaje de alelos compartidos para todos los pares de individuos usando los datos derivados de la genotipificación de SSR y SNPs. Para el *set* de datos SSR, las distancias variaron de 0,012 a 0,78, con un promedio de 0,47. Para el *set* de datos SNP, se obtuvieron distancias en el rango de 0,003 a 0,45 con un promedio de 0,28. Se observó correlación significativa entre las estimas de distancia génica basadas en SSR y SNPs, observándose un coeficiente de correlación de Spearman de $r = 0,419$ (Test de Mantel, $p < 0,05$).

Finalmente, las relaciones entre las 137 líneas endocriadas que conforman la PMA fueron graficadas utilizando el método de *Neighbor-Joining* (Figura 2.10, A-C). A fin de facilitar las comparaciones, se colorearon los individuos de acuerdo a la salida obtenida para el mismo *set* de marcadores moleculares utilizando el *software* STRUCTURE. A diferencia de los otros métodos utilizados (DAPC, STRUCTURE) no se observa en este caso una clara agrupación de líneas.

A



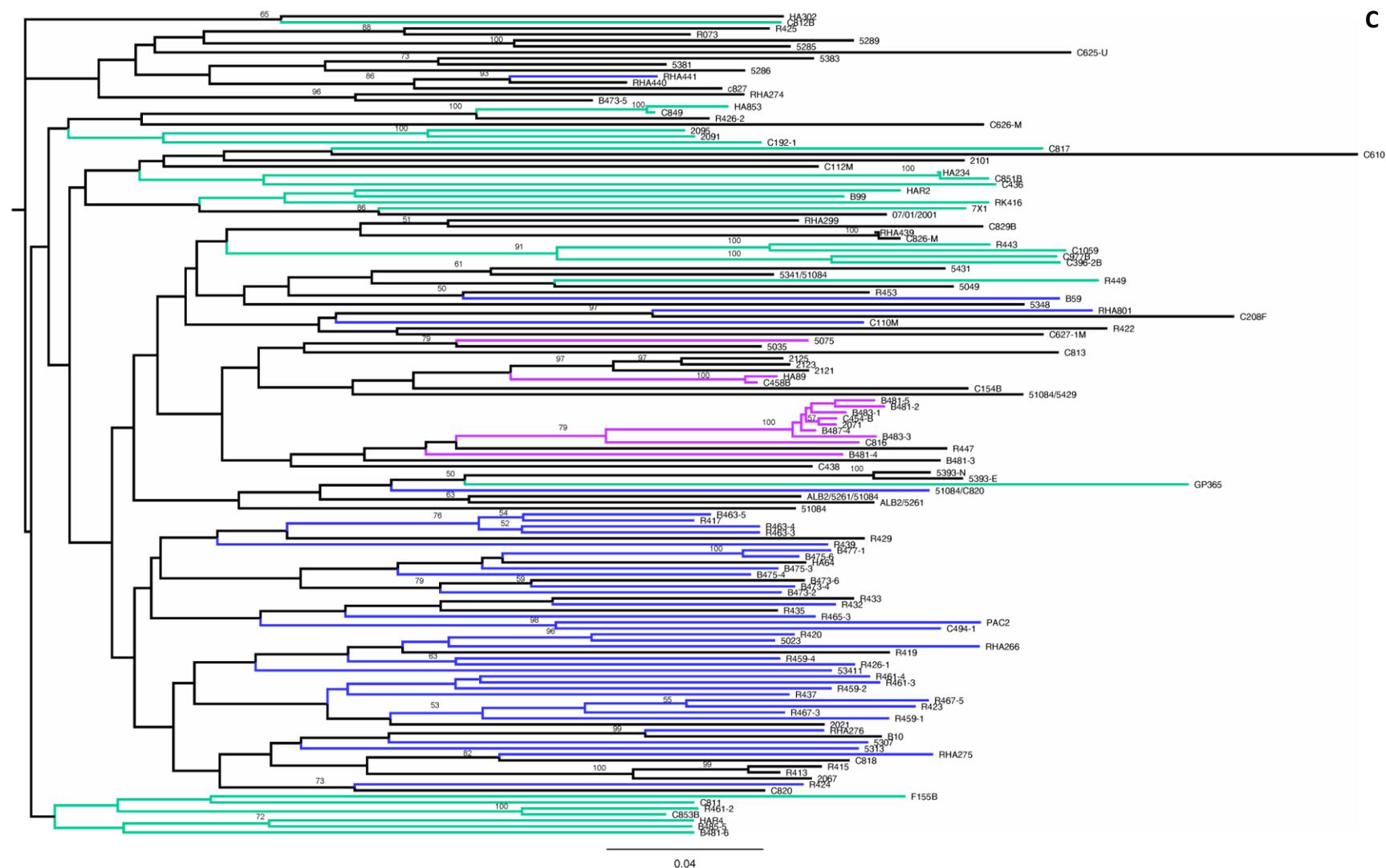


Figura 2.10 Filogramas de Neighbor-Joining para las 137 líneas endocriadas que conforman la PMA. Los genotipos fueron coloreados en base a los resultados de STRUCTURE para $K = 3$. (A) Derivados de SSR; (B) Derivados de SNP; (C) Derivados de SSR+SNP. Se indican los valores de *bootstrap* mayores a 50.

DISCUSIÓN

La Argentina tiene una larga tradición en mejoramiento de girasol, y su germoplasma es un recurso genético invaluable en el mundo, ya que muchas de las líneas de características diferenciales obtenidas a nivel internacional derivan de variedades argentinas (Gulya & Masirevic 1995; Bertero de Romano 2003).

Las líneas endocriadas incluidas en el presente trabajo son una parte esencial del programa de mejoramiento de girasol de INTA, ya que constituyen la Población de Mapeo por Asociación para caracteres complejos. En este estudio también se incluyó una selección de OP y CP relacionadas, a fin de comparar los niveles de diversidad genética contenidos en la PMA con la diversidad potencial presente en el germoplasma preservado en el BAG-IM.

En la mayoría de los estudios genético-poblacionales, las estimas de estructura genética y diversidad suelen hacerse utilizando información derivada de marcadores tipo SSR. Los SSR genómicos (gSSR) son marcadores atractivos para estudios poblacionales debido a su abundancia relativa, reproducibilidad y altos niveles polimórficos. Recientemente, el uso de EST-SSR se ha visto incrementado, debido a que pueden ser fácilmente obtenidos mediante búsquedas *in silico* en bases de datos de EST. Al pertenecer a regiones codificantes del ADN, se espera que este tipo de marcadores estén más conservados que los gSSR. Sin embargo, pocos estudios han documentado las diferencias entre ellos para la obtención de estimas de variabilidad genética (Gadaleta *et al.* 2011; Hu *et al.* 2011; Missio *et al.* 2011; Nandha & Singh 2014). Las comparaciones realizadas entre gSSR y EST-SSR sobre el panel completo de entradas de girasol y sobre la PMA, confirman que los marcadores tipo gSSR capturan mayores niveles de diversidad que los EST-SSR (medidos como número total de alelos, H_e y PIC). Estos resultados son consistentes con los reportados por Hu *et al.* (2011) al comparar el desempeño de gSSR y EST-SSR para obtener estimas de diversidad genética en otra especie no modelo, como lo es el coco.

Todos los marcadores SSR seleccionados para este análisis fueron exitosamente amplificados en el panel de entradas de girasol evaluadas. En el caso de los SNPs, el 68 % de los *loci* representados en el panel de Illumina resultaron exitosos. Las fallas en el 32 % remanente pueden ser atribuidas al origen de los SNPs incluidos en el panel, ya que fueron obtenidos principalmente mediante búsquedas *in silico* a partir de bases de datos de EST (Zubrzycki *et al.* 2012, Filippi *et al.* 2013). Estudios previos en otras especies (Wang *et al.*

2008; Lepoittevin *et al.* 2010), han descrito también fallas frecuentes en la genotipificación de SNPs derivados de EST, siendo las mismas el resultado de tres causas principales: (1) errores de secuenciación y consecuente identificación de falsos positivos; (2) baja calidad de las secuencias que flanquean al SNP; y (3) la existencia de uniones intrón-exón en las proximidades del SNP seleccionado. Sin embargo, el porcentaje de SNPs de buena calidad obtenidos en este trabajo (68 %, al considerar SNP polimórficos y monomórficos) no es bajo, si se compara con otros paneles de SNPs desarrollados para especies no modelo a través de búsquedas *in silico* (42 % en pino marítimo, Lepoittevin *et al.* 2010, y 66,1 % en eucalipto, Grattapaglia *et al.* 2011).

El análisis de los polimorfismos en el conjunto de entradas de girasol evaluado mostró que tanto los marcadores microsatélites como los SNPs resultaron informativos. Para evaluar la capacidad discriminante de los paneles de marcadores, se estimaron las PI y PI_{sibs} . Dentro de la PMA, la PI para los 42 loci SSR fue $3,5 \times 10^{-27}$, y la $PI_{sibs} = 3,3 \times 10^{-12}$. Para los 182 SNPs polimórficos, la PI fue $1,0 \times 10^{-46}$, y la $PI_{sibs} = 3,3 \times 10^{-24}$, demostrando que ambos paneles de marcadores permiten una gran capacidad discriminante para ser usados sobre colecciones de germoplasma de girasol, siendo el conjunto de marcadores SNP el más informativo. Yu *et al.* (2009) plantearon que se necesitan 10 veces más SNPs que SSR para obtener el mismo nivel de información genética, mientras que Van Inghelandt *et al.* (2010) propusieron un rango de entre 7 y 11 veces. En este estudio, un total de 109 SNPs elegidos al azar fueron suficientes para obtener el mismo PI que los 42 SSR, sugiriendo que si bien se necesita un mayor número de SNPs que de SSR para obtener el mismo contenido informativo, la relación entre el número de SNPs y SSR es dependiente de las características de los marcadores y de las especies consideradas.

De acuerdo a lo esperado, la distribución de frecuencias alélicas difirió entre SSR y SNPs. Se observó una mayor presencia de alelos de SSR a bajas frecuencias, mientras que los SNPs mostraron más alelos a frecuencias intermedias. Este espectro de distribución de frecuencias alélicas es consistente con lo reportado en estudios previos (Hamblin *et al.* 2007; Yang *et al.* 2011; Würschum *et al.* 2013), dado que los SSR están comúnmente dominados por alelos raros.

La brecha entre las estimas de H_e obtenidas en este estudio para SSR y SNP (0,52 y 0,29, respectivamente) parece ser menor a la brecha observada en otros cultivos, como uva (0,81-0,34, Emanuelli *et al.* 2013), maíz (0,80-0,32, Hamblin *et al.* 2007); y soja (0,77- 0,35, Li

et al. 2010). Estas diferencias están mayormente causadas por las estimas de He relativamente bajas obtenidas para los datos derivados de marcadores tipo SSR en este trabajo, probablemente subestimadas como consecuencia de haber incluido marcadores EST-SSR.

A la fecha, se han reportado resultados contradictorios respecto a la existencia de correlación entre las estimas de distancia genética obtenidas con SSR y SNP. Jones *et al.* (2007) y Hamblin *et al.* (2007) no encontraron correlación significativa entre las medidas de distancia genética obtenidas en poblaciones de maíz, excepto para aquellos individuos estrechamente relacionados. Por su parte, Wurschum *et al.* (2013) reportaron correlación significativa en trigo, independientemente del rango de distancias consideradas. En el presente trabajo, se pudo observar una correlación entre las distancias genéticas derivadas de los marcadores SSR y SNP, sugiriendo que ambos tipos de marcadores son igualmente apropiados para el relevamiento y la clasificación de la variación genética en el girasol cultivado.

En general, las estimas de diversidad genética obtenidas para la PMA son moderadas y un poco menores a las detectadas para las 271 líneas de la NCRPIS y el INRA que componen el panel de mapeo por asociación usado por Mandel *et al.* (2011, 2013). Sin embargo, es necesario tener presente la imposibilidad de hacer comparaciones directas, dada la diferencia en el número de líneas analizadas en cada caso y el hecho de que, a diferencia del trabajo de Mandel *et al.* (2011), casi no se han incluido materiales confiteros en la PMA de INTA.

Los análisis de los niveles de diversidad en el panel completo de entradas y en la PMA mostraron que la última no incluye todos los alelos que están presentes en OP y CP. Esto sugiere que nuevas líneas endocriadas podrían ser incluidas en nuestro panel de MA para capturar totalmente la diversidad alélica preservada en el BAG-IM. Esta reducción, o pérdida aparente de la diversidad genética, es una consecuencia muy común de la estrategia de muestreo, donde los alelos de menor frecuencia tienen pocas chances de ser capturados. Del mismo modo, el panel de MA usado por Mandel *et al.* (2013) tampoco incluyó el total de alelos detectados en las colecciones provenientes del NCRPIS y el INRA de las cuales deriva (Mandel *et al.* 2011).

Al comparar las distintas categorías en las que se subdividió la PMA, se detectaron diferencias en el número total de alelos entre ellas, al tiempo que se identificaron alelos

privados, pero, notablemente, no se observaron diferencias en términos de heterocigosis esperada entre dichas categorías. Diversos estudios han evaluado los niveles y distribución de la diversidad genética en diferentes entradas de girasol (Dong *et al.* 2007; Mandel *et al.* 2011; Moreno *et al.* 2013). En concordancia con lo observado en el presente trabajo, Mandel *et al.* (2011) no detectaron diferencias en términos de diversidad alélica entre las diferentes categorías delimitadas por dichos autores (*i.e.*: HA, RHA, aceitera, confitera).

El análisis molecular de la varianza (AMOVA) permitió describir los niveles de diferenciación entre los grupos establecidos para las entradas aquí estudiadas. A pesar de ser moderada, se detectó diferenciación significativa entre la PMA y el conjunto de OP+CP, e incluso entre las distintas subclases en las cuales la PMA fue subdividida, confirmando hipótesis previas acerca de la existencia de diferencias en la constitución genética del germoplasma argentino y aquel de diferentes orígenes (Rusia, Israel, Europa, Estados Unidos) (Moreno *et al.* 2013).

Adicionalmente, dos métodos estadísticos diferentes fueron utilizados para la identificación de grupos genéticos y asignación de individuos, STRUCTURE (Pritchard *et al.* 2000a) y DAPC (Jombart *et al.* 2010). STRUCTURE es ampliamente utilizado para la identificación de subdivisión poblacional, aunque debe tenerse en cuenta que ha sido desarrollado para poblaciones naturales, ya que presume equilibrio de Hardy-Weinberg, condición que es violada por la mayoría de los materiales de programas de mejoramiento, entre ellos las líneas endocriadas. Por su parte, DAPC resulta un método de mayor validez para paneles de MA, debido a que relaja el principio de equilibrio Hardy-Weinberg (Jombart *et al.* 2010).

La aplicación de ambos métodos reveló la existencia de subestructuración genética dentro de la PMA para todos los *sets* de datos considerados (SSR, SNP, SSR+SNP). Bajo la aproximación bayesiana, los resultados fueron casi idénticos independientemente del *set* de datos usado. No obstante, pudo observarse que el número de individuos clasificados como mixtos resultó mucho mayor usando el *set* de datos de SNP, en comparación con los *sets* de SSR o SSR+SNP (Tabla 2.8). Adicionalmente, al aumentar los valores de K, el porcentaje de asignación decayó para el *set* de datos de SNPs. Diferencias en los porcentajes de asignación entre SSR y SNPs fueron también reportadas en varios estudios (Hamblin *et al.* 2007; Van Inghelandt *et al.* 2010; Yang *et al.* 2011; Emanuelli *et al.* 2013) y fueron atribuidas al mayor contenido de información genética de los SSRs (Lu *et al.* 2009; Emanuelli *et al.* 2013).

Continuando con la comparación de los distintos *sets* de marcadores bajo el método bayesiano, el análisis de las correlaciones de Spearman reveló que para aquellos individuos

que excedían el límite de pertenencia arbitrario delimitado en 0,70, la asignación a cada grupo fue consistente a través de todos los *sets* de marcadores, siendo el grupo 1 el más conservado y mejor definido. Es más, inspeccionando las distancias entre los grupos definidos por STRUCTURE puede observarse que el grupo 1 resultó ser el más diferenciado de los tres. Del mismo modo, la constitución de los tres grupos obtenidos utilizando la aproximación DAPC para los distintos *sets* de datos de marcadores moleculares fue similar pero no idéntica (Apéndice, Tabla A2.6). Al comparar los resultados de ambas metodologías, se observó una gran correspondencia entre los grupos de DAPC y aquellos generados por STRUCTURE a $K = 3$, aunque con diferencias menores (Tabla 2.8). Estas discrepancias pueden deberse al hecho de que bajo el algoritmo DAPC todas las líneas deben ser clasificados en un grupo, aun cuando se detecte cierto nivel de mezcla. Este no es el caso del análisis realizado por STRUCTURE, donde las líneas se asignan a un grupo solo cuando sobrepasar un umbral de pertenencia, arbitrariamente delimitado, en este caso, en 0,70.

Además de la detección de los grupos genéticos dentro de la PMA, DAPC puede ser utilizado también para identificar aquellos alelos que contribuyen en forma mayoritaria a las funciones discriminantes, como una aproximación para detectar patrones putativos entre los genes responsables de la diferenciación entre grupos (Jombart *et al.* 2010). Se realizó una gráfica de las contribuciones de los alelos de SNPs para identificar aquellos de contribución mayoritaria, e interesantemente la mayoría de ellos correspondían a genes asignados a procesos metabólicos, mediante anotación GO. A pesar de que se necesitan más estudios para determinar la naturaleza de los procesos metabólicos involucrados y cómo pueden relacionarse con la diferenciación entre líneas, estos resultados sirven para resaltar el potencial que posee el método DAPC para profundizar en los factores subyacentes a la delimitación de grupos.

Mientras que los grupos generados por cada conjunto de datos utilizando STRUCTURE y DAPC podían separar con facilidad individuos en grupos similares, los métodos basados en distancia fueron menos capaces de identificar grupos reproducibles entre los diferentes *sets* de datos. Al comparar con los resultados obtenidos por STRUCTURE para $K = 3$, los tres filogramas de NJ generados (uno para cada *set* de marcadores) fueron consistentes en la delimitación del grupo 1 de STRUCTURE, aunque no pudieron apreciarse diferencias genéticas discernibles entre individuos pertenecientes a los dos grupos remanentes. Esto se condice con los resultados reportados previamente (Bracco *et al.* 2009; Lia *et al.* 2009) y con

las propuestas de Rosenberg *et al.* (2001) quienes argumentaron que STRUCTURE es capaz de utilizar los datos genotípicos individuales en forma más eficiente que los filogramas basados en matrices de distancia genética.

En resumen, al considerar aquellas líneas que fueron consistentemente asignadas al mismo grupo, sin importar el método de análisis ni el *set* de marcadores utilizados, es posible distinguir un patrón general de relaciones entre las líneas que componen la PMA (Apéndice, Tabla A2.6). El grupo 1 (violeta) está compuesto principalmente por líneas mantenedoras, estando fuertemente influenciado por la contribución de la línea pública HA89, la cual está involucrada en el origen de las líneas 2071, 2125 y C454B. El grupo 2 (verde) está dominado por la presencia de líneas restauradoras, entre ellas las líneas públicas RHA801 RHA276. La gran mayoría de las líneas argentinas incluidas en este grupo fueron desarrolladas como parte del programa de mejoramiento para resistencia a estrés hídrico de INTA, teniendo sus progenitores contribuciones de especies salvajes de *Helianthus*. Finalmente, el grupo 3 (celeste) fue el mayor y más diverso, con una alta proporción de líneas mantenedoras. Los materiales incluidos en este grupo son derivados de recursos públicos de EE.UU, como así también de variedades argentinas tradicionales tales como Impira INTA, Saenz Peña y RusoxKlein (Apéndice, Tabla A2.6). Estas líneas se caracterizan por poseer contribución de germoplasma ruso diferente a aquel involucrado en el origen de HA89.

Asimismo, en términos generales, los patrones de estructura poblacional detectados para la PMA de INTA concuerdan con aquellos reportados por Mandel *et al.* (2013) y Cadic *et al.* (2013) para las colecciones de NCRPIS e INRA, en donde la condición de mantenedora o restauradora resultó ser la característica prevalente en relación con la delimitación entre grupos. En concordancia con los resultados obtenidos en este trabajo, los estudios de Mandel *et al.* (2013) y Cadic *et al.* (2013) también identificaron tres grupos, dos de los cuales estaban compuestos por líneas mantenedoras o “B” y el tercero compuesto por líneas restauradoras o “R”. A pesar de que las afinidades entre grupos que emergen de esos diferentes trabajos todavía deben ser determinadas, es interesante notar que mientras que en los estudios de Mandel *et al.* (2013) y Cadic *et al.* (2013) los grupos de líneas mantenedoras parecen estar más cercanos uno a otro que el que contiene a las líneas restauradoras, nuestros resultados de STRUCTURE sugieren la existencia de una relación más cercana entre el grupo que contiene líneas restauradoras, grupo 2, y el grupo 3, que contiene líneas mantenedoras. Un patrón similar, aunque no tan claro, surge de la

inspección de las gráficas de DAPC. En suma, todo pareciera indicar que la distribución de la diversidad genética del girasol cultivado en todo el mundo sigue in patrón común dominado principalmente por la condición de mantenedora/restauradora.

El alcance con el cual un grupo de marcadores moleculares es capaz de capturar la estructura poblacional puede tener implicancias prácticas y económicas a la hora de genotipificar un gran número de individuos. Por ejemplo, los tres grupos detectados por Mandel *et al.* (2013) utilizando cerca de 5500 SNP no pudieron ser identificados al usar 34 EST-SSR sobre el mismo *set* de entradas (Mandel *et al.* 2011). Del mismo modo, casi 6000 SNP fueron incluidos en el análisis de las 384 líneas endocriadas usadas por Cadic *et al.* (2013), mientras que los 136 SNP usados por Talukder *et al.* (2014) sobre un panel de 260 líneas endocriadas sólo lograron identificar dos grupos mediante el análisis de STRUCTURE. Al respecto, los 42 SSR y los 384 SNP examinados en este trabajo, usados en forma independiente o conjunta, permitieron una clara diferenciación entre grupos. Si bien es cierto que la capacidad discriminante es claramente dependiente del conjunto de entradas evaluado, el uso de los *sets* de marcadores aquí analizados en diferentes colecciones de germoplasma de girasol puede proveer una alternativa económica para la genotipificación, con el potencial para alcanzar un alto nivel resolutivo.

La consistencia de los niveles y patrones de estructuración poblacional para las líneas que componen la PMA de girasol de INTA dan indicio de la robustez de estas estimas para su inclusión en los estudios de asociación que se detallarán en el Capítulo 4.

CAPÍTULO 3

BÚSQUEDA, SELECCIÓN Y GENOTIPIFICACIÓN DE GENES CANDIDATO DE RESPUESTA A ESTRÉS BIÓTICO

INTRODUCCIÓN

Los niveles de rendimiento obtenidos para el cultivo de girasol suelen verse seriamente afectados debido a enfermedades fúngicas, siendo la podredumbre húmeda del capítulo causada por *Sclerotinia sclerotiorum* una de las más perjudiciales (Boland & Hall 1994; Talukder *et al.* 2014).

Dada la necesidad de desarrollar materiales con resistencia a este patógeno, numerosos estudios de mapeo de QTL han sido conducidos en los últimos 16 años (Gentzbittel *et al.* 1998; Mestries *et al.* 1998; Bert *et al.* 2002, 2004; Micic *et al.* 2004, 2005a; 2005b), resultando en la identificación de más de 20 regiones genómicas involucradas en la resistencia a PHC en girasol. A pesar de que los resultados de estos estudios podrían resultar promisorios, dado que la resolución de los mismos suele estar en el orden de 10 centimorgans (cM) o más, distancia que puede corresponderse con miles de bases nucleotídicas en girasol, no se cuenta aún con una real evaluación de los genes individuales involucrados en el proceso.

Para dar respuesta a esta problemática, los estudios de mapeo por asociación han surgido como una herramienta alternativa para la búsqueda de genes involucrados en variaciones cuantitativas (Yu & Buckler 2006; Ehrenreich *et al.* 2009), ya que alcanzan mayor poder resolutivo, al tiempo que tienen el potencial de evaluar un mayor rango de alelos (Buckler & Thornsberry 2002; Flint-Garcia *et al.* 2003; Wilson & Whitt 2004). A pesar de que los estudios de asociación de tipo *genome-wide* han ganado prominencia en los últimos años (Zhu *et al.* 2008), la aproximación por medio de genes candidato resulta clave para estudios en los que la atención se centra en uno o pocos caracteres de interés (Tabor *et al.* 2002): mientras los estudios de tipo *genome-wide* escanean un gran número de marcadores ampliamente distribuidos en todo el genoma, los estudios a través de genes candidato se centran en genes blanco seleccionados por su función ya conocida o descrita en el carácter de interés, con la expectativa que de este modo se enriquezca el número de caracteres con asociación significativa.

La estrategia de genes candidato comienza con la selección de un conjunto de genes potencialmente relacionados con el carácter o el mecanismo de interés. Esto es seguido por un paso de estudio de la diversidad, para identificar polimorfismos en los mismos, principalmente de tipo SNP o InDels, los cuales podrían tener consecuencias funcionales, ya

sea afectando la regulación génica o el producto proteico. Finalmente, las distintas variantes génicas son evaluadas para su asociación con el carácter de interés (Patnala *et al.* 2013). La elección de los genes candidato suele estar asistida por: (a) recursos genómicos, como bases de datos de EST o estudios de perfiles de expresión; (b) conocimiento previo de función de los genes de interés en organismos modelo; o (c) información de referencia sobre la fisiología, bioquímica o genética molecular para el carácter de interés (Haseneyer *et al.* 2010).

El primer antecedente exitoso de aplicación de esta estrategia en plantas fue reportado por Thornsberry *et al.* (2001), quienes detectaron asociación entre el gen *dwarf8* y variaciones en el tiempo de floración en maíz. Numerosos estudios fueron reportados posteriormente, principalmente para caracteres de interés agronómico en especies de importancia económica, destacándose pino, papa, raigrás perenne, cebada, álamo (Wilson & Whitt 2004; Weber *et al.* 2007; González-Martínez *et al.* 2007; Skøt *et al.* 2007; Li *et al.* 2008; Ingvarsson 2008; Pajerowska-Mukhtar *et al.* 2009; Haseneyer *et al.* 2010), y más recientemente, girasol (Fusari *et al.* 2012; Mandel *et al.* 2013; Cadic *et al.* 2013; Talukder *et al.* 2014).

Una de las metodologías más utilizadas para la detección de genes involucrados en procesos biológicos es la de microarreglos de ADN. Éstos permiten observar de forma casi instantánea la dinámica de todos los genes del genoma de un organismo. El Grupo de Genómica de Girasol del Instituto de Biotecnología de INTA Castelar ha diseñado un microarreglo de óligos sintéticos basados en las secuencias de 40000 genes de girasol (Fernandez *et al.* 2012b). Utilizando esta herramienta, se analizaron los perfiles de ARNm de tres genotipos de girasol (RK416, HA853 y HA89), bajo dos condiciones (control, e inoculado con esporas del hongo), a tres tiempos (0, 4 y 8 días post inoculación). De este trabajo, numerosos genes sobre y subexpresados surgieron al comparar plantas inoculadas y control de los distintos genotipos (Ehrenbolger *et al.* 2013), convirtiéndose en interesantes candidatos para el estudio de asociación.

Una segunda fuente importante de información resulta de los análisis funcionales de perfiles de expresión relacionados con defensa a estrés desarrollados en especies modelo como *Arabidopsis thaliana*, ya que proveen una fuente potencial de genes candidato para el estudio de los procesos de resistencia en girasol y otros cultivos (Zhou 1999; Hernández-Blanco *et al.* 2007; Tang *et al.* 2007; Guo & Stotz 2007; Talukder *et al.* 2014). En el trabajo de Fusari *et al.* (2012), a partir del análisis de perfiles de expresión de genes de *Brassica napus*

en respuesta a la infección con *S. sclerotiorum* contra un microarreglo de oligos sintéticos basados en las secuencias de 26.090 genes de *Arabidopsis* (Zhao *et al.* 2007b), se seleccionaron 28 *loci* cuyos ARNm se incrementaban en el genotipo resistente a 24 horas post infección. Para identificar los genes ortólogos en girasol, se estableció un protocolo de reconstrucción filogenética y se utilizó la información resultante para el diseño de iniciadores especie específicos. Estos genes fueron evaluados con éxito para asociación con resistencia a PHC en girasol, utilizando una PMA conformada por 94 líneas endocriadas. Dada la especificidad de estos genes para respuesta a *S. sclerotiorum*, resulta de interés profundizar su estudio y caracterización por mapeo por asociación en girasol en una población de mapeo mayor y un mayor número de evaluaciones de la enfermedad. Una estrategia similar a la de Fusari *et al.* (2012) fue aplicada por Talukder *et al.* (2014), para la identificación de 8 ortólogos en girasol relacionados con respuesta a podredumbre de base de tallo causada por *S. sclerotiorum*.

La búsqueda bibliográfica de estudios relacionados con resistencia a distintos estreses puede aportar información sobre genes candidato aún no identificados en girasol, pero con potencial en otras especies. Ejemplos de esto lo constituyen las Proteínas tipo Germina (GLPs, del inglés *Germin-like Proteins*), que se ha reportado poseen actividad oxalato oxidasa y/o superóxido dismutasa, y podrían participar en procesos de defensa (Carter & Thornburg, 2000; 1995; Dumas *et al.* 1993; Hu *et al.* 2003). Al respecto, Rietz *et al.* (2012) destacaron la acción de GLPs en *Brassica napus*, como iniciadoras de respuesta oxidativa frente al ataque del patógeno *S. sclerotiorum*. En girasol se ha reportado la existencia de 9 GLPs, aunque se estima que el número real asciende a más de 16 (Verónica Lia, *comunicación personal*).

Adicionalmente, el mapeo por asociación puede ser también utilizado para validar los resultados obtenidos mediante mapeo de QTL en poblaciones biparentales (Cadic *et al.* 2013). Ejemplos de validaciones han sido reportados recientemente por Brachi *et al.* (2010), Famoso *et al.* (2011), Mir *et al.* (2012) y Cadic *et al.* (2013). Dentro del programa de mejoramiento de girasol se han realizado ensayos de mapeo de QTL sobre una población biparental derivada del cruzamiento de dos líneas públicas, PAC2 (moderadamente resistente, MR) x RHA266 (susceptible, S), y se han detectado QTL relacionados con menor incidencia de la enfermedad (Zubrzycki *et al.* 2014). Dado el alto contenido informativo de los marcadores que colocalizaron con dichos QTL, y la necesidad de validar por otra técnica

los resultados obtenidos, resulta de interés su inclusión en estudios de asociación para el mismo carácter.

Una vez seleccionado el conjunto de genes candidato (GC) a ser analizados para asociación, debe definirse la metodología de identificación de polimorfismos y genotipificación. Si bien la secuenciación directa es una metodología ampliamente utilizada debido a su rapidez y eficiencia, la genotipificación de InDels (*insertion/deletion*) mediante electroforesis capilar con detector de fluorescencia inducida por láser (ECF) y la detección de SNPs y/o InDels mediante técnicas de detección de moléculas heteroduplex (CEL1CH y dHPLC) son más económicas y alcanzan niveles de precisión comparables (Giordano *et al.* 1999; Comai *et al.* 2004). En girasol, estas metodologías han sido puestas a punto con éxito, demostrando ser robustas y de alta capacidad de discriminación (Fusari *et al.* 2010; Paniego *et al.* 2015).

En este capítulo se describe el proceso de selección, identificación de polimorfismos y genotipificación de los genes candidato que serán incluidos, junto con los SNPs derivados del panel de Illumina presentado en el Capítulo 2, en el estudio de mapeo por asociación que se presenta en el Capítulo 4. Como punto de partida, se utilizó la información provista por: (1) los perfiles transcripcionales de tres líneas públicas de girasol inoculadas con *S. sclerotiorum*, obtenidos mediante el análisis de microarreglos (Ehrenbolger *et al.* 2013); (2) los perfiles transcripcionales de *B. napus* en respuesta a la infección con *S. sclerotiorum* obtenidos a partir del análisis de microarreglos de secuencias de *Arabidopsis* (Zhao *et al.* 2007b); (3) genes descritos en la literatura como partícipes en el proceso de defensa frente al ataque por patógenos fúngicos (Fernández *et al.* 2003; Fusari *et al.* 2012; Talukder *et al.* 2014); y (4) marcadores moleculares de tipo SSR que colocalizaron con QTLs detectados para resistencia a *S. sclerotiorum* en estudios de mapeo biparental en girasol (Zubrzycki *et al.* 2014).

MATERIALES Y MÉTODOS

Líneas de la PMA

Los análisis que se describen a continuación fueron desarrollados sobre las 137 líneas que conforman la PMA de girasol de INTA para caracteres complejos. Las extracciones de ADN genómico y cuantificación del mismo se efectuaron de acuerdo a lo descrito en el Capítulo 2.

Selección de genes candidatos derivados de un análisis de expresión diferencial en respuesta a *S. sclerotiorum* en girasol

El grupo Genómica de Girasol realizó un ensayo de expresión diferencial sobre un microarreglo de ADN, a partir de ARNm de tres líneas públicas, HA853, HA89 y RK416, artificialmente inoculadas con *S. sclerotiorum*, evaluándose muestras a 0, 4 y 8 dpi (Ehrenbolger *et al.* 2013).

Del total de genes sobre y subexpresados, se decidió tomar como candidatos aquellos que mostraran un aumento o disminución de expresión significativo ($p < 0,05$, tasa de cambio $> |2|$), y que además cumplieran con al menos uno de los siguientes requisitos: (1) su función molecular inferida estuviera relacionada con respuesta a estrés, (2) sus niveles de expresión hubieran sido validados por qPCR (PCR de tipo cuantitativa) o (3) se tratara de genes ya descritos en la literatura por su rol activo en procesos de defensa contra patógenos fúngicos (*i.e.*, HeAn_12562, homólogo al gen Ve1 de resistencia a *Verticillium dahliae* en tomate) (Tabla 3.1).

Tabla 3.1 Genes candidato seleccionados a partir de un microarreglo de girasol

ID ^a	Clase	Función probable (GO)	Línea ^b	Dpi ^c	Fc ^d	p-valor	Longitud (pb) ^e
Unión a ATP							
HeAn_28738		Proteína de choque térmico (<i>Heat shock protein</i>) <i>Hsp70</i>	RK416	8	2,95	0,001	722
Desconocida							
HeAn_36825		Desconocida	HA853	8	2,95	0	867
HeAn_1756		Guanilil ciclasa	RK416	8	-5,26	0,045	686
HeAn_41022		Glicoproteína 1- Unión a polen	HA863	8	-4,55	0	383
HeAn_36589		Chaperona similar a HSP20	RK416	4	-3,70	0,055	711

ID ^a	Clase	Función probable (GO)	Línea ^b	Dpi ^c	Fc ^d	p-valor	Longitud (pb) ^e
HeAn_315		Chaperona similar a HSP20	RK416	4	-2,50	0,052	777
HeAn_6452		ATPasa	HA89	4	-6,67	0	713
HeAn_35791		Desconocida	HA89	4	-2,78	0,001	312
Unión a calcio							
HeAn_18494		Dominio similar al brazo EF	HA89	4	-2,63	0,005	762
HeAn_38637		Dominio similar al brazo EF	HA89	8	2,97	0,003	763
Metabolismo de lípidos							
HeAn_10191		Lipidasa	HA89	4	-5,26	0,001	884
Proteínas de defensa							
HeAn_41021		Gamma tionina	HA853	8	12,50	0	368
Respuesta a <i>Verticillium dahliae</i>							
HeAn_12562		Gen mayoritario de respuesta a Vd	RK416	4	2,36	0,01	930
Biosíntesis de ATP							
HeAn_12761		Factor de virulencia	HA89	4	0,31	0,001	1475

a Acrónimo designado para denominar el gen candidato en base al código de identificación de la base de datos Unigenes Girasol de IB-INTA

b Línea evaluada con el microarreglo de girasol donde se detectó la expresión diferencial del gen seleccionado

c Días post inoculación a los que fue detectada la expresión diferencial del gen seleccionado

d Tasa de cambio para el nivel de expresión del gen

e Longitud de la secuencia seleccionada en la base de datos de Unigenes de Girasol de IB-INTA

Selección de genes candidato derivados de estudios de expresión diferencial en *B. napus* sobre un microarreglo de *Arabidopsis* (Zhao *et al.* 2007b)

El trabajo de Zhao *et al.* (2007b) de perfiles de expresión de dos genotipos de *B. napus* de respuesta diferencial a *S. sclerotiorum* sobre un microarreglo de *A. thaliana* constituye una fuente valiosa de genes candidato para ser utilizados en estudios de asociación en otras especies vegetales para la resistencia a dicho hongo. La identificación en girasol de genes ortólogos a aquellos diferencialmente expresados en la respuesta de *B. napus* frente al ataque de *S. sclerotiorum* a partir de un microarreglo de *Arabidopsis* (Zhao *et al.* 2007b) se llevó a cabo mediante una estrategia basada en reconstrucción filogenética según lo propuesto por Fusari *et al.* (2012). Brevemente, se realizaron búsquedas tBLASTn contra diversas bases de datos (GenBank *Viridiplantae EST data base*, KEGG *Helianthus annuus data base*, DFCI *Sunflower Gene Index*) usando para la interrogación la secuencia proteica del gen de *A. thaliana* con aumento o disminución significativa de la expresión frente a la infección. Por cada búsqueda, un grupo de EST con *E-values* < 10⁻²⁰ y alta similitud (≥50 % de identidad en la secuencia aminoacídica) fue seleccionado como representante de las

diferentes familias de plantas, y usado para determinar la posición y la longitud mínima de las regiones a incluir en la reconstrucción filogenética. Fueron descartados del análisis aquellos ESTs que presentaron hits con dos o más regiones, con diferentes marcos abiertos de lectura (ORF, *Open Reading Frames*), como así también aquellos alineamientos que incluían menos de 150 aminoácidos. Para evitar sobre-representación de grupos particulares, se permitió un máximo de cuatro ESTs por especie, con excepción de aquellas especies del género *Helianthus* y *Lactuca*, para las cuales todos los *hits* detectados fueron considerados en el análisis.

Todas las secuencias seleccionadas fueron transcritas en el ORF apropiado y alineadas utilizando la rutina L-INS-i del *software* MAFFT (Katoh & Standley 2013). Los alineamientos proteicos fueron utilizados para obtener las relaciones filogenéticas utilizando el método de máxima parsimonia, incluido en el *software* TNT (Goloboff *et al.* 2008). Se realizaron búsquedas tradicionales mediante métodos heurísticos para encontrar el conjunto de árboles óptimos, es decir, aquellos de menor longitud o número de pasos. Los valores de soporte de rama se obtuvieron mediante *jackknife* (Farris *et al.* 1996), utilizando una probabilidad de 33, para 100 réplicas. En la Tabla A3.1 del Apéndice se presenta la lista con el género y especie de las secuencias de GenBank utilizadas en la construcción de los árboles filogenéticos.

Finalmente, las secuencias ortólogas y parálogas en girasol fueron determinadas en base a las relaciones filogenéticas para cada locus. Dado que la información de secuencias de la especie *H. annuus* es limitada, todas las secuencias del género *Helianthus*, *Lactuca* (principalmente) y otros géneros de la familia *Asteraceae* (en menor proporción) que se encontraron dentro del mismo clado que la secuencia de *Arabidopsis* de interés, fueron consideradas “ortólogas” a la de *Arabidopsis* y co-ortólogas (*in-paralogs*) entre sí, cuando se observó más de una secuencia del mismo género. En contraposición, se consideraron “parálogas” las secuencias de Asteráceas que quedaron ubicadas por fuera del clado donde se ubicó la secuencia de *Arabidopsis*. Las secuencias seleccionadas para cada *loci* fueron usadas como molde para el diseño de oligonucleótidos (Tabla 3.2).

Tabla 3.2 Genes candidato seleccionados a partir del análisis de expresión a las 24 horas post inoculación en plantas de *B. napus* resistentes a *S. sclerotiorum* (Zhao *et al.* 2007b)

ID	Clase	Descripción	Locus	Nivel de expresión ^a	p-valor ^b
Modificación de pared celular					
G21		Familia de proteínas inhibidoras Invertasa/pectin metilesterasa	At3g17130	18,74	4,49E-05
Genes de respuesta a hormonas					
Vía relacionada con auxinas					
G22		Familia de proteínas GH3 de respuesta a auxinas	At4g03400	18,03	4,82E-06
Vía de transducción de señales					
Proteínas de unión a calcio					
G26		Calmodulina	At3g10190	14,92	9,72E-06
Metabolismo secundario					
G28		Región de unión a ATP, dominio ATP-asa	At1g19100	14,23	8,31E-07
G33		Proteína conteniendo dominio radical SAM	At3g19630	12,17	1,20E-07
Proteínas relacionadas con degradación					
G34		Glicoproteína P	At3g28390	12,16	1,20E-05
Genes relacionados con defensa/antimicrobianos					
G30		Proteína tipo traumatina	At5g02140	13,89	3.97E-07

a Indica el número de veces que aumentó la expresión de los transcritos en las plantas inoculadas con el patógeno versus plantas inoculadas con agua.

b El p-valor indica la probabilidad de que el cambio en el nivel de expresión sea estadísticamente significativo (corrección FDR y significación del 0,05).

Selección de genes candidato descritos en literatura como partícipes en los procesos de defensa frente a patógenos fúngicos

La nómina de genes candidato seleccionados a partir de la literatura incluyó: (1) dos proteínas tipo-germina de girasol (HaGLPs) (Ehrenbolger *et al.* 2012); (2) una peptidasa y una proteína relacionada con respuesta a estrés abiótico (Fernández *et al.* 2003); (3) dos genes candidato que fueron reportados como asociados con menor incidencia de podredumbre de base de tallo causada por *S. sclerotiorum* (HaCOI1 y HaCOI2) (Talukder *et al.* 2014); y (4) siete genes seleccionados del estudio de mapeo por asociación para la resistencia a PHC en girasol de Fusari *et al.* (2012). Evidencias experimentales avalan la elección de estos genes, ya que podrían estar implicados en la respuesta a varios estreses ambientales (Tabla 3.3).

Tabla 3.3 Genes candidato seleccionados a partir de literatura por su participación en procesos de defensa

ID	Clase	Función	Origen	Largo (pb)
Modificación de la pared celular				
Proteína de pared (similar a <i>A. thaliana</i>)				
HaWP		cwx_arath	Fusari <i>et al.</i> (2012)	425
HaTRP		Proteína relacionada con la formación de tumores	Fusari <i>et al.</i> (2012)	266
Síntesis de fenilpropanoides				
HaPAL		Fenilalanina-amonio-liasa	Fusari <i>et al.</i> (2012)	296
Factores de transcripción				
HaWRKY7		Factor de transcripción WRKY7	Fusari <i>et al.</i> (2012)	1721
HaWRKY5		Factor de transcripción WRKY5	Fusari <i>et al.</i> (2012)	667
Estrés abiótico				
HaDRP		Proteína RD22 con respuesta a deshidratación	Fernández <i>et al.</i> (2003)	843
Desconocida				
RhoBP		Proteína con dominio p21-rho-binding	Fusari <i>et al.</i> (2012)	1424
Peptidasas				
HaCP		Cisteín-proteinasa	Fernández <i>et al.</i> (2003)	578
Proteínas tipo germina				
HaGLP3		Proteína tipo germina 3	Ehrenbolger <i>et al.</i> (2012)	621
HaGLP4		Proteína tipo germina 4	Fusari <i>et al.</i> (2012)	761
HaGLP5		Proteína tipo germina 5	Ehrenbolger <i>et al.</i> (2012)	1710
Respuesta a Podredumbre de base de tallo				
HaCO11		Reguladoras de la inducción de Jasmonato	Talukder <i>et al.</i> (2014)	668
HaCO12		Reguladoras de la inducción de Jasmonato	Talukder <i>et al.</i> (2014)	660

Selección de marcadores identificados mediante análisis de QTL implicados en procesos de resistencia a *S. sclerotiorum* en poblaciones biparentales de girasol

A partir de la población biparental derivada del cruzamiento de las líneas públicas PAC2 (MR) x RHA266 (S) (Gentzbittel *et al.* 1999), se identificaron 39 regiones asociadas a menor PHC (LOD \geq 3, Zubrzycki *et al.* 2014). Se decidió la inclusión en este trabajo de MA de 11 marcadores tipo SSR que colocalizaran con los QTL detectados, a fin de intentar validar a los mismos (Tabla 3.4).

Tabla 3.4 Marcadores que colocalizaron con QTL de resistencia a *S. sclerotiorum* en poblaciones biparentales de girasol

ID	GL	Motivo	Min	Max	Iniciador <i>Forward</i>	Iniciador <i>Reverse</i>
HA77	1	GA	247	262	HEX-TGTAATCTGTATCACTCCACC	GTTGTTCTGTTAGGTCGTTCCG
HA1848	7	ATT	230	268	HEX-CATCCCTTCTGAATAGAAA	TAGGCTCGTTAAACTTACGG
HA2989	14	ATCCTC	127	142	HEX-GCCTGTGTCCTCCTCAGAGTC	GCCTGTGTCCTCCTCAGAGTC
HA3039	10	GA	184	174	FAM-GGTTGATTAAGTGTGGGTC	GGTTGATTAAGTGTGGGTC
HA3103	7	GA	184	174	NED-CCTTAAAAGGCAAACGATTCTCC	CCTTAAAAGGCAAACGATTCTCC
HA3847	10	GT	148	143	NED-CATCACTTCAACATGCCTCC	CATCACTTCAACATGCCTCC
HA4264	14	GT	181	179	HEX-TCCCAAACCTACACCTCTAA	TCCCAAACCTACACCTCTAA
HA3950	14	CGA	129	137	NED-CATTCTCACCAACATAACA	CATTCTCACCAACATAACA
ORS331	7	CT-AC	283	290	FAM-TGAAGAAGGGTTGTTGATTACAAG	TGAAGAAGGGTTGTTGATTACAAG
ORS460	7	GA	302	300	FAM-ACTCGGCTACCACCTCACAC	ACTCGGCTACCACCTCACAC
ORS1041	7	CT	274	282	FAM-AAACAAACCTTAATGGGGTCGTA	ATATTGGGCTGGTTGATGCTGAT

Construcción de Colecciones Núcleo (CN)

Las colecciones núcleo (CN) o *core sets* son muestreos de una colección grande de germoplasma, conformados por un número reducido de entradas elegidas con el fin de representar la mayor parte posible de la diversidad genética contenida en la colección mayor (Emanuelli *et al.* 2013). Para la búsqueda de polimorfismos en las secuencias de los GC para los cuales no se contaba con información previa, se procedió a la identificación de una colección núcleo de 8 líneas a partir de la PMA de girasol, y posterior re-secuenciación de los GC sobre ADN genómico de las mismas. Para la construcción de esta CN de ocho líneas, se utilizó la estrategia *simulated annealing algorithm*, implementada en el *software* PowerMarker versión 3.51 (Liu & Muse 2005) sobre la base de los datos genotípicos obtenidos utilizando el panel de 42 SSR en la PMA detallado en el Capítulo 2. El número de iteraciones realizadas por corrida fue de 1000, con 10 repeticiones por muestreo.

Diseño de oligonucleótidos

Para la amplificación de los genes seleccionados por su expresión diferencial a partir del análisis del microarreglo de girasol, se recuperaron las secuencias expresadas diferencialmente de la base de datos *Sunflower Unigenes* (www.atgc-sur.inta.gob.ar), y con las mismas se realizaron búsquedas por BLAST en la base de datos de Heliagene (www.heliagene.org). La totalidad de secuencias recuperadas se alinearon usando el

software BioEdit (Hall 1999) y se diseñaron oligonucleótidos sobre las secuencias más conservadas, intentando abarcar el segmento de mayor tamaño posible para su posterior secuenciación.

Para el caso de los genes candidato derivados de *Arabidopsis*, el punto de partida para la amplificación de los mismos fueron las secuencias nucleotídicas recuperadas de GenBank para el género *Helianthus*. Nuevamente, las secuencias recuperadas fueron alineadas usando el *software* BioEdit (Hall 1999) y se diseñaron oligonucleótidos sobre las regiones más conservadas.

En todos los casos el diseño de los iniciadores para la amplificación por PCR de los genes candidato se realizó mediante el programa Primer3 (Rozen & Skaletsky 2000).

Aquellas amplificaciones que dieron productos de PCR múltiples, fueron purificadas a partir del taco de agarosa y secuenciadas (este paso se detalla en la próxima sección). A partir de las secuencias obtenidas, se diseñaron nuevamente oligonucleótidos específicos.

Para la amplificación de los genes seleccionados del trabajo de Fusari *et al.* (2012) y Talukder *et al.* (2014), como así también de los SSR que colocalizaron con QTL (Zubrzycki *et al.* 2014), se utilizaron los iniciadores reportados en dichos trabajos. En la Tabla A3.2 del Apéndice se presenta la lista completa de los iniciadores utilizados para amplificar, secuenciar y genotipificar los genes candidato en la PMA.

Reacciones de PCR

Las reacciones de PCR se llevaron a cabo en un volumen final de 12 μ L de acuerdo a lo especificado en el Capítulo 2. Cada par de iniciadores se probó sobre ADN genómico de dos líneas de la PMA seleccionadas arbitrariamente: HA89 y RHA801. Para garantizar alta concentración de producto de PCR, la reacción se realizó con un programa que incluye una disminución progresiva de la temperatura de hibridación (*touchdown*) en un rango de 60 °C a 55 °C. El ciclado utilizado fue el siguiente: 2 min a 94 °C de desnaturalización inicial, cinco ciclos de (a) 40 s a 94 °C, (b) 40 s a 60 °C (-1°C/ciclo) y (c) 2 min a 72 °C; seguido por 32 ciclos de (a) 40 s a 94 °C, (b) 40 s a 55 °C y (c) 1 min a 72 °C; con una extensión final de 10 min a 72 °C. Aquellos pares de iniciadores que no produjeron amplificación, se evaluaron posteriormente a una temperatura de hibridación menor (50 °C). Los pares de iniciadores

que dieron múltiples productos de amplificación se ensayaron a una temperatura de hibridación de 60 °C, para descartar amplificaciones inespecíficas.

Los productos de PCR obtenidos fueron visualizados mediante cromatografía en gel de agarosa 1,5 %. En aquellos casos en los cuales se utilizó iniciadores marcados con fluoróforos, los geles de agarosa fueron visualizados utilizando un escáner de fluorescencia (Typhoon Trio, GE Healthcare Bio-Sciences, Reino Unido), a fin de corroborar la correcta incorporación de la marca fluorescente en los mismos.

Purificación de los productos de PCR, clonado y secuenciación automática

Los productos de PCR que dieron más de una banda de amplificación fueron purificados a partir de tacos de agarosa utilizando el sistema QIAEX II (QIAGEN, Alemania), de acuerdo a las especificaciones del fabricante y se clonaron en el vector pGEMT-easy (Promega, EUA). Se secuenciaron con los iniciadores universales T7 y SP6 entre cinco y diez clones, provenientes de cada una de las reacciones de transformación llevadas a cabo (Sambrook *et al.* 1989). Aquellos productos de PCR de amplificación única fueron purificados con las enzimas Exo I-SAP (Thermo Fisher Scientific y Affymetrix, EUA) y secuenciados directamente como se describe en el apartado “secuenciación”.

Confirmación de la identidad de los genes candidato derivados de *Arabidopsis*

La validación de las secuencias obtenidas en girasol a partir de la amplificación por PCR se realizó por homología mediante búsquedas BLAST (Altschul *et al.* 1990). En primer lugar, se realizó una búsqueda BLASTx contra proteínas de *Arabidopsis* para evidenciar si la región obtenida para girasol se correspondía con la secuencia de aminoácidos de la proteína candidata de partida. Luego, se realizó una búsqueda tBLASTx contra ESTs de la familia *Asteraceae* para determinar si la región amplificada se correspondía con las secuencias de partida utilizadas como template en el diseño de iniciadores.

Identificación de SNPs y análisis de variabilidad

Las regiones codificantes y no codificantes de cada gen candidato se identificaron a través de búsquedas BLASTx (Altschul *et al.* 1990). Los niveles de variación genética se estimaron a través de los parámetros de polimorfismo nucleotídico (θ_w , Watterson 1975) y de diversidad

nucleotídica (π , Tajima 1983). El valor θ de Watterson se basa en el número de sitios segregantes, mientras que el valor π de Tajima se basa en las diferencias entre pares de secuencias de la muestra total. Estos parámetros fueron obtenidos usando el paquete de programas DNAsp 5.10.01 (Rozas 2009).

Selección del método de genotipificación de los genes candidato en la PMA

Para la selección de la metodología más adecuada para la genotipificación de los distintos genes candidato, se procedió al análisis de cada caso en forma independiente (Figura 3.1).

Para aquellos genes que presentaron únicamente dos o tres alelos o haplotipos en la colección CN, o que pudieron ser acotados a una región donde se observaran dos o tres haplotipos únicos, la genotipificación de la PMA se realizó mediante la detección de moléculas heterodúplex, ya sea mediante corte con endonucleasa CEL1 o cromatografía líquida de alta *performance* en condiciones de desnaturalización parcial (dHPLC).

Los GC que presentaron polimorfismos que generaron diferencias en el tamaño de los fragmentos (InDels o microsatélites) se genotipificaron utilizando la técnica de electroforesis capilar con detector de fluorescencia inducida por láser (ECF).

En aquellos casos en los que los patrones hallados no se correspondieron con ninguno de los identificados en las líneas control de la CN secuenciadas, los genes candidato se amplificaron con iniciadores no marcados y se secuenciaron para determinar los haplotipos nuevos.

Finalmente, se seleccionaron para secuenciación directa aquellos genes que presentaron varios SNPs y más de tres haplotipos entre las 8 líneas del GC, o aquellos casos en que no se pudo poner a punto la genotipificación mediante ninguna de las técnicas precedentes.

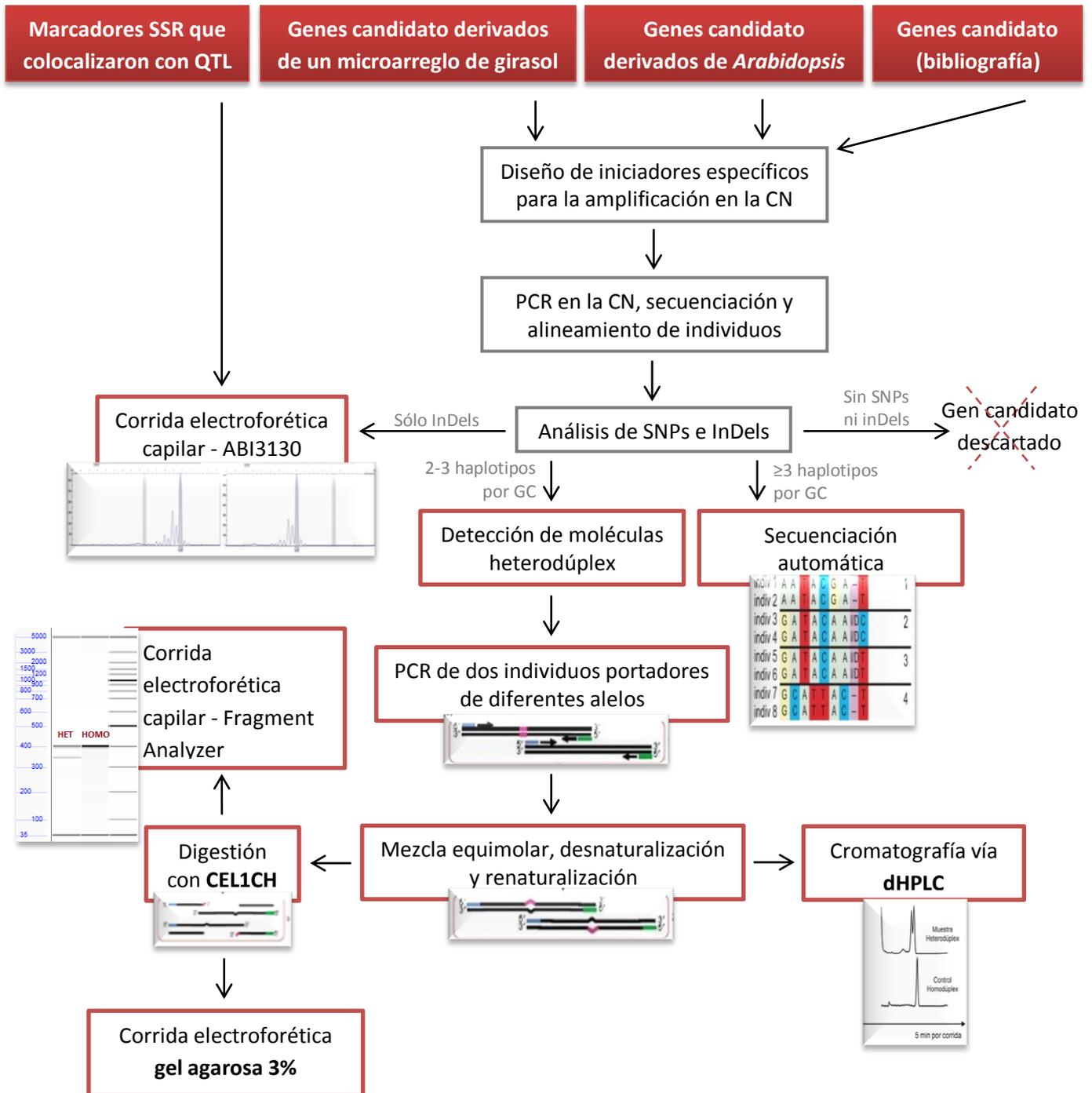


Figura 3.1 Diagrama de trabajo para la selección de la metodología de genotipificación de los GC

Genotipificación por métodos de detección de moléculas heterodúplex

La estrategia de detección de moléculas heterodúplex implica la comparación de dos o más alelos de un *locus* dado (A y B), cuyo producto de amplificación específico se mezcla y somete a un proceso de desnaturalización y lenta renaturalización de las hebras de ADN. A partir de este proceso, se producen por apareamiento aleatorio de las hebras complementarias dos tipos de moléculas: (1) las moléculas homodúplex, que son idénticas a las moléculas originales, y (2) las moléculas heterodúplex, que resultan del apareamiento de una hebra proveniente del alelo A y la complementaria del alelo B, y viceversa. Estas últimas, poseen apareamiento imperfecto (*mismatch*) en los sitios variables (SNPs o InDels). La distinción entre estos dos tipos de moléculas puede obtenerse mediante corte específico con la enzima CEL1 (CEL1CH), y posterior detección de los fragmentos correspondientes a las moléculas heterodúplex digeridas y a las moléculas homodúplex sin digerir (Perry *et al.* 2003; Till *et al.* 2007; Wang *et al.* 2008; Cooper *et al.* 2008; Xin *et al.* 2008; Uauy *et al.* 2009; Fusari *et al.* 2010), o mediante elución de las moléculas heterodúplex a tiempo diferencial respecto de las moléculas homodúplex, utilizando la técnica de dHPLC (*denaturing High Performance Liquid Chromatography*) (Giordano *et al.* 1999; Wagner *et al.* 1999; Kota *et al.* 2004, 2008; Lai *et al.* 2005; Randall *et al.* 2005; Fusari *et al.* 2010).

Para la genotipificación de las 137 líneas que conforman la PMA, el producto de PCR de cada línea se mezcló en forma equimolar con un producto de PCR de genotipo conocido obtenido por secuenciación de la CN.

La cuantificación de los productos de PCR para realizar las mezclas heterodúplex se llevó a cabo mediante espectrometría de fluorescencia usando el reactivo de Hoechst o por comparación de los datos de volumen e intensidad de las bandas visualizadas en el gel de agarosa, con un patrón de peso molecular de concentración conocida (100pb, New England Biolabs, EUA), utilizando el programa Image Quant (Amersham, EUA). La cuantificación con el reactivo Hoechst 33258 (Intercalante de ADN, PolySciences, EUA) se realizó por espectrofluorimetría en placas de 96 pocillos (Fluorómetro Spectra MAX GEMINI EM, Molecular Devices, EUA). Para la construcción de las curvas de calibración, se utilizó un *stock* de ADN de timo de cabra de concentración conocida, y mediante diluciones seriadas del mismo se construyeron los puntos de: 22,5; 45; 90; 180; 360 y 720 ng/ μ L. Cada uno de los puntos de la curva de calibración, así como de las distintas muestras, se mezclaron con la

solución de trabajo TNE-Hoechst (Hoechst 100 ng/mL y TNE 1X: Tris 10 mM, NaCl 200 mM, EDTA 1 mM, pH 7,4), en una relación 1:100 (ADN: solución de trabajo). El espectro se ajustó a una longitud de onda de excitación (λ) de 365 nm y una longitud de onda de emisión (λ) de 468 nm.

Una vez cuantificadas las muestras, se procedió a la mezcla de las mismas en forma equimolar. El proceso de desnaturalización y lenta renaturalización de las mezclas equimolares de productos de PCR se realizó en el termociclador Mastercycler *epgradient S* (Eppendorf, Alemania), mediante un programa de disminución gradual de la temperatura de 95 °C a 25 °C.

a. Corte del Heterodúplex con CEL1

La enzima CEL1 (endonucleasa 1 de apio, *Apium graveolens* L.) posee la capacidad de reconocer apareamientos imperfectos (*mismatches*) en las dobles hebras de ADN, y cortar ambas hebras en el extremo 3' de la región desapareada. El análisis de moléculas heterodúplex mediante corte específico con la enzima CEL1 (CEL1CH), fue realizado siguiendo el protocolo descrito por Fusari *et al.* (2010).

Brevemente, 150 ng de producto de PCR de una línea de genotipo conocido, fueron mezclados con 150 ng de producto de PCR de una muestra incógnita, y sometidos a un proceso de desnaturalización y renaturalización, tal como se describe en el inciso anterior. Esta mezcla fue posteriormente incubada con 1 μ L de extracto crudo de enzima (CJE) CEL1CH, 2 μ L de buffer 10x y H₂O *Chromasol* para llevar a volumen final de 20 μ L. La incubación se realizó a 45°C, temperatura óptima de acción de la enzima, durante 15 minutos.

La detección de los fragmentos generados, se realizó mediante corridas electroforéticas en gel de agarosa 3 %. Para el caso del gen HaGLP5, las moléculas heterodúplex observadas en gel de agarosa fueron confirmadas mediante detección de fragmentos de diferente tamaño utilizando el equipo Fragment Analyzer (Advanced Analytical Technologies, EUA), y el kit *Mutation discovery* (DNF-910-1000T Mutation Discovery 910 Gel Kit, 35 bp - 5000 pb, Advanced Analytical Technologies, EUA). Este equipo realiza electroforesis capilar automática, permitiendo la discriminación de fragmentos de diferente tamaño sin la necesidad de utilizar iniciadores portadores de marca, ya que utiliza agentes intercalantes que pueden ser excitados utilizando LEDs de alta densidad. Este instrumento provee alta

capacidad de proceso (hasta 96 muestras en una inyección simple), tiempos de corrida muy cortos (15 a 30 minutos en la mayoría de las aplicaciones), y alta resolución en la separación de fragmentos (hasta 2 pb en fragmentos de menos de 300 pb). Las salidas del mismo fueron analizadas con el *software* PROSize™, que asigna automáticamente el tamaño de los fragmentos obtenidos por comparación con un marcador de peso molecular, y corregidas manualmente.

b. Método de dHPLC

Para la realización de este método, se utilizó un equipo Agilent Serie 1100 (Agilent Technologies, EUA) equipado con un sistema de bio-compatibilidad, una unidad de degasificación de solvente, una bomba binaria, un módulo con control de temperatura para la inyección de muestra automática, un horno para la columna y un detector UV-Visible de arreglo de diodos. Las separaciones fueron realizadas utilizando la columna para ADN Helix™ Varian (3X50 mm, Varian, EUA), tal como se describe en Fusari *et al.* (2010). La cromatografía fue realizada con un caudal de 0,9 mL/min y un gradiente binario lineal de las Soluciones Helix A y B (Varian, EUA). La detección de las moléculas homo- y heterodúplex de ADN se realizó a $\lambda = 260$ nm.

Las combinaciones óptimas de gradiente y temperatura para la separación de las muestras homo- y heterodúplex fueron obtenidas empíricamente al ensayar las distintas posibles combinaciones predichas por el programa *dHPLC Melt Program* (disponible en el sitio de Internet de la Universidad de Stanford, <http://insertion.stanford.edu/melt.html>), sobre muestras heterodúplex obtenidas al mezclar dos líneas control de haplotipo distinto por cada gen candidato.

De acuerdo a las especificaciones del fabricante de la columna, se utilizaron como estándares el plásmido pUC18 digerido con Hae III (Varian, EUA ó Sigma Aldrich, EUA) y el estándar de detección de heterodúplex DYS271 (Varian Argentina LTD, Argentina).

El seguimiento de las corridas de dHPLC y la detección y análisis de los picos obtenidos se realizaron utilizando el *software* HPCHEM (Agilent Technologies Inc., EUA). Se consideraron los parámetros de integración por defecto del programa.

c. Asignación de genotipos

Dado que el producto de PCR de cada línea de genotipo incógnita se mezcló en cantidades equimolares con producto de PCR de cada una de las líneas de genotipo conocido detectado

por secuenciación sobre la CN, las mezclas control/incógnita solo revelaron un perfil de tipo homodúplex en aquellos casos en que la línea incógnita tuviera el mismo haplotipo que la control con la que se realizó la mezcla. De esta manera, cuando la mezcla control/incógnita mostró un perfil de tipo homodúplex, se asignó al genotipo incógnita el mismo genotipo que la línea control. Por su parte, aquellas que mostraron perfiles de tipo heterodúplex fueron mezcladas con producto de PCR de cada una de las otras variantes haplotípicas, hasta encontrar aquella coincidente.

La presencia de haplotipos distintos a los presentes en las líneas control se determinó cuando las mezclas control/incógnita no mostraron un patrón homodúplex para ninguno de los haplotipos ensayados. En este caso, se procedió a la secuenciación de los mismos para identificar nuevas variantes.

Secuenciación

Los productos de PCR de los genes candidato se purificaron para su posterior secuenciación tratando 10 μ L de la reacción de PCR con 5 U de Exo-I (Exonuclease I, Thermo Fisher Scientific, EUA) y 0,5U SAP (Shrimp Alkaline Phosphatase, Affymetrix, EUA). Las secuencias fueron obtenidas con el analizador automático de ADN ABI 3130xl (Applied Biosystems®, EUA). Cuando la credibilidad de las lecturas fue menor al 98%, se repitió la reacción de secuenciación. Los cromatogramas se analizaron con el programa BioEdit Sequence Alignment Editor v7.0.9.0 (Hall 1999).

Construcción del mapa genético

Con el propósito de testear las distintas metodologías de genotipificación utilizadas para otros usos, se seleccionaron 3 GC, que presentaron polimorfismos entre las líneas endocriadas PAC2 y RHA266, parentales de una población de mapeo de 90 RILS, y que cumplieran con las características de ser aptos para su genotipificación mediante las diferentes metodologías aquí utilizadas (dHPLC, corrida electroforética capilar, y digestión con endonucleasa CEL1), para ser genotipificados en dicha población biparental y así ser agregados al mapa disponible (Poormohammad Kiani *et al.* 2006). La construcción del mapa de ligamiento se realizó con el programa CarthaGene 0.999 (Schiex & Gaspin 1997). El análisis de la segregación de cada *locus* se efectuó utilizando la prueba de *Chi*-cuadrado. La función de mapeo de Kosambi (1944) se utilizó para calcular las distancias del mapa (cM) a

partir de las frecuencias de recombinación. Las frecuencias de recombinación estimadas en la población de mapeo de RILs fueron corregidas por el factor que considera meiosis múltiples $r = R / (2-2R)$, donde $R = nr / (nr + nn)$ es la frecuencia de recombinación, y nr y nn 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*) de 4,0 y una frecuencia de recombinación máxima de 0,35. Una vez definidos los grupos de ligamiento que contenían los genes candidato, el orden específico de cada marcador dentro de cada grupo de ligamiento fue analizado de manera individual usando el programa Mapmaker 3.0 (Lander *et al.* 1987). El programa Mapchart 2.1 se usó para realizar la representación gráfica de los grupos de ligamiento y las ubicaciones de los marcadores funcionales (Voorrips 2002).

RESULTADOS

Amplificación de genes candidato derivados de un microarreglo de EST de girasol

Del total de 14 GC seleccionados a partir del análisis de expresión diferencial en microarreglos de girasol (Tabla 3.1), se logró la amplificación de un producto único para los genes HeAn_41021, HeAn_41022, HeAn_36589, HeAn_315, HeAn_12761, HeAn_12562, HeAn_28738, HeAn_6452 y HeAn_35791, con un tamaño entre 287 y 793 pb. Para el gen HeAn_1751 se obtuvo un producto de amplificación de muy baja concentración, y de un tamaño mucho mayor a lo esperado (~2000 pb), lo cual impidió el posterior paso de secuenciación. Para los genes restantes, no se observó amplificación a ninguna de las temperaturas probadas. Los nueve GC resultantes fueron amplificados en las líneas que conforman la CN para la búsqueda de SNPs e InDels.

Identificación de ortólogos y amplificación de genes candidato derivados de estudios de expresión diferencial en *B. napus* (Zhao *et al.* 2007)

La identificación de genes de girasol ortólogos a los tomados del trabajo de Zhao *et al.* (2007) (Tabla 3.2) resultó exitosa en todos los casos. El número de secuencias recuperadas mediante tBLASTn para cada locus fue variable, oscilando entre 55 secuencias para el gen AT1G19100, hasta 111 para el gen AT3G28390. En cuanto a la cantidad de géneros representados, los valores extremos fueron 34 (para el *locus* AT1G19100) y 84 (para el *locus* AT4G03400) (Tabla 3.5).

Tabla 3.5 Identificación de ortólogos mediante reconstrucción filogenética

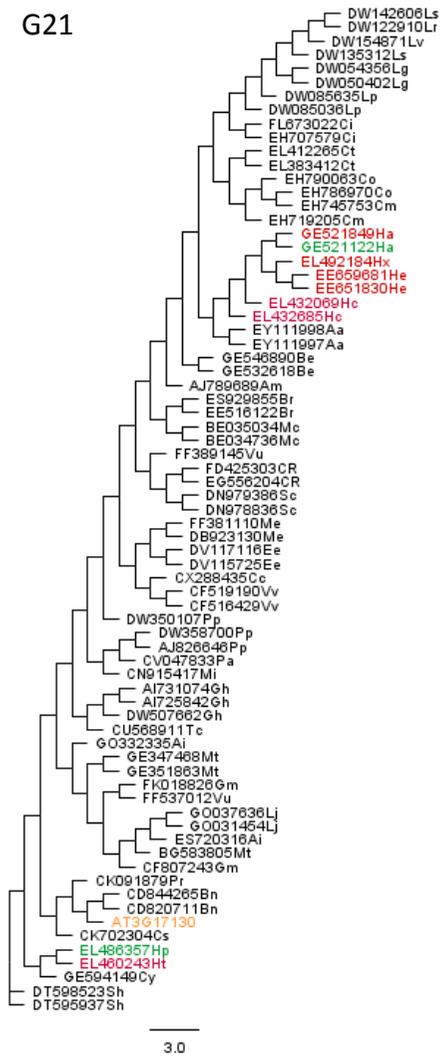
ID Girasol	<i>Locus</i> de interés	#Secuencias ^a (# Géneros)	# Árboles de máxima parsimonia	Secuencias utilizadas para el diseño de iniciadores ^b
G21	AT3G17130	73 (43)	12	EL486357HxA, GE521122HaB
G22	AT4G03400	92 (84)	2024	EE636970HeA, BQ917026Ha
G26	AT3G10190	73 (42)	39	DY919352Ha,
G28	AT1G19100	55 (34)	2	EL438232Ht
G30	AT5G02140	103 (65)	2100	DY943567Hp, EL430860Hc
G33	AT3G19630	100 (77)	2637	GE499412Ha
G34	AT3G28390	111 (73)	3192	BQ916932Ha, BQ969360Ha

a Número de secuencias recuperadas mediante las búsquedas tBLASTn (número de géneros representados).

b Números de Accesos del GenBank o KEGG de las secuencias seleccionadas para el diseño de oligonucleótidos de acuerdo al análisis de los árboles obtenidos.

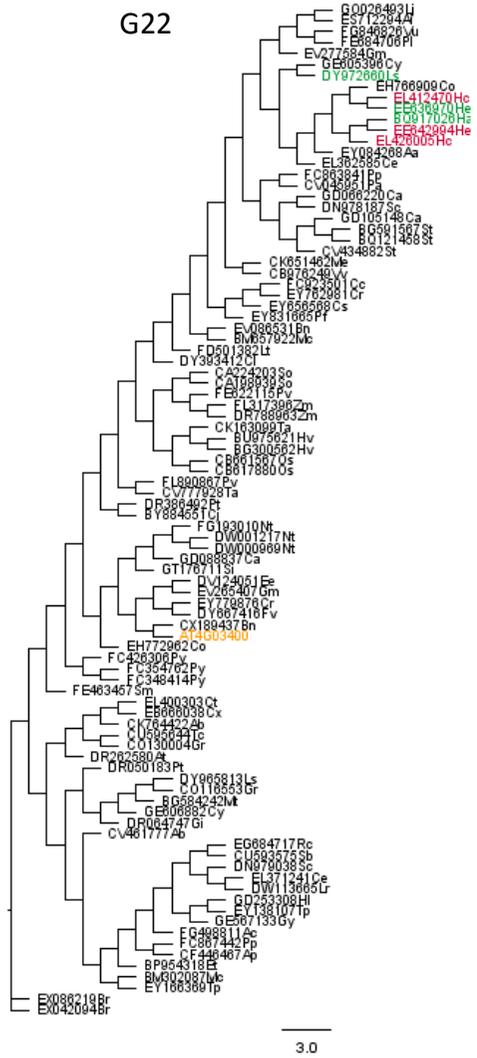
Las reconstrucciones filogenéticas basadas en MP permitieron obtener las relaciones entre las secuencias recuperadas para todos los genes analizados. En la Figura 3.2 se presentan los árboles consenso obtenidos para los siete *loci*. En algunos casos, al menos una secuencia del género *Helianthus* quedó dentro del mismo clado que la secuencia de *Arabidopsis*, facilitándose de este modo la identificación de ortólogos (Figura 3.2, D, E). Por el contrario, para algunos *loci* la información contenida en las secuencias analizadas no fue suficiente para obtener árboles con alto nivel de resolución, debiendo elegirse al menos una secuencia de cada uno de los grupos involucrados en las distintas politomías (*i.e.* Figura 3.2 A, G). Para otros casos, si bien el/los árboles encontrados alcanzaron un alto grado de resolución, la topología del árbol parece indicar la existencia de un evento de duplicación génica en el grupo de representantes de la familia *Asteraceae*, haciendo difícil la elección de un único conjunto de secuencias para el diseño de los iniciadores. Los genes identificados como posibles ortólogos o parálogos se utilizaron para el diseño de iniciadores, con los cuales se amplificaron las regiones correspondientes a partir de ADN genómico de girasol.

G21



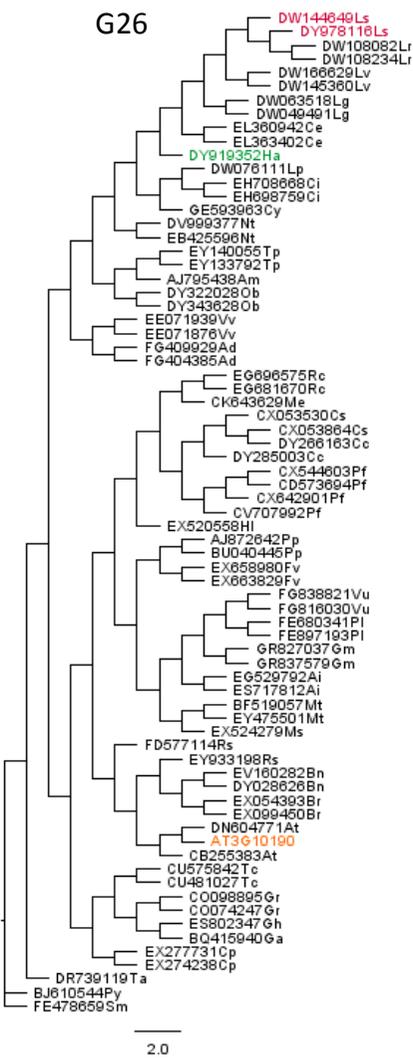
A

G22



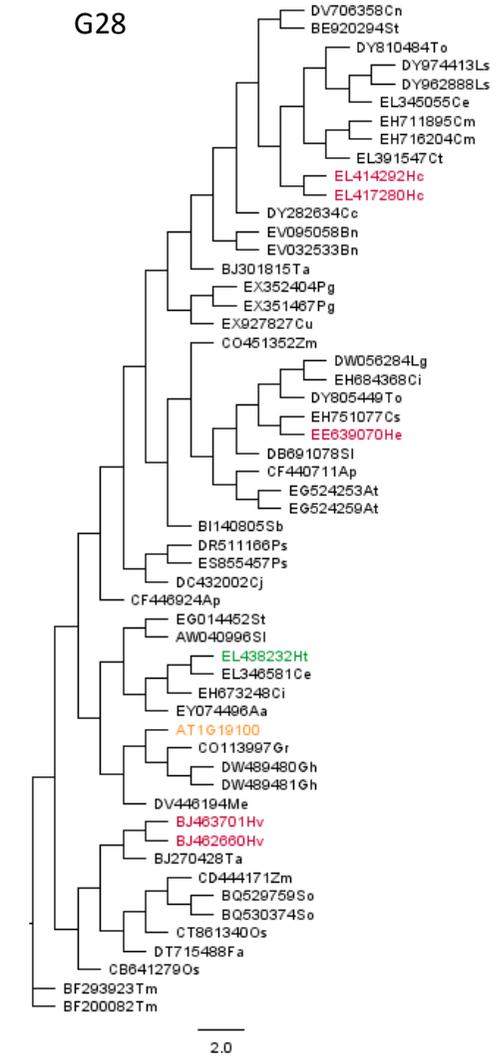
B

G26



C

G28



D

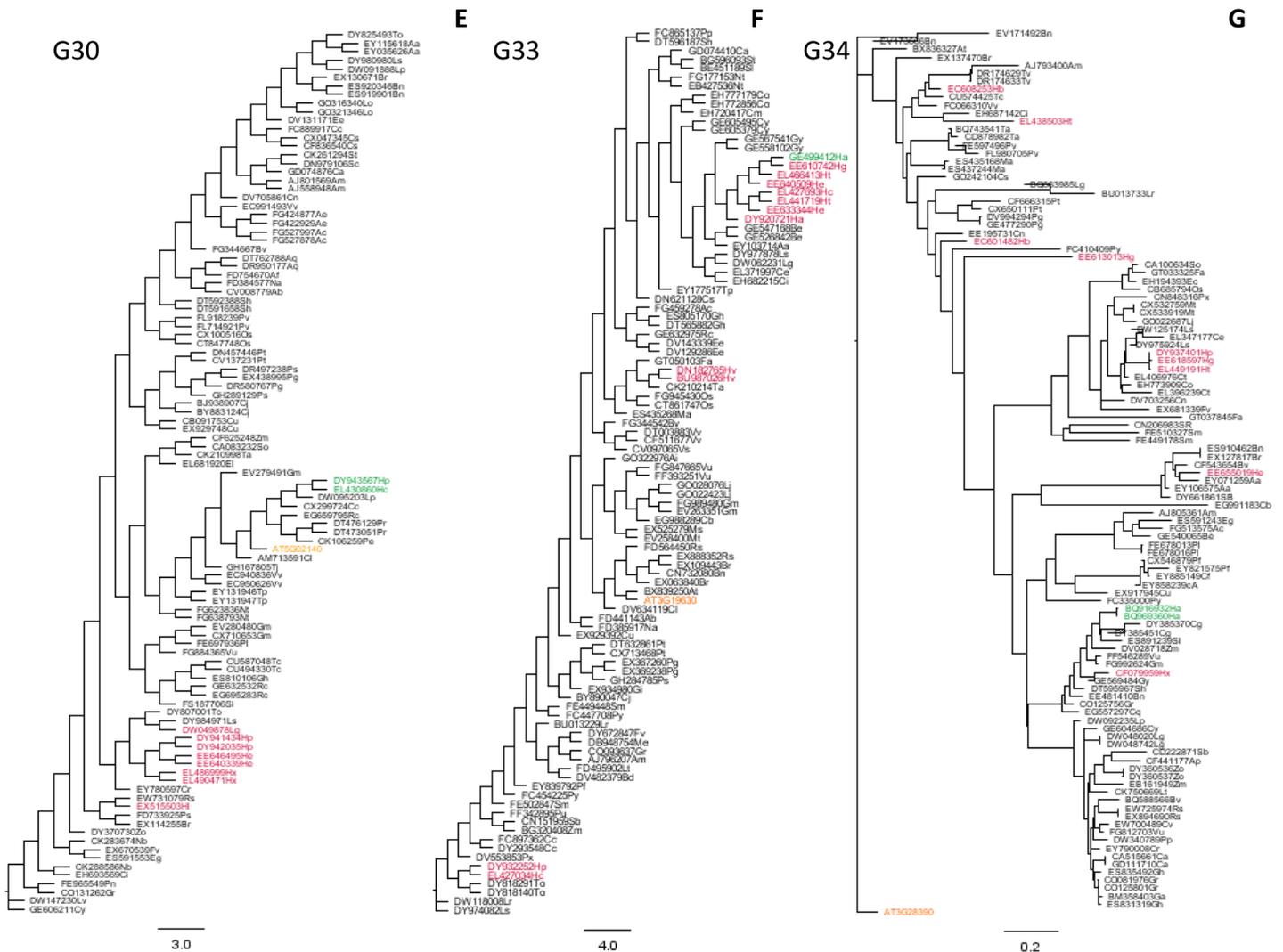


Figura 3.2 Consenso de los Árboles de Máxima Parsimonia (MP) obtenidos para los genes candidato tomados de Zhao *et al.* (2007) (A) Consenso de 12 árboles de MP para el locus AT3G17130 (G21); (B) Consenso de 204 árboles de MP para el locus AT4G03400 (G22); (C) Consenso de 39 árboles de MP para el locus AT3G10190 (G27); (D) Consenso de 2 árboles de máxima parsimonia para el locus AT1G19100 (G28); (E) Consenso de 2100 árboles de máxima parsimonia para el locus AT5G02140 (G30); (F) Consenso de 2637 árboles de máxima parsimonia para el locus AT3G19630 (G33); (G) Consenso de 3192 árboles de máxima parsimonia para el locus AT3G28390 (G34). Se indican las secuencias originales de *Arabidopsis* tomadas de Zhao *et al.* (2007) (naranja), y las secuencias ortólogas o parálogas, de los géneros *Helianthus* y *Lactuca* identificadas (rojo). En verde se señalan las secuencias seleccionadas para el diseño de iniciadores. Taxones identificados con el número de acceso del GenBank, seguido por un identificador de género y especie (detallados en el Apéndice, Tabla A3.1).

Al realizar las reacciones de amplificación sobre ADN genómico de girasol, se observó una sola banda de amplificación para todos los genes. La única excepción fue el gen G28 para el cual se observaron dos bandas de amplificación de distinto tamaño. Ambas fueron purificadas a partir de taco de agarosa, clonadas y secuenciadas con iniciadores universales.

El gen G26 solo pudo ser amplificado en una de las dos líneas ensayadas (RHA801), sugiriendo la posible existencia de un polimorfismo en el sitio de pegado de iniciadores.

Las secuencias obtenidas para los siete GC fueron cotejadas mediante búsquedas BLASTx contra base de datos de *Arabidopsis* para confirmar su similitud con la secuencia de partida (Tabla 3.6).

Debido a que los genes habían sido amplificados a partir de ADN genómico, la mayoría presentó regiones no codificantes, correspondientes a intrones de dichos genes, que contribuyeron a disminuir los E-valores de los resultados de las búsquedas mediante BLAST (Tabla 3.6).

Una vez confirmada su homología con la secuencia origen, la secuencias de G28 y G26 fueron utilizadas como templatado para el diseño de nuevos iniciadores específicos que amplifiquen una región única, para ser amplificados en la CN a fin de identificar SNPs e InDels factibles de ser incorporados en el estudio de mapeo por asociación.

Tabla 3.6 Resultados de las búsquedas BLASTx de las secuencias de girasol correspondientes a los genes candidato tomados de Zhao *et al.* (2007)

Gen candidato	ID Girasol	Longitud (pb)	Resultado BASTx contra <i>A. thaliana</i>	E-valor
At3g17130 Familia de proteínas inhibidoras Invertasa/pectina metilesterasa	G21	425	NP_188339.1 Plant invertase/pectin methylesterase inhibitor superfamily protein	1 E-11
At4g03400 Familia de proteínas GH3 de respuesta a auxinas	G22	876	NP_001078069.1 jasmonic acid-amido synthetase JAR1	5 E-64
At3g10190 Calmodulina	G26	1025	NP_187630.1 putative calcium-binding protein CML36	4 E-10
At1g19100 Región de unión a ATP, dominio ATP-asa	G28	447	NP_173344.2 Microrchidia adenosine triphosphatase (ATPase) family protein	3 E-4
At5g02140 Proteína tipo thaumatina	G30	606	NP_195834.1 pathogenesis-related thaumatin-like protein	1 E-99
At3g19630 Proteína conteniendo dominio radical SAM	G33	249	NP_188597.1 radical SAM domain-containing protein	3 E-16
At3g28390 Glicoproteína P	G34	640	BAF00235.1 P-glycoprotein	1 E-88

Amplificación de GC descriptos como partícipes en procesos de defensa y de marcadores que colocalizaron con QTL de resistencia a *Sclerotinia*

Dado que los 7 GC seleccionados a partir del trabajo de Fusari *et al.* (2012) ya habían sido caracterizados sobre un subconjunto de líneas endocriadas de la PMA descrita en este trabajo, no fue necesario diseñar nuevos iniciadores, contándose además con información previa los niveles de polimorfismo de dichos GC.

Para la amplificación de los dos genes candidato seleccionados por su participación en defensa contra PBT a partir del trabajo de Talukder *et al.* (2014), se tomaron los iniciadores diseñados por los autores. A pesar de haberse ensayado tres programas de ciclado diferentes, sólo uno de ellos (HaCOI_1) permitió amplificación en concentración suficiente como para permitir la secuenciación del producto amplificado en las 8 líneas que integran la CN.

Para los cuatro genes restantes seleccionados por bibliografía, los iniciadores diseñados permitieron la amplificación de todas las regiones genómicas.

En el caso de los once marcadores tipo SSR que colocalizaron con QTL, se descartaron dos por no amplificar (debido probablemente a la antigüedad de los iniciadores), quedando nueve para análisis posteriores.

Construcción de colecciones núcleo

Con el fin de abarcar la mayor diversidad posible para la exploración de los polimorfismos en los GC y de minimizar los costos de secuenciación, se construyó una CN constituida por 8 líneas utilizando el método propuesto por Liu y Muse (2005). De las diez réplicas generadas a partir de la información derivada de SSR, se seleccionó arbitrariamente una de las que incluyó el mayor número de alelos: 132 del total de 172 detectados para el conjunto total de líneas de la PMA (Capítulo 2). La colección núcleo definitiva quedó conformada por las líneas endocriadas 2091, 2121, 51084/5429, B99, C154B, C625U, R419 y R449 (Tabla 3.7).

Tabla 3.7 Posibles conformaciones de la Colección Núcleo de 8 líneas

Repetición	# Alelos	Líneas
1	131	2091-51084/5429-C110M-C154B-C208F-C454B-R449-R461-2
2	132	2091-51084/5429-C110M-C154B-R449-R461-2-R463-4-RHA801
3	131	2091-5313-B99-C154B-C208F-R449-R459-1-R461-2

Repetición	# Alelos	Líneas
4	132	51084/5429-5286-B99-C110M-C154B-C625U-R449-R461-2
5	131	5286-7X1-B99-C110M-C154B-C625U-R449-R461-2
6	129	51084/5429-5341.1-C110M-C154B-C625U-R424-R449-R463-4
7	132	2091-2121-51084/5429-B99-C154B-C625U-R419-R449
8	132	2091-51084/5429-5393E-B99-C154B-C625U-R449-R461-2
9	131	2121-5313-5393E-B10-C1059-C154B-R449-R459-4
10	129	2091-51084/5429-5313-B99-C154B-R449-RHA274-RHA440

En negrita se indica el conjunto de líneas seleccionado

Caracterización de las regiones candidatas en la Colección Núcleo de líneas de girasol

De los 34 GC seleccionados para asociación, 27 pudieron ser correctamente secuenciados en la CN, obteniéndose de este modo un total de 16438 pb secuenciadas por individuo, de las cuales 9752 pb corresponden a regiones codificantes y las 6686 pb restantes a regiones no codificantes. La región génica HeAn_6452, obtenida a partir del microarreglo de girasol, presentó problemas para la secuenciación, de modo que debió ser descartada del análisis. La longitud de las regiones amplificadas varió entre 189 y 1762 pb, excluyendo los InDels, con un promedio de 608,8 pb por región (Tabla 3.8). Al observar la Tabla 3.8 se evidencia la ocurrencia de al menos un SNP en 25 de los 27 genes evaluados, con un total de 238 cambios nucleotídicos identificados en el conjunto de genes e individuos analizados. En promedio, se detectó un SNP cada 69 pb (excluyendo los InDels) y 8,8 SNPs por región. El número de sustituciones sinónimas encontrado fue de 85, 1,7 veces más abundantes que las sustituciones no sinónimas (50). El número de SNPs varió también entre las regiones codificantes y no codificantes: la frecuencia de SNPs varió de 1 SNP/72 pb en zonas codificantes a 1/65 pb en zonas no SNPs codificantes.

El número de InDels fluctuó entre 0 y 10, contabilizándose un total de 66 InDels en el panel completo. La frecuencia de InDels encontrada fue de uno cada 249 pb, con un promedio de 2,44 InDels por región analizada. El tamaño de estas inserciones/deleciones osciló entre un único nucleótido hasta 12 pb en el caso del gen HaWRKY7.

Los niveles de polimorfismo detectados fueron moderados (Tabla 3.8). La variación genética a nivel nucleotídico fue estimada como la diversidad nucleotídica en promedio para todos los genes evaluados ($\pi_T = 0,007$), y como el número de sitios segregantes ($\theta_W = 0,006$). Los valores de diversidad promedio para sitios silenciosos, es decir sitios sinónimos y no-

codificantes ($\pi_{\text{sil}} = 0,015$), y para sitios sinónimos únicamente ($\pi_{\text{sin}} = 0,013$) fueron mayores que los de los cambios no sinónimos ($\pi_{\text{no-sin}} = 0,005$). En 21 de las 27 regiones examinadas $\pi_{\text{no-sin}}$ fue o bien cero (0) o menor al valor de π_{sin} . Sin embargo, las regiones G28, HeAn_12761 y HaTRP mostraron valores de $\pi_{\text{no-sin}}$ mayores a los de π_{sin} (Figura 3.3 y Tabla 3.8). Para G28 $\pi_{\text{no-sin}}$ ascendió a 0,021, mientras que el valor de π_{sin} fue cero (0); la diferencia en estos valores se debió a dos sustituciones nucleotídicas: la primera en la posición 116 que genera un cambio de isoleucina por valina y la segunda en la posición 175, que genera un cambio de ácido aspártico a glutámico. Algo similar ocurrió con la región HeAn_12761, donde $\pi_{\text{no-sin}}$ ascendió a 0,00091, mientras que el valor de π_{sin} fue cero (0); en este caso, la diferencia en estos valores se debió a una sustitución nucleotídica en la posición 332, que generó un cambio de isoleucina a treonina. Finalmente, para el gen HaTRP, $\pi_{\text{no-sin}}$ ascendió a 0,012, mientras que el valor de π_{sin} fue 0,004. En este caso, la diferencia se debió a cinco sustituciones: la primera en la posición 39, (leucina por glicina), la segunda en posición 69 (leucina por valina), la tercera en posición 168, (asparragina por ácido aspártico), la cuarta en posición 333 (serina por alanina) y la quinta en posición 394 (glicina por ácido aspártico). Estos resultados se presentan gráficamente en la Figura 3.3. En la misma, se observa que la mayor parte de los GC (puntos azules) presentan un π_{sin} mayor a $\pi_{\text{no-sin}}$, con la excepción de G28 (rojo), HeAn_12761 (verde) y HaTRP (naranja).

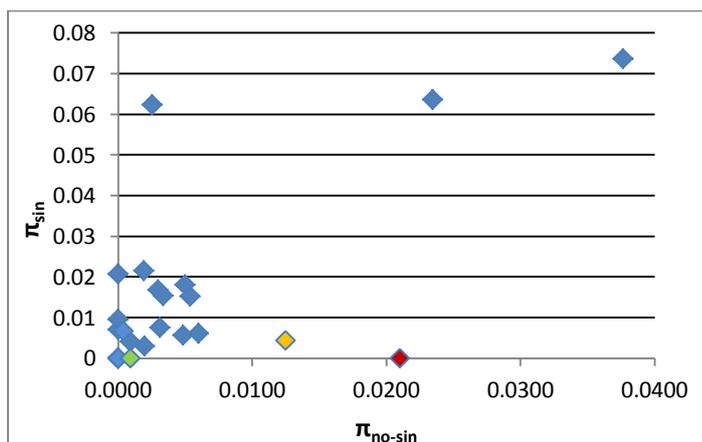


Figura 3.3 SNPs sinónimos (π_{sin}) y no sinónimos ($\pi_{\text{no-sin}}$) en las 27 regiones candidatas analizadas

Se observa la diversidad nucleotídica de SNPs sinónimos y no sinónimos. Se señalan aquellas regiones candidatas que mostraron un valor de $\pi_{\text{no-sin}}$ mayor que π_{sin} : G28 (rojo), HeAn_12761 (verde) y HaTRP (naranja).

Tabla 3.8 Características de los 27 genes candidato seleccionados para realizar mapeo por asociación en girasol

ID	Tamaño (pb)	S _i ^a	N° InDel	θ_w	π_T	π_{sil}	π_{sin}	π_{no-sin}	$\frac{\pi_{no-sin}}{\pi_{sin}}$	N° haplotipos	Diversidad Haplotípica ^b
HeAn_12562	504	21	0	0,02	0,02	0,02	0,06	0,023	0,369	2	0,536
HeAn_12761	606	1	0	0	0	0	0	0,001		2	0,429
HeAn_28738	677	0	3	0	0	0	0	0		2	0,286
HeAn_315	322	5	1	0,01	0,01	0	0,02	0,005	0,352	3	0,556
HeAn_35791	793	0	0	0	0	0	0	0		1	0
HeAn_36589	397	7	3	0,01	0,01	0	0,01	0,005	0,858	3	0,464
HeAn_41021	509	13	0	0,01	0,01	0,02	0,02	0,003	0,219	3	0,607
HeAn_41022	287	3	0	0	0	0	0	0		3	0,464
G21	392	5	1	0	0	0	0	9E-04	0,225	2	0,133
G22	836	9	1	0	0	0	0,01	0,006	0,984	3	0,514
G26	980	9	5	0	0	0	0	0		2	0,476
G28	429	16	9	0,01	0,01	0,01	0	0,021		5	0,362
G30	586	2	0	0	0	0	0,01	0	0	2	0,514
G33	225	3	8	0	0	0	0	0,002	0,609	5	0,133
G34	621	15	2	0,01	0,01	0	0,02	0,005	0,291	4	0,619
HaGLP3	621	10	0	0,01	0,01	0,02	0,02	0,002	0,089	3	0,607
HaGLP4	782	11	5	0,01	0	0,01	0,01	0	0,058	5	0,576
HaGLP5	1762	2	4	0	0	0	0	0		3	0,618
HaCOI_1	659	14	0	0,01	0,01	0,02	0,02	0		5	0,857
HaCP	542	18	1	0,02	0,02	0,02	0,07	0,038	0,512	3	0,833
HaDRP	597	22	2	0,01	0,02	0,05	0,06	0,003	0,041	6	0,778
HaPAL	343	2	0	0	0	0,01	0,01	0		2	0,467
HaTRP	676	16	5	0,01	0,01	0,02	0	0,012	2,843	5	0,722
HaWIP	376	4	3	0	0,01	0,01	0,02	0,003	0,177	2	0,533
HaWRKY5	249	7	2	0,01	0,01	0,21	0	0		4	0,821
HaWRKY7	201	1	1	0	0	0	0	0		2	0,222
RhoBP	1466	22	10	0,01	0,01	0	0,01	0,003	0,422	4	0,764
Promedio	608,81	8,8	2,44	0,006	0,007	0,015	0,013	0,005	0,503	3,185	0,514

a Sitios informativos (SI) usados en las medidas de diversidad nucleotídica.

b El número de haplotipos y la diversidad haplotípica se obtuvieron teniendo en cuenta los polimorfismos de tipo InDel.

Los espacios vacíos en la relación π_{no-sin}/π_{sin} corresponden a aquellos casos en los que π_{sin} es igual a cero.

La cantidad de haplotipos detectados varió de uno a seis dependiendo de la región analizada, exhibiendo la mayoría de los genes entre dos y tres haplotipos (Figura 3.4). El promedio de diversidad haplotípica fue de 0,514; pero los valores variaron dependiendo del número de haplotipos presentes y de las frecuencias encontradas para cada uno.

Aunque la secuencia del gen candidato HeAn_28738 no mostró polimorfismos de tipo SNP, los InDels identificados permitieron determinar dos haplotipos, obteniendo un valor de

diversidad haplotípica de 0,286. Por último, la región HeAn_35791 no presentó polimorfismos de ningún tipo, es decir, se observó un único haplotipo (Figura 3.4).

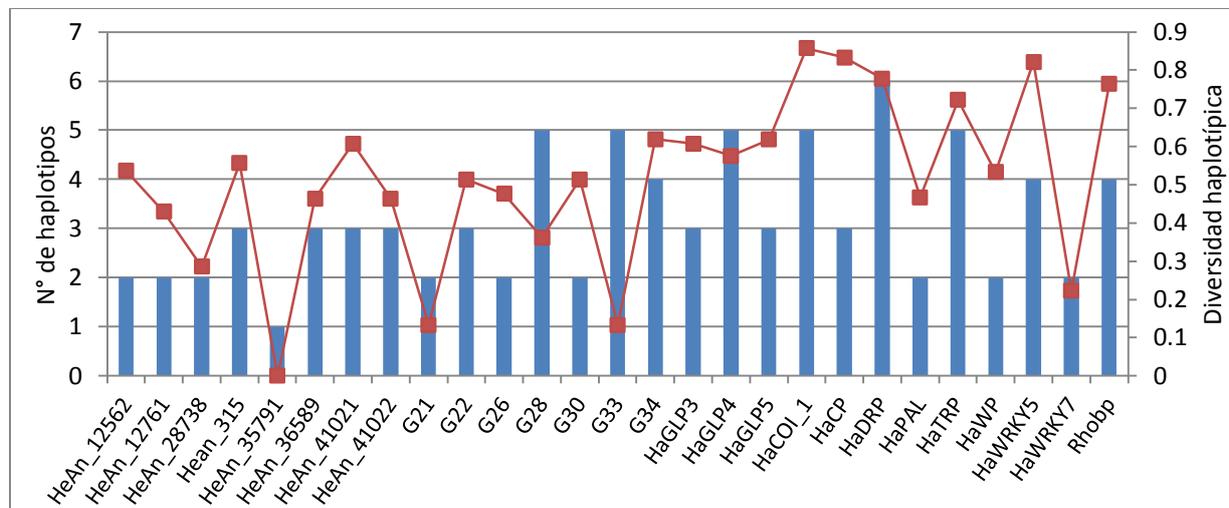


Figura 3.4 Número de haplotipos encontrados y diversidad haplotípica en los 27 genes de girasol analizados

Las barras indican el número de haplotipos de cada gen (eje izquierdo); los puntos señalan los respectivos valores de diversidad haplotípica (eje derecho).

Genotipificación de genes candidato para análisis de asociación

Dada la alta estabilidad genotípica por línea (ver Capítulo 4), los genes candidato se genotipificaron sobre un único individuo por línea. Para la selección de los métodos de genotipificación a utilizar en la caracterización de la PMA se siguió el protocolo establecido en la Figura 3.1, teniendo en consideración el costo general del ensayo y las posibilidades técnicas disponibles. De este modo, cuando los genes presentaron únicamente dos haplotipos en la CN, o cuando el gen pudo acotarse a una región donde se establecieran dos haplotipos únicos, se realizó la genotipificación de la PMA mediante la detección de moléculas heterodúplex. Los genes candidato que cumplieron esta condición fueron: G21, G22, G26, G28, G30, G34, HaGLP3, HaGLP5, HACP, HADRP, HAPAL, HeAn_315, HeAn_41022, HeAn_12761.

De estos 14 genes candidato, HaGLP5 y HeAn_315 fueron puestos a punto para la identificación mediante corte con endonucleasa CEL1. El revelado de ambos GC se realizó mediante corridas electroforéticas en gel de agarosa al 3 %, dado que los productos de digestión presentaban tamaños fácilmente diferenciables. Para el gen HaGLP5, se confirmaron los fragmentos heterodúplex detectados mediante corridas electroforéticas automáticas con el equipo Fragment Analyzer (Figura 3.5).

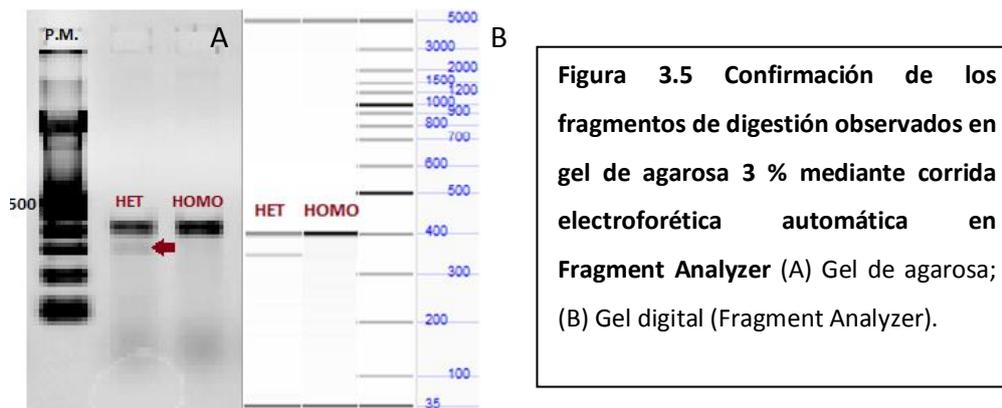


Figura 3.5 Confirmación de los fragmentos de digestión observados en gel de agarosa 3 % mediante corrida electroforética automática en **Fragment Analyzer** (A) Gel de agarosa; (B) Gel digital (Fragment Analyzer).

Si bien los 12 GC seleccionados para genotipificación mediante detección de heterodúplex cumplían con las condiciones para el uso de dHPLC, no siempre fue posible adoptar este método. Ejemplo de esto lo constituyen los GC G28, HeAn_12761 y HeAn_41022, para los cuales ninguna de las condiciones preestablecidas por el programa *dHPLC MeltProgram* permitió la separación de las moléculas heterodúplex de las homodúplex. Por su parte, sí pudieron ser genotipificados utilizando la metodología dHPLC los GC G21, G22, G26, G28, G30, HaGLP3, HaCP, HaDRP, HaPAL y G34. Este último, si bien contaba con tres haplotipos, fueron fácilmente separables utilizando esta metodología, aunque debieron optimizarse las condiciones de gradiente y temperatura para la detección de las moléculas heterodúplex, debido a que las propuestas por el programa *dHPLC MeltProgram* no lograban una separación clara.

En la Tabla 3.9 se detallan los GC, el tamaño de los fragmentos utilizados para genotipificación por detección de heterodúplex y las variantes alélicas que conforman cada haplotipo. Se observa que tanto CEL1CH como dHPLC fueron capaces de discriminar haplotipos que diferían entre sí en un número de bases nucleotídicas que fue desde uno, como fue el caso de HaGLP5, hasta a once, como fue el caso de HaCP.

Tabla 3.9 Características de los genes candidato genotipificados por la metodología de detección de moléculas heterodúplex

Método	ID	Tamaño (pb)	Haplotipos ^a
CEL1CH	HaGLP5	401	a... t... 291
	HeAn_315	322	t... a... a... g... C... -... C... t... g... a... g... g... 23 25 116 172 175 314

Método	ID	Tamaño (pb)	Haplotipos ^a											
dHPLC	HaDRP	200	c....	t....	g....	g....	c....							
			t....	a....	a....	a....	a....							
			33	108	114	138	162							
	HaPAL	343	a....	g....										
			g....	c....										
			43	256										
	HaGLP3	505	a....	t....	g....	g....	t....	t....	a....					
			g....	c....	a....	a....	a....	c....	g....					
			21	63	88	234	316	324	483					
	HaCP	319	t....	c....	c....	g....	t....	g....	a....	g....	g....	c....	a....	
			c....	g....	t....	a....	c....	c....	g....	t....	a....	g....	g....	
			49	68	93	107	114	128	159	185	228	229	245	
	G21	392	c....	a....	t....	t....	t....							
			g....	g....	a....	c....	c....							
			40	141	223	235	246							
	G22	555	c....	a....	g....									
			t....	g....	a....									
			32	50	283									
	G26	448	t....	t....										
			c....	c....										
			137	144										
	G30	383	g....	g....										
			a....	a....										
			26	29										
	G34	374	a....	-....	t....	t....	g....	g....	g....	t....	a....			
			-....	c....	c....	c....	a....	t....	g....	c....	g....			
			a....	-....	t....	t....	g....	g....	a....	t....	a....			
			62	64	121	139	156	197	204	206	297			

^a Los números debajo de las bases nucleotídicas indican la posición de los sitios polimórficos en el alineamiento de todos los haplotipos evaluados.

Al aplicar el método de dHPLC para los genes candidato HaDRP, G26, G30 y HaCP sobre las 137 líneas que conforman la PMA, se identificaron líneas cuyo genotipo no coincidió con ninguno de los representantes de los haplotipos encontrados en la colección núcleo. Al amplificar y secuenciar dichos GC en las líneas diferenciales, logró identificarse un nuevo haplotipo para los genes HaDRP y G30, dos nuevos haplotipos para el G26, y tres nuevos haplotipos para el gen HaCP (Figura 3.6).

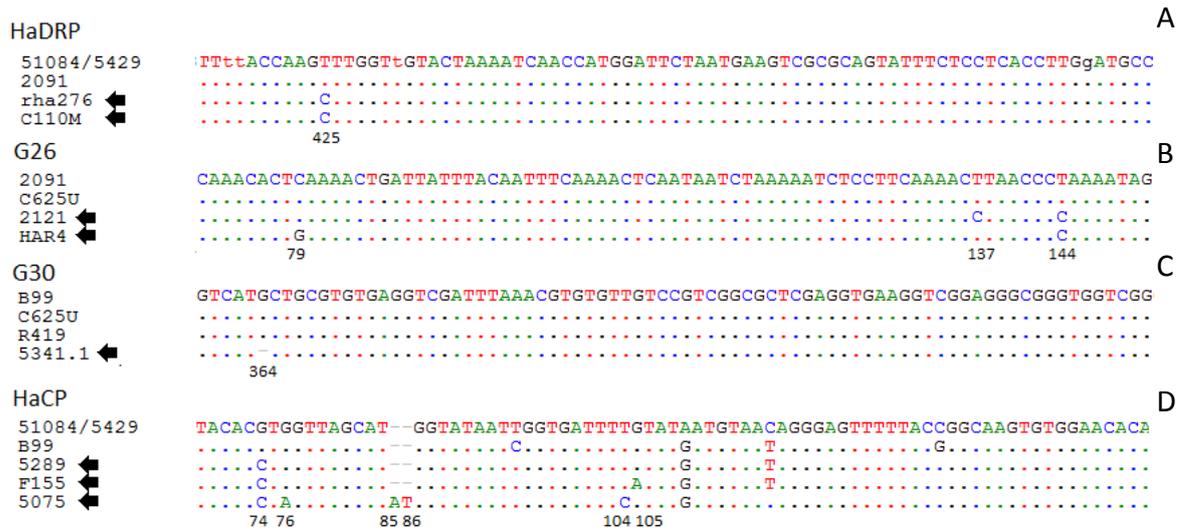


Figura 3.6 Nuevos haplotipos detectados con dHPLC y confirmados por secuenciación directa Genotipos de las líneas control para los genes (A) HaDRP; (B) G26; (C) G30; (D) HaCP, junto con las líneas que evidenciaron un haplotipo distinto a los conocidos. Los puntos identifican posiciones idénticas o secuencias conservadas. Se destacan las posiciones variables, aquellos SNPs no presentes entre las líneas control corresponden a los nuevos sitios encontrados. Debajo de cada conjunto de secuencias se señala la posición afectada. Las flechas señalan las líneas portadoras de los nuevos genotipos.

Cuando las regiones presentaron InDels o microsatélites, se realizó la genotipificación de la PMA por diferencia en el tamaño del fragmento amplificado mediante ECF. Los genes seleccionados para este método de genotipificación fueron G33, RhoBP, HaWP, HaTRP, HaWRKY5, HaWRKY7 y HaGLP4, y los SSR HA77, HA1848, HA2989, HA3103, HA3847, HA4264, ORS331, ORS460 y ORS1041.

Finalmente, se seleccionaron para secuenciación directa aquellos genes que presentaron varios SNPs y más de dos haplotipos entre las 10 líneas de la CN, siendo HeAn_41021, HeAn_36589, HeAn_126562 y HaCOI_1 los que cumplieron esta condición. La secuenciación directa de productos de PCR se realizó con un único iniciador, que fue seleccionado por obtener el mejor rendimiento en cuanto a cantidad de pb secuenciadas y calidad de la secuencia.

Al caracterizar las 137 líneas de la PMA para las 31 regiones candidatas (22 GC y 9 loci SSR), se obtuvo un alto porcentaje de éxito, registrando sólo un 3 % de datos faltantes para el total de los genes (Tabla 3.10). Se identificaron ocho haplotipos como máximo para el gen HaCOI_1, y ocho genes presentaron únicamente dos haplotipos en la PMA (G21, G22, HaGLP5, HaPAL, HaTRP, HaWP, HaWRKY7 y HeAn_315).

Todos los iniciadores diseñados y utilizados para la genotipificación de las 137 líneas de la PMA se detallan en la Tabla A3.2 del Apéndice y los resultados de la genotipificación de los GC se presentan en la Tabla A3.3 del Apéndice.

Tabla 3.10 Genotipificación de GC en la PMA

Gen ID	Tamaño (pb)	Nº Haplotipos ^a	Método	% genotipificación	Comentarios ^b
HeAn_12562	504	5	Secuenciación	93,3	Iniciador <i>Forward</i>
Hean_315	322	2	CEL1	100	
HeAn_36589	397	3	Secuenciación	98,5	Iniciador <i>Reverse</i>
HeAn_41021	509	6	Secuenciación	91,1	Iniciador <i>Reverse</i>
G21	392	2	dHPLC	100	56°C// 50-55,6-61
G22	555	2	dHPLC	100	59°C// 50-55,6-62,3
G26	448	3	dHPLC	100	58°C // 50-56,6-63,1
G30	383	4	dHPLC	100	58°C// 50-55,6-61
G33	252	4	ECF	90,5	
G34	374	3	dHPLC	100	56°C// 50-55,6-61 48°C // 50-52-59
HaGLP3	505	3	dHPLC	100	57°C// 50-54,9-60,3
HaGLP4	209	4	ECF	85,2	
HaGLP5	401	2	CEL1	100	
HaCOI_1	659	8	Secuenciación	95,6	Iniciador <i>Reverse</i>
HaCP	319	5	dHPLC	100	56°C// 50-55,6-61
HaDRP	200	3	dHPLC	100	56°C// 50-55,6-61
HaPAL	343	2	dHPLC	100	58°C// 50-55,6-61
HaTRP	234	2	ECF	96,3	
HaWP	177	2	ECF	99,3	
HaWRKY5	252	5	ECF	96,3	
HaWRKY7	174	2	ECF	98,5	
RhoBP	237	3	ECF	97,8	
HA77	259	3	ECF	97,8	
HA1848	259	8	ECF	85,4	
HA2989	142	4	ECF	97	
HA3103	183	4	ECF	94,1	
HA3847	147	5	ECF	93,3	
HA4264	180	7	ECF	100	
ORS331	211	4	ECF	97,8	
ORS460	302	5	ECF	97	
ORS1041	282	5	ECF	97,8	

a Número de haplotipos totales encontrados al genotipificar los 137 individuos que conforman la PMA.

b Características de los métodos de genotipificación. Para dHPLC, se indica la temperatura de la columna y el gradiente de buffer B utilizado para la separación de las moléculas heterodúplex. Para secuenciación directa, se indica el iniciador utilizado.

Mapeo de SNPs

Tres genes candidato polimórficos para las líneas PAC2 y RHA266, uno puesto a punto para su genotipificación utilizando dHPLC (HaGLP3), el segundo mediante ECF (HaGLP4) y el tercero mediante digestión con CEL1 (HaGLP5) se genotipificaron en una población de mapeo biparental conformada por 90 RILs derivadas del cruzamiento de las líneas públicas PAC2 x RHA266, y se incorporaron como marcadores funcionales al mapa saturado de girasol cultivado descrito previamente (Poormohammad Kiani *et al.* 2006). Estos tres *loci* se ubicaron en distintos grupos de ligamiento (GL), como se muestra en la Figura 3.7. HaGLP3 se localizó en el GL12, a 1,9 y 2,5 cM de los marcadores flanqueantes HA1155 y HA3059, respectivamente. El *locus* HaGLP4 se pudo asociar al GL10, a 8,3 y 1,6 cM de los marcadores ORS691 y TC15366. Finalmente, HaGLP5 se ancló en el GL6, entre dos marcadores bastante distantes entre sí: E40M50_5 y E38M60_13, ubicándose a 25 y 12,7 cM de los mismos.

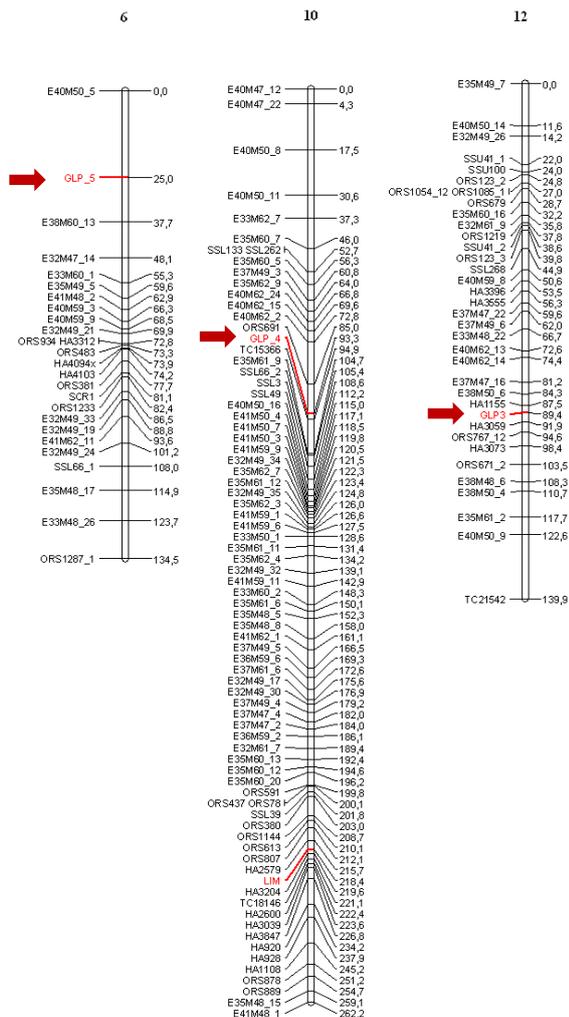


Figura 3.7 Incorporación de marcadores funcionales al mapa saturado de girasol RHA266 X PAC2 Se indican los tres marcadores polimórficos que fueron incorporados al mapa (flechas rojas). HaGLP5 se ubicó en el GL6, en el GL10 lo hizo el marcador HaGLP4 y en el GL12 HaGLP3.

DISCUSIÓN

La resistencia a PHC causada por *S. sclerotiorum* es compleja y poligénica, involucrando numerosos genes, cada uno con un efecto modesto sobre el fenotipo (Jan & Seiler. 2006). Si bien la identificación de un gen dominante que controle la resistencia a PHC es poco probable, la estrategia de mapeo por asociación se presenta como una alternativa a la identificación de QTL asociados a PHC en poblaciones bi-parentales (Micic *et al.* 2004, 2005a; b; Maringolo 2007). Al presente, sólo un estudio de mapeo por asociación utilizando la estrategia de genes candidato ha sido aplicado exitosamente en girasol para la resistencia a la podredumbre húmeda del capítulo causada por *S. sclerotiorum* (Fusari *et al.* 2012). Dada la complejidad del carácter, es necesario profundizar en la identificación de alelos con implicancias en el fenotipo.

En este estudio, un panel de 34 genes candidato y 11 SSR fueron seleccionados a partir de diversas fuentes: (1) los perfiles transcripcionales de tres líneas públicas de girasol de comportamiento diferencial ante la enfermedad PHC obtenidos mediante el análisis de microarreglos; (2) los perfiles transcripcionales de *B. napus* obtenidos sobre microarreglos de *Arabidopsis* (Zhao *et al.* 2007); (3) genes descritos en la literatura como partícipes en el proceso de defensa frente al ataque por patógenos fúngicos y (4) marcadores que colocalizaron con QTL de resistencia a *Sclerotinia*.

El microarreglo de girasol constituye una fuente de información precisa para la selección de genes candidato. Dado el gran número de transcriptos expresados diferencialmente que fueron identificados (más de 2500 considerando todas las líneas), la selección se restringió a transcriptos cuyos niveles de sobre y subexpresión pudieron ser validados por la técnica de qPCR, excepto para aquellos casos en los que el gen se tratara de un candidato fuerte, como es el caso del gen HeAn_126562, homólogo a Ve1 de tomate, gen mayor de respuesta a *Verticillium* (Diwan *et al.* 1999). La amplificación de estos transcriptos a partir de ADN genómico de girasol fue exitosa en 9/14 genes. La escasa o nula amplificación de los genes HeAn_18494, HeAn_36825, HeAn_38637 y HeAn_10191, aún utilizando diferentes temperaturas de hibridización (*annealing*) en la PCR, pudo deberse a que las secuencias de girasol sobre las que se diseñaron los iniciadores provinieron de bases de datos de EST no curadas, lo que pudo haber llevado a la generación de iniciadores con errores que no hibridaron correctamente con el ADN molde. Otra posibilidad es que los iniciadores hayan

sido diseñados en la región de *splicing* del ARNm, provocando que éstos no hibriden con el gen correspondiente. De hecho, para el gen HeAn_41021 se obtuvo un fragmento mayor al esperado (304 pb vs 509 pb), sugiriendo que se amplificó una región intrónica de 205 pb. Finalmente, la observación de un producto de amplificación mayor a 2000 pb para el gen HeAn_1756, a instancias de la presencia de probables intrones, y con una baja concentración, hizo que el mismo debiera ser descartado para los pasos posteriores, mientras que el gen HeAn_6452, si bien pudo ser amplificado correctamente, presentó severos errores en los patrones de secuenciación que imposibilitaron los pasos posteriores de identificación de polimorfismos para genotipificación. De este modo, de los 14 genes seleccionados, sólo pudieron ser evaluados 8/14 genes en la colección núcleo.

La rutina de identificación de ortólogos implicó la reconstrucción de árboles de genes con la información disponible para el género *Helianthus* y *Lactuca*, especies cuyo genoma completo aún no está secuenciado, en un contexto general de secuencias de distintas especies y géneros vegetales con similitud a la secuencia de *Arabidopsis* de partida. Se construyeron 7 árboles de MP correspondientes a los 7 *loci* candidato correspondientes, logrando de este modo identificar secuencias ortólogas y/o parálogas al gen de *Arabidopsis*, pertenecientes a géneros y especies de la familia *Asteraceae*. Para 5 *loci* las secuencias utilizadas para el diseño de iniciadores pertenecieron a *H. annuus* (G21, G22, G26, G33, G34), mientras que para 2 *loci* se seleccionaron secuencias de otras especies del género *Helianthus* (G28, G30).

Dentro de la diversidad de soluciones encontradas para cada gen, se puede mencionar, por ejemplo, los casos del gen 28, para el cual se encontraron sólo dos árboles de máxima parsimonia, poniendo de manifiesto la ausencia de contradicción en la matriz básica de datos. Para otros dos *loci* el número de árboles de longitud mínima fue menor a cien (G21 y G22). Finalmente, en los cuatro *loci* restantes la cantidad de árboles igualmente parsimoniosos superó la centena, conformando las soluciones más conflictivas. Esto último podría deberse a la insuficiente información con que se contaba para definir los grupos, es decir, alineamientos muy cortos, demasiadas indeterminaciones en las secuencias y/o a la sobrerepresentación de los géneros *Helianthus* y *Lactuca* respecto de los demás. Sin embargo, a pesar de estas distintas situaciones, el hecho de haber recurrido a la reconstrucción filogenética permitió explorar alternativas que finalmente condujeron a identificar secuencias potencialmente ortólogas a la secuencia de *Arabidopsis*. Esta

estrategia de búsqueda de ortólogos en girasol de secuencias derivadas de *Arabidopsis* para ser utilizada en estudios de asociación ya había sido utilizada con éxito por Fusari *et al.* (2012) y Talukder *et al.* (2014).

En líneas generales, la inclusión de GC a partir de fuentes bibliográficas y de marcadores tipo SSR que colocalizaban con QTL de resistencia a PHC en girasol no representó gran dificultad, ya que muchas de las regiones seleccionadas ya habían sido evaluadas en girasol para sus niveles de amplificación y polimorfismos. Sólo el gen HaCOI_2 no pudo ser amplificado a las temperaturas probadas, pero sí lo hizo su parálogo HaCOI_1.

Del total de 34 genes seleccionados, 27 pudieron ser secuenciados en las 8 líneas endocriadas de girasol que conforman la colección núcleo (CN) revelando diferentes patrones de polimorfismos. Es destacable que 26/27 genes hayan presentado SNPs y/o InDels potenciales para la genotipificación de la PMA. La ocurrencia de SNPs (1/69 pb) resultó un tanto menor, aunque bastante cercana a la reportada por Fusari *et al.* (2008) para un conjunto de líneas endocriadas (1/61 pb). Es probable que estas diferencias se deban a la cantidad de individuos analizados en uno u otro caso, es decir el tamaño de la muestra. Mientras que en el trabajo de Fusari *et al.* (2008) este análisis se realizó con alineamientos de al menos 15 individuos por gen, el análisis para los genes aquí evaluados se realizó sobre la CN formada por 8 individuos. Sin embargo, la ocurrencia de SNPs resulta baja si se compara con otros estudios realizados en girasol, como 1/48,25 pb observado tras el análisis de 15 regiones en 30 líneas endocriadas (Moreno 2011), 1 SNP/45,7 pb en 10 líneas endocriadas (Kolkman *et al.* 2007), 1 SNP/39 pb para 16 cultivares primitivos y mejorados de girasol (Liu & Burke 2006) y 1 SNP/34 pb en 8 líneas endocriadas (Giordani *et al.* 2011). La diferencia entre estos valores y las estimas obtenidas para las 8 líneas de girasol que conforman la CN puede explicarse por: (a) las diferencias esperadas en la diversidad genética de los accesos analizados (por ejemplo, al comparar las estimas para accesos de girasol cultivados primitivos y mejorados con líneas endocriadas); y (b) las diferencias en la cantidad de regiones o en el criterio de selección de las mismas. Con respecto a este último punto, para la selección de las regiones a analizar en el presente capítulo, se dio prioridad a las regiones codificantes: del total de 16438 pb secuenciadas por individuo, 9752 pb (59 %) corresponden a regiones codificantes. Dado que la ocurrencia de variaciones es superior en regiones intrónicas que exónicas, resulta esperable que se observe un nivel de polimorfismo menor.

En comparación con otros cultivos de interés agronómico, los valores de polimorfismo hallados en este trabajo fueron superiores a lo observado en sorgo (*Sorghum bicolor*) con 1 SNP/123 pb (Hamblin 2004), soja (*Glycine max*) con 1 SNP/328 y 1 SNP/536 pb (Zhu *et al.* 2003; Van *et al.* 2005), arroz (*Oryza sativa*) que mostró 1 SNP/113 pb y 1 SNP/100 pb (Garris *et al.* 2003; Monna *et al.* 2006), y garbanzo (*Cicer arietinum*) con 1SNP/138,7 pb (Deokar *et al.* 2014), lo cual resulta interesante porque indicaría un mayor nivel de polimorfismo en la especie en estudio.

Por su parte, la ocurrencia de eventos de inserción/delección (1/249 pb) fue levemente superior a lo reportado por Fusari *et al.* (2008) (1/288 pb), Moreno (2010) (1/332,12 pb), Kolkman *et al.* (2007) (1/277 pb), y Liu y Burke (2006) (1/264,7 pb). Esto pudo deberse a la definición de InDels, ya que en este trabajo fueron contabilizados todas las regiones identificadas, desde 1 pb.

El análisis de los SNPs en regiones codificantes (1/72 pb) y regiones no codificantes (1/65 pb) evidencia la mayor conservación de secuencia de las regiones codificantes con respecto a las otras. Más aún, del total de SNPs en regiones codificantes el 63 % (85 SNPs) fueron cambios sinónimos. Además, los valores promedio de la diversidad nucleotídica de sitios silenciosos ($\pi_{\text{sil}} = 0,015$) y de sitios sinónimos ($\pi_{\text{sin}} = 0,013$) fueron mayores que el promedio de la diversidad de sitios no-sinónimos ($\pi_{\text{no-sin}} = 0,005$), sin embargo, tres *loci* mostraron un $\pi_{\text{no-sin}}$ mayor que el π_{sin} . Estos genes son: G28 ($\pi_{\text{no-sin}} = 0,021$, $\pi_{\text{sin}} = 0$), HeAn_12761 ($\pi_{\text{no-sin}} = 0,00091$, $\pi_{\text{sin}} = 0$) y HaTRP ($\pi_{\text{no-sin}} = 0,012$, $\pi_{\text{sin}} = 0,004$). En los tres casos, las sustituciones se dieron en sitios informativos que cambian la secuencia de aminoácidos. Este tipo de cambios ocurren frecuentemente en líneas endocriadas que estuvieron sujetas a selección artificial (Sato *et al.* 2001).

Estos resultados fueron similares a los obtenidos en otros trabajos Kolkman *et al.* (2007) ($\pi_{\text{sil}} = 0,0135$, $\pi_{\text{sin}} = 0,0176$ y $\pi_{\text{no-sin}} = 0,0028$), Fusari *et al.* (2008) ($\pi_{\text{sil}} = 0,0140$, $\pi_{\text{sin}} = 0,0174$, y $\pi_{\text{no-sin}} = 0,0013$), aunque superiores a los reportados por Moreno (2010) ($\pi_{\text{sil}} = 0,00558$, $\pi_{\text{sin}} = 0,00508$ y $\pi_{\text{no-sin}} = 0,00076$), haciendo evidente que si bien moderadas, los patrones de diversidad observados son representativos del girasol cultivado. La tendencia de reducción de los valores de π al comparar sustituciones sinónimas y no sinónimas resultó coincidente con todos los trabajos revisados, reflejando la conservación de sitios desde una frecuencia de cambios mayor en regiones no codificantes y sustituciones sinónimas hacia frecuencias menores para las sustituciones no sinónimas. Al comparar los valores de diversidad

haplotípica hallados en el conjunto de 8 líneas que conforman la CN (0,514) con respecto al reportado por otros autores para girasol cultivado, se observa que los mismos son similares (0,51, Fusari *et al.* 2008, y 0,47, Moreno 2010).

Para la genotipificación de los GC en la totalidad de la PMA se aplicaron diferentes técnicas de mediana procesividad. La selección de la metodología de genotipificación más adecuada se realizó *a priori* siguiendo ciertas reglas: si las variantes alélicas diferían en tamaño, la espectrometría capilar fluorescente fue la técnica de elección; si en cambio las secuencias presentaban pocos haplotipos (2-3), se intentó la separación de moléculas heterodúplex mediante dHPLC y/o digestión con endonucleasa CEL1; finalmente, si ningún método de genotipificación mencionado resultó eficiente, la opción elegida fue la secuenciación directa. Los tres métodos fueron exitosos, con una tasa de datos faltantes levemente superior al 3 % para el total de genes. La técnica de dHPLC fue una de las más exitosas, siendo el método que permitió caracterizar nueve GC. Sin embargo, esta técnica no pudo utilizarse para la genotipificación de tres genes que *a priori* cumplirían con las condiciones necesarias para su aplicación (HeAn_41022, HeAn_12761 y G28), ya que no se encontraron las condiciones óptimas de temperatura y gradiente para poder resolver las moléculas heterodúplex formadas por mezcla de los dos haplotipos encontrados en la CN de líneas secuenciadas.

Es interesante destacar que la metodología de dHPLC permitió la detección de nuevos haplotipos, diferentes a los dos haplotipos inicialmente identificados en la CN, dado que no se observó un patrón de tipo homodúplex para ninguna de las mezclas con los haplotipos de las líneas control, mientras que las muestras homodúplex se comportaron de acuerdo a lo esperado para individuos homocigotas (un único pico). La secuenciación de los productos correspondientes a estos individuos reveló nuevos sitios polimórficos para los genes HaDRP, HaCP, G26 y G30. Si bien la técnica de dHPLC se aplica generalmente a SNPs conocidos, el hecho de haber detectado nuevos haplotipos ofrece una perspectiva promisoriosa en la utilización de la técnica para la identificación de SNPs *de novo* (Han *et al.* 2004). Si bien la genotipificación mediante dHPLC tiene una etapa laboriosa que implica mezclar los productos de PCR en forma equimolecular (manualmente), la misma se ve compensada por la automatización posterior, pudiendo contar con la información completa para un gen en menos de una semana. Además, el costo del experimento por muestra es mucho menor que el de los otros dos sistemas (dHPLC: 1 USD, ECF: 3,5 USD, Secuenciación: 4 USD),

convirtiéndola en una técnica excepcional para el escrutinio de al menos dos haplotipos en un conjunto de individuos, para luego restringir la secuenciación sólo a los individuos portadores de los nuevos haplotipos detectados.

La técnica de detección de moléculas heterodúplex por corte con CEL1 permitió la genotipificación rápida y a bajo costo de dos GC, HaGLP5 y HeAn_315, siendo la metodología más económica de todas las utilizadas en este trabajo de Tesis. La posibilidad de obtener la enzima CEL1 endonucleasa a partir de tallo de apio (Till *et al.* 2004) reduce considerablemente el costo de la técnica CEL1CH, sin pérdida de sensibilidad del método. Por otro lado, la detección de los fragmentos generados mediante corrida electroforética en gel de agarosa 3 % o electroforesis en Fragment Analyzer contribuye a la reducción de los costos, ya que se elimina la necesidad de utilizar iniciadores marcados.

La técnica ECF fue la más empleada, utilizándose en la genotipificación de 13 regiones candidatas, aunque el porcentaje de datos perdidos utilizando esta metodología fue mayor (4,7 %) al alcanzado utilizando dHPLC. Esta técnica tiene grandes ventajas en cuanto a la facilidad en el procesamiento de grandes volúmenes de muestras, una alta precisión en la adjudicación de los alelos, la posibilidad de automatización del proceso, el uso de programas de análisis de datos sencillos y la posibilidad de disminuir los costos por muestra al realizar corridas electroforéticas de conjuntos de muestras de diferente tamaño, o bien portadores de diferentes marcas (Chen & Sullivan 2003).

Muchos de los trabajos de mapeo por asociación utilizan la secuenciación como método de genotipificación por excelencia (Ingvarsson 2008; Haseneyer *et al.* 2010; Krill *et al.* 2010). La mayor ventaja de este método, respecto de ECF y dHPLC, radica en que los haplotipos no detectados en la CN utilizada para el desarrollo de los SNPs son identificados directamente. Por el contrario, dHPLC y ECF deberán pasar por una instancia adicional de secuenciación, si se desea determinar la secuencia de los polimorfismos determinantes de los nuevos haplotipos hallados. Sin embargo, el método de secuenciación no siempre resulta eficiente, como se demuestra en este trabajo, ya que los genes aquí secuenciados mostraron la mayor cantidad de datos faltantes obtenidos (5,37 % para el conjunto de cuatro GC secuenciados). Finalmente, las técnicas de genotipificación utilizadas permitieron conocer la ubicación de 3 GC en el mapa genético saturado de girasol cultivado, aportando de este modo a la caracterización de los mismos, y a la posibilidad de la evaluación de su colocalización con los QTLs identificados para PHC (Gentzbittel *et al.* 1998; Bert *et al.* 2002; Micic *et al.* 2004,

2005a; 2005b) en el caso que resultaran asociados significativamente con menores niveles de enfermedad. De estas metodologías, ECF es una de las más utilizadas para la saturación de mapas de ligamiento, mientras que el uso de dHPLC en girasol ya ha sido reportado (Lexer *et al.* 2004; Lai *et al.* 2005; Fusari *et al.* 2010). Sin embargo, no se habían obtenido reportes al momento del uso de CEL1CH para agregar marcadores a mapas de ligamiento de girasol, de modo que este resultado agrega una nueva potencialidad a las que ya posee la técnica.

En este capítulo se expuso el proceso de selección, amplificación, caracterización y genotipificación de genes candidato para mapeo por asociación, y, debido a que PHC es un carácter complejo, la cantidad final de 22 genes y 9 marcadores exitosamente genotipificados, resulta promisorio. A su vez, las fuentes de selección (ensayos de expresión diferencial en líneas de girasol de comportamiento contrastante ante la enfermedad, ensayos en *B. napus* infectadas con *S. sclerotiorum*, genes seleccionados de literatura como partícipes de procesos de defensa y marcadores derivados de mapeo de QTL) resultaron informativas dada la escasa información con que se cuenta sobre los genes implicados en la determinación de resistencia a PHC en girasol. Finalmente, además de la implicancia que estos genes puedan tener sobre el carácter, la caracterización de estos genes permitió aumentar la información sobre la variabilidad intraespecífica y la estructura génica en girasol. El estudio de asociación entre estos genes, junto con el panel de 384 SNPs del sistema BeadXpress de Illumina presentado en el Capítulo 2, y las variables fenotípicas evaluadas para resistencia a PHC se presenta en el Capítulo 4.

CAPÍTULO 4

**MAPEO POR ASOCIACIÓN PARA LA RESISTENCIA A LA
PODREDUMBRE HÚMEDA DEL CAPÍTULO EN GIRASOL**

INTRODUCCIÓN

El mapeo por asociación (MA) es un campo relativamente reciente de la genética que tiene el potencial de identificar los determinantes de características complejas a nivel de genes individuales, mediante la observación de asociaciones que son atribuidas a la fuerza del desequilibrio de ligamiento entre marcadores moleculares y fenotipos en una colección de germoplasma. En los últimos años, ha surgido un gran interés en aplicar estrategias de MA para el estudio de características de importancia agronómica y evolutiva, justificado principalmente por los avances recientes en tecnologías de genotipificación y secuenciación y el desarrollo de métodos estadísticos robustos.

Los estudios de asociación involucran cinco grandes etapas: (a) la **selección del germoplasma** que conformará la población de mapeo (PMA), (b) la **fenotipificación** de la PMA para el carácter de interés, (c) la **genotipificación** de la PMA con marcadores en genes candidato o a lo largo de todo el genoma, (d) el estudio de la existencia de **estructura poblacional y/o relaciones de parentesco** entre las líneas que componen la PMA, y (e) el **análisis estadístico para detectar asociaciones fenotipo-genotipo** (Ersoz *et al.* 2007). Las cuatro primeras etapas ya fueron detalladas ampliamente en los capítulos anteriores, restando sin discutir aún el análisis de las asociaciones.

El **modelo de asociación** a utilizar para el análisis estadístico de las asociaciones genotipo-fenotipo dependerá de la PMA seleccionada, de sus características y del número y tipo de marcadores genotipificados. Bajo una situación ideal, estimaciones estadísticas clásicas como las regresiones lineales, análisis de varianza (ANOVA), pruebas de *t* o de *chi*-cuadrado bastarían para realizar la inferencia de asociación. Sin embargo, dado que la presencia no contemplada de estructura poblacional puede generar asociaciones fenotipo-genotipo espurias, es necesario aplicar un enfoque estadístico que considere también la posible presencia de estructura en la población. Los métodos más comúnmente usados para abordar el estudio de estructura en las muestras que conforman una PMA son control genético (GCo, *Genetic Control*) y asociación estructurada (SA, *Structured Association*). El método GCo utiliza un conjunto de marcadores aleatorios para estimar el grado en que las pruebas estadísticas están “sobre-estimadas” por la presencia de estructura en la población, asumiendo que la estructura tiene un efecto similar sobre todos los *loci* (Devlin *et al.* 2001). Por su parte, el análisis de SA involucra dos pasos, primero la estimación de la estructura

poblacional (Q) utilizando un conjunto de marcadores aleatorios, seguido de la incorporación de esta estimación en el análisis estadístico de asociación (Pritchard & Rosenberg 1999; Pritchard *et al.* 2000b; Falush *et al.* 2003). Un modelo más complejo, que da cuenta de los múltiples niveles de relación, fue desarrollado por Yu *et al.* (2006), en el cual los marcadores aleatorios son utilizados para estimar Q y una matriz de parentesco (K), para luego utilizar ambas matrices en el marco de un modelo lineal mixto para las pruebas de asociaciones fenotipo-genotipo, proporcionando un poderoso complemento de los métodos disponibles en la actualidad para el mapeo de asociación (Zhao *et al.* 2007a). La disponibilidad de tecnologías de genotipificación masiva ha generado que en los estudios de asociación el número de datos genotípicos a asociar supere ampliamente el número de datos fenotípicos. La consecuente sobre-parametrización hace que sea imposible analizar los datos mediante una regresión multivariada tradicional, de modo que los modelos utilizados generalmente analizan un SNP o variante genotípica por vez, y mediante un ajuste por comparaciones múltiples, se obtienen luego los niveles de significación para todos los genes detectados (McCarthy *et al.* 2008). Estas estrategias basadas en el análisis independiente de los SNPs han sido exitosas en la detección de genes implicados en caracteres complejos (Donnelly 2008). Sin embargo, presentan dos serias desventajas: (a) debido a que la mayoría de los caracteres complejos son poligénicos, un análisis independiente de los SNPs podrá solo detectar una pequeña porción de la variación genética, perdiéndose aquellos genes que hacen contribuciones menores al carácter de interés (Hoggart *et al.* 2008); y (b) dado que muchos genes podrían estar interactuando para generar una red más compleja de interacciones genéticas, éstas no podrían ser identificadas ni caracterizadas utilizando el análisis de SNPs independientes.

Para superar estas limitaciones, una propuesta alternativa para los análisis de asociación es el uso de haplotipos, en vez de sitios SNP individuales. Se denominan haplotipos a los bloques polimórficos en desequilibrio de ligamiento. Debido a las correcciones que deben hacerse por comparaciones múltiples, los análisis de asociación basados en haplotipos suelen ser más poderosos en detectar asociaciones que al utilizar los sitios por separado (Rafalski 2010).

Por otra parte, más recientemente se han propuesto aproximaciones principalmente basadas en enfoques bayesianos para el análisis simultáneo de múltiples SNPs en ensayos de asociación (Satagopan *et al.* 1996; Sillanpää & Arjas 1998, 1999; Sillanpää &

Bhattacharjee 2005; Wu *et al.* 2009; Logsdon *et al.* 2010). Los métodos de regresión bayesianos combinan información previa sobre el efecto k de los marcadores con los n fenotipos observados para el carácter de interés para estimar todos los efectos de los marcadores. Si bien estos métodos fueron propuestos originalmente para predicción genómica (Meuwissen *et al.* 2001), su uso también ha sido reportado en estudios de asociación (Sun *et al.* 2011; Fan *et al.* 2011). Entre ellos se destacan las regresiones *ridge* y *bridge* (Frank & Friedman 2012), el LASSO bayesiano (*least absolute shrinkage and selection operator*, Tibshirani 1996), el enfoque bayesiano de *elastic net* (Zou & Hastie 2005) y la penalización *smoothly clipped absolute deviation* (SCAD) (Fan & Li 2001) como métodos paramétricos, mientras que la regresión RKHS (*Reproducing Kernel Hilbert Space*) es uno de los métodos de regresión semiparamétricos más ampliamente utilizados (Gianola & van Kaam 2008; Crossa *et al.* 2010).

Ejemplos exitosos de aplicación de la estrategia de MA en plantas incluyen mapeo de genes de resistencia al tizón tardío de la papa (Gebhardt *et al.* 2004), oídio en la cebada (Ivandic *et al.* 2003), marchitez por *Verticillium* en papa (Simko *et al.* 2004a; 2004b), y muerte regresiva en lechuga (Simko *et al.* 2009); así como mapeo por asociación para encontrar genes relacionados con tiempo de floración en maíz (Thornsberry *et al.* 2001), cebada (Saïdou *et al.* 2009) y mijo perla (Stracke *et al.* 2009); caracteres relacionados con contenido de almidón en maíz (Wilson & Whitt 2004); y la biosíntesis de antocianinas en maíz (Szalma *et al.* 2005). Si bien los modelos estadísticos utilizados varían, los modelos lineales mixtos (MLM) son los más populares actualmente, ya sea a través de las aproximaciones QK o PK. Aproximaciones estadísticas bayesianas han sido reportadas principalmente mediante el uso de simulaciones (*i.e.* Logsdon *et al.* 2010), y para estudios de asociación de tipo *genome-wide* en especies modelo, como tamaño y forma de granos de arroz (Iwata *et al.* 2007) y acumulación de sodio en hoja en *Arabidopsis thaliana* (Segura *et al.* 2012).

En sus comienzos, el abordaje de los estudios de asociación solía clasificarse en: (1) MA a través de la selección y genotipificación de genes candidato para un determinado carácter, y (2) MA a través de la genotipificación de marcadores homogéneamente distribuidos en el genoma o GWAS. En girasol, como en muchas otras especies no modelo, a pesar de que la información sobre marcadores moleculares funcionales de tipo SNP ha aumentado notablemente en los últimos cinco años (Bachlava *et al.* 2012; Pegadaraju *et al.* 2013; Cadic *et al.* 2013), la cantidad total de polimorfismos reportados (~10000 SNPs) no es suficiente

para realizar una cobertura total del genoma. Dado el tamaño del genoma de girasol (~3500 Mb) y el alcance del desequilibrio de ligamiento ($r^2 < 0,1$ hacia las 100 Kb, Fusari *et al.* 2008), se estima que la cantidad de marcadores requeridos para implementar una estrategia de tipo GWAS debería superar los 50.000 SNPs. De este modo, surge una nueva categoría en la escala de los estudios de asociación que engloba aquellos estudios en los que se analizó un alto número de marcadores, distribuidos en regiones candidatas o aleatorias del genoma, pero que cuyo número no es suficiente como para que alcancen la clasificación de GWAS. Esta estrategia suele estar apoyada en metodologías de genotipificación de mediana a alta procesividad, como los paneles de SNPs de Illumina o Affymetrix que suelen resultar altamente informativas, permitiendo la identificación de múltiples polimorfismos de efectos moderados sobre el carácter de interés (Zhou *et al.* 2014; Elmer *et al.* 2015; Kumar *et al.* 2015).

En este capítulo se presenta la conducción de un estudio de mapeo por asociación para la resistencia a la podredumbre húmeda del capítulo en girasol, a través de una estrategia que complementa genes candidato con el uso de metodologías de genotipificación de mediana a alta procesividad, considerando tanto modelos lineales mixtos que contemplan la existencia de estructuración y relaciones de parentesco entre los individuos de la población de mapeo, como modelos bayesianos de interrogación simultánea de *loci*. Se evaluará la *performance* de cada método en las estimas de asociación para el carácter resistencia a PHC en girasol.

MATERIALES Y MÉTODOS

El análisis de las asociaciones genotipo-fenotipo se realizó en base a los datos tomados sobre la Población de Mapeo por Asociación (PMA) para caracteres complejos de INTA, conformada por 137 líneas endocriadas de diversos orígenes geográficos, pertenecientes al Banco Activo de Germoplasma de INTA Manfredi, según se detalló en el Capítulo 1 de este trabajo de Tesis.

Estudio de la homogeneidad genética de las líneas que constituyen la PMA

A fin de corroborar la estabilidad genética año a año de las distintas líneas que constituyen la PMA se estableció un protocolo de control basado en la genotipificación de 8 marcadores gSSR, seleccionados del panel de 42 SSR presentado en el Capítulo 2 de esta Tesis. Estos SSRs se seleccionaron por su ubicación en diferentes grupos de ligamiento (GL) del mapa de referencia de girasol (Poormohammad Kiani *et al.* 2006), su reproducibilidad y facilidad de interpretación de los cromatogramas (Moreno *et al.* 2013) y su alto poder discriminante ($PI = 5,7 \times 10^{-6}$, estimado utilizando el *software* Genalex v6.1, Peakall & Smouse 2006). Las características de estos SSRs utilizados para su amplificación se muestran en la Tabla 4.1.

Tabla 4.1 Características de los SSRs utilizados para la evaluación de la estructura poblacional

Marcador	Rango pb	N° alelos	He	PIC
HA77	247 – 262	3	0,37	0,33
HA991	137 – 148	3	0,49	0,38
HA293	113 – 123	4	0,66	0,59
HA2063	167 – 180	6	0,68	0,62
HA2920	163 – 187	5	0,46	0,42
HA3239	134 – 144	4	0,67	0,60
HA928	217 – 227	4	0,62	0,54
HA4103	141 – 152	3	0,51	0,39
Promedio		4	0,56	0,48

Los iniciadores utilizados para la amplificación de estas regiones se detallan en la Tabla 2.1 (Capítulo 2).

En cada campaña, se tomaron muestras de hoja de planta joven de tres a cinco individuos por línea, haciendo un total de 2055 individuos genotipificados mediante electroforesis capilar con detector de fluorescencia inducida por láser (ECF). La recolección de muestras, extracción de ADN y genotipificación se realizaron según lo detallado en el Capítulo 2.

A fin de obtener estimas confiables para los estudios de asociación, fueron eliminados los datos fenotípicos de aquellas líneas de la PMA que presentaron diferencias en dos o más loci gSSR a lo largo de las campañas.

Evaluación fenotípica

Las líneas de la PMA fueron evaluadas a campo con inoculación asistida con esporas del hongo durante cinco campañas (2009-10, 2010-11, 2011-12, 2012-13 y 2013-14). Tal como se detalla en el Capítulo 1 de este trabajo de Tesis, los datos fenotípicos utilizados para inferir asociación corresponden a los valores de incidencia y severidad a tiempo final (28 días post inoculación), período de incubación de la enfermedad y área bajo la curva de progreso de la enfermedad (AUDPC) para incidencia. Las medias ajustadas de los mismos fueron obtenidas utilizando el modelo estadístico [1.2] descrito en el Capítulo 1, el cual considera la presencia de interacción aleatoria Genotipo x Ambiente (GxA). Para las medidas fenotípicas incidencia, AUDPCinc y severidad los datos se ajustaron a las distribuciones determinadas en el Capítulo 1. Para la medida período de incubación, los datos se ajustaron a una distribución normal.

Marcadores utilizados para el establecimiento de asociaciones

Para los estudios de asociación, la información de partida fue la derivada del panel de 384 SNPs, cuyo desarrollo y genotipificación se detallan en el Capítulo 2 y el conjunto de 22 genes candidato y 9 marcadores SSR que colocalizaron con QTL detallados en el Capítulo 3. Luego de descartar aquellos SNPs con frecuencia del alelo minoritario (MAF) menor a 0,05 y aquellos alelos de marcadores multialélicos que no superaran dicho valor, restaron 168 regiones para ser incluidas en los análisis de asociación.

Al estar todos los SNPs y/o InDels dentro de un gen en DL, es posible analizarlos en forma conjunta, como parte de bloques polimórficos, o haplotipos. De este modo, en aquellos casos en los que el *software* estadístico permitió el ingreso de múltiples alelos por *locus*, los GC fueron analizados como haplotipos, mientras que en aquellos casos en los que el *software* sólo permitió el ingreso de marcadores bialélicos se utilizaron los SNPs y los GC y SSR que sólo presentaron dos alelos.

Análisis estadístico

Con la nómina final de datos fenotípicos (Y), las matrices de estructura (Q) y relaciones de parentesco (K), los datos genotípicos derivados del panel de Illumina y de los GC (G), se construyó la estrategia final de mapeo por asociación para continuar con el análisis mediante un modelo lineal mixto y un modelo bayesiano de interrogación simultánea de *loci* (Figura 4.1).

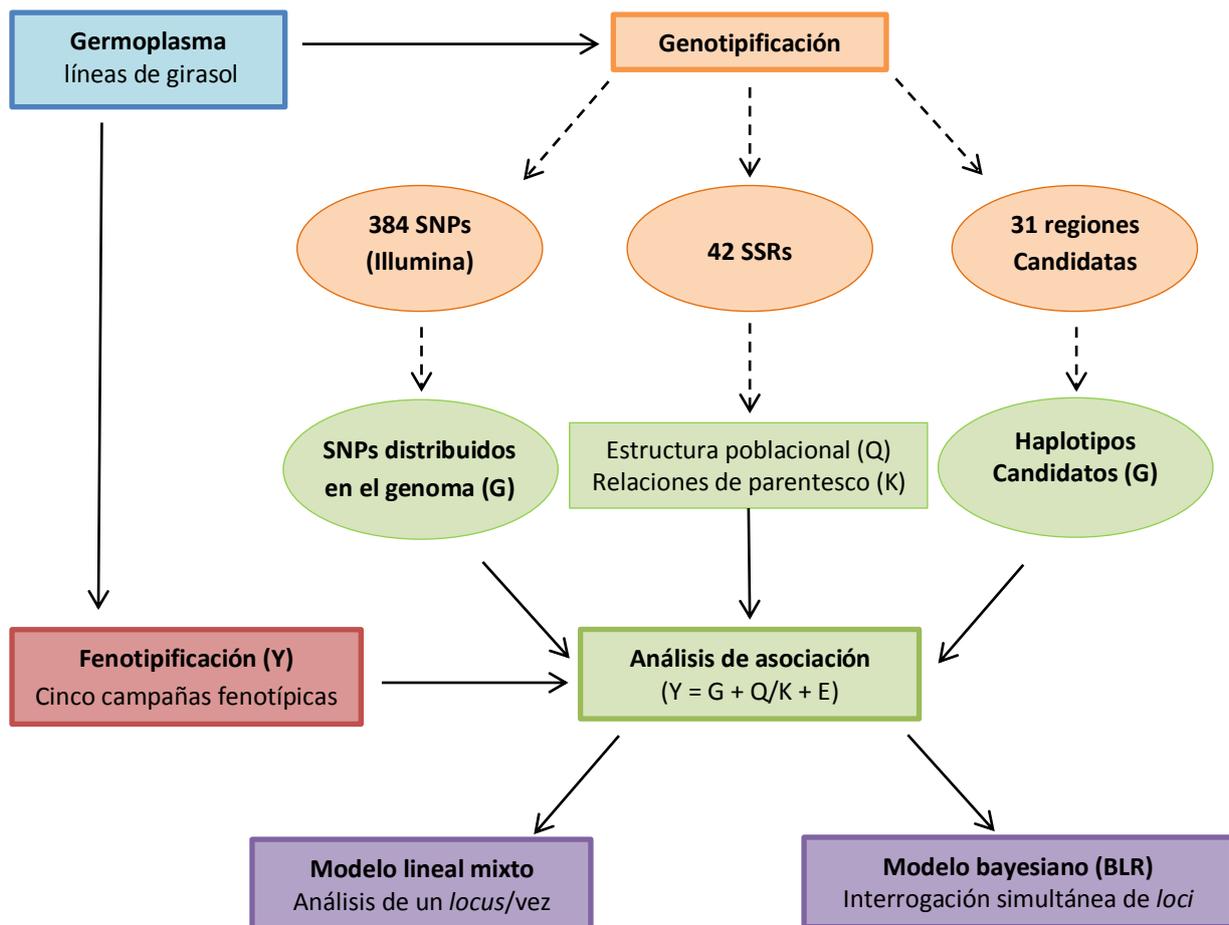


Figura 4.1 Metodología de trabajo para la conducción de los análisis de asociación en girasol para la resistencia a la PHC

1. Análisis estadístico mediante el uso de modelos lineales mixtos

Existe una gran variedad de paquetes estadísticos disponibles para el análisis de datos de mapeo por asociación. En este trabajo, se utilizó la aproximación de dos pasos propuesta por Stich *et al.* (2008), la cual se ajustó con la función `lme` de la librería `nlme` (Pinheiro *et al.* 2014) del *software* de uso libre R (R core team 2011), dada la facilidad de esta función para aceptar una matriz de incidencia para los efectos aleatorios.

El primer paso de este método involucra la estimación del efecto genotípico (línea) sobre el fenotipo, o media ajustada, lo cual fue realizado en la sección “evaluación fenotípica”.

La segunda etapa consiste en ajustar un modelo para el efecto genotípico en función de los marcadores moleculares (uno a la vez) descontando la estructura poblacional (Q) y la estructura de parentesco (K). El modelo estadístico Q+K utilizado puede describirse de la siguiente manera:

$$Y_{ij} = \mu + M_i + \mathbf{Q}\boldsymbol{\beta} + \mathbf{K}\mathbf{k} + e_{ij} \quad [4.1]$$

donde Y_{ij} es el vector de medias ajustadas obtenidas en el paso uno de longitud igual a L (número de líneas), que se modelan de acuerdo a los alelos del marcador M_i , \mathbf{Q} es una matriz de regresoras de longitud LxQ que representa la estructura poblacional estimada por el software STRUCTURE, $\boldsymbol{\beta}$ es vector de dimensión Q que contiene los coeficientes de regresión que acompañan a \mathbf{Q} , y \mathbf{K} es la matriz de incidencia de los efectos aleatorios representados por el vector \mathbf{k} , que se incluyen para modelar la estructura de parentesco de las líneas. Finalmente, e_{ij} es un vector de dimensión L conteniendo los términos de error que se suponen normales con esperanza cero y matriz de covarianzas $\sigma^2_{\epsilon}\mathbf{I}$.

Dado que este paquete permite el ingreso de marcadores multialélicos, los genes candidato fueron analizados como haplotipos, ya que, según se detalló previamente, los análisis de asociación basados en haplotipos suelen ser más poderosos a la hora de detectar asociaciones (Rafalski 2010).

Los p-valores críticos para determinar la significancia de las asociaciones fueron calculados en base a sus correspondientes q-valores (Storey 2002). Siguiendo la propuesta de Elmer *et al.* (2015), se utilizó un valor de $q = 0,2$ como límite para considerar una asociación significativa.

2. Análisis bayesiano de interrogación simultánea de loci

Los datos genómicos fueron convertidos a código numérico de tipo {0, 1, 2}, correspondiente al número de copias del alelo minoritario. Con este panel de datos, y partiendo del modelo [4.1], se utilizó la aproximación bayesiana *least absolute shrinkage and selection operator* (LASSO) (Park & Casella, 2008) para estimar el efecto de los marcadores mediante el análisis simultáneo de *loci* (*i.e.* la media posterior de los efectos de los marcadores). Esta aproximación fue implementada utilizando el paquete BLR (regresión lineal bayesiana, o *Bayesian linear regression*, Perez *et al.* 2010) de R (R core team 2013).

Siguiendo la propuesta del Dr. Eduardo Cappa (*comunicación personal*), los valores de las hipervarianzas para las varianzas genética y residual de cada carácter fenotípico fueron aquellos estimados utilizando el *software* TASSEL versión 5.0 (<http://www.maizegenetics.net>), utilizando la información del conjunto de 42 marcadores SSR. Los grados de libertad fueron establecidos en 3.

El muestreo de Gibbs implementado en BLR (Pérez *et al.* 2010) se utilizó para obtener un total de 10000 muestras, 2000 de las cuales fueron descartados como *burn-in*. Se evaluó la convergencia de las cadenas de Markov (MCMC) mediante la inspección visual de las gráficas obtenidas para los parámetros varianzas genética y residual, utilizando el paquete BOA (Smith 2007) de R (R core team 2011).

La capacidad predictiva del modelo fue evaluada realizando un diseño de validación cruzada de tipo *Leave-one-out* (LOO, dejando uno afuera), esquema que implica separar los datos de forma que para cada iteración se tenga una sola muestra para los datos de prueba (validación) y todo el resto conformando los datos de entrenamiento (Figura 4.2). Esta metodología fue repetida N veces (siendo N el número de líneas que conforman la PMA), de modo que cada línea fue utilizada para validación una sola vez.

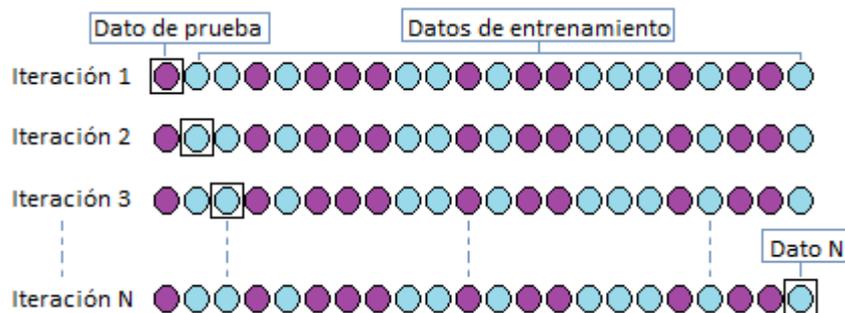


Figura 4.2 Esquema de validación cruzada dejando uno afuera (LOO)

La proporción de la varianza genética explicada por los marcadores se obtuvo como $(2p_i(1-p_i)\hat{\alpha}^2)$, donde p_i es la frecuencia de un alelo en un locus y $\hat{\alpha}^2$ es la estima del efecto del marcador (Cappa *et al.* 2013). Con estos datos, se generó un ranking de marcadores de acuerdo a su contribución al fenotipo. Se reportaron los marcadores que explicaron al menos 1% de la varianza fenotípica total.

Establecimiento de la estructura poblacional y relaciones de parentesco

Dado el alto valor informativo y la concordancia de las estimas obtenidas utilizando distintos paneles de marcadores moleculares en el Capítulo 2, se utilizaron los datos genotípicos derivados de los marcadores tipo SSR para el establecimiento de la estructura poblacional. Las medidas de estructura poblacional consideradas fueron obtenidas utilizando el método bayesiano implementado en el *software* STRUCTURE (Pritchard *et al.* 2000b). La matriz Q generada por STRUCTURE es una matriz $n \times p$, donde n es el número de individuos y p es el número de subpoblaciones definidas. Se realizó un análisis de la varianza con el *software* Infostat (Di Rienzo, 2011) para determinar si las medias ajustadas de las variables utilizadas para la evaluación de resistencia a PHC (incidencia, AUDPCinc, severidad y período de incubación), presentaban una distribución diferencial en relación con la estructura poblacional subyacente determinada utilizando STRUCTURE.

Para el análisis de asociación realizado utilizando la aproximación de Stich de dos pasos (Stich *et al.* 2008), la estructura de parentesco entre las líneas fue modelada en el mismo paso en el cual se establecieron las asociaciones. Siguiendo una nueva aproximación propuesta por el prof. MSc. J. Di Rienzo (*comunicación personal*), esta estructura de parentesco se obtuvo a partir del conjunto de 42 marcadores SSR, con el cual se generó una matriz S de proporción de alelos compartidos entre líneas. A partir de esta matriz se calculó una matriz de distancias $D = \sqrt{1 - S}$, y con ella se generó un dendrograma utilizando el algoritmo descrito como *ward.D* de la función *hclust* del *software* de uso libre R. A partir de este dendrograma fue posible obtener, mediante la función *cutree*, un conjunto de entre 2 y 134 grupos (número de líneas menos 1) que corresponden a la estructura jerárquica del árbol. Pudo establecerse, se acuerdo al criterio de menor AIC, que 60 efectos aleatorios anidados eran suficientes para modelar la estructura de parentesco. Para una mayor precisión en la estima, se imputaron los datos faltantes de la tabla de SSR utilizando la herramienta “Imputación de datos faltantes” del submenú “Análisis multivariado” del *software* estadístico InfoStat (Di Rienzo *et al.* 2008), en base al coeficiente de similitud de Jaccard.

Para el análisis de asociación realizado utilizando el paquete BLR (Pérez *et al.* 2010), las relaciones de parentesco a considerar fueron estimadas con el *software* SpaGeDi (Hardy & Vekemans 2002). El mismo se utiliza para estimar K entre los individuos (Hardy & Vekemans

2002), la cual es una matriz de dimensiones $n \times n$ con elementos fuera de la diagonal; siendo F_{ij} , una estima de la probabilidad de que los individuos sean idénticos por descendencia, basados en los datos derivados de marcadores moleculares. Los elementos diagonales de K son 1 para individuos endogámicos y $0,5 \times (1 + F_x)$ para los individuos no endogámicos, donde F_x es el coeficiente de endogamia. Para poder ser ingresada en BLR, la matriz de parentesco recuperada de SpaGEDi fue convertida a positiva definida utilizando la función `neadPD` del paquete `Matrix` (Bates *et al.* 2014) implementado en el *software* de uso libre R (R core team, 2011).

Caracterización de los GC asociados

Con el fin de contribuir a la caracterización de los genes asociados con tolerancia a PHC, se inspeccionaron los niveles de expresión por tejido analizando los datos de transcriptómica de girasol desarrollados en el *Institut National de la Recherche Agronomique*, INRA y depositados en la base de datos Heliagene. Este portal presenta la información sobre los niveles de transcritos detectados en distintos tejidos correspondientes a las líneas HA412 y XQR (INRA) (Langlade 2012). Los niveles de expresión se expresan en RPKM, medida que contempla la cantidad de lecturas obtenidas para cada unigen (*counts*), el número de lecturas totales (*nb. reads*) y el tamaño del unigen (o *contig size*), según la siguiente relación:

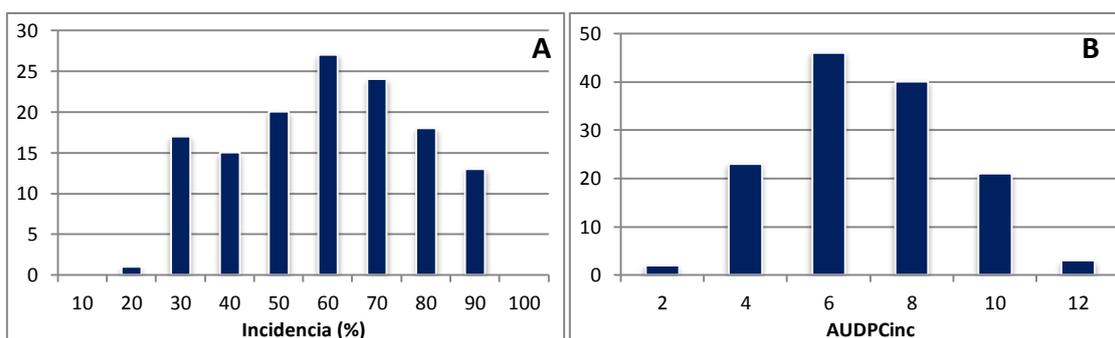
$$\text{RPKM} = \text{counts} / (\text{nb reads} * \text{contig size})$$

De este modo, ingresando en dicho portal la anotación del unigen correspondiente depositado en el repositorio ATGC-SUR, es posible recuperar los niveles de expresión del mismo discriminados por tejido (hoja, tallo, raíz, brácteas, lígulas, ovario estéril, estilo, estámen, corola, polen y semilla).

RESULTADOS

La caracterización de las líneas que conforman la PMA utilizando marcadores SSRs permitió estudiar la estabilidad genética de las mismas, campaña a campaña. El genotipo de referencia utilizado fue el de la campaña 2011-12, ya que de estos materiales fue extraído el ADN utilizado para la genotipificación de gSSR, EST-SSR y el panel de 384 SNPs de Illumina en los estudios de diversidad y estructura poblacional desarrollados el Capítulo 2, como así también de los genes candidato detallados en el Capítulo 3. Utilizando el criterio de homogeneidad descrito en los Materiales y Métodos, se eliminaron dos líneas de la PMA (HA302 y R417), por mostrar inconsistencias entre los distintos años. Estas inconsistencias estuvieron dadas por exceso de heterocigotas, y por la aparición de alelos “nuevos” en homocigosis. Dado que la composición de la PMA cambió de 137 a 135 líneas, se estimó nuevamente la estructura de la población, y las medias ajustadas de las medidas fenotípicas (incidencia, AUDPC incidencia, severidad y período de incubación).

Para la obtención de las medias ajustadas de cada medida fenotípica, el análisis estadístico fue realizado utilizando el modelo 1.2 descrito en el Capítulo 1, el cual contempla la existencia de interacción aleatoria GxA. Nuevamente, el mismo permitió la discriminación entre genotipos susceptibles (S) y moderadamente resistentes (MR) en la PMA, para las cuatro medidas fenotípicas seleccionadas. En la Figura 4.3 se muestran los histogramas de distribución de las nuevas medias ajustadas por línea por carácter. El listado de medias ajustadas se presenta en la Tabla A4.1.



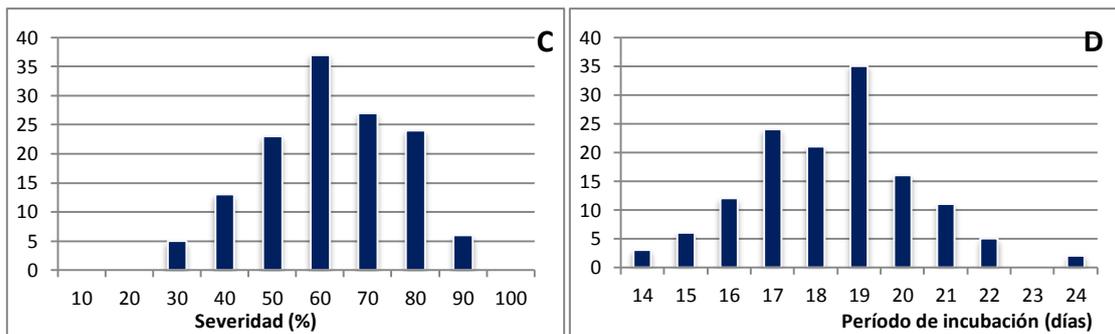


Figura 4.3 Medias ajustadas de las cuatro medidas fenotípicas descriptoras de la enfermedad PHC en girasol en las 135 líneas analizadas de la PMA (A) Incidencia de PHC; (B) severidad de PHC; (C) AUDPC de incidencia; (D) Período de incubación. En el eje Y se presenta el número de líneas endocrizadas.

Por su parte, el análisis de estructura poblacional realizado sobre el panel de 135 líneas que restaron en la PMA, utilizando la información de 42 marcadores tipo SSR y evaluado con el método de Evano *et al.* (2005) reconfirmó la presencia de tres subpoblaciones en la PMA (Figura 4.4), tal como se había observado para el conjunto total de líneas en el Capítulo 2.

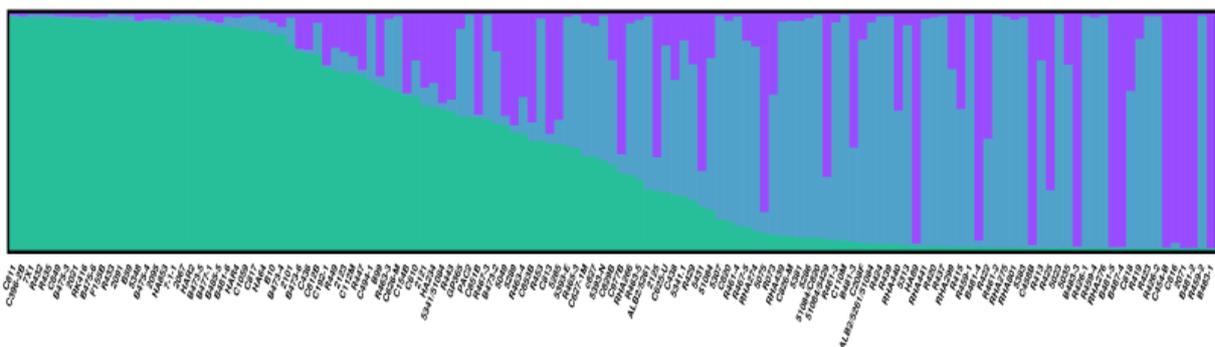


Figura 4.4 Resultado de la inferencia de estructura poblacional utilizando el programa STRUCTURE. Representación de las tres poblaciones identificadas para el panel de 135 líneas de la PMA que fueron evaluadas para asociación.

Se estudió si los niveles de enfermedad, medidos como incidencia, AUDPCinc, severidad y período de incubación presentaban una distribución diferencial en relación con la estructura poblacional subyacente mediante análisis de la varianza. No se observaron diferencias significativas entre los niveles de resistencia de las líneas contenidas en cada una de las tres subpoblaciones determinadas utilizando el *software* STRUCTURE ($p = 0,1185$ para incidencia, $p = 0,0655$ para AUDPCinc, $p = 0,3076$ para severidad y $p = 0,5589$ para período de incubación).

Modelo lineal mixto: análisis estadístico de un *locus* por vez

Los análisis estadísticos llevados a cabo utilizando el modelo de dos pasos propuesto por Stich *et al.* (2008) y contemplando la presencia de estructura en la PMA y las relaciones de parentesco entre las líneas, permitieron la identificación de alelos asociados con aumento de la resistencia a PHC evaluada a través de las variables incidencia máxima (IEmax), AUDPCinc, severidad y período de incubación.

La evaluación de la distribución del porcentaje acumulado de p-valores vs. los p-valores observados para el análisis de asociación de cada variable, permitió distinguir que de las cuatro medidas fenotípicas analizadas, la que presentó mejor ajuste a lo esperado fue período de incubación (Figura 4.5, puntos rojos), mientras que para las otras tres se observó un alto número de marcadores moleculares con p-valores bajos.

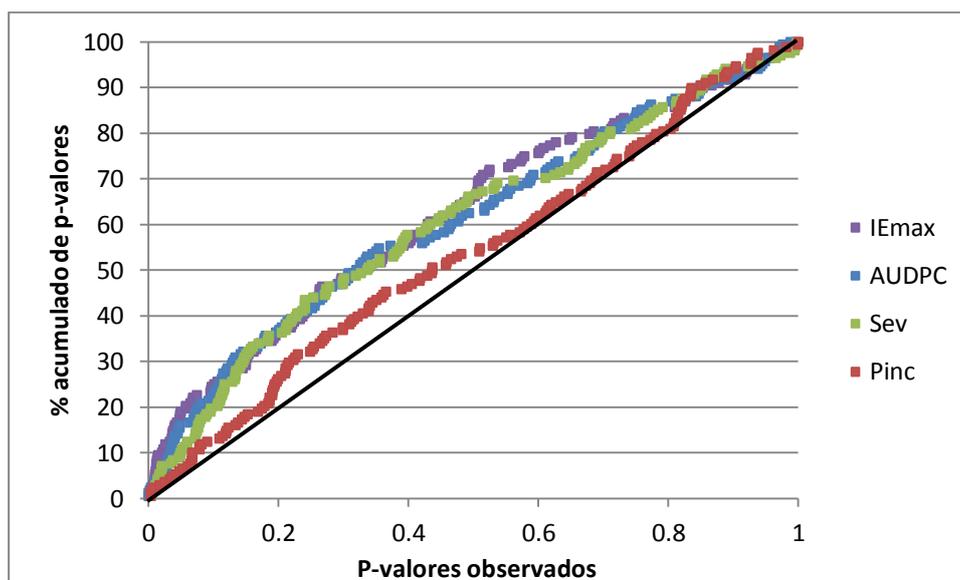


Figura 4.5 Distribución acumulada de los p-valores obtenidos para las cuatro características fenotípicas La diagonal corresponde al valor esperado si H_0 : el marcador no tiene efecto sobre el fenotipo, es verdadera.

Incidencia de PHC

Tal como se observa en la Figura 4.5, al utilizar el modelo Q+K para establecer asociación entre las distintas regiones candidatas y menor incidencia de PHC, se encontró una gran proporción (19 %) de regiones asociadas con p-valores < 0,05. Por su parte, la proporción de regiones con p-valores < 0,01 fue igual al 5 % del total de regiones evaluadas (Apéndice A4.2).

Debido al número de marcadores evaluados, y para establecer un FDR (*false discovery rate*, o proporción esperada de falsos positivos) razonable, se estimaron q-valores (Storey 2002) para el conjunto total de p-valores. Utilizando un límite de q-valor igual a 0,2, se identificaron 32 marcadores asociados con menor incidencia de la enfermedad que ajustaron a este criterio (Apéndice A4.2). El marcador SNP117 mostró la asociación con el p-valor más bajo (p-valor = 0,0007, q-valor = 0,08). Por su parte, la segunda asociación más importante se observó con el marcador HA1848 (p-valor = 0,0030, q-valor = 0,11).

Utilizando el mismo modelo, pero esta vez considerando los marcadores como aleatorios, fue posible estimar la proporción de la varianza explicada por cada uno de ellos, determinado como efecto de la sustitución alélica. Cabe destacar que cada alelo de cada marcador se toma como una clase separada, de modo que el efecto marcador resultante no se descompone en efectos aditivos y de dominancia, sino que se evalúa la significancia global. Del conjunto de alelos asociados, el que mayor porcentaje de la varianza explica es el SNP144 (27,72 %), seguido por SNP117 (23,89 %).

En la Tabla 4.2 se detallan los marcadores asociados con menor incidencia que se ajustaron al criterio de FDR establecido y que, para ser aún más restrictivos en el proceso de selección de marcadores asociados, obtuvieron valores de $p \leq 0,01$.

Tabla 4.2 Análisis estadístico Marcadores asociados con menor incidencia de PHC (p-valor $\leq 0,01$, q-valor $< 0,2$)

Marcador	p-valor	q-valor	Prop. de varianza explicada	media alelo 1	media alelo 2	media alelo 3	media alelo 4	media alelo 5	media alelo 6	media alelo 7	media alelo 8
SNP117	0,0007	0,08	23,89	43±4,0	58±2,0						
HA1848	0,0030	0,11	6,79	63±11	55±5,0	55±7,0	51±5,0	56±7,0	43±5,0	55±5,0	73±5,0
SNP136	0,0032	0,11	11,39	52±2,0	63±3,0						
SNP144	0,0041	0,11	27,72	53±2,0	69±5,0						
SNP128	0,0069	0,11	13,94	56±2,0	46±4,0						
SNP80	0,0094	0,11	12,48	62±4,0	52±3,0						
SNP60	0,0095	0,11	8,30	49±3,0	58±2,0						
SNP125	0,0098	0,11	10,19	51±3,0	61±3,0						
G34	0,0102	0,11	12,17	57±2,0	42±5,0	51±6,0					

En negrita se señala el alelo que presentó menor media de incidencia.

Dado que cada alelo asociado hace una contribución moderada al fenotipo, a fin de identificar aquellos materiales con potencial para ser utilizado en los programas de mejoramiento para la resistencia a PHC, se seleccionaron aquellas líneas portadoras de la mayor cantidad de alelos asociados con menor enfermedad. Para el caso de la medida

fenotípica IEmax, se identificó una línea portadora de ocho de los nueve alelos asociados (51084/C820), una línea portadora de siete (5381) y quince líneas portadoras de seis de los nueve alelos asociados (Tabla 4.3).

Tabla 4.3 Líneas de la PMA portadoras de al menos seis de los nueve alelos de los marcadores asociados con menor incidencia de la enfermedad

Línea	SNP117	HA1848	SNP136	SNP128	SNP144	SNP125	SNP80	G34	SNP60	IEmax
51084	X		X		X		X	X	X	0,32
51084/C820	X	X	X		X	X	X	X	X	0,46
5381	X		X	X	X	X	X		X	0,27
5383	X		X	X		X	X		X	0,18
ALB2/5261/51084	X		X		X	X	X		X	0,29
B481-3	X		X		X	X	X		X	0,86
C192-1		X	X	X	X		X		X	0,41
C208F	X			X	X		X	X	X	0,31
C438			X		X	X	X	X	X	0,24
C610				X	X	X	X	X	X	0,23
C627-1M	X		X		X	X	X		X	0,37
C818			X	X	X	X	X	X		0,51
C826-M	X	X	X		X		X		X	0,62
HA853		X	X		X		X	X	X	0,27
RHA439	X	X	X		X		X		X	0,38
RHA441			X	X	X	X	X	X		0,22
RHA801	X			X	X	X	X		X	0,32

AUDPC incidencia de PHC

Al igual que para la medida fenotípica anterior, para establecer asociaciones se evaluó un modelo estadístico que contempló la presencia de estructura en la población y las relaciones de parentesco entre las líneas (Q+K). Nuevamente, una gran proporción (16 %) de regiones mostraron p-valores por debajo de 0,05, mientras que la proporción de regiones con p-valores inferiores a 0,01 fue de 3,5 %, como se observa en la Figura 4.5.

Utilizando un límite de q igual a 0,2, se identificaron seis marcadores asociados con menor AUDPCinc que ajustaron a este criterio (Apéndice A4.2). El marcador cuyo p-valor resultó más bajo fue HA1848 (p-valor = 0,0003, q-valor = 0,02). Por su parte, la segunda asociación más importante se observó con el marcador SNP117 (p-valor = 0,0004, q-valor = 0,02). Cabe destacarse que éstos también fueron los dos marcadores que presentaron las asociaciones más fuertes para el carácter incidencia de PHC, y que todos los marcadores asociados con menor

AUDPCinc también estuvieron asociados con menor incidencia, destacando de este modo la importancia de los niveles de enfermedad a tiempo final en la determinación del AUDPC.

Al igual que en el caso anterior, la utilización de este modelo, pero considerando a los marcadores como aleatorios, permitió estimar la proporción de la varianza explicada. Del conjunto de marcadores asociados, el que mayor aporte hace a la varianza fenotípica es el SNP117 (24,99%).

En la Tabla 4.4 se detallan los marcadores asociados con menor incidencia que se ajustaron al criterio de FDR establecido y que obtuvieron p-valores $\leq 0,01$.

Tabla 4.4 Análisis estadístico Marcadores asociados con menor AUDPCinc de PHC (p-valor $\leq 0,01$, q-valor $< 0,2$)

Marcador	p-valor	q-valor	Prop. Varianza explicada	media alelo 1	media alelo 2	media alelo 3	media alelo 4	media alelo 5	media alelo 6	media alelo 7	media alelo 8
HA1848	0,0003	0,02	10,15	6,3±1,4	5,9±0,6	6,2±0,8	5,3±0,6	5,8±0,8	4,3±0,6	6,2±0,6	8,2±0,6
SNP117	0,0004	0,02	24,99	4,5±0,5	6,3±0,2						
HACOI	0,0022	0,08	11,08	6,7±0,3	5,2±0,3	6,1±0,7					
SNP136	0,0038	0,11	12,24	5,6±0,3	6,8±0,4						
SNP144	0,0089	0,18	22,42	5,7±0,3	7,4±0,6						
SNP128	0,0093	0,18	12,31	6,1±0,6	4,9±0,4						

En negrita se señala el alelo que presentó menor media de AUDPCinc.

Con el fin de identificar los genotipos portadores de alelos asociados que podrán ser utilizados en los programas de mejoramiento, se señalan las líneas portadoras de al menos cuatro de los seis alelos asociados (Tabla 4.5).

Tabla 4.5 Líneas de la PMA portadoras de al menos cuatro de los seis alelos de los marcadores asociados con menor AUDPCinc de la enfermedad

Línea	SNP117	HA1848	SNP136	SNP128	SNP144	HaCOI_1	AUDPCinc
5286			X	X	X	X	6,60
5381	X		X	X	X		4,13
51084	X		X		X	X	3,81
51084/C820	X	X	X		X		4,03
B10		X	X		X	X	2,90
C112M			X	X	X	X	5,12
C192-1		X	X	X	X		3,49
C208F	X			X	X	X	4,48
C627-1M	X		X		X	X	4,57
C826-M	X	X	X		X		7,58
C849		X	X		X	X	4,87
HA853		X	X		X	X	8,35

Línea	SNP117	HA1848	SNP136	SNP128	SNP144	HaCOI_1	AUDPCinc
R429			X	X	X	X	5,27
R449	X		X	X	X		3,47
RHA274			X	X	X	X	3,04
RHA439	X	X	X		X	X	3,85
RHA441			X	X	X	X	2,52
RHA801	X			X	X	X	3,75
RK416		X	X	X	X		2,37

Severidad

Al igual que para las medidas fenotípicas anteriores, la evaluación de un modelo Q+K para establecer asociaciones fenotipo-genotipo para el carácter severidad permitió la identificación de un gran porcentaje (10) de regiones asociadas con p-valores < 0,05; y de un 2,4 % de regiones con p-valores < 0,01.

Utilizando un límite de q-valor igual a 0,2, se identificaron doce marcadores asociados con menor severidad de la enfermedad que se ajustaron a este criterio (Apéndice A4.2). El marcador cuyo p-valor resultó más bajo fue HaCOI_1 (p-valor = 0,004, q-valor = 0,19).

Nuevamente, la utilización de este modelo estadístico, pero considerando a los marcadores como aleatorios, permitió estimar la proporción de la varianza explicada. Del conjunto de marcadores asociados, el que mayor aporte hace a la varianza fenotípica es HA1848 (10,32 %).

En la Tabla 4.6 se detallan los marcadores asociados con menor incidencia que se ajustaron al criterio de FDR establecido y, para ser más restrictivos en el proceso de selección, que obtuvieron p-valores ≤0,01.

Tabla 4.6 Análisis estadístico Marcadores asociados con menor severidad de PHC (p-valor ≤ 0,01, q-valor < 0,2)

Marcador	p-valor	q-valor	Prop. Varianza explicada	media alelo 1	media alelo 2	media alelo 3	media alelo 4	media alelo 5	media alelo 6	media alelo 7	media alelo 8
HACOI	0,004	0,19	9,15	61,4±2,9	51,9±3,0		56,6±5,3				
SNP136	0,005	0,19	9,86	54,7±2,3		62,9±3,0					
HA1848	0,006	0,19	10,32	58,2±9,3	53,7±4,4	55,8±5,5	53,3±4,0	57,2±5,7	48,2±4,4	57,2±4,1	69.8±4.4
G33	0,009	0,19	9,87	58,9±2,5		50,4±3,3					
SNP44	0,010	0,19	7,35	61,1±2,9		54,2±2,6					

En negrita se señala el alelo que presentó menor media de severidad.

Para la medida severidad de la enfermedad, se identificaron quince líneas que llevaban cuatro de los cinco alelos asociados (Tabla 4.7).

Tabla 4.7 Líneas de la PMA portadoras de al menos cuatro de los cinco alelos de los marcadores asociados con menor severidad de la enfermedad

Línea	HaCOI_1	HA1848	SNP136	G33	SNP44	Severidad (%)
2071	X		X	X	X	54,8
2095	X		X	X	X	60,4
51084	X		X	X	X	45,6
51084/C820		X	X	X	X	31,4
B10	X	X	X			41,8
B477-1	X	X		X	X	30,3
B481-2	X		X	X	X	41,4
B481-4	X		X	X	X	66,8
B481-5	X		X	X	X	77,9
B483-1	X		X	X	X	55,3
B483-3	X		X	X	X	56,0
C454-B	X		X	X	X	55,8
C820	X		X	X	X	59,3
HA853	X	X	X	X		42,9
HA89		X	X	X	X	58,1

Período de incubación

Tal como se observa en la Figura 4.5, la medida fenotípica período de incubación es la que presenta una distribución acumulada de p-valores vs. p-valores observados más uniforme: sólo el 6 % de las regiones candidatas presentaron p-valores < 0,05, mientras que 2,35% de las regiones presentaron p-valores < 0,01.

Utilizando un límite de q-valor igual a 0,2, se identificaron 3 marcadores asociados con menor severidad de la enfermedad que se ajustaron a este criterio (Apéndice A4.2). Los marcadores cuyo p-valor resultaron más bajos fueron los marcadores SNP23 y SNP177 (ambos con p-valor = 0,003, q-valor = 0,17). Del conjunto de marcadores asociados, el que mayor aporte hace a la varianza fenotípica es SNP23 (25,31%).

En la Tabla 4.8 se detallan los marcadores asociados con menor incidencia que se ajustaron al criterio de FDR establecido y que obtuvieron valores de $p \leq 0,01$.

Tabla 4.8 Análisis estadístico Marcadores asociados con mayor período de incubación (p-valor $\leq 0,01$, q-valor < 0,2)

Marcador	p-valor	q-valor	Prop. Varianza explicada	media alelo 1	media alelo 2	media alelo 3
SNP23	0,003	0,17	25,31	20,4±0,5	18,7±0,2	
SNP117	0,003	0,17	20,51	19,9±0,4	18,6±0,2	
HACOI	0,004	0,17	9,11	18,3±0,3	19,5±0,3	18,9±0,6

En negrita se señala el alelo que presentó mayor media de período de incubación.

Para la medida período de incubación se identificó una línea portadora de los tres alelos asociados y cuatro líneas portadoras de dos de los tres alelos asociados (Tabla 4.9), y, todas ellas con valores de período de incubación superiores a la media global (18,89 días).

Tabla 4.9 Líneas de la PMA portadoras de al menos dos de los tres alelos de los marcadores asociados con mayor período de incubación de la enfermedad

Línea	SNP117	HaCOI_1	SNP23	Pinc (días)
51084/C820	X		X	21,2
ALB2/5261	X		X	24,8
C817		X	X	20,1
RHA441		X	X	21,5
RHA801	X	X	X	20,8

Método bayesiano: análisis estadístico con interrogación simultánea de *loci*

Luego de descartar los marcadores multialélicos, y aquellos que presentaban una MAF < 5 %, un total de 157 marcadores fueron considerados para el análisis simultáneo de *loci* para asociación. Los análisis estadísticos llevados a cabo utilizando el *software* TASSEL sobre la información de los 42 marcadores SSR (Bradbury *et al.* 2007), utilizando un modelo que contempla la presencia de estructura en la población y las relaciones de parentesco entre las líneas, fueron utilizados para obtener los valores de las hipervarianzas (S_u y S_e) para las varianzas genética (V_u) y residual (V_e) de cada carácter fenotípico (Tabla 4.10).

Tabla 4.10 Estimación de las hipervarianzas para las distribuciones *a priori* de las varianzas genética y residual

Carácter	V_u (σ_u^2 Tassel)	$S_u^c = V_u (df^a + 2) / \bar{a}^b$	V_e (σ_e^2 Tassel)	$S_e^c = V_e (df^a + 2)$
Incidencia	0,134	0,67	0,042	0,21
AUDPCinc	2,139	10,697	1,479	7,395
Severidad	0,306	1,533	0,433	2,165
P. de Incubación	0,539	2,697	0,895	4,477

a Tomando $df = 3$

b $\bar{a} = 1$ en ausencia de *inbreeding*

c Según Pérez *et al.* (2010)

La cantidad de iteraciones realizadas, y las descartadas por *burn-in* permitieron la convergencia de las cadenas de Markov, lo cual fue evaluado mediante inspección visual de los gráficos diagnósticos de convergencia (Figura 4.6). Se observa que el aprendizaje bayesiano tuvo lugar, ya que las distribuciones posteriores se alejan de las distribuciones *a priori*.

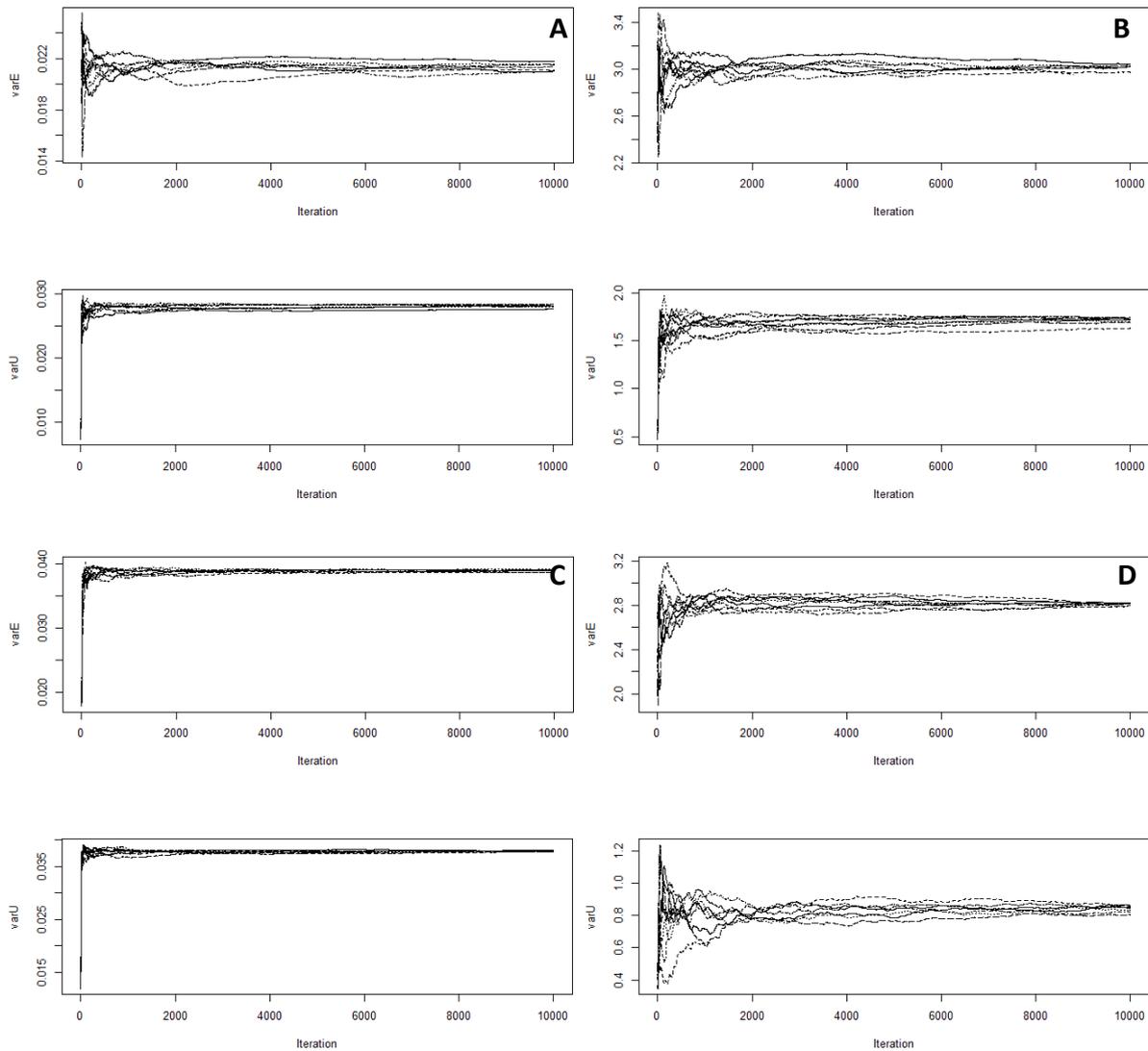


Figura 4.6 Gráficos diagnóstico de convergencia de MCMC (A) Incidencia; (B) AUDPCinc; (C) Severidad; (D) Período de incubación. Arriba, varianza de E; Abajo, varianza de U.

El modelo bayesiano implementado en el paquete BLR permitió estimar el efecto de los marcadores sobre las medidas fenotípicas de interés. Este tipo de aproximaciones mediante el análisis simultáneo de *loci* son muy poderosas, ya que permiten la detección de marcadores que hacen contribuciones infinitesimales al fenotipo. En la Tabla 4.11 se presentan aquellos marcadores cuya contribución resultó mayor o igual al 1% de la varianza fenotípica total, junto con el carácter al cual está asociado, y la frecuencia del alelo minoritario. Los resultados completos obtenidos se presentan en la Tabla A4.2.

Tabla 4.11 Análisis Bayesiano Marcadores que explican al menos el 1% de la resistencia a PHC

Incidencia			AUDPCinc			Severidad			Período de incubación		
Marcador	Efecto (%)	MAF	AUDPCinc	Efecto (%)	MAF	Marcador	Efecto (%)	MAF	Marcador	Efecto (%)	MAF
SNP38	8,32	0,39	SNP44	11,52	0,35	SNP44	7,07	0,35	SNP60	9,15	0,38
SNP44	6,98	0,35	SNP60	4,52	0,38	SNP64	4,54	0,39	SNP44	5,06	0,35
SNP111	3,99	0,20	SNP147	3,74	0,22	SNP89	3,79	0,35	SNP17	3,77	0,33
SNP60	3,58	0,38	SNP38	3,34	0,39	SNP82	3,77	0,46	SNP157	3,34	0,46
SNP66	2,95	0,36	SNP143	2,89	0,39	SNP68	3,49	0,36	SNP175	3,31	0,22
SNP80	2,70	0,24	SNP117	2,88	0,18	SNP150	3,41	0,38	SNP117	3,28	0,18
SNP109	2,61	0,47	SNP48	2,70	0,44	SNP97	3,20	0,43	SNP58	3,27	0,43
SNP45	2,51	0,30	SNP80	2,60	0,24	SNP149	2,67	0,36	SNP82	3,17	0,46
SNP82	2,23	0,46	SNP34	2,31	0,35	SNP112	2,61	0,29	SNP89	2,52	0,35
SNP124	2,15	0,19	SNP136	2,20	0,27	SNP143	2,40	0,39	SNP64	2,15	0,39
SNP188	2,14	0,19	SNP175	2,16	0,22	SNP113	2,24	0,32	SNP143	1,96	0,39
SNP7	2,12	0,25	SNP17	2,01	0,33	SNP72	1,99	0,37	SNP104	1,79	0,33
SNP125	1,91	0,26	SNP111	1,91	0,20	SNP48	1,98	0,44	SNP172	1,78	0,36
SNP117	1,86	0,18	SNP82	1,82	0,46	SNP165	1,95	0,38	SNP97	1,78	0,43
HaGLP3	1,81	0,34	SNP188	1,73	0,19	SNP38	1,92	0,39	SNP150	1,75	0,38
SNP136	1,80	0,27	SNP7	1,72	0,25	SNP60	1,92	0,38	SNP100	1,72	0,27
SNP143	1,74	0,39	SNP115	1,55	0,10	SNP111	1,81	0,20	HAPAL	1,67	0,21
SNP156	1,68	0,22	SNP66	1,39	0,36	HaWRKY7	1,78	0,39	SNP152	1,62	0,49
SNP140	1,67	0,10	SNP140	1,38	0,10	SNP191	1,75	0,34	SNP163	1,54	0,15
SNP128	1,63	0,20	SNP124	1,29	0,19	G33	1,73	0,26	G26	1,47	0,19
HaDRP	1,56	0,27	SNP64	1,29	0,39	SNP30	1,65	0,31	SNP63	1,43	0,34
SNP81	1,50	0,39	SNP109	1,26	0,47	SNP128	1,61	0,20	SNP77	1,42	0,45
SNP127	1,30	0,49	G33	1,21	0,26	SNP194	1,54	0,27	SNP168	1,32	0,35
SNP70	1,21	0,19	SNP194	1,18	0,27	SNP168	1,51	0,35	G33	1,28	0,26
SNP175	1,16	0,22	SNP112	1,17	0,29	SNP136	1,40	0,27	SNP34	1,27	0,35
SNP160	1,13	0,44	SNP113	1,13	0,32	SNP175	1,24	0,22	G22	1,21	0,27
SNP89	1,09	0,35	SNP128	1,12	0,20	SNP80	1,23	0,24	SNP147	1,18	0,22
SNP163	1,08	0,15	SNP163	1,08	0,15	SNP142	1,15	0,43	SNP70	1,09	0,19
SNP194	1,03	0,27	SNP156	1,05	0,22	SNP34	1,06	0,35	SNP158	1,08	0,36
			SNP19	1,05	0,17	SNP85	1,06	0,19	SNP85	1,05	0,19
			SNP187	1,04	0,46	SNP114	1,00	0,38	HeAn_315	1,02	0,33
			SNP89	1,00	0,35						

Efecto (%): porcentaje de la varianza explicada por el marcador; MAF: frecuencia del alelo minoritario. En negrita se señalan los marcadores asociados utilizando el método de dos pasos.

Resumen conjunto de los análisis de asociación

Por cada medida fenotípica, se construyó una gráfica de p-valores obtenidos para el método de dos pasos de Stich *et al.* (2008) vs. el porcentaje de la varianza explicado por cada marcador estimado utilizando el método bayesiano implementado en el paquete BLR (Pérez

et al. 2010) (Figura 4.7). Los marcadores que más interés generan para ser utilizados en programas de mejoramiento son aquellos que se ubican hacia la derecha sobre el eje de las abscisas, es decir, aquellas que presentan bajos p-valores y alto porcentaje de varianza fenotípica explicada. Del total de marcadores que cumplen con esta condición, se colorearon aquellos que además pasaron las correcciones de Storey *et al.* (2002).

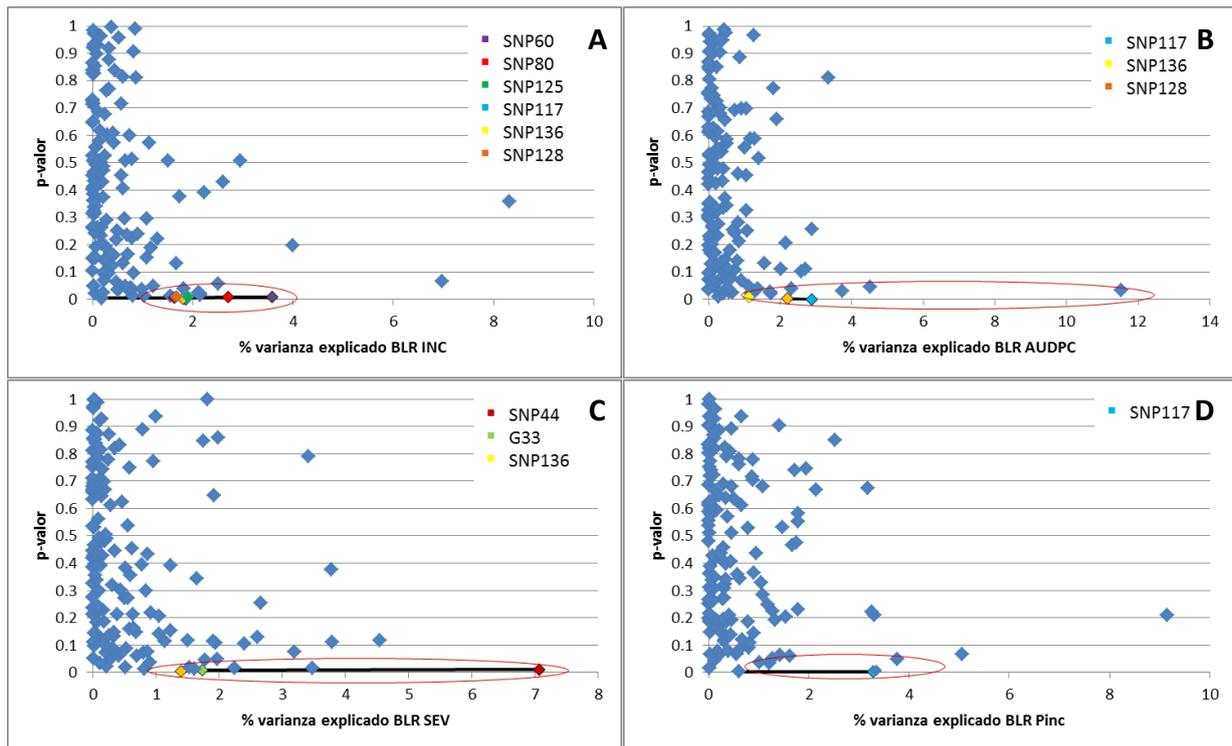


Figura 4.7 Relacion entre los resultados obtenidos mediante la implementación de los métodos de análisis unilocus y multilocus (A) Incidencia; (B) AUDPCinc; (C) Severidad; (D) Período de incubación. La elipse en rojo señala la región que incluye a los marcadores con p-valores menores a 0,05 y porcentaje de varianza fenotípica explicada mayor al 1 %. Se colorearon los marcadores seleccionados como asociados por ambos métodos.

Haciendo un análisis global de los resultados obtenidos, se observa que con excepción de los SNPs 23 y 144, todos los marcadores bialélicos que fueron seleccionados como asociados con el método de dos pasos también fueron seleccionados por el método multilocus, por explicar más del 1% de la varianza fenotípica. Todos los marcadores coincidentes entre ambos métodos estadísticos tienen la particularidad de presentar frecuencias alélicas intermedias, mientras que los marcadores que sólo resultaron asociados con el método de dos pasos presentaron una MAF entre 0,05 y 0,1 (Apéndice A4.2).

Dada esta concordancia de resultados, y que el método de Stich *et al.* (2008) resulta más abarcativo que el multilocus de BLR, ya que permite el análisis de marcadores multialélicos, se consideraron los marcadores asociados por este método. Del total de marcadores, algunos resultaron asociados con más de un carácter, mientras que otros fueron específicos de una única medida fenotípica. De este modo, para las cuatro determinaciones fenotípicas evaluadas, se encontraron 13 alelos de regiones candidatas asociados con diferentes factores de resistencia: cuatro de ellos son compartidos entre tres características, dos entre dos características, cuatro resultaron específicos para incidencia, dos específicos para severidad, y dos para período de incubación. En la Tabla 4.12 se detallan los marcadores, su ubicación en el mapa de ligamiento de girasol cultivado (Zubrzycki *et al.* 2014), el origen de cada marcador, los caracteres a los cuáles están asociados, la concordancia con los resultados del BLR, y sus anotaciones GO (*Gene Ontology*).

Del panel de líneas de girasol seleccionadas por ser las que portaban la mayor cantidad de alelos asociados, las líneas 51084, HA853, RHA441 y RHA801 se destacan por haber sido seleccionadas para tres de las cuatro características fenotípicas evaluadas, mientras que la línea 51084/C820 fue seleccionada para las cuatro características fenotípicas.

Tabla 4.12 Marcadores asociados con resistencia a PHC en girasol ($p \leq 0,01$)

Marcador	GL	Origen	Carácter asociado	Confirmado BLR	Función molecular inferida	GOs
SNP117	10	EST	IEmax,AUDPCinc,Pinc	Si	progesterona 5- β -reductasa	P:cellular metabolic process; F:coenzyme binding; F:catalytic activity
HaCOI_1	14*	Talukder <i>et al.</i> (2014)	sev,AUDPCinc,Pinc	-	Insensible a coronatina	
HA1848	7	Zubrzycki (2014)	sev,IEmax,AUDPCinc	-		
SNP136	9	EST	sev,IEmax,AUDPCinc	Si	β -n-acetilhexosaminidasa	P:carbohydrate metabolic process; C:cytoplasmic membrane-bounded vesicle; F:beta-N-acetylhexosaminidase activity; F:cation binding
SNP144		EST	IEmax, AUDPCinc	No	Tirosina Aminotransferasa	F:pyridoxal phosphate binding; P:cellular amino acid and derivative metabolic process; F:1-aminocyclopropane-1-carboxylate synthase activity; P:vitamin E biosynthetic process; F:L-tyrosine:2-oxoglutarate aminotransferase activity
SNP128	12	Fusari (2010)	IEmax, AUDPCinc	Si	Proteína tipo germina 3 (HaGLP3)	F:superoxide dismutase activity; C:plant-type cell wall; F:oxalate oxidase activity; F:nutrient reservoir activity; P:oxidation reduction; P:auxin mediated signaling pathway; F:oxidoreductase activity;
G33		Zhao <i>et al.</i> (2007)	Severidad	Si	Proteína con dominio radical SAM	
SNP125	2	EST	IEmax	Si	Xilulosa quinasa	phosphate metabolic process;transferase activity, transferring phosphorus-containing groups;phosphotransferase activity, alcohol group as acceptor;primary metabolic process;carbohydrate metabolic process
SNP80		EST	IEmax	Si	Embryonic flower 2	P:genetic imprinting; P:negative regulation of flower development; F:transcription factor activity; C:nucleus
SNP60	10	Kolkman <i>et al.</i> (2007)	IEmax	Si	Serin/treonin quinasa	P: protein phosphorylation, F: Protein serine/threonine kinase activity, F: ATP binding.
G34		Zhao <i>et al.</i> (2007)	IEmax	-	Glicoproteina P	
SNP23		EST	Pinc	No	MAP-quinasa	P:protein amino acid phosphorylation; F:ATP binding; F:protein serine/threonine kinase activity
SNP44	6	Kolkman <i>et al.</i> (2007)	Severidad	Si	Proteína familia GRAS	

* Ubicación definida en el mapa de Tang *et al.* (2002).

Análisis de los perfiles de expresión por tejido

Del total de trece regiones asociadas con resistencia a PHC en girasol, tres son genes candidato fuertes cuya selección se detalló en el Capítulo 3 de este trabajo de Tesis (G33, G34 y HaCOI_1), uno es un marcador que colocalizó con un QTL para incidencia de PHC en una población biparental (HA1848, Zubrzycki, 2014), mientras que los nueve restantes corresponden a SNPs presentes en el ensayo de genotipificación de 384 SNPs presentado en el Capítulo 2. De estas nueve regiones, tres fueron caracterizadas por resecuenciación de GC ya reportados como relacionados con resistencia a enfermedades fúngicas: SNP128 (HaGLP3), SNP60 (serin-treonin quinasa) y SNP44 (SCR 1), mientras que las seis restantes fueron caracterizadas *in silico* a partir del análisis de variantes de tipo de tipo SNP en secuencias disponibles para la especie en bases de datos de EST, a las que se le asignó una función inferida por BLASTx (NCBI, base de datos *Viridiplantae*) y anotación usando *Gene Ontology* relacionada con respuesta a distintos tipos de estrés (biótico-abiótico, atgc-sur.inta.gob.ar). De este modo, si bien se hizo una pre-selección de los SNPs derivados de EST a incluir en el ensayo de Illumina, no existe evidencia acerca de su participación en procesos de defensa. Es por ello que, para brindar mayor soporte a la asociación detectada entre los mismos y la resistencia a PHC, enfermedad limitada al capítulo, resulta de interés evaluar la expresión en este tejido de los unigenes asociados.

Utilizando la base de datos Heliagene (www.heliagene.org), se observaron los niveles de expresión por tejido de todas las regiones asociados con menor enfermedad, con excepción de HA1848. Se observa que 11/12 regiones asociadas se expresan en tejido asociado a capítulo (ya sea en brácteas, polen, estilo, estigma, lígulas, etc) (Figura 4.8). La única región que no muestra expresión en capítulo es aquella portadora del SNP128, la cual se corresponde con HaGLP3. Sin embargo, en otros estudios se ha reportado que esta región se expresa en capítulo en estadíos tardíos (estadío R5.6 de la escala de Schneider & Miller) (Ehrenbolger *et al.* 2012).

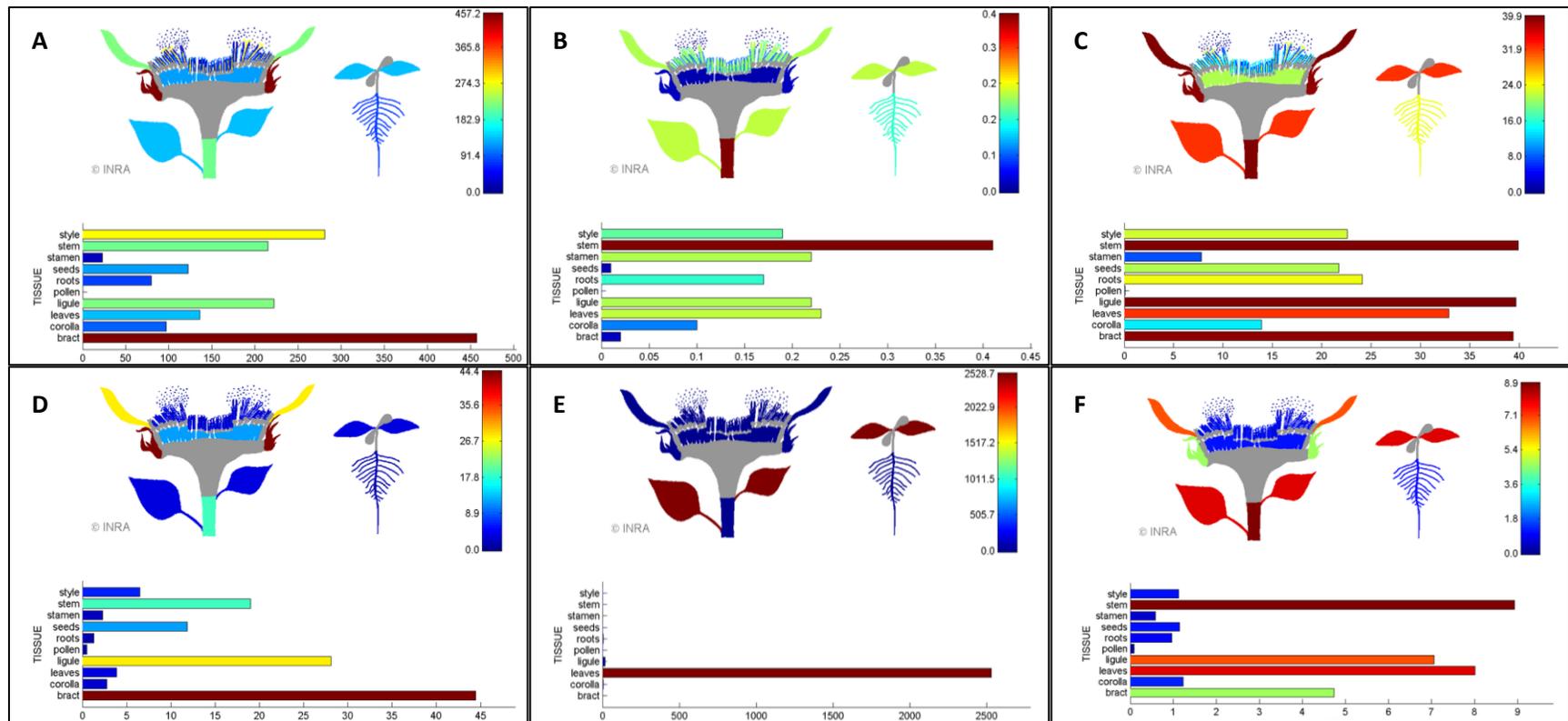


Figura 4.8 Niveles de expresión por tejido de los distintos genes candidato A la derecha de cada gráfica se presenta una escala colorimétrica que se corresponde con los niveles de expresión, siendo los colores rojizos los de más alto nivel y los azulados los de expresión más baja. Los valores del eje x están expresados en RPKM de acuerdo a lo descrito en la sección Materiales y Métodos. (A) SNP117; (B) HaCOI_1; (C) SNP136; (D) SNP144; (E) SNP128; (F) G33. Imágenes tomadas de www.heliagene.com.

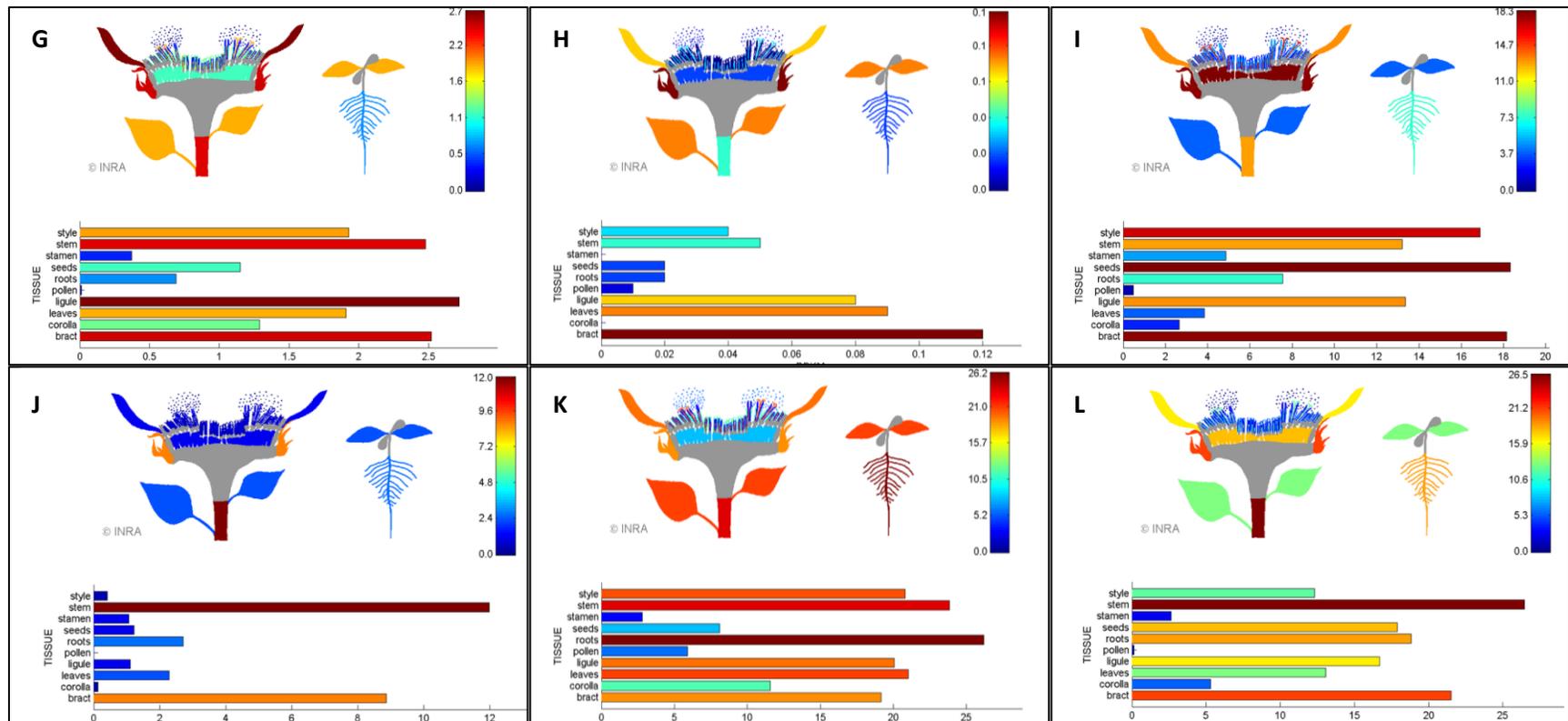


Figura 4.8 Niveles de expresión por tejido de los distintos genes candidato (continuación) A la derecha de cada gráfica se presenta una escala colorimétrica que se corresponde con los niveles de expresión, siendo los colores rojizos los de más alto nivel y los azulados los de expresión más baja. Los valores del eje x están expresados en RPKM de acuerdo a lo descrito en la sección Materiales y Métodos. (G) SNP125; (H) SNP60; (I) SNP80; (J) G34; (K) SNP23; (L) SNP44. Imágenes tomadas de www.heliogene.com.

DISCUSIÓN

El mapeo por asociación enriquece el repertorio de herramientas disponibles para la disección de caracteres complejos y se aplica tanto a especies modelo como a cultivos de importancia económica (Rafalski 2010). Este enfoque comenzó a desarrollarse en especies vegetales a comienzos del siglo XXI, gracias al salto de tecnología producido en la biología molecular en términos de información genómica disponible en bases de datos y de técnicas de genotipificación de alto rendimiento (*high-throughput*).

Este trabajo representa un ejemplo de aplicación de genética de asociación en una especie no modelo, el girasol, como enfoque complementario al tradicional mapeo de QTL en poblaciones biparentales, y como herramienta para dilucidar los componentes genéticos involucrados en caracteres agronómicos complejos, como la resistencia a la Podredumbre Húmeda del Capítulo (PHC) causada por *S. sclerotiorum*.

La población utilizada para tal fin, compuesta por líneas endocriadas pertenecientes al programa de mejoramiento de INTA y preservadas en el Banco Activo de Germoplasma de la EEA INTA Manfredi (BAG-IM, INTA) constituye un insumo fundamental para este estudio, dada su gran variabilidad morfológica y de respuesta frente al ataque del patógeno (presentadas en el Capítulo 1 de esta Tesis) y sus niveles de diversidad (estudiados en el Capítulo 2). Estudios comparativos de diversidad de la población de mapeo por asociación (PMA) frente a un muestreo de poblaciones de polinización abierta y poblaciones compuestas del BAG-IM, confirmaron que los materiales seleccionados para conducir el estudio de MA albergan gran parte de la diversidad presente en dicho Banco (Capítulo 2). De este modo, a pesar de que el número de líneas que conforman la PMA es inferior a lo sugerido por algunos autores (*i.e.* Rafalski 2010; Wang *et al.* 2011), esto se compensa por la variabilidad observada entre ellas, y aún se mantiene en el rango de muchos estudios de asociación reportados a la fecha en especies vegetales, como soja (101) (Elmer *et al.* 2015), cebada (145) (Mezaka *et al.* 2013), trigo (96) (Neumann *et al.* 2010) y (198) (Kulwal *et al.* 2012), durazno (105) (Picañol *et al.* 2012) y *Arabidopsis thaliana* (117) (Yano *et al.* 2013).

La presencia de estructura poblacional fue corregida utilizando el panel de 42 marcadores tipo SSR, mediante la aproximación bayesiana implementada en el *software* STRUCTURE (Pritchard *et al.* 2000b). En el Capítulo 2, la información aportada por este conjunto de

marcadores y el método estadístico seleccionado, fue comparada con la obtenida utilizando otros conjuntos de marcadores (SNPs y SSR+SNPs), y otras aproximaciones estadísticas (DAPC, Jombart *et al.* 2010), alcanzándose siempre el mismo resultado ($K = 3$). A pesar de la eliminación de dos líneas por falta de homogeneidad, en este capítulo nuevamente se obtuvo un valor de subpoblaciones igual a tres. De este modo, se llega a la conclusión de que la corrección realizada utilizando el panel de marcadores SSR y el método implementado en STRUCTURE es acertada, pese a las limitaciones que supone la utilización de un método estadístico diseñado para poblaciones panmícticas, en un conjunto de líneas de endocría avanzada.

Los genes candidato a asociar se obtuvieron luego de un riguroso proceso de selección, focalizando el interés en aquellos genes específicos de respuesta a *Sclerotinia* en plantas, obtenidos por estudios de perfiles de expresión diferencial (Zhao *et al.* 2007b; Ehrenbolger *et al.* 2013), y un conjunto de genes reportados como partícipes en procesos de defensa de girasol al ataque de hongos (Ehrenbolger *et al.* 2012; Fusari *et al.* 2012; Talukder *et al.* 2014). Con el objetivo de validar QTL reportados como relacionados con aumento de la resistencia al patógeno, se incluyó en el trabajo un conjunto de marcadores de tipo SSR que colocalizaron con QTL de respuesta a PHC (Zubrzycki *et al.* 2014). Finalmente, se incluyó también el conjunto de 182 marcadores de tipo SNP genotipificados utilizando la tecnología BeadXpress de Illumina, que resultaron polimórficos en la PMA (Capítulo 2), 59 de ellos caracterizados por resecuenciación de GC y los 123 restantes caracterizados *in silico* y seleccionados a partir de la función molecular inferida sobre el unigen correspondiente, relacionada con respuestas a distintos tipos de estrés (biótico-abiótico). El número de marcadores utilizado para asociación es superior al reportado para los dos trabajos de MA publicados a la fecha para el estudio de la resistencia a enfermedades en girasol (16 marcadores, Fusari *et al.* 2012; 8 marcadores, Talukder *et al.* 2014), pero inferior al utilizado para asociación con otros caracteres de interés agronómico como días a floración y número de ramificaciones (~5000 marcadores, Cadic *et al.* 2013 y Mandel *et al.* 2013).

La estratificación poblacional, también conocida como subestructura poblacional, hace referencia a la presencia de diferencias sistemáticas en las frecuencias alélicas entre subpoblaciones dentro de una muestra, posiblemente debido a diferente ancestría, las cuales resultan del apareamiento no aleatorio entre grupos seguido de deriva génica de las frecuencias alélicas en cada grupo. La presencia de estratificación no considerada, supone

un problema en los estudios de MA, ya que aumenta el error de tipo I generando resultados espurios (Stich *et al.* 2008). En esta PMA, los niveles de enfermedad no mostraron correlación directa con la estructura poblacional subyacente, aunque sí se observó una correlación marginalmente significativa ($p = 0,0655$) entre la estructura y AUDPCinc. De este modo, a fin de evitar asociaciones espurias, en este estudio se utilizó una estrategia basada en modelos mixtos (Yu *et al.* 2006; Zhao *et al.* 2007a; Kang *et al.* 2008), que posibilitó considerar diferentes niveles de parentesco y estructura.

Los análisis de MA de un marcador por vez resultan apropiados cuando la extensión del desequilibrio de ligamiento (DL), que en girasol se hace despreciable a las 100 kb (Fusari *et al.* 2008), es comparable o menor que la separación entre marcadores (Hayes 2013). Cuando la extensión del DL es mayor, múltiples marcadores dentro de una misma región pueden estar asociados con el carácter de interés, debido a la existencia de desequilibrio de ligamiento con el *locus* causal. De este modo, la aplicación de un método multilocus es más acertada para este tipo de situaciones, aunque los métodos de análisis de MA de un marcador por vez continúan siendo utilizados (*i.e.* Mandel *et al.* 2013, Elmer *et al.* 2015). Todos los SNPs pueden ser ajustados simultáneamente utilizando modelos propuestos originariamente para predicción genómica (*i.e.* Meuwissen *et al.* 2001). Para esto deben considerarse a los SNPs como efectos aleatorios (*i.e.* derivados de una distribución), asumiendo diferentes *priors* sobre la distribución de los efectos posibles de los SNPs.

En los análisis que consideran el efecto de un SNP por vez, bajo la expectativa de que un SNP tomado al azar no está asociado con el carácter fenotípico de interés (H_0 : el SNP no tiene efecto), las aproximaciones estadísticas que controlan apropiadamente el error de tipo I deberían mostrar una distribución uniforme de p-valores (una diagonal en la gráfica de porcentaje acumulado de p-valores vs. p-valores observados) (Yu *et al.* 2006). En este trabajo, dado que los SNPs y regiones a asociar no fueron elegidos al azar, sino que pertenecen a regiones candidatas seleccionadas por su función en procesos de defensa, es de esperar que, aun trabajando con un modelo que contemple no sólo la presencia de estructura en la población de mapeo (Q) sino también las relaciones de parentesco entre líneas (K), se observe un mayor número de SNPs asociados con el carácter que lo que se esperaría por azar. Esto se hizo evidente para las características IEmax, AUDPCinc y severidad, y en menor medida para período de incubación.

Al analizar un gran número de marcadores para asociación, resulta importante definir un adecuado valor de α , a fin de evitar el exceso de falsos positivos obtenidos al realizar múltiples pruebas (Hayes 2013). Una opción es ajustar el nivel de significancia por el número de marcadores utilizados, mediante la corrección de Bonferroni. Sin embargo, esta corrección no tiene en cuenta que las pruebas llevadas a cabo sobre un mismo cromosoma podrían no ser independientes, ya que los marcadores pueden estar en desequilibrio de ligamiento no sólo con el QTL, sino también entre ellos mismos. Como resultado, la corrección de Bonferroni tiende a ser demasiado conservadora, o requiere cierta decisión acerca de cuántas regiones independientes del genoma están siendo evaluadas (Hayes 2013). Una alternativa para evitar falsos positivos es monitorear el número de falsos positivos relativos al número de resultados positivos verdaderos (Fernando & Garrick 2013). El investigador puede seleccionar un nivel de significancia con una proporción aceptable de falsos positivos: la tasa de falsos descubrimientos (FDR, *false discovery rate*) constituye la proporción esperada de QTL detectados que son en realidad falsos positivos (Benjamini & Hochberg 1995; Weller *et al.* 1998). Siguiendo las recomendaciones de Elmer *et al.* (2015), en este trabajo se utilizó la corrección por q-valor propuesta por Storey *et al.* (2002), con un punto de corte en q-valor = 0,2. El q-valor de una hipótesis individual, análogo FDR del p-valor, se define como el mínimo FDR con el cual la prueba puede ser considerada significativa. Siguiendo este criterio, se encontraron seis regiones asociadas con menor AUDPC, tres con mayor período de incubación, doce con menor severidad y 32 con menor incidencia de PHC. Dada la gran cantidad de regiones asociadas, y que el objetivo ulterior de este trabajo es la transferencia de las mismas a los programas de mejoramiento locales, sólo se consideraron como verdaderas aquellas asociaciones que además de ajustarse al q-valor propuesto, también obtuvieron p-valores $\leq 0,01$. De este modo, para las medidas fenotípicas incidencia y severidad de la enfermedad, el número de regiones asociadas disminuyó a nueve y cinco, respectivamente, mientras que las asociadas con menor AUDPCinc y mayor Pinc, se mantuvieron invariables en seis y tres, respectivamente.

Estos marcadores asociados no fueron específicos de cada carácter, sino que muchos de ellos fueron compartidos entre más de un carácter fenotípico, confirmando la existencia de correlación significativa observada entre estas medidas fenotípicas en el Capítulo 1. De este modo, cuatro marcadores se asociaron con tres características, dos marcadores fueron compartidos entre dos características y se encontraron siete marcadores específicos de un

carácter: el SNP44 y el gen G33 fueron específicos de severidad, los SNPs 125, 80 y 60 y el gen G34 fueron específicos de incidencia y el SNP23 fue específico de Pinc.

Se observa que las distintas fuentes de información utilizadas para la búsqueda de GC y de SNPs resultaron fructíferas, ya que se encontraron regiones asociadas a partir de todas ellas: SNPs derivados de bases de datos de EST (SNPs 117, 136, 144, 125, 80 y 23 del ensayo de Illumina), SNPs caracterizados por resecuenciación de regiones de interés (SNP 44, 128 y 60 del ensayo de Illumina), genes candidato derivados de estudios de perfiles de expresión (G33 y G34), genes derivados de búsquedas bibliográficas (HaCOI_1) y marcadores que colocalizaron con QTL en estudios de mapeo biparental (HA1848). Esto confirma la importancia de la selección de genes candidato a partir de distintas fuentes de información. Con el desarrollo de las tecnologías de genotipificación masiva, ha surgido la necesidad de métodos que permitan el análisis de grandes volúmenes de datos, en tiempos razonables. Dado que los marcadores de preferencia son los SNPs, no resulta extraño que muchos de los paquetes y programas para la conducción de ensayos de asociación estén diseñados para admitir solamente marcadores bialélicos (*i.e.* PLINK, Purcell *et al.* 2007; BLR, Pérez *et al.* 2010). Esto supone un problema para análisis combinados de distintos tipos de marcadores moleculares, ya que limita el ingreso de marcadores multialélicos, como es el caso de los SSR, o de múltiples haplotipos de GC. La inclusión en estudios de MA de aquellos marcadores que colocalizaron con QTL es una estrategia interesante para enriquecer el número de regiones asociadas, y de validar los resultados obtenidos por otra metodología. En girasol, la mayor parte de los trabajos relacionados con búsquedas de fuentes de resistencia a *S. sclerotiorum* corresponden a estudios de mapeos de QTL en poblaciones biparentales, muchos de ellos realizados en base a mapas saturados con marcadores de tipo SSR (*i.e.* Micic *et al.* 2004, 2005a; Maringolo 2007; Yue *et al.* 2008). Por su parte, la evaluación de múltiples haplotipos de GC, en lugar de SNPs individuales, aporta una gran ventaja en los estudios de MA ya que los haplotipos observados en una región genómica potencialmente contendrían más información acerca del *locus* causal, por estar en mayor desequilibrio de ligamiento con los múltiples alelos de un QTL que un marcador único (Hayes 2013). A modo de ejemplo, en el estudio de asociación conducido por Pryce *et al.* (2010) para el carácter fertilidad en ganado, sólo se detectaron efectos significativos, que fueron posteriormente validados en otra PMA, cuando se utilizaron haplotipos en lugar de SNPs individuales.

En el análisis simultáneo de *loci* para asociación llevado a cabo con el *software* BLR, sólo fueron incluidos los SNPs y aquellos GC o marcadores SSR que presentaron 2 alelos. Siguiendo las recomendaciones de Perez *et al.* (2010) y De los Campos *et al.* (2013), sólo se estimaron *priors* para los parámetros aleatorios, mientras que para los efectos fijos se utilizaron *priors* de tipo *flat*, o no informativos. Las estimas de los *priors* σ^2 y σ^2_u , escala del error y de la varianza genética, respectivamente, se determinaron siguiendo las recomendaciones de Perez *et al.* (2010). Para la selección de los grados de libertad, dado que BLR considera distribuciones de tipo *Chi-cuadrado* invertidas, cuando más grados de libertad (*df*) se seleccionan (*i.e.*: más de cinco), más se concentra la distribución alrededor del parámetro de escala, dándole al *prior* más peso. Al revés, cuanto más cercano a cero es este número más vago es el *prior*. Siguiendo las recomendaciones de Perez *et al.* (2010) y De los Campos *et al.* (2013), se seleccionó un valor intermedio, $df = 3$. Tras 10000 iteraciones, 2000 descartadas por *burn-in*, se observó la convergencia de las cadenas de Markov, permitiendo obtener de este modo, a partir de una única distribución a *priori* para todos los marcadores, una distribución a *posteriori* para cada marcador. Con estos valores se generó un *ranking* de marcadores de acuerdo a cuánto explican del fenotipo. De acuerdo a lo esperado, los resultados demuestran que la mayor parte de los marcadores no hacen contribuciones significativas al fenotipo. En total, se encontraron 29 regiones que explican más del 1 % de la varianza fenotípica para el carácter incidencia de PHC (con un máximo de 8,31 % para el marcador SNP38), 32 para AUDPCinc (con un máximo de 11,52 % para el marcador SNP44), 30 para severidad (con un máximo de 7,07 % para el marcador SNP44) y 31 para Pinc (con un máximo de 9,14 % para el marcador SNP60). Cabe destacar que los SNPs 44 y 60 habían resultado asociados con resistencia a PHC utilizando el método de dos pasos.

Al comparar los resultados obtenidos utilizando ambos métodos estadísticos para cada característica fenotípica, puede observarse que todos los marcadores bialélicos seleccionados como asociados utilizando el método de dos pasos se encontraron entre los seleccionados utilizando el método de análisis simultáneo de *loci* por explicar más del 1 % de la varianza, con excepción de los SNPs 144 (asociado con menor incidencia y AUDPCinc de PHC) y 23 (asociado con mayor Pinc). Mientras que los marcadores que fueron seleccionados por ambas metodologías como asociados con menor enfermedad (SNPs 134, 44, 117, 125, 128, 60, 80 y G33) presentan frecuencias alélicas intermedias, los SNPs no

coincidentes 144 y 23 se caracterizaron por tener una MAF entre 0,05 y 0,1. Esta relación entre la repetitividad de resultados y los valores de MAF ya había sido reportada por Vukcevic (2009), quien observó que los *loci* con alelos en frecuencias intermedias eran más propensos a repetirse en ambas metodologías estadísticas, mientras que ninguno de los *loci* con MAF menor al 0,1 replicó con éxito.

De este modo, se observa que para este conjunto de datos, los dos métodos estadísticos utilizados son complementarios: en aquellos casos en los que se trabaja con marcadores bialélicos a frecuencias alélicas intermedias, los resultados obtenidos utilizando el método de análisis simultáneo de *loci* le dan soporte a los obtenidos mediante el método de análisis de un marcador por vez.

La única evidencia de que una asociación significativa detectada mediante MA es “real” (es decir, que un marcador está verdaderamente asociado con el carácter de interés), es la validación en una población independiente. A pesar de los esfuerzos puestos para controlar la estructura poblacional, y del uso de correcciones de los p-valores obtenidos, los falsos positivos pueden ocurrir en los estudios de MA dado el número de marcadores testeados. Esto significa que el diseño de un estudio de MA involucra tanto el descubrimiento como la validación de regiones asociadas con el carácter de interés (Hayes 2013).

Siguiendo esta premisa, del conjunto de regiones asociadas, HA1848 y HaCOI_1 ya podrían ser transferidas a programas de mejoramiento, ya que ambas han sido validadas mediante otras aproximaciones: HA1848 mediante mapeo de QTL en una población biparental de RILs (Zubrzycki *et al.* 2014) y HaCOI_1 por mapeo de asociación para la resistencia a podredumbre de base de tallo (PBT) en girasol causada también por *S. sclerotiorum*, en una población de MA completamente diferente (Talukder *et al.* 2014). Si bien la respuesta de genotipos de girasol a PBT ha demostrado ser diferente a la de PHC, sugiriendo la existencia de diferencias genéticas subyacentes (Gulya 1989; Talukder *et al.* 2014), contar con genes que confieran resistencia contra ambas enfermedades constituye un valioso aporte al cultivo de girasol.

Para todos estos marcadores asociados, con excepción quizás de HaCOI_1 y HA1848, aún resta el paso de validación. Si bien las asociaciones encontradas entre los mismos y menor PHC no pueden ser consideradas espurias, resulta necesaria la validación en otras poblaciones, como así también evaluar la funcionalidad y el patrón de expresión de estas regiones génicas en girasol en condiciones normales y de infección para poder determinar si

efectivamente son éstos los genes que determinan la resistencia a la enfermedad, o existen otros genes en desequilibrio de ligamiento que lo están haciendo. De este modo, experimentos de genómica funcional serán necesarios para desentrañar el rol que puedan tener en la respuesta a la infección de *S. sclerotiorum*.

Entre las regiones asociadas con menor enfermedad que aún resta validar, cinco son candidatas fuertes, seleccionados por su participación en procesos de defensa: G33, G34, SNP128, SNP80 y SNP44; mientras que las otras seis fueron seleccionadas por su función putativa relacionada con respuesta a estrés por *Gene Ontology*. A modo de reforzar la evidencia acerca de la participación de estas regiones génicas en la respuesta a PHC, una aproximación exploratoria útil es el uso de fuentes de información y recursos bioinformáticos públicos para el estudio de los niveles de expresión, que permiten identificar si en condiciones normales, los genes asociados tienen expresión preferencial en tejido asociado a capítulo (brácteas, estilo, estamen, entre otros), lo cual resulta interesante ya que las flores constituyen la principal forma de ingreso de este patógeno a la planta (Pereyra & Escande 1994) y, como su nombre lo indica, la podredumbre húmeda del capítulo es una enfermedad que afecta principalmente al capítulo de girasol. De modo que es de esperarse que los genes implicados en los procesos de defensa se expresen en dicho tejido. Las búsquedas bioinformáticas llevadas a cabo utilizando los repositorios públicos ATGC-SUR y Heliagene, permitieron corroborar que todos los genes asociados se expresan en tejidos relacionados con capítulo, presentando cuatro de ellos (representados por los SNPs 117, 136, 144 y 80) expresión preferencial en capítulo.

Además del estudio de los patrones de expresión, las búsquedas bibliográficas y el estudio de las vías de señalización y respuesta en plantas pueden ayudar a ampliar la comprensión acerca de los procesos en los que podrían estar involucradas las regiones que resultaron asociadas significativamente con aumento de la resistencia a PHC en girasol.

Las fitohormonas son importantes moduladores de la respuesta inmune en plantas (Robert-Seilaniantz *et al.* 2007). El ácido salicílico (SA) y la combinación de ácido jasmónico (JA) y etileno (ET) regulan las defensas contra patógenos biotróficos y necrotóxicos, respectivamente (Glazebrook 2005). El gen **COI1** de *A. thaliana* codifica una proteína que contiene una repetición rica en leucinas y un motivo F-box degenerado que se asocia físicamente con AtCUL1, AtRbx1, y las proteínas Skp1-like ASK1 o ASK2 para formar el complejo ubiquitin-ligasa SCF(COI1) (Xu *et al.* 2002), el cual es esencial en la vía de

señalización de JA. Los mutantes funcionales de COI1 son insensibles, o tienen sensibilidad reducida, a JA (Xie *et al.* 1998): una única sustitución aminoacídica en el motivo *F-box* impide la formación del complejo SCF(COI1) (Xu *et al.* 2002). Mediante el uso de mutantes de receptores COI1 en el patosistema *Sclerotinia-Arabidopsis*, pudo determinarse que las vías de señalización que involucran JA regulan las respuestas de defensa contra el hongo necrotófico *S. sclerotiorum* (Guo & Stotz 2007; Perchepped *et al.* 2010). El gen HaCOI1 genotipificado en este trabajo de Tesis es un ortólogo del gen COI1 de *A. thaliana*. Coincidentemente con los hallazgos de Talukder *et al.* (2014), el haplotipo asociado en este trabajo con menor enfermedad presenta una sustitución aminoacídica de importancia en la secuencia final de la repetición altamente conservada rica en leucina, según lo reportado por Xie *et al.* (1998): presenta una asparagina, en vez de una arginina, como se observó en COI1 de *Arabidopsis* (Xie *et al.* 1998), la cual constituye un residuo funcional diferente en tamaño, pH, carga e hidrofobicidad. Tras el análisis filogenético de este gen incluyendo 52 especies vegetales, la presencia de asparagina en dicha posición solo se encontró en *Helianthus annuus*, llevando a sugerir que esta mutación podría corresponder a una ganancia de función a partir del gen ancestral (Talukder *et al.* 2014).

Para el caso del marcador **HA1848**, resulta imposible con las herramientas actuales determinar qué gen estaría implicado en la resistencia, dado que se encuentra a 3,3 y 7,7 cM de los marcadores flanqueantes en el mapa de ligamiento de Zubrzycki *et al.* (2014), lo cual podrían significar millones de pares de bases en girasol. Sin embargo, el grupo de ligamiento donde fue localizado, GL7, ya ha sido mencionado en diversos trabajos de mapeo de QTL para resistencia *S. sclerotiorum* en girasol, sugiriendo que allí se localiza una región de importancia para el carácter. Yue *et al.* (2008) han reportado un QTL relacionado con menor incidencia de PHC; Bert *et al.* (2002) reportaron un QTL relacionado con menor severidad y dos relacionados con mayor índice de latencia (similar a período de incubación); mientras que Bert *et al.* (2004) reportaron un QTL relacionado con menor severidad en yemas, y dos QTL relacionados con menor severidad e índice de latencia en capítulo. De este modo, si bien el marcador ya puede ser transferido al programa de mejoramiento, resultaría de interés continuar con la caracterización de la región para lograr identificar qué gen o grupo de genes estarían implicados en la resistencia a PHC.

El **SNP117**, que resultó asociado con tres características fenotípicas relacionadas con menor PHC, pertenece a un gen inédito no estudiado aún en girasol, y poco en plantas en general:

un gen que codifica para una proteína análoga a progesterona-5- β -reductasa. El metabolismo de esteroides en plantas resulta de interés, no solo debido a que está involucrado en la síntesis de compuestos esenciales para el desarrollo y crecimiento de las plantas (*i.e.* brasinoesteroides) sino también debido a que los esteroides vegetales incluyen productos económicamente importantes, como cardenolinas, glicoalcaloides y saponinas (Bauer *et al.* 2010). La enzima progesterona-5- β -reductasa (P5 β R) cataliza la reducción de 5- β -progesterona a 5- β -pregnana-3,20-diona, lo cual constituye el primer paso en la vía de síntesis de cardenolinas (Bauer *et al.* 2010). Diversos estudios han reportado la ocurrencia de ortólogos a P5 β R en especies productoras y no productoras de cardenolinas en diversas familias de angiospermas (*Brassicales*, *Gentianales*, *Lamiales*, *Malvales* y *Solanales*), observándose alta conservación de la secuencia aminoacídica, y de la capacidad de reducción de las dobles uniones C-C mediante un mecanismo de adición 1,4 (Pérez-Bermúdez *et al.* 2010, Bauer *et al.* 2010). Asimismo, diferentes hipótesis han surgido para justificar la existencia de estas enzimas en especies no productoras de cardenolinas. Recientemente, Pérez-Bermúdez *et al.* (2010) postularon que P5 β R podría ser de un gen de defensa, dado que sus niveles de expresión muestran variaciones en respuesta a distintas condiciones de estrés, al tiempo que es regulado por los niveles de etileno y peróxido de hidrógeno, aunque su mecanismo de acción no está definido (Pérez-Bermúdez *et al.* 2010). Por su parte, estudios de expresión diferencial en líneas de trigo infectadas con roya vs. control no infectado, mostraron aumento en la expresión de P5 β R (Hulbert *et al.* 2007). Cabe destacar que este trabajo de Tesis constituye el primer reporte sobre la ocurrencia de ortólogos a P5 β R en girasol.

Existen mecanismos de defensa innatos en las plantas que se disparan tras el reconocimiento de características microbianas conservadas, denominadas patrones moleculares asociados a patógenos (PAMPs, *pathogen-associated molecular patterns*). Un ejemplo de PAMP fúngico es la quitina, componente principal de las paredes celulares de los hongos. El reconocimiento de PAMPs puede inducir respuesta hipersensible, muerte celular y síntesis de moléculas antimicrobianas (*i.e.* fitoalexinas; Heath 2000). La activación de estas vías también resulta en la síntesis de una variedad de proteínas relacionadas con patogénesis (PR, *pathogeneses related*) que juegan un rol fundamental en la defensa contra patógenos. El marcador **SNP136**, asociado con menor incidencia y AUDPCinc, constituye un ejemplo de esto, ya que pertenece a una quitinasa β -n-acetilglucosaminidasa, enzima que

cataliza la hidrólisis de los polímeros de quitina, compuestos por uniones β -1 \rightarrow 4 *N*-acetilglucosamina (Shibuya & Minami 2001). La rápida activación de cascadas de señalización mediada por quinasas, como aquellas portadoras de los SNPs **60** y **125**, asociados con menor incidencia de PHC, y MAP-kinasas, como **SNP23**, asociado con menor período de incubación de la enfermedad, es importante para respuesta provocada por quitina (Libault & Wan 2007). El hecho de que la quitina gatille la expresión *de novo* de genes involucrados en respuesta a patógenos, sugiere que en este proceso también están involucrados diversos factores de transcripción (FT). Estudios realizados utilizando microarreglos de ADN han permitido identificar una larga lista de FT que responden a la inoculación con patógenos. Entre estos, los factores de transcripción de la familia GRAS (Day *et al.* 2003, 2004; Libault & Wan 2007). El acrónimo GRAS fue acuñado luego de la identificación de los miembros que la conforman: GAI (*gibberellin acid insensitive*), RGA (*repressor of GA1*) y SCR (*Scarecrow*) (Pysh *et al.* 1999). Se ha demostrado que dos miembros de la familia GRAS SCR, CIGR1 y CIGR2 (Inducibles por quitina, de respuesta a giberelina, *chitin-inducible, gibberellin-responsive*), actúan como reguladores transcripcionales en los eventos tempranos de la respuesta inducida por activadores, entre ellos los monómeros de quitina (Day *et al.* 2003, 2004). Este podría ser el rol putativo de la región que contiene al **SNP44**, asociada con menor severidad, en el proceso de defensa contra PHC.

Como se mencionara previamente, *S. sclerotiorum* es un ascomicete capaz de secretar oxalato, lo cual es fundamental para su éxito patogénico. El papel del oxalato en la patogenicidad ha sido objeto de numerosas investigaciones (Godoy *et al.* 1990; Cessna 2000; Kim *et al.* 2008; Williams *et al.* 2011; Liang *et al.* 2014; Kabbage *et al.* 2015). Este ácido dicarboxílico multifuncional contribuye en numerosos procesos fisiológicos que facilitan la colonización de hongos en las plantas hospederas: el oxalato es capaz de inducir una respuesta similar a muerte celular apoptótica en el huésped, un proceso que depende en gran medida de la generación de especies reactivas de oxígeno (ROS, *reactive oxygen species*). Aunque la inducción de ROS y la apoptosis parecen ser críticas para el éxito patogénico de *S. sclerotiorum* en las etapas posteriores de la infección, durante las primeras etapas, la secreción de oxalato amortigua el estallido oxidativo del huésped, una respuesta temprana asociada con la defensa de la planta. Por lo tanto, este ácido orgánico sencillo parece tener funciones duales opuestas: la inhibición de las defensas de la planta

incluyendo ROS en la primera etapa de la infección, como preludeo a la posterior generación de ROS e inducción de la muerte celular en el huésped. Esto sugiere una mediación del medio ambiente *redox* del huésped por parte *S. sclerotiorum* y más específicamente del ácido oxálico. Esta supresión de la respuesta inicial del huésped mediante la secreción de ácido oxálico, se ve contrarrestada por la función oxalato oxidasa y/o superóxido dismutasa de proteínas germinas o tipos germinas en las plantas, como la HaGLP3, o **SNP128**, que en este trabajo se encontró asociado con menor incidencia y AUDPCinc (Dumas *et al.* 1993; Carter & Thornburg 2000; Hu *et al.* 2003).

El marcador **G34**, perteneciente a una región que codifica para una glicoproteína-p, se asoció significativamente con menor incidencia de PHC. Las plantas utilizan las glicoproteínas-p para el transporte de auxinas (Geisler & Murphy 2006): ya ha sido demostrado que la interrupción de la expresión del gen que codifica para glicoproteína-p en *Arabidopsis thaliana* (Noh *et al.* 2001), maíz y sorgo (Multani *et al.* 2003) resulta en una reducción del transporte de auxinas y un fenotipo enano. Si bien la función mayormente asociada con la fitohormona auxina es la de reguladora del crecimiento en plantas, también ha sido asociada con la defensa a patógenos (Yamada 1993; Robert-Seilaniantz *et al.* 2007; Navarro *et al.* 2008). Estudios realizados en *Arabidopsis* sugieren que la señalización mediada por auxinas es requerida para la resistencia contra hongos necrotróficos como *S. sclerotiorum* (Tiryaki & Staswick 2002; Llorente *et al.* 2008; Cuzick *et al.* 2009), pero genera susceptibilidad contra patógenos fúngicos biotróficos (Tor 2002). También se ha observado que las vías de señalización mediadas por JA y auxinas comparten muchas características en común e interactúan positivamente (Kazan & Manners 2009). Las señales mediadas por auxinas y JA son percibidas y transmitidas a través de las vías de señalización en las cuales las proteínas F-box (TIR1 y **CO11** -*coronatine-insensitive1*- para auxinas y JA, respectivamente) y la ubiquitinación de represores (AUX/IAA y JAZ -*jasmonate zim domain*- proteínas para la señalización mediada por auxinas y JA, respectivamente) son necesarias para la activación de las respuestas posteriores (Mockaitis & Estelle 2008).

El **SNP144** pertenece a una región que codifica para una tirosina aminotransferasa (TAT). Se ha reportado que TAT puede ser inducido por octadecanoides, como el JA y su precursor ácido 12-oxofitodienoico (OPDA) (Farmer & Ryan 1992; Weiler *et al.* 1994; Lopukhina & Dettenberg 2001), así como también por heridas (Laudert & Weiler 1998), exceso de luz, luz UV y el herbicida oxi-fluoreno (Holländer-Czytko *et al.* 2005). La enzima TAT, que cataliza la

transaminación de tirosina a p-hidroxifenilpiruvato, es la primera enzima de una vía que conduce a través de ácido homogentísico a plastoquinona y tocoferoles, siendo los últimos conocidos por su actividad antioxidante en plantas. Se estima que los tocoferoles protegen el aparato fotosintético mediante la captación de oxígeno (Trebst *et al.* 2002), ROS y radicales lipídicos (Fryer 1992; Munné-Bosch & Alegre 2010). De una manera alternativa, un papel más directo de TAT en el ataque de patógenos podría ser que los compuestos fenólicos formados a partir de p-hidroxifenilpiruvato pueden servir como agentes de reticulación en la pared celular, fortaleciendo de ese modo esta barrera física (Schmidt *et al.* 1998; Lopukhina & Dettenberg 2001). Por su parte, el análisis comparativo de los perfiles metabólicos de dos líneas endocriadas de girasol en respuesta a la infección con *S. sclerotiorum* condujo a la identificación de cambios notorios en los niveles del aminoácido tirosina a con el transcurso de los días post inoculación, particularmente para la línea susceptible (Peluffo *et al.* 2010).

El marcador **G33**, perteneciente a una región que codifica para una proteína con dominio radical SAM, se asoció significativamente con menor severidad de la enfermedad. Muchos metabolitos secundarios son sometidos a procesos de metilación a partir de S-adenosil-L-metionina, entre ellos salicilato y jasmonato. Las enzimas que catalizan la metilación de salicilato aún no han sido identificadas, pero un gen que codifica para una ácido-jasmónico-carboximetiltransferasa ha sido clonado a partir de *Arabidopsis* y sus propiedades han sido descritas por Seo *et al.* (2001). El ácido jasmónico, en comparación con el metil-jasmonato volátil, ha sido históricamente propuesto como el compuesto responsable de la señalización intracelular en plantas. Sin embargo, la sobre-expresión de ácido-jasmónico-carboximetiltransferasa muestra que el metil-jasmonato constituye en sí mismo una señal intracelular importante, que actúa en las respuestas de defensa de las plantas a patógenos (Seo *et al.* 2001; Moffatt & Weretilnyk 2001).

Tal como se describió en el Capítulo 1 de esta Tesis, dado que los materiales incluidos en la PMA pertenecen al programa de mejoramiento, de encontrarse alelos asociados al carácter de interés, se verá allanado el camino para su introgresión con vistas al mejoramiento genético del cultivo (Myles *et al.* 2009). Si bien son portadoras de diferentes fuentes de resistencia, las líneas 51084, HA853, RHA441 y RHA801 se destacan por haber sido seleccionadas para tres de las cuatro características fenotípicas evaluadas, mientras que la línea 51084/C820 fue seleccionada para las cuatro. Este panel de líneas, conformado por

dos líneas pertenecientes al Banco Activo de Germoplasma de INTA Manfredi y tres líneas públicas, puede ser utilizado en los programas de mejoramiento para el establecimiento de nuevos materiales comerciales con resistencia a PHC, como así también para el desarrollo de poblaciones biparentales para continuar con el estudio de la resistencia a la enfermedad. El establecimiento de la estrategia de mapeo por asociación para la resistencia a PHC ha sido producto de un esfuerzo multidisciplinario, que incluyó biólogos moleculares, genetistas, fitopatólogos, mejoradores y estadísticos. Este abordaje resultó exitoso en la identificación de trece marcadores asociados con resistencia a PHC, medida como incidencia, AUDPCinc, severidad y período de incubación. Estos resultados no sólo incrementan la información acerca de los factores genéticos en girasol involucrados en la defensa contra PHC, sino que además aportan herramientas para el mejoramiento asistido por marcadores.

CONCLUSIONES y PERSPECTIVAS

CONCLUSIONES FINALES y PERSPECTIVAS

Este trabajo de Tesis ha contribuido a la identificación y caracterización de nuevas fuentes de resistencia genética a la Podredumbre Húmeda del Capítulo (PHC) en girasol. La línea de trabajo en poblaciones no estructuradas de girasol de INTA fue iniciada en el año 2008 por el grupo de Genómica de Girasol, IB, INTA, Castelar, en colaboración con el grupo de Fitopatología de la EEA Balcarce y el programa de mejoramiento y Banco de Germoplasma EEA Manfredi (BAG-IM) a través del desarrollo del primer estudio de mapeo por asociación (MA) para caracteres complejos en el cultivo, basado en la evaluación de un panel de 16 genes candidato en una población de mapeo compuesta por 94 líneas endocriadas (Fusari *et al.* 2012). Tomando como base estos primeros esfuerzos, y haciendo uso de las nuevas tecnologías de tipificación disponibles, este trabajo de Tesis ha dado lugar al establecimiento de una nueva y más amplia población de mapeo, exhaustivamente caracterizada tanto a nivel molecular como fenotípico. Más importante aún, la información así generada, junto con la incorporación de herramientas informáticas y estadísticas robustas que han sido desarrolladas en los últimos años, ha permitido la identificación de polimorfismos novedosos asociados con la resistencia a la PHC, contribuyendo a la generación no sólo de conocimiento original en el área de biotecnología vegetal, sino también de nuevas estrategias y herramientas para estudios genómicos orientados al mejoramiento de girasol en particular y de cultivos en general. Las conclusiones principales de este trabajo son:

- I La constitución y caracterización de la Población de Mapeo por Asociación (PMA), a partir de materiales provenientes del programa de mejoramiento y de líneas públicas, constituye un insumo importante para el mejoramiento del cultivo en nuestro país. Además de ser utilizada para el carácter resistencia a Podredumbre Húmeda del Capítulo (PHC), actualmente constituye la base de los estudios de asociación de girasol de INTA para diversos caracteres complejos de relevancia agronómica.
- II Como resultado del diseño y caracterización del panel de 384 SNPs (Polimorfismos de Nucleótido Simple) para girasol (GoldenGate, Illumina, EUA), una nueva tecnología capaz de disminuir los tiempos y costos de genotipificación se encuentra disponible para futuros estudios poblacionales, de diversidad y asociación.

-
- III La generación de paneles de marcadores moleculares capaces de discriminar entre distintas entradas de girasol ha permitido incrementar el conocimiento acerca de los niveles de diversidad genética y patrones de estructura poblacional de los materiales conservados en el BAG-IM. Estos paneles constituyen una valiosa herramienta para ser transferidos tanto al BAG-IM como al Instituto Nacional de Semillas (INASE) para la caracterización de materiales de girasol, la creación de colecciones núcleo y la certificación de pureza genética de las entradas con el fin de complementar los procedimientos de protección de obtenciones vegetales.
 - IV Los métodos de análisis de *locus* individuales y de análisis simultáneo de *loci* demostraron ser complementarios, especialmente para aquellos casos en los que se trabaja con marcadores bialélicos de frecuencias intermedias, aportando nuevas evidencias a un área de desarrollo activo, como son las aproximaciones de mejoramiento molecular basadas en genética de poblaciones y los métodos estadísticos para determinar las asociaciones fenotipo-genotipo.
 - V El trabajo permitió la identificación de 13 regiones asociadas con menor PHC, medida como incidencia, área bajo la curva de progreso (AUDPC) de incidencia, severidad y período de incubación, bajo un modelo que contempla no sólo la estructura poblacional sino también las relaciones de parentesco entre las líneas. De este conjunto de regiones asociadas, dos validan los resultados hallados por otras metodologías, convirtiéndose en candidatas robustas para su incorporación a corto plazo en los programas de mejoramiento.
 - VI Los resultados obtenidos permitieron la identificación de cinco líneas endocriadas (dos pertenecientes al BAG-IM y tres públicas) que expresan de manera estable un fenotipo resistente a PHC evaluado mediante inoculación asistida durante cinco campañas sucesivas y presentan combinaciones de alelos asociados con la resistencia. Las mismas podrán ser utilizadas como materiales semilla para el establecimiento de nuevos materiales experimentales y avanzados con resistencia a PHC a partir de aproximaciones de mejoramiento asistido, permitiendo así acelerar la creación de nuevos híbridos comerciales y el desarrollo de poblaciones biparentales para continuar con el estudio de la resistencia a la enfermedad.

Los resultados obtenidos demuestran que los estudios de asociación son herramientas útiles a la hora de identificar genes implicados en caracteres complejos. Del total de regiones asociadas con menor PHC, se distinguen HaCOI_1 y Ha1848 como insumos de uso inmediato en planes de mejoramiento, dado que ya han sido descritas previamente como asociadas con resistencia a *Sclerotinia* en girasol y validadas en los estudios conducidos en esta Tesis. Mientras tanto, para los genes inéditos asociados con resistencia a PHC se proponen futuras instancias de validación para ser luego incorporados en una segunda etapa a los programas de mejoramiento. Los métodos de validación disponibles están basados en el estudio de líneas introgresadas, estrategias de mapeo de QTL y/o mutagénesis dirigida (TILLING), siendo las principales regiones candidatas las portadoras de los SNPs 117, 128, 136 y 144 por estar asociadas con múltiples medidas fenotípicas.

Lic. Carla V. Filippi
Doctorando

Dra. Norma B. Paniego
Directora Asistente

Dra. Verónica V. Lia
Directora

BIBLIOGRAFÍA

BIBLIOGRAFÍA

- Agrama HA, Eizenga GC, Yan W (2007) Association mapping of yield and its components in rice cultivars. *Molecular Breeding*, **19**, 341–356.
- Aguero ME, Pereyra VR, Escande AR (2001) Efecto de la pudrición húmeda del capítulo de girasol [*Sclerotinia sclerotiorum* (Lib.) de Bary] sobre el contenido de impurezas del producto cosechado, y el contenido y la acidez del aceite. *The Journal of agriculture of the University of Puerto Rico*, **85**, 177-186.
- Akaike H (1974) A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716–723.
- Akhunov E, Nicolet C, Dvorak J (2009) Single nucleotide polymorphism genotyping in polyploid wheat with the Illumina GoldenGate assay. *TAG. Theoretical and applied genetics*, **119**, 507–517.
- Al-Chaarani GR, Gentzbittel L, Huang XQ, Sarrafi A (2004) Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *TAG. Theoretical and applied genetics*, **109**, 1353–1360.
- Al-Chaarani GR, Roustae A, Gentzbittel L *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). *TAG. Theoretical and applied genetics*, **104**, 490–496.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. *Journal of molecular biology*, **215**, 403–410.
- Altshuler D, Daly MJ, Lander ES (2008) Genetic mapping in human disease. *Science*, **322**, 881–888.
- Álvarez D (2002) Germoplasma de Girasol: Un Recurso Irreemplazable. In: *Girasol*, pp. 148–151.
- Arbelbide M, Bernardo R (2006) Mixed-model QTL mapping for kernel hardness and dough strength in bread wheat. *TAG. Theoretical and applied genetics*, **112**, 885–890.
- Baack EJ, Whitney KD, Rieseberg LH (2005) Hybridization and genome size evolution: timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *The New phytologist*, **167**, 623–630.
- Bachlava E, Taylor CA, Tang S *et al.* (2012) SNP discovery and development of a high-density genotyping array for sunflower. *PLoS one*, **7**, e29814.
- Baldini M, Turi M, Vischi M, Vannozzi GP, Olivieri AM (2002) Evaluation of genetic variability for *Sclerotinia sclerotiorum* Lib. de Bary resistance in sunflower and utilization of associated molecular markers. *Helia*, **26**, 177–190.
- Balzarini M, Bruno C, Peña A, Teich I, Di Rienzo J. (2010) *Estadística en Biotecnología. Aplicaciones en Info-Gen*. Grupo de Editores. Córdoba, Argentina.
- Balzarini MG, Di Rienzo JA (2011) InfoGen: Software para análisis estadístico de datos genéticos. Facultad de Ciencia Agropecuarias. Universidad Nacional de Córdoba. Argentina.
- Bates D, Mächler M, Bolker BM, Walker SC (2014) Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*.

- Bauer P, Munkert J, Brydziun M *et al.* (2010) Highly conserved progesterone 5 beta-reductase genes (P5 beta R) from 5 beta-cardenolide-free and 5 beta-cardenolide-producing angiosperms. *Phytochemistry*, **71**, 1495–1505.
- Bazzalo ME, Dimarco P, Martinez F, Daleo GR (1991) Indicators of resistance of sunflower plant to basal stalk rot (*Sclerotinia sclerotiorum*): Symptomatological, biochemical, anatomical, and morphological characters of the host. *Euphytica*, **57**, 195–205.
- Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society Series B Methodological*, **57**, 289–300.
- Berger RD, Luke HH (1979) Spatial and temporal spread of oat crown rust. *Phytopathology*, **69**, 1199–1201.
- Berry ST, Leon AJ, Challis P *et al.* (1996) Construction of a high density, composite RFLP linkage map for cultivated sunflower *Helianthus annuus*. *Proceedings of the 14th International Sunflower Conference - Beijing, China.*, **2**, 1150–1160.
- Berry ST, Leon AJ, Hanfrey CC *et al.* (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. *Theoretical and Applied Genetics*, **91**, 195.199.
- Berry ST, Leon AJ, Peerbolte R *et al.* (1997) Presentation of the Advanta sunflower RFLP linkage map for public research. In: *Proc. 19th Sunflower Res. Workshop, Fargo, ND.*, pp. 113–118.
- Bert PF (2002) Etude génétique et marquage moléculaire des facteurs de résistance aux principaux champignons pathogènes du tournesol (*Helianthus annuus* L.). Université Clermont Ferrand II.
- Bert PF, Dechamp-Guillaume G, Serre F *et al.* (2004) Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.) 3. Characterisation of QTL involved in resistance to *Sclerotinia sclerotiorum* and *Phoma macdonaldi*. *TAG. Theoretical and applied genetics*, **109**, 865–874.
- Bert PF, Jouan I, De Labrouhe DT *et al.* (2002) Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.) 1. QTL involved in resistance to *Sclerotinia sclerotiorum* and *Diaporthe helianthi*. *Tag Theoretical And Applied Genetics*, **105**, 985–993.
- Bertero de Romano A (2003) Origin of the Argentine sunflower. *Helia*, **26**, 127–136.
- Birhman RK, Singh BP (1995) Path-coefficient analyses and genetic parameters of the components of field resistance of potatoes to late blight. *Annals of Applied Biology*, **127**, 353–362.
- Bock CH, Poole GH, Parker PE, Gottwald TR (2010) Plant Disease Severity Estimated Visually, by Digital Photography and Image Analysis, and by Hyperspectral Imaging. *Critical Reviews in Plant Sciences*, **29**, 59–107.
- Boland GJ, Hall R (1994) Index of plant hosts of *Sclerotinia sclerotiorum*. *Canadian Journal of Plant Pathology*, **16**, 93–108.
- Bolker BM (2008) *Ecological Models and Data in R*. Princeton University Press.

- Bowers JE, Nambeesan S, Corbi J *et al.* (2012) Development of an Ultra-Dense Genetic Map of the Sunflower Genome Based on Single-Feature Polymorphisms. (S-B Wu, Ed). *PLoS ONE*, **7**, e51360.
- Bracco M, Lia V V, Gottlieb AM (2009) Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina. *Genetica*, **135**, 39–49.
- Brachi B, Faure N, Horton M *et al.* (2010) Linkage and Association Mapping of Arabidopsis thaliana Flowering Time in Nature (TFC Mackay, Ed). *PLoS Genetics*, **6**, e1000940.
- Bradbury PJ, Zhang Z, Kroon DE *et al.* (2007) TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics (Oxford, England)*, **23**, 2633–2635.
- Breseghele F, Sorrells ME (2006) Association Analysis as a Strategy for Improvement of Quantitative Traits in Plants. *Crop Science*, **46**, 1323.
- Broers LHM, Subias XC, Atilano RML (1996) Field assessment of quantitative resistance to yellow rust in ten spring bread wheat cultivars. *Euphytica*, **90**, 9–16.
- Bruniard JM (2002) Presente y Futuro de la Investigación en Girasol en Argentina. *Idia XXI*, **2**, 116–118.
- Buckler ES, Thornsberry JM (2002) Plant molecular diversity and applications to genomics. *Current opinion in plant biology*, **5**, 107–111.
- Burke JM, Tang S, Knapp SJ, Rieseberg LH (2002) Genetic analysis of sunflower domestication. *Genetics*, **161**, 1257–1267.
- Cadic E, Coque M, Vear F *et al.* (2013) Combined linkage and association mapping of flowering time in Sunflower (*Helianthus annuus* L.). *TAG. Theoretical and applied genetics*, **126**, 1337–1356.
- Campbell CL, Madden L V. (1990) *Introduction to plant disease epidemiology*. John Wiley & Sons.
- Cappa EP, El-Kassaby YA, Garcia MN, Acuña C, Borralho NM Grattapaglia D & Poltri SN (2013) Impacts of population structure and analytical models in genome-wide association studies of complex traits in forest trees: a case study in *Eucalyptus globulus*. *PLoS one*, **8**, e81267.
- Carson ML (1995) Inheritance of latent period length in maize infected with *Exserohilum turcicum*. *Plant disease*, **79**, 581–585.
- Carter JF, Putt ED (1978) *Sunflower Science and Technology*. American Society of Agronomy, Crop Science Society of America, Soil Science Society of America.
- Carter C, Thornburg RW (2000) Tobacco nectarin I. Purification and characterization as a germin-like, manganese superoxide dismutase implicated in the defense of floral reproductive tissues. *The Journal of biological chemistry*, **275**, 36726–33.
- Casa AM, Pressoir G, Brown PJ *et al.* (2008) Community Resources and Strategies for Association Mapping in Sorghum. *Crop Science*, **48**, 30.
- Castaño F, Vear F, Labrouhe DT (1993) Resistance of sunflower inbred lines to various forms of attack by *Sclerotinia sclerotiorum* and relations with some morphological characters. *Euphytica*, **68**, 85–98.

- Cessna SG (2000) Oxalic Acid, a Pathogenicity Factor for *Sclerotinia sclerotiorum*, Suppresses the Oxidative Burst of the Host Plant. *The plant cell*, **12**, 2191–2200.
- Chapman MA, Leebens-Mack JH, Burke JM (2008) Positive selection and expression divergence following gene duplication in the sunflower CYCLOIDEA gene family. *Molecular biology and evolution*, **25**, 1260–1273.
- Chen X, Sullivan PF (2003) Single nucleotide polymorphism genotyping: biochemistry, protocol, cost and throughput. *The pharmacogenomics journal*, **3**, 77–96.
- Chester KS (1950) Plant disease losses: their appraisal and interpretation. *Plant Disease Reporter*, **193**, 190–362.
- Cholastova T, Knotova D (2012) Using Morphological and Microsatellite (SSR) Markers to Assess the Genetic Diversity in Alfalfa (*Medicago sativa* L.). In: *Proceedings of World Academy of Science*, p. 69.
- Close TJ, Bhat PR, Lonardi S *et al.* (2009) Development and implementation of high-throughput SNP genotyping in barley. *BMC genomics*, **10**, 582.
- Cobb NA (1892) Contributions to an economic knowledge of the Australian rusts (Uredineae). *Agricultural Gazette of New South Wales*, **3**, 44–48.
- Collard BCY, Vera Cruz CM, McNally KL, Virk PS, Mackill DJ (2008) Rice molecular breeding laboratories in the genomics era: Current status and future considerations. *International journal of plant genomics*, **2008**, 524847.
- Collett D (1991) Statistical inference for binary data. In: *Modelling Binary Data.*, pp. 17–42. Springer US.
- Comai L, Young K, Till BJ *et al.* (2004) Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *The Plant Journal*, **37**, 778–786.
- Conles MY, Cragolini CI, Yossen VE, Balzarini M, Macchiavelli RE (2011) incidencia de podredumbre blanca (*Sclerotium cepivorum* Berk) en cultivos de ajo mediante un modelo no lineal mixto. *Agriscientia*, **28**, 61–74.
- Cooper JL, Till BJ, Laport RG *et al.* (2008) TILLING to detect induced mutations in soybean. *BMC plant biology*, **8**, 9.
- Coque M, Mesnildrey S, Romestant M *et al.* (2008) Sunflower nested core collections for association studies and phenomics. In: *Proceeding ASTA Conference, Cordoba*
- Crites (1993) Domesticated Sunflower in Fifth Millennium B.P. Temporal Context: New Evidence from Middle Tennessee. *American Antiquity*, **58**, 146–148.
- Crossa J, Campos G de L, Pérez P *et al.* (2010) Prediction of genetic values of quantitative traits in plant breeding using pedigree and molecular markers. *Genetics*, **186**, 713–24.
- Cuzick A, Maguire K, Hammond-Kosack KE (2009) Lack of the plant signalling component SGT1b enhances disease resistance to *Fusarium culmorum* in *Arabidopsis* buds and flowers. *The New phytologist*, **181**, 901–12.
- Dalloul RA, Long JA, Zimin A V *et al.* (2010) Multi-platform next-generation sequencing of the domestic turkey (*Meleagris gallopavo*): genome assembly and analysis. *PLoS biology*, **8**, e1000475.

- Damicone JP, Snow JP, Berggren GT (1990) Spatial and temporal spread of soybean stem canker from an inoculum point source. *Phytopathology*, **80**, 571–578.
- Das MK, Rajaram S, Mundt CC, Kronstad WE (1992) Inheritance of Slow-Rusting Resistance to Leaf Rust in Wheat. *Crop Science*, **32**, 1452.
- Day RB, Shibuya N, Minami E (2003) Identification and characterization of two new members of the GRAS gene family in rice responsive to N-acetylchitooligosaccharide elicitor. *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression*, **1625**, 261–268.
- Day RB, Tanabe S, Koshioka M *et al.* (2004) Two rice GRAS family genes responsive to N - acetylchitooligosaccharide elicitor are induced by phytoactive gibberellins: evidence for cross-talk between elicitor and gibberellin signaling in rice cells. *Plant molecular biology*, **54**, 261–272.
- Dehmer KJ, Friedt W (1998) Evaluation of different microsatellite motifs for analysing genetic relationships in cultivated sunflower (*Helianthus annuus* L.). *Plant Breeding*, **117**, 45–48.
- Dempewolf H, Rieseberg LH, Cronk QC (2008) Crop domestication in the Compositae: a family-wide trait assessment. *Genetic Resources and Crop Evolution*, **55**, 1141–1157.
- Deokar AA, Ramsay L, Sharpe AG *et al.* (2014) Genome wide SNP identification in chickpea for use in development of a high density genetic map and improvement of chickpea reference genome assembly. *BMC genomics*, **15**, 708.
- Devlin B, Roeder K, Wasserman L (2001) Genomic control, a new approach to genetic-based association studies. *Theoretical population biology*, **60**, 155–66.
- Diwan N, Fluhr R, Eshed Y, Zamir D, Tanksley SD (1999) Mapping of Ve in tomato: a gene conferring resistance to the broad-spectrum pathogen, *Verticillium dahliae* race 1. *TAG Theoretical and Applied Genetics*, **98**, 315–319.
- Don RH, Cox PT, Wainwright BJ, Baker K, Mattick JS (1991) “Touchdown” PCR to circumvent spurious priming during gene amplification. *Nucleic acids research*, **19**, 4008.
- Dong GJ, Liu GS, Li KF (2007) Studying genetic diversity in the core germplasm of confectionary sunflower (*Helianthus annuus* L.) in China based on AFLP and morphological analysis. *Russian Journal of Genetics*, **43**, 627–635.
- Donnelly P (2008) Progress and challenges in genome-wide association studies in humans. *Nature*, **456**, 728–31.
- Dorrell DG, Vick BA (1997) Properties and processing of oilseed sunflower. In: *Schneiter, A.A. (Eds), Sunflower technology and production American Society of Agronomy, Madison, Wisconsin*, pp. 709–744.
- Dreisigacker S, Zhang P, Warburton ML *et al.* (2005) Genetic Diversity among and within CIMMYT Wheat Landrace Accessions Investigated with SSRs and Implications for Plant Genetic Resources Management. *Crop Science*, **45**, 653–661.
- Dumas B, Sailland A, Cheviet JP, Freyssinet G, Pallett K (1993) Identification of barley oxalate oxidase as a germin-like protein. *Comptes rendus de l'Académie des sciences. Série III, Sciences de la vie*, **316**, 793–8.

- Earl D, VonHoldt B (2011) STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, **4**, 359–361.
- Ebrahimi A, Maury P, Berger M *et al.* (2008) QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes. *Genome / National Research Council Canada = Génome / Conseil national de recherches Canada*, **51**, 599–615.
- Ehrenbolger F, Beracochea V, Peluffo L *et al.* (2012) Sunflower germin-like proteins: evolution, gene structure and functional characterization. In: *Proceedings 18 International Sunflower Conference*
- Ehrenbolger F, Fass M, Peluffo L *et al.* (2013) Análisis de los perfiles de expresión génica de girasol el respuesta al patógeno *Sclerotinia sclerotiorum*. In: *Proceeding of the VIII Encuentro Latinoamericano y del Caribe de Biotecnología REDBIO*
- Ehrenreich IM, Hanzawa Y, Chou L *et al.* (2009) Candidate gene association mapping of *Arabidopsis* flowering time. *Genetics*, **183**, 325–35.
- Elmer I, Humira S, François B (2015) Association mapping of QTLs for sclerotinia stem Rot resistance in a collection of soybean plant introductions using a genotyping by sequencing (GBS) approach. *BMC plant biology*, **15**, 5.
- Emanuelli F, Lorenzi S, Grzeskowiak L *et al.* (2013) Genetic diversity and population structure assessed by SSR and SNP markers in a large germplasm collection of grape. *BMC plant biology*, **13**, 39.
- Ersoz ES, Yu J, Buckler ES (2007) Chapter 5: Applications of linkage disequilibrium and association mapping in crop plants. In: *Genomics Approaches and Platforms*, pp. 97–119. Springer Netherlands.
- Escande AR, Laich FS, Pedraza M V. (2002) Field testing of honeybee-dispersed *Trichoderma* spp. to manage sunflower head rot (*Sclerotinia sclerotiorum*). *Plant Pathology*, **51**, 346–351.
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology*, **14**, 2611–2620.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of Molecular Variance Inferred From Metric Distances Among DNA Haplotypes: Application. *Genetics*, **131**, 479–491.
- Falconer DS, Mackay TFC (1996) Introduction to Quantitative Genetics. *Longman*, **19**, 4th edn.
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics*, **164**, 1567–1587.
- Famoso AN, Zhao K, Clark RT *et al.* (2011) Genetic architecture of aluminum tolerance in rice (*Oryza sativa*) determined through genome-wide association analysis and QTL mapping. *PLoS genetics*, **7**, e1002221.
- Fan, Chee, Gunderson (2006) Highly parallel genomic assays. *Nature Reviews Genetics*, **7**, 632–644.
- Fan J, Li R (2001) Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, **96**, 1348–1360.

- Fan B, Onteru SK, Du Z-Q *et al.* (2011) Genome-wide association study identifies Loci for body composition and structural soundness traits in pigs. *PLoS one*, **6**, e14726.
- Farmer EE, Ryan CA (1992) Octadecanoid Precursors of Jasmonic Acid Activate the Synthesis of Wound-Inducible Proteinase Inhibitors. *The Plant cell*, **4**, 129–134.
- Farris JS, Albert VA, Kallersjo M, Lipscomb D, Kluge AG (1996) Parsimony jackknifing outperforms neighbor-joining. *Cladistics*, **12**, 99–124.
- Felsenstein J (2005) PHYLIP (Phylogeny Inference Package) v 3.68.
- Fernández P, Paniego N, Lew S, Hopp HE, Heinz RA (2003) Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. *BMC Genomics*, **4**, 40.
- Fernandez P, Di Rienzo J, Soria M *et al.* (2012a) Development and validation of a high density sunflower microarray for functional studies on biotic and abiotic stresses. In: *Proceeding of the 18 International Sunflower Conference*. Mar del Plata, Argentina.
- Fernandez P, Soria M, Blesa D *et al.* (2012b) Development, characterization and experimental validation of a cultivated sunflower (*Helianthus annuus* L.) gene expression oligonucleotide microarray. *PLoS one*, **7**, e45899.
- Fernando RL, Garrick D (2013) Bayesian methods applied to GWAS. *Methods in molecular biology (Clifton, N.J.)*, **1019**, 237–74.
- Filippi CV, Aguirre N, Rivas JG *et al.* (2015) Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biology*, **15**, 52.
- Filippi C, Zubryzcki J, Aguirre N *et al.* (2013) Caracterización de un microarreglo de SNPs para estudios genómicos en girasol utilizando la plataforma BeadXpress. In: *Proceeding of the VIII Encuentro Latinoamericano y del Caribe de Biotecnología REDBIOVIII Encuentro Latinoamericano y del Caribe de Biotecnología REDBIO*. Mar del Plata, Argentina.
- Flint-Garcia S a, Thornsberry JM, Buckler ES (2003) Structure of linkage disequilibrium in plants. *Annual review of plant biology*, **54**, 357–374.
- Flint-Garcia SA, Thuillet A-C, Yu J *et al.* (2005) Maize association population: a high-resolution platform for quantitative trait locus dissection. *The Plant journal : for cell and molecular biology*, **44**, 1054–1064.
- Flores Berríos E, Gentzbittel L, Kayyal H, Alibert G, Sarrafi A (2000) AFLP mapping of QTLs for in vitro organogenesis traits using recombinant inbred lines in sunflower (*Helianthus annuus* L.). *TAG Theoretical and Applied Genetics*, **101**, 1299–1306.
- Frank Ildiko E, Friedman JH (2012) A Statistical View of Some Chemometrics Regression Tools. *Technometrics*, **35**, 109-135.
- Fryer MJ (1992) The antioxidant effects of thylakoid Vitamin E (alpha-tocopherol). *Plant, Cell and Environment*, **15**, 381–392.
- Furbank RT, Tester M (2011) Phenomics--technologies to relieve the phenotyping bottleneck. *Trends in plant science*, **16**, 635–644.

- Fusari CM, Lia V V, Hopp HE, Heinz RA, Paniego NB (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.
- Fusari CM, Lia V V, Nishinakamasu V *et al.* (2010) Single nucleotide polymorphism genotyping by heteroduplex analysis in sunflower (*Helianthus annuus* L.). *Molecular Breeding*, **28**, 73–89.
- Fusari CM, Di Rienzo JA, Troglia C *et al.* (2012) Association mapping in sunflower for sclerotinia head rot resistance. *BMC plant biology*, **12**, 93.
- Fusari CM (2010) Mapeo por asociación en girasol: diversidad nucleotídica, desequilibrio de ligamiento e identificación de genes involucrados en la resistencia a la podredumbre húmeda del capítulo. Tesis doctoral. Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires. Gadaleta A, Giancaspro A, Zacheo S *et al.* (2011) Comparison of genomic and EST-derived SSR markers in phylogenetic analysis of wheat. *Plant Genetic Resources*, **9**, 243–246.
- Garris AJ, McCouch SR, Kresovich S (2003) Population Structure and Its Effect on Haplotype Diversity and Linkage Disequilibrium Surrounding the xa5 Locus of Rice (*Oryza sativa* L.). *Genetics*, **165**, 759–769.
- Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schüler K (2004) Assessing genetic potential in germplasm collections of crop plants by marker-trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type. *Molecular Breeding*, **13**, 93–102.
- Gedil MA, Wye C, Berry S *et al.* (2001) An integrated restriction fragment length polymorphism - amplified fragment length polymorphism linkage map for cultivated sunflower. *Genome*, **44**, 213–221.
- Geisler M, Murphy AS (2006) The ABC of auxin transport: the role of p-glycoproteins in plant development. *FEBS letters*, **580**, 1094–102.
- Gentzbittel L, Mestries E, Mouzeyar S *et al.* (1999) A composite map of expressed sequences and phenotypic traits of the sunflower (*Helianthus annuus* L.) genome. *TAG Theoretical and Applied Genetics*, **99**, 218–234.
- Gentzbittel L, Mouzeyar S, Badaoui S *et al.* (1998) Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L. *TAG Theoretical and Applied Genetics*, **96**, 519–525.
- Gentzbittel L, Vear F, Zhang YX, Bervill A, Nicolas P (1995) Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics*, **90**, 1079–1086.
- Gianola D, van Kaam JBCHM (2008) Reproducing kernel hilbert spaces regression methods for genomic assisted prediction of quantitative traits. *Genetics*, **178**, 2289–303.
- Giordani T, Buti M, Natali L *et al.* (2011) An analysis of sequence variability in eight genes putatively involved in drought response in sunflower (*Helianthus annuus* L.). *TAG. Theoretical and applied genetics*, **122**, 1039–49.
- Giordano M, Oefner PJ, Underhill PA *et al.* (1999) Identification by denaturing high-performance liquid chromatography of numerous polymorphisms in a candidate region for multiple sclerosis susceptibility. *Genomics*, **56**, 247–253.

- Glazebrook J (2005) Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annual review of phytopathology*, **43**, 205–27.
- Godoy G, Steadman JR, Dickman MB, Dam R (1990) Use of mutants to demonstrate the role of oxalic acid in pathogenicity of *Sclerotinia sclerotiorum* on *Phaseolus vulgaris*. *Physiological and Molecular Plant Pathology*, **37**, 179–191.
- Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. *Cladistics*, **24**, 774–786.
- González-Martínez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB (2007) Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics*, **175**, 399–409.
- González-Pérez S, van Koningsveld G a, Vereijken JM *et al.* (2005) Emulsion properties of sunflower (*Helianthus annuus*) proteins. *Journal of agricultural and food chemistry*, **53**, 2261–2267.
- Grattapaglia D, Silva-Junior OB, Kirst M *et al.* (2011) High-throughput SNP genotyping in the highly heterozygous genome of *Eucalyptus*: assay success, polymorphism and transferability across species. *BMC plant biology*, **11**, 65.
- Groth J V., Ozmon EA, Busch RH (1999) Repeatability and Relationship of Incidence and Severity Measures of Scab of Wheat Caused by *Fusarium graminearum* in Inoculated Nurseries. *Plant Disease*, **83**, 1033–1038.
- Gulya TJ (1989) *Sclerotinia* Head Rot of Sunflower in North Dakota: 1986 Incidence, Effect on Yield and Oil Components, and Sources of Resistance. *Plant Disease*, **73**, 504.
- Gulya T, Rashid KY, Marisevic SM (1997) Sunflower technology and Production. In: *Sunflower diseases* (ed Berglund DR), pp. 263–379. American Society of Agronomy.
- Gulya, T. J., & Masirevic S (1995) Proposed methodologies for inoculation of sunflower with *Puccinia helianthi* and for disease assessment. Studies on Common Methodologies of Artificial Inoculation and Population Dynamics of Sunflower Pathogens. In: *The FAO European Research Network on Sunflower*, pp. 3–47. Rome, Italy.
- Guo X, Stotz HU (2007) Defense against *Sclerotinia sclerotiorum* in *Arabidopsis* is dependent on jasmonic acid, salicylic acid, and ethylene signaling. *Molecular plant-microbe interactions : MPMI*, **20**, 1384–1395.
- Gupta PK, Rustgi S, Kulwal PL (2005) Linkage disequilibrium and association studies in higher plants: present status and future prospects. *Plant molecular biology*, **57**, 461–485.
- Gupta PK, Rustgi S, Mir RR (2008) Array-based high-throughput DNA markers for crop improvement. *Heredity*, **101**, 5–18.
- Haddadi P, Ebrahimi a., Langlade NB *et al.* (2011) Genetic dissection of tocopherol and phytosterol in recombinant inbred lines of sunflower through quantitative trait locus analysis and the candidate gene approach. *Molecular Breeding*, **29**, 717–729.
- Haddadi P, Yazdi-samadi B, Langlade NB *et al.* (2010) Genetic control of protein, oil and fatty acids content under partial drought stress and late sowing conditions in sunflower (*Helianthus annuus*). *African Journal of Biotechnology*, **9**, 6768–6782.

- Hahn V (2002) Genetic variation for resistance to Sclerotinia head rot in sunflower inbred lines. *Field Crops Research*, **77**, 153–159.
- Haldane JB, Waddington CH (1931) Inbreeding and Linkage. *Genetics*, **16**, 357–374.
- Hall A, Feoli C, Ingaramo J, Balzarini M (2010) Proyecto brechas ASAGIR: variación interzonal e interanual de las diferencias entre rendimientos de girasol alcanzables y logrados para el período 1999-2007. In: *5to. Congreso Argentino de Girasol: con la mirada en la competitividad*, p. 364. ASAGIR, Buenos Aires.
- Hamblin MT (2004) Comparative Population Genetics of the Panicoid Grasses: Sequence Polymorphism, Linkage Disequilibrium and Selection in a Diverse Sample of Sorghum bicolor. *Genetics*, **167**, 471–483.
- Hamblin MT, Warburton ML, Buckler ES (2007) Empirical comparison of Simple Sequence Repeats and single nucleotide polymorphisms in assessment of maize diversity and relatedness. *PLoS one*, **2**, e1367.
- Han W, Yip SP, Wang J, Yap MKH (2004) Using denaturing HPLC for SNP discovery and genotyping, and establishing the linkage disequilibrium pattern for the all-trans-retinol dehydrogenase (RDH8) gene. *Journal of human genetics*, **49**, 16–23.
- Hardy OJ, Vekemans X (2002) spagedi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. *Molecular Ecology Notes*, **2**, 618–620.
- Harlan JR (1987) *Plant Genetic Resources: A Conservation Imperative* (Yeatman, Kafton, Wilkes, Eds). Westview, Boulder, CO.
- Harter A V, Gardner K a, Falush D *et al.* (2004) Origin of extant domesticated sunflowers in eastern North America. *Nature*, **430**, 201–205.
- Hartman GL, Wang TC (1993) Resistance in Lycopersicon species to black leaf mold caused by Pseudocercospora Fuligena. *Euphytica*, **71**, 125–130.
- Haseneyer G, Stracke S, Piepho H-P *et al.* (2010) DNA polymorphisms and haplotype patterns of transcription factors involved in barley endosperm development are associated with key agronomic traits. *BMC plant biology*, **10**, 5.
- Hayes B (2013) *Genome-Wide Association Studies and Genomic Prediction* (C Gondro, J van der Werf, B Hayes, Eds). Humana Press, Totowa, NJ.
- Headrick JM, Pataky JK (1988) Spatial and temporal development of common rust in susceptible and partially resistant sweet corn hybrids. *Phytopathology*, **78**, 227–233.
- Heath MC (2000) Nonhost resistance and nonspecific plant defenses. *Current Opinion in Plant Biology*, **3**, 315–319.
- Heesacker A, Kishore VK, Gao W *et al.* (2008) SSRs and INDELS mined from the sunflower EST database: abundance, polymorphisms, and cross-taxa utility. *TAG. Theoretical and applied genetics*, **117**, 1021–1029.
- Hegedus DD, Rimmer SR (2005) Sclerotinia sclerotiorum: when “to be or not to be” a pathogen? *FEMS microbiology letters*, **251**, 177–184.
- Heiser, Schilling (1981) Infrageneric Classification of Helianthus (Compositae). *Taxon*, **30**, 393–403.

- Heiser C, Smith D, Clevenger S, Martin WC (1969) The North American sunflowers (*Helianthus*). Torrey Bot Club Mem.
- Hernández-Blanco C, Feng DX, Hu J *et al.* (2007) Impairment of cellulose synthases required for Arabidopsis secondary cell wall formation enhances disease resistance. *The Plant cell*, **19**, 890–903.
- Heywood VH (1993) *Flowering plants of the world*. Mayflower Books, New York.
- Hoggart CJ, Whittaker JC, De Iorio M, Balding DJ (2008) Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. (PM Visscher, Ed). *PLoS genetics*, **4**, e1000130.
- Holländer-Czytko H, Grabowski J, Sandorf I, Weckermann K, Weiler EW (2005) Tocopherol content and activities of tyrosine aminotransferase and cystine lyase in Arabidopsis under stress conditions. *Journal of plant physiology*, **162**, 767–770.
- Holleley CE & Geerts PG (2009) Multiplex Manager 1.0: a crossplatform computer program that plans and optimizes multiplex PCR. *BioTechniques*, **46**, 511–517
- Hu X, Bidney DL, Yalpani N *et al.* (2003) Overexpression of a gene encoding hydrogen peroxide-generating oxalate oxidase evokes defense responses in sunflower. *Plant Physiology*, **133**, 170–181.
- Hu J, Vick BA (2003) Target Region Amplification Polymorphism: A Novel Marker Technique for Plant Genotyping. *Plant Molecular Biology Reporter*, **21**, 289–294.
- Hu J, Wang L, Li J (2011) Comparison of genomic SSR and EST-SSR markers for estimating genetic diversity in cucumber. *Biologia Plantarum*, **55**, 577–580.
- Hulbert S, Bai J, Fellers J (2007) Gene expression patterns in near isogenic lines for wheat rust resistance gene Lr34/Yr18. *Am Phytopath Society*, **97**, 1083–1093.
- Hyten DL, Song Q, Choi I-Y *et al.* (2008) High-throughput genotyping with the GoldenGate assay in the complex genome of soybean. *TAG. Theoretical and applied genetics*, **116**, 945–952.
- Van Inghelandt D, Melchinger AE, Lebreton C, Stich B (2010) Population structure and genetic diversity in a commercial maize breeding program assessed with SSR and SNP markers. *TAG. Theoretical and applied genetics*, **120**, 1289–1299.
- Ingvarsson PK (2008) Multilocus patterns of nucleotide polymorphism and the demographic history of *Populus tremula*. *Genetics*, **180**, 329–340.
- Ivandi V, Thomas WTB, Nevo E, Zhang Z, Forster BP (2003) Associations of simple sequence repeats with quantitative trait variation including biotic and abiotic stress tolerance in *Hordeum spontaneum*. *Plant Breeding*, **122**, 300–304.
- Iwata H, Uga Y, Yoshioka Y, Ebana K, Hayashi T (2007) Bayesian association mapping of multiple quantitative trait loci and its application to the analysis of genetic variation among *Oryza sativa* L. germplasms. *TAG. Theoretical and applied genetics*, **114**, 1437–1449.
- James W., Teng PS (1979) *The quantification of production constraints associated with plant diseases*. CIMMYT.
- Jan CC, Seiler. GJ (2006) *Genetic Resources, Chromosome Engineering, and Crop Improvement* (R Singh, Ed). CRC Press.

- Jan CC, Vick BA, Miller JF, Kahler AL, Butler ET (1998) Construction of an RFLP linkage map for cultivated sunflower. *TAG Theoretical and Applied Genetics*, **96**, 15–22.
- Jeger MJ, Jones DG, Griffiths E (1983) Disease spread of non-specialised fungal pathogens from inoculated point sources in intraspecific mixed stands of cereal cultivars. *Annals of Applied Biology*, **102**, 237–244.
- Jombart T (2008) Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics (Oxford, England)*, **24**, 1403–1405.
- Jombart T, Devillard S, Balloux F (2010) Discriminant analysis of principal components: a new method for the analysis of genetically structured populations. *BMC genetics*, **11**, 94.
- Jones ES, Sullivan H, Bhatramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). *TAG. Theoretical and applied genetics*, **115**, 361–371.
- Kabbage M, Yarden O, Dickman MB (2015) Pathogenic attributes of *Sclerotinia sclerotiorum*: Switching from a biotrophic to necrotrophic lifestyle. *Plant Science*, **233**, 53–60.
- Kane NC, Gill N, King MG *et al.* (2011) Progress towards a reference genome for sunflower. *Botany*, **437**, 429–437.
- Kang MS (1998) Using genotype-by-environment interaction for crop cultivar development. *Advances in Agronomy*, **62**, 199–252.
- Kang HM, Zaitlen N a, Wade CM *et al.* (2008) Efficient control of population structure in model organism association mapping. *Genetics*, **178**, 1709–1723.
- Katoh K, Standley DM (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**, 772–780.
- Kazan K, Manners JM (2009) Linking development to defense: auxin in plant-pathogen interactions. *Trends in plant science*, **14**, 373–382.
- Kiani SP, Grieu P, Maury P *et al.* (2007) Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.). *TAG. Theoretical and applied genetics*, **114**, 193–207.
- Kim KS, Min J, Dickman MB (2008) Oxalic Acid Is an Elicitor of Plant Programmed Cell Death during., **21**, 605–612.
- Kolkman JM, Berry ST, Leon AJ *et al.* (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. *Genetics*, **177**, 457–468.
- Kosambi DD (1944) The estimation of map distances from recombination values. *Ann Eugen*, **12**, 172–175.
- Kota R, Varshney RK, Prasad M *et al.* (2008) EST-derived single nucleotide polymorphism markers for assembling genetic and physical maps of the barley genome. *Functional & integrative genomics*, **8**, 223–233.
- Kota R, Varshney RK, Thiel T, Dehmer KJ, Graner A (2004) Generation and Comparison of EST-Derived SSRs and SNPs in Barley (*Hordeum Vulgare* L.). *Hereditas*, **135**, 145–151.

- Kraakman ATW, Martínez F, Mussiraliev B, Eeuwijk FA, Niks RE (2006) Linkage Disequilibrium Mapping of Morphological, Resistance, and Other Agronomically Relevant Traits in Modern Spring Barley Cultivars. *Molecular Breeding*, **17**, 41–58.
- Kranz J (1988) Measuring Plant Disease. In: *Experimental Techniques in Plant Disease Epidemiology*, pp. 35–50. Springer Berlin Heidelberg.
- Kranz J (1990) Tansley Review No. 28 Fungal diseases in multispecies plant communities. *New Phytologist*, **116**, 383–405.
- Krill AM, Kirst M, Kochian L V, Buckler ES, Hoekenga OA (2010) Association and linkage analysis of aluminum tolerance genes in maize. *PLoS one*, **5**, e9958.
- Kruglyak L (2008) The road to genome-wide association studies. *Nature reviews. Genetics*, **9**, 314–318.
- Kulwal P, Ishikawa G, Benscher D *et al.* (2012) Association mapping for pre-harvest sprouting resistance in white winter wheat. *TAG. Theoretical and applied genetics*, **125**, 793–805.
- Kumar V, Singh A, Mithra SVA *et al.* (2015) Genome-wide association mapping of salinity tolerance in rice (*Oryza sativa*). *DNA research : an international journal for rapid publication of reports on genes and genomes*, dsu046.
- Lai Z, Livingstone K, Zou Y *et al.* (2005) Identification and mapping of SNPs from ESTs in sunflower. *TAG. Theoretical and applied genetics*, **111**, 1532–1544.
- Lander ES, Green P, Abrahamson J *et al.* (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. *Genomics*, **1**, 174–181.
- Langar K, Lorieux M, Desmarais E *et al.* (2003) Combined mapping of DALP and AFLP markers in cultivated sunflower using F9 recombinant inbred lines. *TAG. Theoretical and applied genetics*, **106**, 1068–1074.
- Laudert D, Weiler EW (1998) Allene oxide synthase: a major control point in Arabidopsis thaliana octadecanoid signalling. *The Plant Journal*, **15**, 675–684.
- Leclercq P (1966) Une stérilité mâle utilisable pour la production d'hybrides simples de tournesol. *Annales de l'Amélioration des Plantes*, **16**, 135–144.
- Leclercq P (1969) Une stérilité mâle cytoplasmique chez le tournesol. *Annales de l'Amélioration des Plantes*, **19**, 99–106.
- Legendre, P., and Legendre L (1998) *Numerical Ecology*. Elsevier, Amsterdam.
- Lepoittevin C, Frigerio J-M, Garnier-Géré P *et al.* (2010) In vitro vs in silico detected SNPs for the development of a genotyping array: what can we learn from a non-model species? *PLoS one*, **5**, e11034.
- Lexer C, Lai Z, Rieseberg LH (2004) Candidate gene polymorphisms associated with salt tolerance in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. *The New phytologist*, **161**, 225–233.
- Li Y-H, Li W, Zhang C *et al.* (2010) Genetic diversity in domesticated soybean (*Glycine max*) and its wild progenitor (*Glycine soja*) for simple sequence repeat and single-nucleotide polymorphism loci. *The New phytologist*, **188**, 242–253.

- Li L, Paulo M-J, Strahwald J *et al.* (2008) Natural DNA variation at candidate loci is associated with potato chip color, tuber starch content, yield and starch yield. *TAG. Theoretical and applied genetics*, **116**, 1167–1181.
- Lia V V, Poggio L, Confalonieri VA (2009) Microsatellite variation in maize landraces from Northwestern Argentina: genetic diversity, population structure and racial affiliations. *TAG. Theoretical and applied genetics*, **119**, 1053–1067.
- Liang X, Liberti D, Li M *et al.* (2014) Oxaloacetate acetylhydrolase gene mutants of *Sclerotinia sclerotiorum* do not accumulate oxalic acid, but do produce limited lesions on host plants. *Molecular plant pathology*.
- Libault M, Wan J (2007) Identification of 118 *Arabidopsis* transcription factor and 30 ubiquitin-ligase genes responding to chitin, a plant-defense elicitor. *Molecular plant-microbe interactions*, **20**, 900–911.
- Liu A, Burke JM (2006) Patterns of nucleotide diversity in wild and cultivated sunflower. *Genetics*, **173**, 321–330.
- Liu K, Muse S V (2005) PowerMarker: an integrated analysis environment for genetic marker analysis. *Bioinformatics (Oxford, England)*, **21**, 2128–2129.
- Llorente F, Muskett P, Sánchez-Vallet A *et al.* (2008) Repression of the auxin response pathway increases *Arabidopsis* susceptibility to necrotrophic fungi. *Molecular plant*, **1**, 496–509.
- Logsdon BA, Hoffman GE, Mezey JG (2010) A variational Bayes algorithm for fast and accurate multiple locus genome-wide association analysis. *BMC bioinformatics*, **11**, 58.
- Lopukhina A, Dettenberg M (2001) Cloning and characterization of a coronatine-regulated tyrosine aminotransferase from *Arabidopsis*. *Plant physiology*, **126**, 1678–1687.
- De Los Campos G, Pérez P, Vazquez AI, Crossa J (2013) Genome-enabled prediction using the BLR (Bayesian Linear Regression) R-package. *Methods in molecular biology (Clifton, N.J.)*, **1019**, 299–320.
- Lovell D, Hunter T, Powers S (2004) Effect of temperature on latent period of septoria leaf blotch on winter wheat under outdoor conditions. *Plant Pathology*, **53**, 170–181.
- Lu Y, Yan J, Guimarães CT *et al.* (2009) Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms. *TAG. Theoretical and applied genetics*, **120**, 93–115.
- Luke HH, Berger RD (1982) Slow Rusting in Oats Compared With the Logistic and Gompertz Models. *Phytopathology*, **72**, 400–402.
- Ma H, Singh RP, Mujeeb-Kazi A (1995) Resistance to stripe rust in *Triticum turgidum*, *T. tauschii* and their synthetic hexaploids. *Euphytica*, **82**, 117–124.
- Mackay TF (2001) The genetic architecture of quantitative traits. *Annual review of genetics*, **35**, 303–39.
- Madden L V., Campbell CL (1990) Nonlinear Disease Progress Curves. *Epidemics of Plant Diseases*, **13**, 181–229.
- Madden L V., Hughes G (1995) Plant disease incidence: distributions, heterogeneity, and temporal analysis. *Annual Review of Phytopathology*, **33**, 529–564.

- Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA (2007) A mixed-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato. *Genetics*, **175**, 879–889.
- Mandel JR, Dechaine JM, Marek LF, Burke JM (2011) Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *TAG. Theoretical and applied genetics*, **123**, 693–704.
- Mandel JR, Nambeesan S, Bowers JE *et al.* (2013) Association Mapping and the Genomic Consequences of Selection in Sunflower (NM Springer, Ed). *PLoS Genetics*, **9**, e1003378.
- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in genetics : TIG*, **24**, 133–141.
- Maringolo C (2007) Regiones Cromosómicas Asociadas a Resistencia a Podredumbre Húmeda del Capítulo de Girasol (*Sclerotinia sclerotiorum* (Lib.) de Bary). Posgrado en Producción Vegetal. Universidad Nacional de Mar del Plata. Facultad de Ciencias Agrarias, Unidad Integrada Balcarce Argentina.
- McCarthy MI, Abecasis GR, Cardon LR *et al.* (2008) Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nature reviews. Genetics*, **9**, 356–69.
- McCouch SR, McNally KL, Wang W, Sackville Hamilton R (2012) Genomics of gene banks: A case study in rice. *American journal of botany*, **99**, 407–23.
- McCouch SR, Zhao K, Wright M *et al.* (2010) Development of genome-wide SNP assays for rice. *Breeding Science*, **60**, 524–535.
- McMullen MD, Kresovich S, Villeda HS *et al.* (2009) Genetic properties of the maize nested association mapping population. *Science (New York, N.Y.)*, **325**, 737–740.
- Mestries E, Gentzbittel L, Labrouhe DT De *et al.* (1998) Analyses of quantitative trait loci associated with resistance to *Sclerotinia sclerotiorum* in sunflowers (*Helianthus annuus* L.) using molecular markers. *Molecular Breeding*, **4**, 215–226.
- Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, **157**, 1819–1829.
- Mezaka I, Legzdina L, Waugh R, Close TJ, Rostoks N (2013) Genetic Diversity in Latvian Spring Barley Association Mapping Population. In: *Advance in Barley Sciences*, pp. 25–35. Springer Netherlands, Dordrecht.
- Micic Z, Hahn V, Bauer E *et al.* (2004) QTL mapping of *Sclerotinia* midstalk-rot resistance in sunflower. *Tag Theoretical And Applied Genetics*, **109**, 1474–1484.
- Micic Z, Hahn V, Bauer E *et al.* (2005a) Identification and validation of QTL for *Sclerotinia* midstalk rot resistance in sunflower by selective genotyping. *Tag Theoretical And Applied Genetics*, **111**, 233–242.
- Micic Z, Hahn V, Bauer E, Schön CC, Melchinger AE (2005b) QTL mapping of resistance to *Sclerotinia* midstalk rot in RIL of sunflower population NDBLOSsel x CM625. *TAG. Theoretical and applied genetics*, **110**, 1490–1498.
- Miller, Fick (1997) The genetics of sunflower, in *Sunflower Technology and Production*, Agronomy Series. ASA, CSSA and SSSA, Madison, WI, USA, 441–496.

- Mir RR, Kumar N, Jaiswal V *et al.* (2012) Genetic dissection of grain weight in bread wheat through quantitative trait locus interval and association mapping. *Molecular Breeding*, **29**, 963–972.
- Missio RF, Caixeta ET, Zambolim EM *et al.* (2011) Genetic characterization of an elite coffee germplasm assessed by gSSR and EST-SSR markers. *Genetics and molecular research: GMR*, **10**, 2366–2381.
- Mockaitis K, Estelle M (2008) Auxin receptors and plant development: a new signaling paradigm. *Annual review of cell and developmental biology*, **24**, 55–80.
- Moffatt BA, Weretilnyk EA (2001) Sustaining S-adenosyl-L-methionine-dependent methyltransferase activity in plant cells. *Physiologia Plantarum*, **113**, 435–442.
- Molinero-Ruiz ML, Cerdón-Torres MM, Martínez-Aguilar J, Melero-Vara JM, Domínguez J (2008) Resistance to metalaxyl and to metalaxyl-M in populations of *Plasmopara halstedii* causing downy mildew in sunflower. *Canadian Journal of Plant Pathology*, **30**, 97–105.
- Monna L, Ohta R, Masuda H, Koike A, Minobe Y (2006) Genome-wide searching of single-nucleotide polymorphisms among eight distantly and closely related rice cultivars (*Oryza sativa* L.) and a wild accession (*Oryza rufipogon* Griff.). *DNA research: an international journal for rapid publication of reports on genes and genomes*, **13**, 43–51.
- Morariu VI, Srinivasan B V, Raykar VC, Duraiswami R, Davis LS (2008) Automatic online tuning for fast Gaussian summation. *Advances in Neural Information Processing Systems (NIPS)*.
- Moreno MV (2010) Diversidad genética en girasol cultivado: análisis de una colección de germoplasma local para su aplicación en programas de mejoramiento. Tesis Doctoral. Facultad de Ciencias Exactas y Naturales - Universidad de Buenos Aires.
- Moreno M V, Nishinakamasu V, Loray M *et al.* (2013) Genetic characterization of sunflower breeding resources from Argentina: assessing diversity in key open-pollinated and composite populations. *Plant Genetic Resources*, **11**, 238–249.
- Multani DS, Briggs SP, Chamberlin MA *et al.* (2003) Loss of an MDR transporter in compact stalks of maize br2 and sorghum dw3 mutants. *Science (New York, N.Y.)*, **302**, 81–84.
- Munné-Bosch S, Alegre L (2010) The Function of Tocopherols and Tocotrienols in Plants. *Critical Reviews in Plant Sciences*, **21**, 31-57.
- Myles S, Peiffer J, Brown PJ *et al.* (2009) Association mapping: critical considerations shift from genotyping to experimental design. *The Plant cell*, **21**, 2194–2202.
- Nandha P, Singh J (2014) Comparative assessment of genetic diversity between wild and cultivated barley using gSSR and EST-SSR markers. *Plant Breeding.*, **133**, 28–35.
- National Sunflower Association (2013) <http://www.sunflowernsa.com/>
- Navarro L, Bari R, Achard P *et al.* (2008) DELLAs control plant immune responses by modulating the balance of jasmonic acid and salicylic acid signaling. *Current biology*, **18**, 650–655.
- Nei M (1978) Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, **89**, 583–590.
- Neumann K, Kobiljski B, Denčić S, Varshney RK, Börner A (2010) Genome-wide association mapping: a case study in bread wheat (*Triticum aestivum* L.). *Molecular Breeding*, **27**, 37–58.

- Noh B, Murphy AS, Spalding EP (2001) Multidrug Resistance-like Genes of Arabidopsis Required for Auxin Transport and Auxin-Mediated Development. *PLANT CELL*, **13**, 2441–2454.
- Nordborg M, Tavaré S (2002) Linkage disequilibrium: what history has to tell us. *Trends in Genetics*, **18**, 83–90.
- Nordborg M, Weigel D (2008) Next-generation genetics in plants. *Nature*, **456**, 720–723.
- Nutter F., Gleason M., Jenco J., Christians N. (1993) Assessing the accuracy, intra-rater repeatability, and inter-rater reliability of disease assessment systems. *Phytopathology*, **83**, 806–812.
- Oraguzie, Oraguzie NC, Rikkerink EHA, Gardiner SE, Silva HN (2007) *Association mapping in plants*. Springer-Verlag GmbH.
- Pajerowska-Mukhtar K, Stich B, Achenbach U *et al.* (2009) Single nucleotide polymorphisms in the allene oxide synthase 2 gene are associated with field resistance to late blight in populations of tetraploid potato cultivars. *Genetics*, **181**, 1115–1127.
- Palaisa K, Morgante M, Tingey S, Rafalski A (2004) Long-range patterns of diversity and linkage disequilibrium surrounding the maize Y1 gene are indicative of an asymmetric selective sweep. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 9885–9890.
- Paniego N, Echaide M, Muñoz M *et al.* (2002) Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). *Genome*, **43**, 34–43.
- Paniego N, Fusari C, Lia V, Puebla A (2015) SNP genotyping by heteroduplex analysis. In: *Methods in molecular biology*, pp. 141–150.
- Paniego N, Heinz R, Fernandez P *et al.* (2007) Sunflower. In: *Oilseed. Genome Mapping and Molecular Breeding in Plants*, pp. 153–177.
- Parker J, Navia O (1991) Estrategias de control químico del tizón tardío (*Phytophthora infestans*). *Informe Anual 1990-91 del Proyecto de Investigación de la Papa PROINPA (IBTA-CIP COTESU)*, 45–74.
- Parlevliet JE (1976) Partial resistance of barley to leaf rust, *Puccinia hordei*. III. The inheritance of the host plant effect on latent period in four cultivars. *Euphytica*, **25**, 241–248.
- Parlevliet JE, Van Ommeren A (1975) Partial resistance of barley to leaf rust, *Puccinia hordei*. II. Relationship between field trials, micro plot tests and latent period. *Euphytica*, **24**, 293–303.
- Patnala R, Clements J, Batra J (2013) Candidate gene association studies: a comprehensive guide to useful in silico tools. *BMC genetics*, **14**, 39.
- Pavy N, Pelgas B, Beauseigle S *et al.* (2008) Enhancing genetic mapping of complex genomes through the design of highly-multiplexed SNP arrays: application to the large and unsequenced genomes of white spruce and black spruce. *BMC genomics*, **9**, 21.
- Peakall R, Smouse PE (2006) Genalex 6: genetic analysis in Excel. Population genetic software for teaching and research. *Molecular Ecology Notes*, **6**, 288–295.
- Peerbolte RP, Paleman J (1998) The CARTISOL sunflower RFLP map (146 loci) extended with 291 RFLP markers. *Proceedings of the 18th sunflower research forum, Fargo*, 11–12.

- Pegadaraju V, Nipper R, Hulke B, Qi L, Schultz Q (2013) De novo sequencing of sunflower genome for SNP discovery using RAD (Restriction site Associated DNA) approach. *BMC genomics*, **14**, 556.
- Peluffo L, Lía V, Troglia C, Maringolo C, Paniego N *et al.* (2010) Metabolic profiles of sunflower genotypes with contrasting response to *Sclerotinia sclerotiorum* infection, *Phytochemistry*, **71**, 70–80
- Pennypacker SP (1980) A Flexible Model for Studying Plant Disease Progression. *Phytopathology*, **70**, 232.
- Perchepped L, Balagué C, Riou C *et al.* (2010) Nitric oxide participates in the complex interplay of defense-related signaling pathways controlling disease resistance to *Sclerotinia sclerotiorum* in *Arabidopsis thaliana*. *Molecular plant-microbe interactions : MPMI*, **23**, 846–860.
- Pereyra V (2002) Taller de Calidad de Semillas de Girasol. In: *Proceedings of the 23 reunión Convenio INTA-ASAGIR*.
- Pereyra, V.R., Escande AR (1994) *Manual de reconocimiento de enfermedades del Girasol en la Argentina* (CERBAS, Ed). Balcarce, Provincia de Buenos Aires, Argentina.
- Pérez P, de los Campos G, Crossa J, Gianola D (2010) Genomic-Enabled Prediction Based on Molecular Markers and Pedigree Using the Bayesian Linear Regression Package in R. *The Plant Genome Journal*, **3**, 106.
- Pérez-Bermúdez P, Moya García AA, Tuñón I, Gavidia I (2010) *Digitalis purpurea* P5 β R2, encoding steroid 5 β -reductase, is a novel defense-related gene involved in cardenolide biosynthesis. *New Phytologist*, **185**, 687–700.
- Perry J, Wang T, Welham T (2003) A TILLING reverse genetics tool and a web-accessible collection of mutants of the legume *Lotus japonicus*. *Plant Physiology*, **131**, 866–871.
- Picañol R, Eduardo I, Aranzana MJ *et al.* (2012) Combining linkage and association mapping to search for markers linked to the flat fruit character in peach. *Euphytica*, **190**, 279–288.
- Pinheiro J, Bates D, DebRoy S, Sarkar D, Team RC (2014) nlme: Linear and Nonlinear Mixed Effects Models.
- PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, **81**, 559-575.
- Poormohammad Kiani S, Maury P, Nouri L *et al.* (2009) QTL analysis of yield-related traits in sunflower under different water treatments. *Plant Breeding*, **128**, 363–373.
- Poormohammad Kiani S, Talia P, Nishinakamasu V *et al.* (2006) Development of a new AFLP and SSR genetic linkage map and identification of QTLs for photosynthesis related traits under well-watered and water-stressed conditions in sunflower (*Helianthus annuus* L.).
- Prioletta S, Pereyra V, Pedraza M, Escande A (1995) Efecto de la exposición del esclerocio a baja temperatura y de las condiciones ambientales durante la carpogénesis sobre la producción de apotecios de *Sclerotinia sclerotiorum*. In: *Proceedings of the Reunión Nacional de Oleaginosos*. Pergamino. Argentina.
- Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. *American journal of human genetics*, **69**, 1–14.

- Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. *American journal of human genetics*, **65**, 220–228.
- Pritchard JK, Stephens M, Donnelly P (2000a) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.
- Pritchard JK, Stephens M, Rosenberg N a, Donnelly P (2000b) Association mapping in structured populations. *American journal of human genetics*, **67**, 170–181.
- Pryce JE, Bolormaa S, Chamberlain AJ *et al.* (2010) A validated genome-wide association study in 2 dairy cattle breeds for milk production and fertility traits using variable length haplotypes. *Journal of dairy science*, **93**, 3331–3345.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Sham PC (2007).
- Purdy LH (1979) Sclerotinia sclerotiorum: history, diseases and symptomatology, host range, geographic distribution, and impact. *Phytopathology*, **69**, 875–880.
- Pysh LD, Wysocka-Diller JW, Camilleri C, Bouchez D, Benfey PN (1999) The GRAS gene family in Arabidopsis: sequence characterization and basic expression analysis of the SCARECROW-LIKE genes. *The Plant Journal*, **18**, 111–119.
- Quiroz F, Pereyra V, Escande A (1998) Comportamiento de cultivares de girasol frente a infecciones de Sclerotinia Sclerotiorum inducidas por pulverización de ascosporas o por inóculo natural. In: *Proceedings of the Reunión Nacional de Oleaginosos*, p. 35. Bahía Blanca.
- R Core Team (2013) R: a language and environment for statistical computing. Vienna, Austria: R foundation for statistical Computing; <http://www.R-project.org/>.
- Rafalski JA (2002) Novel genetic mapping tools in plants: SNPs and LD-based approaches. *Plant Science*, **162**, 329–333.
- Rafalski JA (2010) Association genetics in crop improvement. *Current opinion in plant biology*, **13**, 174–180.
- Randall LP, Coldham NG, Woodward MJ (2005) Detection of mutations in Salmonella enterica gyrA, gyrB, parC and parE genes by denaturing high performance liquid chromatography (DHPLC) using standard HPLC instrumentation. *The Journal of antimicrobial chemotherapy*, **56**, 619–23.
- Reif JC, Zhao Y, Würschum T, Gowda M, Hahn V (2013) Genomic prediction of sunflower hybrid performance (J León, Ed). *Plant Breeding*, **132**, 107–114.
- Di Rienzo, Casanoves, Balzarini *et al.* (2008) Infostat versión 8. Grupo InfoStat, FCA, Universidad Nacional de Córdoba, Argentina
- Rieseberg (1991) Homoploid reticulate evolution in helianthus (asteraceae): evidence from ribosomal genes. Homoploid reticulate evolution in helianthus (asteraceae): evidence from ribosomal genes. *American Journal of Botany*, **78**, 1218–1237.
- Rieseberg LH (1998) Genetic Mapping as a Tool for Studying Speciation. *Molecular Systematics of Plants II*, 459–487.

- Rieseberg LH, Choi H, Chan R, Spore C, Rieseberg (1993) Genomic map of a diploid hybrid species. *Heredity-London*, **70**, 285.
- Rieseberg LH, Seiler GJ (1990) Molecular Evidence and the Origin and Development of the Domesticated Sunflower (*Helianthus annuum*, Asteraceae). *Economic Botany*, **44**, 79–91.
- Rietz S, Bernsdorff FEM, Cai D (2012) Members of the germin-like protein family in *Brassica napus* are candidates for the initiation of an oxidative burst that impedes pathogenesis of *Sclerotinia sclerotiorum*. *Journal of experimental botany*, **63**, 5507–5519.
- Risch N, Merikangas K (1996) The Future of Genetic Studies of Complex Human Diseases. *Science*, **273**, 1516–1517.
- Robert-Seilaniantz A, Navarro L, Bari R, Jones JDG (2007) Pathological hormone imbalances. *Current opinion in plant biology*, **10**, 372–379.
- Ronicke S, Hahn V, Friedt W (2005) Resistance to *Sclerotinia sclerotiorum* of “high oleic” sunflower inbred lines. *Plant Breeding*, **124**, 376–381.
- Rönicke S, Hahn V, Vogler a, Friedt W (2005) Quantitative Trait Loci Analysis of Resistance to *Sclerotinia sclerotiorum* in Sunflower. *Phytopathology*, **95**, 834–839.
- Rosenberg NA (2003) Distruct: a Program for the Graphical Display of Population Structure. *Molecular Ecology Notes*, **4**, 137–138.
- Rosenberg NA, Burke T, Elo K *et al.* (2001) Empirical evaluation of genetic clustering methods using multilocus genotypes from 20 chicken breeds. *Genetics*, **159**, 699–713.
- Rostoks N, Ramsay L, MacKenzie K *et al.* (2006) Recent history of artificial outcrossing facilitates whole-genome association mapping in elite inbred crop varieties. *Proceedings of the National Academy of Sciences of the United States of America*, **103**, 18656–18661.
- Roussel V, Leisova L, Exbrayat F, Stehno Z, Balfourier F (2005) SSR allelic diversity changes in 480 European bread wheat varieties released from 1840 to 2000. *TAG. Theoretical and applied genetics*, **111**, 162–170.
- Royle DJ (1994) Understanding and predicting epidemics: a commentary based on selected pathosystems. *Plant Pathology*, **43**, 777–789.
- Rozas J (2009) DNA sequence polymorphism analysis using DnaSP. *Methods in molecular biology (Clifton, N.J.)*, **537**, 337–350.
- Rozen S, Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods in molecular biology (Clifton, N.J.)*, **132**, 365–386.
- Saidou A-A, Mariac C, Luong V *et al.* (2009) Association studies identify natural variation at PHYC linked to flowering time and morphological variation in pearl millet. *Genetics*, **182**, 899–910.
- Sala CA, Bulos M, Altieri E, Ramos ML (2012) Sunflower: Improving Crop Productivity and Abiotic Stress Tolerance. In: *Improving Crop Resistance to Abiotic Stress, Volume 1 & Volume 2* (eds N. Tuteja, S. S. Gill, A. F. Tiburcio and R. Tuteja), Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor laboratory press. Cold Spring Harbor.

- SAS Institute Inc (2002-2005) SAS OnlineDoc 9.1.3, Cary, NC: SAS Institute Inc. Satagopan JM, Yandell BS, Newton MA, Osborn TC (1996) A Bayesian Approach to Detect Quantitative Trait Loci Using Markov Chain Monte Carlo. *Genetics*, **144**, 805–816.
- Sato Y, Yoshimichi F, Hiro-Yuki H (2001) Mutations that cause amino acid substitutions at the invariant positions in homeodomain of OSH3 KNOX protein suggest artificial selection during rice domestication. *Genes & genetic systems*, **76**, 381–392.
- Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. *Genome research*, **20**, 1165–1173.
- Schiex T, Gaspin C (1997) CARTHAGENE: constructing and joining maximum likelihood genetic maps. *Proceedings International Conference on Intelligent Systems for Molecular Biology ISMB International Conference on Intelligent Systems for Molecular Biology*, **5**, 258–267.
- Schmidt A, Scheel D, Strack D (1998) Elicitor-stimulated biosynthesis of hydroxycinnamoyltyramines in cell suspension cultures of *Solanum tuberosum*. *Planta*, **205**, 51–55.
- Schneiter AA, Miller JF (1981) Description of Sunflower Growth Stages. *Crop Science*, **21**, 3–5.
- Seeb JE, Carvalho G, Hauser L *et al.* (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular ecology resources*, **11**, 1–8.
- Seem RC (1984) Disease Incidence and Severity Relationships. *Annual Review of Phytopathology*, **22**, 133–150.
- Segura V, Vilhjálmsson BJ, Platt A *et al.* (2012) An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature genetics*, **44**, 825–830.
- Seiler GJ, Rieseberg LH (1997) Systematics, origin and germplasm resources of the wild and domesticated sunflower. In: *Sunflower technology and production*, pp. 21–66. Amer. Soc. Agronomy, Madison, Wisconsin.
- Seo HS, Song JT, Cheong JJ *et al.* (2001) Jasmonic acid carboxyl methyltransferase: a key enzyme for jasmonate-regulated plant responses. *Proceedings of the National Academy of Sciences of the United States of America*, **98**, 4788–4793.
- Shaner G, Finney RE (1977) The effect of nitrogen fertilization on the expression of slow-mildewing resistance in knox wheat. *Phytopathology*, **67**, 1051–1056.
- Shearer BL, Zadoks JC (1972) The latent period of *Septoria nodorum* in wheat. 1. The effect of temperature and moisture treatments under controlled conditions. *Netherlands Journal of Plant Pathology*, **78**, 231–241.
- Shibuya N, Minami E (2001) Oligosaccharide signalling for defence responses in plant. *Physiological and Molecular Plant Pathology*, **59**, 223–233.
- Sillanpää MJ, Arjas E (1998) Bayesian Mapping of Multiple Quantitative Trait Loci From Incomplete Inbred Line Cross Data. *Genetics*, **148**, 1373–1388.
- Sillanpää MJ, Arjas E (1999) Bayesian Mapping of Multiple Quantitative Trait Loci From Incomplete Outbred Offspring Data. *Genetics*, **151**, 1605–1619.
- Sillanpää MJ, Bhattacharjee M (2005) Bayesian association-based fine mapping in small chromosomal segments. *Genetics*, **169**, 427–439.

- Simko I, Costanzo S, Haynes KG, Christ BJ, Jones RW (2004a) Linkage disequilibrium mapping of a *Verticillium dahliae* resistance quantitative trait locus in tetraploid potato (*Solanum tuberosum*) through a candidate gene approach. *TAG. Theoretical and applied genetics*, **108**, 217–224.
- Simko I, Haynes KG, Ewing EE *et al.* (2004b) Mapping genes for resistance to *Verticillium albo-atrum* in tetraploid and diploid potato populations using haplotype association tests and genetic linkage analysis. *Molecular genetics and genomics : MGG*, **271**, 522–531.
- Simko I, Pechenick DA, McHale LK *et al.* (2009) Association mapping and marker-assisted selection of the lettuce dieback resistance gene *Tvr1*. *BMC plant biology*, **9**, 135.
- Singh RP (1993) Resistance to Leaf Rust in 26 Mexican Wheat Cultivars. *Crop Science*, **33**, 633.
- Skøt L, Humphreys J, Humphreys MO *et al.* (2007) Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne* (L.). *Genetics*, **177**, 535–547.
- Smith BJ (2007) boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software*, **21**, 1–37.
- Smith BD, Yarnell RA (2009) Initial formation of an indigenous crop complex in eastern North America at 3800 BP. *Proceedings of the National Academy of Sciences*, **106**, 6561–6566.
- Sobrino B, Brión M, Carracedo A (2005) SNPs in forensic genetics: a review on SNP typing methodologies. *Forensic Science International*, **154**, 181–194.
- Soto-Cerda B, Cloutier S (2012) Association mapping in plant genomes. In: *Genetic Diversity in Plants*, pp. 29–54.
- Stevens P (2010) Angiosperm phylogeny Website. Available from <http://www.Mobot.Org/mobot/research/apweb/>.
- Stich B, Möhring J, Piepho H-P *et al.* (2008) Comparison of mixed-model approaches for association mapping. *Genetics*, **178**, 1745–1754.
- Storey J (2002) A direct approach to false discovery rates. *Journal of the Royal Statistical Society*, **64**, 479–498.
- Stracke S, Haseneyer G, Veyrieras J-B *et al.* (2009) Association mapping reveals gene action and interactions in the determination of flowering time in barley. *TAG. Theoretical and applied genetics*, **118**, 259–273.
- Subbarao K V., Koike ST, Hubbard. JC (1996) Effects of deep plowing on the distribution and density of *Sclerotinia minor* sclerotia and lettuce drop incidence. *Plant disease*, **80**, 28–33.
- Sun X, Habier D, Fernando RL, Garrick DJ, Dekkers JC (2011) Genomic breeding value prediction and QTL mapping of QTLMAS2010 data using Bayesian Methods. *BMC proceedings*, **5**, S13.
- Syvänen a C (2001) Accessing genetic variation: genotyping single nucleotide polymorphisms. *Nature reviews. Genetics*, **2**, 930–942.
- Szalma SJ, Buckler ES, Snook ME, McMullen MD (2005) Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. *TAG. Theoretical and applied genetics*, **110**, 1324–1333.

- Tabor HK, Risch NJ, Myers RM (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nature Reviews Genetics*, **3**, 391–397.
- Tajima F (1983) Evolutionary relationship of DNA sequences in finite populations. *Genetics*, **105**, 437–460.
- Talia P, Nishinakamasu V, Esteban H, Heinz RA (2010) Genetic mapping of EST-SSRs, SSR and InDels to improve saturation of genomic regions in a previously developed sunflower map. *Electronic Journal of Biotechnology*, **13**.
- Talukder ZI, Hulke BS, Qi L *et al.* (2014) Candidate gene association mapping of Sclerotinia stalk rot resistance in sunflower (*Helianthus annuus* L.) uncovers the importance of COI1 homologs. *TAG. Theoretical and applied genetics*, **127**, 193–209.
- Tang S, Leon A, Bridges WC, Knapp SJ (2006) Quantitative Trait Loci for Genetically Correlated Seed Traits are Tightly Linked to Branching and Pericarp Pigment Loci in Sunflower. *Crop Science*, **46**, 721.
- Tang D, Simonich MT, Innes RW (2007) Mutations in LACS2, a long-chain acyl-coenzyme A synthetase, enhance susceptibility to avirulent *Pseudomonas syringae* but confer resistance to *Botrytis cinerea* in Arabidopsis. *Plant physiology*, **144**, 1093–1103.
- Tang S, Yu J-K, Slabaugh B, Shintani K, Knapp J (2002) Simple sequence repeat map of the sunflower genome. *TAG. Theoretical and applied genetics*, **105**, 1124–1136.
- Tanksley SD (1997) Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild. *Science*, **277**, 1063–1066.
- Thal WM, Campbell CL (1986) Spatial patterns analysis of disease severity data for alfalfa leaf spot caused primarily by *Leptosphaerulina briosiana*. *Phytopathology*, **76**, 190–194.
- Thomson MJ, Zhao K, Wright M *et al.* (2011) High-throughput single nucleotide polymorphism genotyping for breeding applications in rice using the BeadXpress platform. *Molecular Breeding*, **29**, 875–886.
- Thornsberry JM, Goodman MM, Doebley J *et al.* (2001) Dwarf8 polymorphisms associate with variation in flowering time. *Nature genetics*, **28**, 286–289.
- Tibshirani R (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society*, **58**, 267–288.
- Till BJ, Burtner C, Comai L, Henikoff S (2004) Mismatch cleavage by single-strand specific nucleases. *Nucleic acids research*, **32**, 2632–2641.
- Till BJ, Cooper J, Tai TH *et al.* (2007) Discovery of chemically induced mutations in rice by TILLING. *BMC plant biology*, **7**, 19.
- Tiryaki I, Staswick PE (2002) An Arabidopsis mutant defective in jasmonate response is allelic to the auxin-signaling mutant *axr1*. *Plant physiology*, **130**, 887–894.
- Tor M (2002) Arabidopsis SGT1b Is Required for Defense Signaling Conferred by Several Downy Mildew Resistance Genes. *THE PLANT CELL ONLINE*, **14**, 993–1003.
- Trebst A, Depka B, Holländer-Czytko H (2002) A specific role for tocopherol and of chemical singlet oxygen quenchers in the maintenance of photosystem II structure and function in *Chlamydomonas reinhardtii*. *FEBS Letters*, **516**, 156–160.

- Trogia C (2003) Factores ambientales que afectan la supervivencia y producción del inóculo de *Sclerotinia sclerotiorum* y umbral mínimo de inóculo para la podredumbre húmeda del capítulo de girasol. Tesis de Maestría Universidad Nacional de Mar del Plata, Facultad de Ciencias Agrarias, Unidad Integrada Balcarce Argentina.
- Tung CW, Zhao K, Wright MH *et al.* (2010) Development of a Research Platform for Dissecting Phenotype–Genotype Associations in Rice (*Oryza* spp.). *Rice*, **3**, 205–217.
- Uauy C, Paraiso F, Colasuonno P *et al.* (2009) A modified TILLING approach to detect induced mutations in tetraploid and hexaploid wheat. *BMC plant biology*, **9**, 115.
- Ullstrop AJ (1978) Corn diseases in the United States and their control. United States Department of Agriculture (USDA), Agricultural Handbook 199. Washington, D.C.: USDA Agricultural Research Service.
- Van K, Hwang E-Y, Kim MY *et al.* (2005) Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea. *The Journal of heredity*, **96**, 529–535.
- Vavilov N (1940) The new systematics of cultivated plants. In: *The new systematics*, pp. 549–566.
- Vear F, Serre F, Roche S, Walser P, Tourvieille de Labrouhe D (2007) Improvement of *Sclerotinia sclerotiorum* head rot resistance in sunflower by recurrent selection of a restorer population. *Helia*, **30**, 1–12.
- Vear F, Tourvieille de Labrouche D (1984) Recurrent selection for resistance to *Sclerotinia sclerotiorum* in sunflowers using artificial infections. *Agronomie*, **4**, 789–794.
- Vear F, Tourvieille de Labrouche D (1988) Heredity of resistance to *Sclerotinia sclerotiorum* in sunflowers. II. - Study of capitulum resistance to natural and artificial ascospore infections. *Agronomie*, **8**, 503–508.
- Vigouroux Y, Glaubitz JC, Matsuoka Y *et al.* (2008) Population structure and genetic diversity of New World maize races assessed by DNA microsatellites. *American journal of botany*, **95**, 1240–1253.
- Voorrips RE (2002) MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *Journal of Heredity*, **93**, 77–78.
- Vukcevic D (2009) Bayesian and Frequentist Methods and Analyses of Genome-Wide Association Studies. University of Oxford.
- Wagner T, Stoppa-Lyonnet D, Fleischmann E *et al.* (1999) Denaturing high-performance liquid chromatography detects reliably BRCA1 and BRCA2 mutations. *Genomics*, **62**, 369–376.
- Wang S, Sha Z, Sonstegard TS *et al.* (2008) Quality assessment parameters for EST-derived SNPs from catfish. *BMC genomics*, **9**, 450.
- Wang ML, Sukumaran S, Barkley N a *et al.* (2011) Population structure and marker-trait association analysis of the US peanut (*Arachis hypogaea* L.) mini-core collection. *TAG. Theoretical and applied genetics*, **123**, 1307–1317.
- Watterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.

- Weber A, Clark RM, Vaughn L *et al.* (2007) Major regulatory genes in maize contribute to standing variation in teosinte (*Zea mays* ssp. *parviglumis*). *Genetics*, **177**, 2349–2359.
- Wei X, Croft WB (2006) LDA-based document models for ad-hoc retrieval. In: *Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval*, p. 178. New York, USA.
- Weiler EW, Kutchan TM, Gorba T *et al.* (1994) The *Pseudomonas* phytotoxin coronatine mimics octadecanoid signalling molecules of higher plants. *FEBS Letters*, **345**, 9–13.
- Weller JI, Song JZ, Heyen DW, Lewin HA, Ron M (1998) A New Approach to the Problem of Multiple Comparisons in the Genetic Dissection of Complex Traits. *Genetics*, **150**, 1699–1706.
- Wilcoxson R, Skovmand B, Atif A (1975) Evaluation of wheat cultivars for the ability to retard development of stem rust. *Ann Appl Biol*, **80**, 275–287.
- Williams B, Kabbage M, Kim H-J, Britt R, Dickman MB (2011) Tipping the balance: *Sclerotinia sclerotiorum* secreted oxalic acid suppresses host defenses by manipulating the host redox environment. *PLoS pathogens*, **7**, e1002107.
- Wills DM, Burke JM (2007) Quantitative Trait Locus Analysis of the Early Domestication of Sunflower. *Genetics*, **176**, 2589–2599.
- Wilson L, Whitt S (2004) Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell*, **16**, 2719–2733.
- Wu TT, Chen YF, Hastie T, Sobel E, Lange K (2009) Genome-wide association analysis by lasso penalized logistic regression. *Bioinformatics (Oxford, England)*, **25**, 714–721.
- Würschum T, Langer SM, Longin CFH *et al.* (2013) Population structure, genetic diversity and linkage disequilibrium in elite winter wheat assessed with SNP and SSR markers. *TAG. Theoretical and applied genetics*, **126**, 1477–1486.
- Xiao Y, Cai D, Yang W *et al.* (2012) Genetic structure and linkage disequilibrium pattern of a rapeseed (*Brassica napus* L.) association mapping panel revealed by microsatellites. *TAG. Theoretical and applied genetics*, **125**, 437–447.
- Xie DX, Feys BF, James S, Nieto-Rostro M, Turner JG (1998) COI1: an Arabidopsis gene required for jasmonate-regulated defense and fertility. *Science (New York, N.Y.)*, **280**, 1091–1094.
- Xin Z, Wang ML, Barkley N a *et al.* (2008) Applying genotyping (TILLING) and phenotyping analyses to elucidate gene function in a chemically induced sorghum mutant population. *BMC plant biology*, **8**, 103.
- Xu L, Liu F, Lechner E *et al.* (2002) The SCF(COI1) ubiquitin-ligase complexes are required for jasmonate response in Arabidopsis. *The Plant cell*, **14**, 1919–1935.
- Yamada T (1993) The role of auxin in plant-disease development. *Annual review of phytopathology*, **31**, 253–273.
- Yang X, Xu Y, Shah T *et al.* (2011) Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. *Genetica*, **139**, 1045–1054.
- Yano R, Takebayashi Y, Nambara E, Kamiya Y, Seo M (2013) Combining association mapping and transcriptomics identify HD2B histone deacetylase as a genetic factor associated with seed

- dormancy in *Arabidopsis thaliana*. *The Plant journal : for cell and molecular biology*, **74**, 815–828.
- Yu J, Buckler ES (2006) Genetic association mapping and genome organization of maize. *Current opinion in biotechnology*, **17**, 155–160.
- Yu J, Holland JB, McMullen MD, Buckler ES (2008) Genetic design and statistical power of nested association mapping in maize. *Genetics*, **178**, 539–551.
- Yu J, Pressoir G, Briggs WH *et al.* (2006) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature genetics*, **38**, 203–208.
- Yu J, Tang S, Slabaugh MB *et al.* (2002) Towards a Saturated Molecular Genetic Linkage Map for Cultivated Sunflower. *Crop Science*, **43**, 367–387.
- Yu, Tang S, Slabaugh M, Heesacker A, Cole G (2003) Towards a saturated molecular genetic linkage map for cultivated sunflower. *Crop Science*, **43**, 367–387.
- Yu J, Zhang Z, Zhu C *et al.* (2009) Simulation Appraisal of the Adequacy of Number of Background Markers for Relationship Estimation in Association Mapping. *The Plant Genome Journal*, **2**, 63.
- Yue B, Radi S a, Vick B a *et al.* (2008) Identifying quantitative trait loci for resistance to Sclerotinia head rot in two USDA sunflower germplasms. *Phytopathology*, **98**, 926–931.
- Zhao K, Aranzana MJ, Kim S *et al.* (2007a) An *Arabidopsis* example of association mapping in structured samples. *PLoS genetics*, **3**, e4.
- Zhao J, Wang J, An L *et al.* (2007b) Analysis of gene expression profiles in response to *Sclerotinia sclerotiorum* in *Brassica napus*. *Planta*, **227**, 13–24.
- Zhao Y, Wang H, Chen W, Li Y (2014) Genetic Structure, Linkage Disequilibrium and Association Mapping of Verticillium Wilt Resistance in Elite Cotton (*Gossypium hirsutum* L.) Germplasm Population. *PloS one*, **9**, e86308.
- Zhao K, Wright M, Kimball J *et al.* (2010) Genomic diversity and introgression in *O. sativa* reveal the impact of domestication and breeding on the rice genome. *PloS one*, **5**, e10780.
- Zhou N (1999) *Arabidopsis* PAD3, a Gene Required for Camalexin Biosynthesis, Encodes a Putative Cytochrome P450 Monooxygenase. *The Plant Cell Online*, **11**, 2419–2428.
- Zhou H, Steffenson BJ, Muehlbauer G *et al.* (2014) Association mapping of stem rust race TTKSK resistance in US barley breeding germplasm. *TAG. Theoretical and applied genetics*, **127**, 1293–304.
- Zhu C, Gore M, Buckler ES, Yu J (2008) Status and Prospects of Association Mapping in Plants. *The Plant Genome Journal*, **1**, 5.
- Zhu T, Salmeron J (2007) High-definition genome profiling for genetic marker discovery. *Trends in plant science*, **12**, 196–202.
- Zhu YL, Song QJ, Hyten DL *et al.* (2003) Single-nucleotide polymorphisms in soybean. *Genetics*, **163**, 1123–1134.
- Zilic S, Barac M, Pesic M *et al.* (2010) Characterization of sunflower seed and kernel proteins. *Helia*, **33**, 103–113.

Zou H, Hastie T (2005) Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **67**, 301–320.

Zubrzycki J, Filippi C, Fusari C *et al.* (2012a) Desarrollo e implementación de un ensayo de genotipificación masiva de SNPs en girasol. In: *ALAG XV Congreso Latinoamericano de Genética - Rosario, Argentina*

Zubrzycki J, Fusari C, Maringolo C *et al.* (2012b) Biparental QTL and Association Mapping for Sclerotinia Head Rot Resistance in Cultivated Sunflower. In: *18th International Sunflower Conference*

Zubrzycki J, Filippi C, Aguirre N *et al.* (2014) Análisis de QTL para la resistencia a la podredumbre húmeda del capítulo en girasol causada por sclerotinia sclerotiorum. In: *Proceedings of the VI Congreso Asagir*

APÉNDICE

Tabla A1.1 Características generales de las líneas endodriadas de girasol incluidas en este estudio

Entrada	Pedigree	Origen	Aceitera /Confitera	Mantenedora/ Restauradora	Otras características
2021	Cruzamientos B 21 B83-3/11335/VNIIMK	Argentina	Aceitera	Mantenedora	Autocompatible
2067	B 85-18-7	Argentina	Aceitera	Mantenedora	Semilla blanca
2071	Derivada de HA335/A69	Argentina	Aceitera	Mantenedora	Resistente a roya y a <i>downy mildew</i>
2091	B92—150-315	Argentina	Aceitera	Mantenedora	Alto oleico
2095	B92—150-534	Argentina	Aceitera	Mantenedora	Alto oleico
2101	Derivada de Israel V3	Israel	Confitera	Mantenedora	—
2121	Derivada de HA341	Argentina	Aceitera	Mantenedora	Alto oleico
2123	Derivada de HA342 HA89*2/Perwvenets(alto oleico)	USA	Aceitera	Mantenedora	Resistente a <i>Verticillium dahliae</i>
2125	Derivada de HA343 HA89*2/Perwvenets (alto oleico)	USA	Aceitera	Mantenedora	Alto oleico
5023	R79-13-232-1	Argentina	Aceitera	Restauradora	—
5035	R79-15-631-1	Argentina	Aceitera	Restauradora	—
5049	R79-25-1522-2	Argentina	Aceitera	Restauradora	—
5075	R79-78-132-1	Argentina	Aceitera	Restauradora	—
5285	RG91-460-1/263	Argentina	Aceitera	Restauradora	Alto oleico
5286	RG91460-2/265	Argentina	Aceitera	Restauradora	Alto oleico
5289	RG91-460-7/763	Argentina	Aceitera	Restauradora	Alto oleico
5307	HA361/	Argentina	Aceitera	Restauradora	Baja estatura
5313	R92-160/RN160	Argentina	Aceitera	Restauradora	Resistente a roya
5348	Derivada de RPoman91-12	Argentina	Aceitera	Restauradora	—
5381	RG98-657-3/P6472HO	Argentina	Aceitera	Restauradora	Alto oleico
5383	RG-98-657-3 derived PION6472HO	Argentina	Aceitera	Restauradora	Alto oleico
5431	Derived 91T608/PION6440//R049	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
7-1-1	Derivada de ELLA-1 B	Argentina	Aceitera	Mantenedora	Moderadamente resistente a PHC
51084	Derivada de EN1084	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC
51084/5429	Derivada de EN1084///PION6440/91T608//R307	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC
51084/C820	Derivada de EN1084/RHA801	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC
5341/51084	Derivada de R94-RCO/EN1084	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC
5341-1	R94/RCO	Argentina	Aceitera	Restauradora	—
5393-E	Derivada de RG-00-667-5/MEx18AO Sel E	Argentina	Aceitera	Restauradora	Alto oleico
5393-N	Derivada de RG-00-667-5/MEx18AO Sel N	Argentina	Aceitera	Restauradora	Alto oleico
7X1	Ella-1 Sel E/Solf	Argentina	Aceitera	Mantenedora	Moderadamente resistente a PHC
ALB2/5261	Derived RALB2/RP261	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC
ALB2/5261/51084	Derivada de RALB2/RP261/EN1084	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC

Entrada	Pedigree	Origen	Aceitera /Confitera	Mantenedora/ Restauradora	Otras características
B10	Derivada de PxS81-162	Argentina	Aceitera	Mantenedora	—
B473-2	Derivada de BHA64/417-2_2	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B473-4	Derivada de BHA64/417-2_4	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B473-5	Derivada de BHA64/417-2_5	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B473-6	Derivada de BHA64/417-2_6	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B475-3	Derivada de BHA64/R432_3	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B475-4	Derivada de BHA64/R432_4	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B475-6	Derivada de BHA64/R432_6	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B477-1	Derivada de BHA64/B71_1	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B481-2	Derivada de B71/R423_2	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B481-3	Derivada de B71/R423_3	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B481-4	Derivada de B71/R423_4	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B481-5	Derivada de B71/R423_5	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B481-6	Derivada de B71/R423_6	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B483-1	Derivada de B71/R432_1	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B483-3	Derivada de B71/R432_3	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B485-5	Derivada de BHA64/HAR4_5	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B487-4	Derivada de B71/BHAR4	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
B59	Derivada de B85-9-7	Argentina	Aceitera	Mantenedora	—
B99	Derivada de Po91-123-2	Argentina	Aceitera	Mantenedora	—
C1059	Derivada de HA 224 Armavirski9345/HA8	USA	Aceitera	Mantenedora	Resistente a roya y a <i>Verticillium dahliae</i>
C110M	C-2055-45	Argentina	Aceitera	Restauradora	—
C112M	C-2055-17	Argentina	Aceitera	Restauradora	—
C154B	Derivada de HA 337 B	USA	Aceitera	Mantenedora	—
C192-1	Derivada de 2 B	Romania	Aceitera	Mantenedora	Citoplasma normal
C208F	Derivada de GMS 801	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC
C396-2B	Cruzamiento de poblaciones B MF 83-3/111335.	Argentina	Aceitera	Mantenedora	Autocompatible
C436	Derivada de MP 557	Argentina	Aceitera	Mantenedora	—
C438	Derivada de SP S3	Argentina	Aceitera	Mantenedora	—
C454B	Derivada de HA 335 B	USA	Aceitera	Mantenedora	Resistente a Downy mildew
C458B	Derivada de VNIIMK 8931	USA	Aceitera	Mantenedora	—
C494-1	Derivada de PAC-1	Francia	Aceitera	Mantenedora	Moderadamente resistente a PHC
C610-1	CM307-1	Otra	—	Mantenedora	—
C625-U	Derivada de L-91-3	Argentina	Aceitera	Mantenedora	—
C626-M	Derivada de L-130 Sel E	Argentina	Aceitera	—	Ramificada

Entrada	Pedigree	Origen	Aceitera /Confitera	Mantenedora/ Restauradora	Otras características
C627-M1	Derivada de L-130 Sel N	Argentina	Aceitera	_	Ramificada
C811	Derivada de HA 124 VNIIMK 8883	USA	Aceitera	Mantenedora	_
C812B	Derivada de HA 390	Argentina	Aceitera	Mantenedora	_
C813	B LINE SCL REC. SEL	USA	Aceitera	Mantenedora	Resistente a <i>Verticillium dahliae</i>
C816	Derivada de RHA 391 START	USA	Aceitera	Restauradora	No-Ramificada
C817	Derivada de RHA 392 SELECT	USA	Aceitera	Restauradora	_
C818	Derivada de RHA 408 ROMANIA R-LINE SCL POP-1	USA	Aceitera	Restauradora	_
C820	Derivada de RHA 801	Argentina	Aceitera	Restauradora	Moderadamente resistente a PHC
C826	Derivada de RHA 439 RHA 377/ AS 3211	USA	Aceitera	Restauradora	Moderadamente resistente a PHC
C827	Derivada de RHA 440 RHA 377/ AS 4379	USA	Aceitera	Restauradora	Moderadamente resistente a PHC
C829B	Derivada de HA 441 B	Argentina	Aceitera	Mantenedora	Moderadamente resistente a PHC
C849	Derivada de HA 853 1975 (compuesta de alto rendimiento)	USA	Aceitera	Mantenedora	Resistente a <i>Verticillium dahliae</i>
C851	Derivada de HA 234 2*Smena//HA6/HA8	USA	Aceitera	Mantenedora	Resistente a roya
C853	Derivada de HA 372 H 55 Sel	USA	Aceitera	Mantenedora	_
C977	Derivada de HA 442 B HA 425/87 CAEB//HA434/HA412	USA	Aceitera	Mantenedora	Resistente a Imidazolinona
F155	ACS	Argentina	Aceitera	Mantenedora	_
GP365	VNIIMK 6540 x CA3	Argentina	Aceitera	Mantenedora	Resistente a roya
HA234	2*SMENA//HA6/HA8	Argentina	Aceitera	Mantenedora	_
HA302	Peredovick 304	USA	Aceitera	_	Resistente a roya y a <i>Verticillium dahliae</i> , susceptible a Mildew
HA441	HA412/SD	USA	Aceitera	Mantenedora	Moderadamente resistente a PHC
HA64	Derivada de VNIIMK 1646	USA	Aceitera	Mantenedora	Tolerante a déficit hídrico
HA853	1975 (Compuesta de alto rendimiento)	USA	Aceitera	Mantenedora	_
HA89	"Vniimk 8931"	USA	Aceitera	Mantenedora	_
HAR2	Derivada de IMPIRA INTA	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
HAR4	Derivada de SAENZ PEÑA 74-1-2	Argentina	Aceitera	Mantenedora	Tolerante a déficit hídrico
PAC2	<i>H. petiolaris</i> x HA61	Francia	_	Restauradora	Resistente a PHC y PBT
R073	R79-73-421	Argentina	Aceitera	Restauradora	_
R413	Derivada de RHA340	Argentina	Aceitera	Restauradora	Resistente a Downy mildew
R415	HA89/T+//PAC1	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R417	PAC1/HA89/T-	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R419	HA89/T-//CF9	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R420	HA89/T-//CF9//RHA340	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R422	PNMR651/M734//RHA340	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R423	M734/PNMR651	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R424	M734/PNMR651	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico

Entrada	Pedigree	Origen	Aceitera /Confitera	Mantenedora/ Restauradora	Otras características
R425	91T608	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R426-1	91T608/RHA340_1	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R426-2	91T608/RHA340_2	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R429	PION6440/91T608//R307	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R432	PION6440/91T608//R049	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R433	PION6440/91T608	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R435	PION6440/PUNTA	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R437	PUNTA/PION6440	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R439	derived Pob R ACG	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R443	A21/R049	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R447	A71//B59/R049	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R449	A9/R049//A59/R049	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R453	A59/R049//RACG	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R459-1	R423/417-1_1	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R459-2	R423/417-1_2	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R459-4	R423/417-1_4	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R461-2	R423/417-2_2	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R461-3	R423/417-2_3	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R461-4	R423/417-2_4	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R463-3	R432/417-2_3	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R463-4	R432/417-2_4	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R463-5	R432/417-2_5	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R465-3	R432/423_3	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R467-3	R423/HA64_3	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
R467-5	R423/HA64_5	Argentina	Aceitera	Restauradora	Tolerante a déficit hídrico
RHA266	Wild <i>H. annuus</i> x Peredovik	USA	_	Restauradora	Resistente a Downy mildew
RHA274	CMSPI343765/HA119//HA62-4-5/2/T66006-2-2-11-3-2	Argentina	Aceitera	Restauradora	_
RHA275	CMS PI343765/HA119//HA62-4-5/2/T66006-2-2-11-3-2	USA	_	Restauradora	Ramificada, Resistente a roya
RHA276	CMS PI343765/HA119//HA62-4-5/2/T66006-2-2-11-3-2	USA	_	Restauradora	Ramificada, Resistente a roya
RHA299	cmsP-21 VR1/T70050*.	USA	Aceitera	Restauradora	_
RHA439	Línea restauradora derivada de RHA377/AS 3211	USA	Aceitera	Restauradora	Moderadamente resistente a PHC, Alto linoleico
RHA440	Línea restauradora derivada de RHA377/AS4379	USA	Aceitera	Restauradora	Moderadamente resistente a PHC, Alto linoleico
RHA801	Derivada de múltiples poblaciones	USA	Aceitera	Restauradora	Moderadamente resistente a PHC

Tabla A1.2 Parámetros de los efectos aleatorios para el modelo [1.1] – Incidencia

Efecto	Varianza	% Varianza	Desvio
Fecha.inoc	1,37	45,36	1,17
Repetición	0,1	3,31	0,32
Campaña	1,55	51,32	1,25

Tabla A1.3 Parámetros de los efectos aleatorios para el modelo [1.2] – Incidencia

Efecto	Varianza	% Varianza	Desvio
Línea:fecha.inoc	0,74	18,83	0,86
Fecha.inoc	1,21	30,79	1,1
Repetición	0,13	3,31	0,36
Campaña	1,85	47,07	1,36

Tabla A1.4 Parámetros de los efectos aleatorios para el modelo [1.3] – Incidencia

Efecto	Varianza	% Varianza	Desvio
Línea:fecha.inoc	1,14	23,17	1,07
Línea	0,65	13,21	0,81
Fecha.inoc	1,21	24,59	1,1
Repetición	0,14	2,85	0,37
Campaña	1,78	36,18	1,33

Tabla A1.5 Parámetros de los efectos aleatorios para el modelo [1.1] – AUDPCInc

Efecto	Varianza	% Varianza	Desvio
Fecha.inoc	4,04	30,40	2,01
Repetición	0,26	1,96	0,51
Campaña	8,99	67,64	3

Tabla A1.6 Parámetros de los efectos aleatorios para el modelo [1.2] – AUDPCInc

Efecto	Varianza	% Varianza	Desvio
Línea:fecha.inoc	3,92	23,17	1,98
Fecha.inoc	3,73	22,04	1,93
Repetición	0,3	1,77	0,55
Campaña	8,97	53,01	2,99

Tabla A1.7 Parámetros de los efectos aleatorios para el modelo [1.3] – AUDPCInc

Efecto	Varianza	% Varianza	Desvio
Línea:fecha.inoc	3,91	20,14	1,98
Línea	2,8	14,43	1,67
Fecha.inoc	4,18	21,54	2,05
Repetición	0,3	1,55	0,55
Campaña	8,22	42,35	2,87

Tabla A1.8 Parámetros de los efectos aleatorios para el modelo [1.1] – Severidad

Efecto	Varianza	% Varianza	Desvio
Fecha.inoc	0,03	60	0,18
Repetición	0	0	0
Campaña	0,02	40	0,16

Tabla A1.9 Parámetros de los efectos aleatorios para el modelo [1.2] – Severidad

Efecto	Varianza	% Varianza	Desvio
Línea:Fecha.inoc	0,02	28,54	0,13
Fecha.inoc	0,03	42,81	0,18
Repetición	7,30E-05	0,10	0,01
Campaña	0,02	28,54	0,16

Tabla A1.10 Parámetros de los efectos aleatorios para el modelo [1.3] – Severidad

Efecto	Varianza	% Varianza	Desvio
Línea:Fecha.inoc	0,02	19,99	0,13
Línea	0,02	19,99	0,13
Fecha.inoc	0,04	39,97	0,19
Repetición	6,60E-05	0,07	0,01
Campaña	0,02	19,99	0,16

Tabla A1.11 Parámetros de los efectos aleatorios para el modelo [1.1] - Período de incubación

Efecto	Varianza	% Varianza	Desvio
Fecha.inoc	2,40E-05	81,00	4,90E-03
Repetición	3,30E-07	1,11	5,80E-04
Campaña	5,30E-06	17,89	2,30E-03

Tabla A1.12 Parámetros de los efectos aleatorios para el modelo [1.2] - Período de incubación

Efecto	Varianza	% Varianza	Desvio
Línea:Fecha.inoc	3,10E-05	54,10	1,00E-02
Fecha.inoc	2,10E-05	36,65	4,50E-03
Repetición	1,70E-06	2,97	1,30E-03
Campaña	3,60E-06	6,28	1,90E-03

Tabla A1.13 Parámetros de los efectos aleatorios para el modelo [1.3] - Período de incubación

Efecto	Varianza	% Varianza	Desvio
Línea:fecha.inoc	4,80E-05	55,17	1,00E-02
Línea	9,20E-06	10,57	3,00E-03
Fecha.inoc	2,30E-05	26,44	4,80E-03
Repetición	2,20E-06	2,53	1,50E-03
Campaña	4,60E-06	5,29	2,20E-03

Tabla A1.14 Medias ajustadas por línea para cada carácter

Línea	Incidencia	AUDPCinc	Severidad	Pinc
2021	0,42	3,91	0,42	16,08
2067	0,94	9,8	0,55	13,58
2071	0,44	5,25	0,48	14,69
2091	0,82	9,45	0,58	14,37
2095	0,65	7,39	0,46	13,64
2101	0,62	6,1	0,35	15,34
2121	0,68	6,38	0,42	15,87
2123	0,61	6,01	0,48	14,73
2125	0,59	5,47	0,42	15,12
5023	0,75	7,49	0,46	14,16
5035	0,3	5,31	0,61	14,34
5049	0,48	4,93	0,6	14,08
5075	0,33	4,45	0,51	15,25
5285	0,43	5,12	0,57	15,16
5286	0,53	6,58	0,7	13,4
5289	0,94	11,21	0,78	13
5307	0,62	7,55	0,51	14,34
5313	0,39	4,87	0,42	14,96
5341.1	0,2	3,27	0,26	14,52
5348	0,29	4,38	0,2	15,91
5381	0,17	4,14	0,21	15,72
5383	0,14	2,06	0,13	18,94
5431	0,91	9,73	0,74	14,26
7 – 1 – 1	0,27	1,42	0,15	12,21
51084	0,29	3,82	0,31	16,26
51084/5429	0,34	2,06	0,21	16,83
51084/C820	0,39	4,05	0,31	16,58
5393-E	0,62	7,83	0,5	13,04
5393-N	0,73	7,5	0,65	14,77
5431/51084	0,75	7,33	0,4	14,06
7X1	0,52	5,59	0,49	15,74
ALB2/5261	0,15	0,92	0,17	20,75
ALB2/5261/51084	0,3	4,01	0,33	16,28
B10	0,19	2,9	0,41	16,92
B473-2	0,72	7,24	0,43	14,95
B473-4	0,62	6,53	0,54	15,4
B473-5	0,39	3,99	0,36	16,26
B473-6	0,44	4,61	0,62	15,2
B475-3	0,77	8,32	0,74	14,89
B475-4	0,8	6,77	0,41	16,22
B475-6	0,72	6,72	0,38	15,3
B477-1	0,45	4,18	0,18	17,23
B481-2	0,67	4,88	0,29	16,61
B481-3	0,64	9,38	0,56	13,88
B481-4	0,4	6,17	0,56	13,62
B481-5	0,74	8,79	0,61	13,69
B481-6	0,72	6,73	0,5	15,09
B483-1	0,5	4,59	0,43	16,23
B483-3	0,49	5,59	0,43	15,42
B485-5	0,69	9,32	0,58	13,31
B487-4	0,51	5,74	0,68	14,79
B59	0,44	4,32	0,25	17,06
B99	0,44	5,36	0,46	15,81

Línea	Incidencia	AUDPCinc	Severidad	Pinc
C1059	0,52	5,2	0,4	14,32
C110M	0,78	8,01	0,54	14,86
C112M	0,48	5,15	0,47	14,98
C154B	0,64	7,49	0,6	13,81
C192-1	0,37	3,53	0,37	15,46
C208F	0,22	4,4	0,43	15,17
C396-2B	0,44	5,6	0,51	14,89
C436	0,3	3,92	0,28	16,11
C438	0,2	2,2	0,25	16,4
C454-B	0,56	5,49	0,5	15,2
C458B	0,67	6,54	0,58	14,11
C494-1	0,41	5,98	0,58	
C610	0,19	4,25	0,26	13,94
C625-U	0,48	5,26	0,45	15,6
C626-M	0,55	6,64	0,85	13,94
C627-1M	0,38	4,56	0,62	15,28
C811	0,7	6,28	0,42	15,39
C812	0,52	6,63	0,5	14,6
C813	0,46	5,64	0,43	14,12
C816	0,27	4,38	0,29	15,26
C817	0,87	7,05	0,42	16,21
C818	0,53	6,3	0,61	14,92
C820	0,52	5,2	0,52	15,88
C826-M	0,65	7,59	0,67	13,68
C827	0,43	4,84	0,34	15,94
C829B	0,21	3	0,25	16,44
C849	0,81	4,87	0,35	16,02
C851B	0,65	7,93	0,56	14,06
C853B	0,84	8,36	0,56	13,44
C977B	0,67	7,16	0,73	14,36
F155B	0,38	3,63	0,41	18,73
GP365	0,58	6,01	0,41	15,01
HA234	0,59	8,2	0,74	13,11
HA302	0,25	2,64	0,25	16,05
HA64	0,46	5	0,49	15,08
HA853	0,19	3,66	0,21	16,73
HA89	0,33	4,55	0,31	14,67
HAR2	0,76	7,26	0,42	16,15
HAR4	0,62	6,76	0,59	13,37
PAC2	0,26	2,05	0,33	17,58
R073	0,6	7,84	0,52	13,04
R413	0,59	6,35	0,31	15,41
R415	0,62	6,26	0,47	15,57
R417	0,22	3,62	0,14	16,38
R419	0,62	7,59	0,38	14,06
R420	0,81	8,6	0,57	14,89
R422	0,5	4,76	0,42	15,41
R423	0,54	5,8	0,66	13,97
R424	0,67	8,16	0,68	13,96
R425	0,31	4,9	0,57	14,55
R426-1	0,5	5,83	0,79	14,33
R426-2	0,73	8,48	0,67	13,59
R429	0,51	5,29	0,38	15,79

Línea	Incidencia	AUDPCinc	Severidad	Pinc
R432	0,81	8,93	0,78	12,68
R433	0,67	6,12	0,34	16,18
R435	0,92	9,39	0,68	14,43
R437	0,9	7,44	0,47	16,5
R439	0,95	9,93	0,7	13,84
R443	0,35	4,49	0,6	16,48
R447	0,21	3,06	0,19	15,42
R449	0,39	3,36	0,26	16,38
R453	0,85	9,21	0,55	14,36
R459-1	0,67	6,09	0,35	16,02
R459-2	0,21	2,04	0,23	15,7
R459-4	0,93	10,54	0,79	12,86
R461-2	0,42	5,9	0,54	15,03
R461-3	0,35	6,15	0,46	17,17
R461-4	0,7	8,07	0,62	14,73
R463-3	0,59	9,16	0,75	12,94
R463-4	0,81	7,42	0,47	15,06
R463-5	0,47	5,28	0,54	15,5
R465-3	0,61	8,11	0,7	13,55
R467-3	0,85	10,68	0,74	12,74
R467-5	0,82	9,32	0,58	13,32
RHA266	0,28	3,6	0,43	15,63
RHA274	0,22	2,93	0,21	16,38
RHA275	0,6	6,35	0,44	15,47
RHA276	0,66	6,11	0,28	15,84
RHA299	0,22	3,71	0,39	17,83
RHA439	0,32	3,86	0,42	14,57
RHA440	0,61	6,21	0,3	15,92
RHA441	0,16	2,52	0,24	16,71
RHA801	0,31	3,76	0,42	16
RK416	0,17	2,38	0,23	17,71

Tabla A2.1 Características generales de las poblaciones de polinización abierta y compuestas de girasol incluidas en este estudio

Entrada	Pedigree	Origen	Aceitera /Confitera	Mantenedora/ Restauradora	Otras características	Clasificación
7-1-2	Derivada de ELLA-1 B	Argentina	Aceitera	Mantenedora	Moderadamente resistente a PHC	CP
7-1-21	Derivada de ELLA-1 B	Argentina	Aceitera	Mantenedora	Moderadamente resistente a PHC	CP
C1005-M	Derivada de XRQ/PSC8 F2 L-3	USA	Aceitera	Mantenedora	Moderadamente resistente a PHC	OP
C111-3U	C-2055-35	Argentina	Aceitera	Restauradora	—	CP
C113M	PR I	Argentina	Aceitera	Restauradora	—	CP
C114M	PR I	Argentina	Aceitera	Restauradora	—	CP
C139-1	Derivada de HAR4.Saenz Peña 74-1-2 Sel.N	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C139-2	Derivada de HAR4.Saenz Peña 74-1-2 Sel.E	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C149-1	Derivada de DM-1. Compuesta de líneas B Sel N	USA	Aceitera	Mantenedora	Resistente a Downy mildew	CP
C149-2	Derivada de DM-1. Compuesta de líneas B Sel B	USA	Aceitera	Mantenedora	Resistente a Downy mildew	CP
C155	Derivada de RHA340. HA89*3/H. Argophyllus 415	USA	Aceitera	Restauradora	Resistente a Downy mildew	CP
C166-U	Derivada de WIR 135.	Rusia	—	—	—	CP
C191-1	Derivada de 1 B	Romania	Aceitera	Mantenedora	Citoplasma normal	OP
C397	Derivada de SP 1-4-SR	Argentina	Aceitera	Mantenedora	Resistente a roya	OP
C398	Derivada de Caburé INTA	Argentina	Aceitera	Mantenedora	Resistente a roya	OP
C428	Derivada de SP 74-1-2	Argentina	Aceitera	Mantenedora	Resistente a roya	CP
C430	Derivada de HAR2.Impira INTA Sel.N	Argentina	Aceitera	Mantenedora	Resistente a roya y a <i>Verticillium dahliae</i>	OP
C432-1	Derivada de HAR4.Saenz Peña 74-1-2 Sel.E	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C432-2	Derivada de HAR4 .Saenz Peña 74-1-2 Sel.N	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C433	Derivada de HAR.Guayacán INTA Sel.	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C434	Guayacán INTA Sel.Castelar	Argentina	Aceitera	Mantenedora	Resistente a roya	CP
C437	Derivada de 953-88-3	Argentina	Aceitera	Mantenedora	Resistente a roya	CP
C618-1M	Derivada de ESE UNO Sel E	Argentina	Aceitera	—	Ramificada	OP
C618-1U	Derivada de Pob RIV ESE UNO	Argentina	Aceitera	—	—	OP
C618-2U	Derivada de ESE UNO Sel EN	Argentina	Aceitera	Mantenedora	—	OP
C620-1U	Derivada de S-377 Sel N	Argentina	Aceitera	Mantenedora	—	OP
C620-2U	Derivada de S-377 Sel E	Argentina	Aceitera	Mantenedora	—	OP
C621	Derivada de S-1	Argentina	Aceitera	Mantenedora	—	OP
C628-M1	Derivada de L-149 Sel E	Argentina	Aceitera	—	Ramificada	CP
C-632-1	Derivada de Pob RIV L-196	Argentina	Aceitera	Mantenedora	—	CP
C633-1U	Derivada de L-363 Sel E	Argentina	Aceitera	Mantenedora	—	CP
C998	Derivada de SF270	USA	Aceitera	Mantenedora	Susceptible a PHC	OP
C999	Derivada de NK277	USA	Aceitera	Mantenedora	Moderadamente resistente a PHC	OP
7-1-2	Derivada de ELLA-1 B	Argentina	Aceitera	Mantenedora	Moderadamente resistente a <i>Sclerotinia sclerotiorum</i>	CP

Entrada	Pedigree	Origen	Aceitera /Confitera	Mantenedora/ Restauradora	Otras características	Clasificación
7-1-21	Derivada de ELLA-1 B	Argentina	Aceitera	Mantenedora	Moderadamente resistente a <i>Sclerotinia sclerotiorum</i>	CP
C1005-M	Derivada de XRO/PSC8 F2 L-3	USA	Aceitera	Mantenedora	Moderadamente resistente a <i>Sclerotinia sclerotiorum</i>	OP
C111-3U	C-2055-35	Argentina	Aceitera	Restauradora	—	CP
C113M	PR I	Argentina	Aceitera	Restauradora	—	CP
C114M	PR I	Argentina	Aceitera	Restauradora	—	CP
C139-1	Derivada de HAR4.Saenz Peña 74-1-2 Sel.N	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C139-2	Derivada de HAR4.Saenz Peña 74-1-2 Sel.E	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C149-1	Derivada de DM-1. Compuesta de líneas B Sel N	USA	Aceitera	Mantenedora	Resistente a Downy mildew	CP
C149-2	Derivada de DM-1. Compuesta de líneas B Sel B	USA	Aceitera	Mantenedora	Resistente a Downy mildew	CP
C155	Derivada de RHA340. HA89*3/H. Argophyllus 415	USA	Aceitera	Restauradora	Resistente a Downy mildew	CP
C166-U	Derivada de WIR 135	Rusia	—	—	—	CP
C191-1	Derivada de 1 B	Rumania	Aceitera	Mantenedora	—	OP
C397	Derivada de SP 1-4-SR	Argentina	Aceitera	Mantenedora	Resistente a roya	OP
C398	Derivada de Caburé INTA	Argentina	Aceitera	Mantenedora	Resistente a roya	OP
C428	Derivada de SP 74-1-2	Argentina	Aceitera	Mantenedora	Resistente a roya	CP
C430	Derivada de HAR2.Impira INTA Sel.N	Argentina	Aceitera	Mantenedora	Resistente a roya y a <i>Verticillium dahliae</i>	OP
C432-1	Derivada de HAR4.Saenz Peña 74-1-2 Sel.E	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C432-2	Derivada de HAR4.Saenz Peña 74-1-2 Sel.N	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C433	Derivada de HAR.Guayacán INTA Sel.	Argentina	Aceitera	Mantenedora	Resistente a Downy mildew, roya y <i>Verticillium dahliae</i>	CP
C434	Guayacán INTA Sel.Castelar	Argentina	Aceitera	Mantenedora	Resistente a roya	CP
C437	Derivada de 953-88-3	Argentina	Aceitera	Mantenedora	Resistente a roya	CP
C618-1M	Derivada de ESE UNO Sel E	Argentina	Aceitera	—	Ramificada	OP
C618-1U	Derivada de Pob RIV ESE UNO	Argentina	Aceitera	—	—	OP
C618-2U	Derivada de ESE UNO Sel EN	Argentina	Aceitera	Mantenedora	—	OP
C620-1U	Derivada de S-377 Sel N	Argentina	Aceitera	Mantenedora	—	OP
C620-2U	Derivada de S-377 Sel E	Argentina	Aceitera	Mantenedora	—	OP
C621	Derivada de S-1	Argentina	Aceitera	Mantenedora	—	OP
C628-M1	Derivada de L-149 Sel E	Argentina	Aceitera	—	Ramificada	CP
C-632-1	Derivada de Pob RIV L-196	Argentina	Aceitera	Mantenedora	—	CP
C633-1U	Derivada de L-363 Sel E	Argentina	Aceitera	Mantenedora	—	CP
C998	Derivada de SF270	USA	Aceitera	Mantenedora	Susceptible a <i>Sclerotinia sclerotiorum</i>	OP
C999	Derivada de NK277	USA	Aceitera	Mantenedora	Moderadamente resistente a <i>Sclerotinia sclerotiorum</i>	OP

* T70050 deriva de un cruzamiento que involucra a cmsPI 343765 [HA 119/2] HA 62-4-5 y plantas individuales seleccionadas de T66006-2 realizado por AR-SEA-USDA, College Station, Tex.

Tabla A2.2 Lista y origen de los SNPs incluidos en el ensayo de 384 SNPs de Illumina que resultaron polimórficos en las líneas de la PMA.

SNP	Locus ID	Base de datos	Número de acceso
SNP1	HeAn_C_10779.1	ATGC-Sur	HeAn_C_10779_Contig10779
SNP10	HeAn_C_6243.1	ATGC-Sur	HeAn_C_6243_Contig6243
SNP100	HeAn_C_5453.1	ATGC-Sur	HeAn_C_5453_Contig5453
SNP101	HeAn_C_10469.1	ATGC-Sur	HeAn_C_10469_Contig10469
SNP102	HeAn_R_327.2	GenBank	DQ503813.1
SNP103	HeAn_C_10375.1	ATGC-Sur	HeAn_C_10375_Contig10375
SNP104	HeAn_C_10581.1	ATGC-Sur	HeAn_C_10581_Contig10581
SNP105	HeAn_C_10349.1	ATGC-Sur	HeAn_C_10349_Contig10349
SNP106	HeAn_C_2467.1	ATGC-Sur	HeAn_C_2467_Contig2467
SNP107	HeAn_C_7972.1	ATGC-Sur	HeAn_C_7972_Contig7972
SNP108	HeAn_C_8440.1	ATGC-Sur	HeAn_C_8440_Contig8440
SNP109	HeAn_C_11458.1	ATGC-Sur	HeAn_C_11458_Contig11458
SNP11	HeAn_C_185.1	ATGC-Sur	HeAn_C_185_Contig185
SNP110	HeAn_C_12116.4	ATGC-Sur	HeAn_C_12116_Contig12116
SNP111	HeAn_C_11352.1	ATGC-Sur	HeAn_C_11352_Contig11352
SNP112	HeAn_C_10220.1	ATGC-Sur	HeAn_C_10220_Contig10220
SNP113	HeAn_C_10285.1	ATGC-Sur	HeAn_C_10285_Contig10285
SNP114	HeAn_C_4895.1	ATGC-Sur	HeAn_C_4895_Contig4895
SNP115	HeAn_C_1210.1	ATGC-Sur	HeAn_C_1210_Contig1210
SNP116	HeAn_C_10599.1	ATGC-Sur	HeAn_C_10599_Contig10599
SNP117	HeAn_C_8563.1	ATGC-Sur	HeAn_C_8563_Contig8563
SNP118	HeAn_R_578.2	GenBank	EF560022.1
SNP119	HeAn_C_7195.1	ATGC-Sur	HeAn_C_7195_Contig7195
SNP12	HeAn_C_10556.3	ATGC-Sur	HeAn_C_10556_Contig10556
SNP120	HeAn_C_1739.2	ATGC-Sur	HeAn_C_1739_Contig1739
SNP121	HeAn_C_12165.2	ATGC-Sur	HeAn_C_12165_Contig12165
SNP122	HeAn_C_3008.1	ATGC-Sur	HeAn_C_3008_Contig3008
SNP123	HeAn_C_2096.2	ATGC-Sur	HeAn_C_2096_Contig2096
SNP124	HeAn_C_3646.2	ATGC-Sur	HeAn_C_3646_Contig3646
SNP125	HeAn_C_12362.3	ATGC-Sur	HeAn_C_12362_Contig12362
SNP126	HeAn_C_12556.2	ATGC-Sur	HeAn_C_12556_Contig12556
SNP127	HeAn_C_11104.1	ATGC-Sur	HeAn_C_11104_Contig11104
SNP128	HeAn_R_GLP3.1	TGI	TC17527
SNP129	HeAn_R_325.1	GenBank	DQ504137.1
SNP13	HeAn_C_8817.1	ATGC-Sur	HeAn_C_8817_Contig8817
SNP130	HeAn_C_7075.1	ATGC-Sur	HeAn_C_7075_Contig7075
SNP131	HeAn_C_9388.1	ATGC-Sur	HeAn_C_9388_Contig9388
SNP132	HeAn_C_9582.2	ATGC-Sur	HeAn_C_9582_Contig9582
SNP133	HeAn_C_11554.1	ATGC-Sur	HeAn_C_11554_Contig11554
SNP134	HeAn_R_136.2	GenBank	EF469901.1
SNP135	HeAn_C_1395.1	ATGC-Sur	HeAn_C_1395_Contig1395
SNP136	HeAn_C_4339.1	ATGC-Sur	HeAn_C_4339_Contig4339
SNP137	HeAn_C_6398.1	ATGC-Sur	HeAn_C_6398_Contig6398
SNP138	HeAn_C_803.1	ATGC-Sur	HeAn_C_803_Contig803
SNP139	HeAn_C_9887.1	ATGC-Sur	HeAn_C_9887_Contig9887
SNP14	HeAn_R_316.1	GenBank	DQ504033.1
SNP140	HeAn_R_174.2	GenBank	EF461947.1
SNP141	HeAn_R_204.1	GenBank	EF461564.1
SNP142	HeAn_C_382.1	ATGC-Sur	HeAn_C_382_Contig382
SNP143	HeAn_R_257.1	GenBank	EF461167.1
SNP144	HeAn_C_3689.1	ATGC-Sur	HeAn_C_3689_Contig3689
SNP145	HeAn_R_176.2	GenBank	EF469928.1

SNP	Locus ID	Base de datos	Número de acceso
SNP146	HeAn_C_10491.1	ATGC-Sur	HeAn_C_10491_Contig10491
SNP147	HeAn_C_11018.1	ATGC-Sur	HeAn_C_11018_Contig11018
SNP148	HeAn_C_11680.1	ATGC-Sur	HeAn_C_11680_Contig11680
SNP149	HeAn_C_3959.1	ATGC-Sur	HeAn_C_3959_Contig3959
SNP150	HeAn_C_10672.1	ATGC-Sur	HeAn_C_10672_Contig10672
SNP151	HeAn_C_5449.1	ATGC-Sur	HeAn_C_5449_Contig5449
SNP152	HeAn_R_343.1	GenBank	Z95153.2
SNP153	HeAn_R_109.2	GenBank	EU112809.1
SNP154	HeAn_C_3307.1	ATGC-Sur	HeAn_C_3307_Contig3307
SNP155	HeAn_R_564.1	GenBank	EU112834.1
SNP156	HeAn_R_531.1	GenBank	FJ373820.1
SNP157	HeAn_C_11835.1	ATGC-Sur	HeAn_C_11835_Contig11835
SNP158	HeAn_C_6989.1	ATGC-Sur	HeAn_C_6989_Contig6989
SNP159	HeAn_C_12332.1	ATGC-Sur	HeAn_C_12332_Contig12332
SNP16	HeAn_C_9073.1	ATGC-Sur	HeAn_C_9073_Contig9073
SNP160	HeAn_R_147.1	GenBank	EF469864.1
SNP161	HeAn_R_326.2	GenBank	GQ884979.1
SNP162	HeAn_C_12291.1	ATGC-Sur	HeAn_C_12291_Contig12291
SNP163	HeAn_R_534.1	GenBank	FJ373755.1
SNP164	HeAn_C_11459.1	ATGC-Sur	HeAn_C_11459_Contig11459
SNP165	HeAn_R_248.1	GenBank	EF461227.1
SNP166	HeAn_C_12910.1	ATGC-Sur	HeAn_C_12910_Contig12910
SNP167	HeAn_C_11786.2	ATGC-Sur	HeAn_C_11786_Contig11786
SNP168	HeAn_C_3781.1	ATGC-Sur	HeAn_C_3781_Contig3781
SNP169	HeAn_R_529.1	GenBank	FJ373874.1
SNP17	HeAn_R_244.1	GenBank	EF469893.1
SNP170	HeAn_R_248.3	GenBank	EF461227.1
SNP171	HeAn_C_12630.2	ATGC-Sur	HeAn_C_12630_Contig12630
SNP172	HeAn_R_133.1	GenBank	EF469913.1
SNP173	HeAn_C_2084.1	ATGC-Sur	HeAn_C_2084_Contig2084
SNP174	HeAn_C_186.1	ATGC-Sur	HeAn_C_186_Contig186
SNP175	HeAn_C_11171.1	ATGC-Sur	HeAn_C_11171_Contig11171
SNP176	HeAn_C_9523.1	ATGC-Sur	HeAn_C_9523_Contig9523
SNP178	HeAn_C_11612.1	ATGC-Sur	HeAn_C_11612_Contig11612
SNP179	HeAn_C_2575.1	ATGC-Sur	HeAn_C_2575_Contig2575
SNP18	HeAn_C_4178.1	ATGC-Sur	HeAn_C_4178_Contig4178
SNP180	HeAn_R_15.3	GenBank	AY541451.1
SNP181	HeAn_R_CytP450_C	GenBank	JN231393.1
SNP182	HeAn_R_GLP4	TGI	TC18217
SNP184	HeAn_R_RhoBP_B.1	GenBank	JN231449.1
SNP185	HeAn_C_3174.1	ATGC-Sur	HeAn_C_3174_Contig3174
SNP186	HeAn_C_3933.1	ATGC-Sur	HeAn_C_3933_Contig3933
SNP187	HeAn_C_9713.1	ATGC-Sur	HeAn_C_9713_Contig9713
SNP188	HeAn_C_3646.1	ATGC-Sur	HeAn_C_3646_Contig3646
SNP19	HeAn_R_277.1	GenBank	EF460942.1
SNP190	HeAn_C_9271.1	ATGC-Sur	HeAn_C_9271_Contig9271
SNP191	HeAn_R_154.2	GenBank	EF462122.1
SNP192	HeAn_C_10970.1	ATGC-Sur	HeAn_C_10970_Contig10970
SNP193	HeAn_R_RhoBP_A	GenBank	JN231440.1
SNP194	HeAn_C_7412.1	ATGC-Sur	HeAn_C_7412_Contig7412
SNP195	HeAn_R_531.2	GenBank	FJ373820.1
SNP196	HeAn_R_128.2	GenBank	EF461762.1
SNP198	HeAn_C_10454.1	ATGC-Sur	HeAn_C_10454_Contig10454
SNP199	HeAn_R_FBoxP.1	GenBank	JN231480.1
SNP2	HeAn_R_319.2	GenBank	EU112570.1

SNP	Locus ID	Base de datos	Número de acceso
SNP21	HeAn_R_536.1	GenBank	FJ373664.1
SNP22	HeAn_C_2096.4	ATGC-Sur	HeAn_C_2096_Contig2096
SNP23	HeAn_C_7628.1	ATGC-Sur	HeAn_C_7628_Contig7628
SNP25	HeAn_C_11928.1	ATGC-Sur	HeAn_C_11928_Contig11928
SNP26	HeAn_C_10113.1	ATGC-Sur	HeAn_C_10113_Contig10113
SNP27	HeAn_C_8468.1	ATGC-Sur	HeAn_C_8468_Contig8468
SNP28	HeAn_R_610.1	GenBank	EU674799.1
SNP29	HeAn_R_183.1	GenBank	EF461805
SNP3	HeAn_C_6997.1	ATGC-Sur	HeAn_C_6997_Contig6997
SNP30	HeAn_C_12824.1	ATGC-Sur	HeAn_C_12824_Contig12824
SNP32	HeAn_C_10802.1	ATGC-Sur	HeAn_C_10802_Contig10802
SNP34	HeAn_C_1638.1	ATGC-Sur	HeAn_C_1638_Contig1638
SNP35	HeAn_C_8891.1	ATGC-Sur	HeAn_C_8891_Contig8891
SNP36	HeAn_R_559.1	GenBank	EU112515.1
SNP38	HeAn_C_3789.2	ATGC-Sur	HeAn_C_3789_Contig3789
SNP39	HeAn_C_9982.1	ATGC-Sur	HeAn_C_9982_Contig9982
SNP4	HeAn_C_7785.1	ATGC-Sur	HeAn_C_7785_Contig7785
SNP40	HeAn_R_291.1	GenBank	AM051287.1
SNP41	HeAn_C_10455.1	ATGC-Sur	HeAn_C_10455_Contig10455
SNP42	HeAn_R_508.1	GenBank	FJ855995.1
SNP43	HeAn_R_183.2	GenBank	EF461805
SNP44	HeAn_R_326.1	GenBank	GQ884979.1
SNP45	HeAn_C_1205.1	ATGC-Sur	HeAn_C_1205_Contig1205
SNP46	HeAn_R_444.2	GenBank	U70374.1
SNP47	HeAn_R_5.1	GenBank	EF469924.1
SNP48	HeAn_C_8647.2	ATGC-Sur	HeAn_C_8647_Contig8647
SNP49	HeAn_C_6169.3	ATGC-Sur	HeAn_C_6169_Contig6169
SNP5	HeAn_C_12263.1	ATGC-Sur	HeAn_C_12263_Contig12263
SNP50	HeAn_R_298.1	GenBank	AF251843.1
SNP51	HeAn_R_339.2	GenBank	AY805132.1
SNP52	HeAn_C_11445.1	ATGC-Sur	HeAn_C_11445_Contig11445
SNP53	HeAn_C_3979.1	ATGC-Sur	HeAn_C_3979_Contig3979
SNP55	HeAn_C_5628.1	ATGC-Sur	HeAn_C_5628_Contig5628
SNP56	HeAn_R_137.1	GenBank	EF461342.1
SNP58	HeAn_R_145.4	GenBank	EF460923.1
SNP59	HeAn_C_10263.1	ATGC-Sur	HeAn_C_10263_Contig10263
SNP6	HeAn_C_12116.1	ATGC-Sur	HeAn_C_12116_Contig12116
SNP60	HeAn_R_135.1	GenBank	EF461525.1
SNP62	HeAn_R_267.1	GenBank	EF469878.1
SNP63	HeAn_R_149.1	GenBank	EF462172.1
SNP64	HeAn_C_4316.1	ATGC-Sur	HeAn_C_4316_Contig4316
SNP65	HeAn_R_176.1	GenBank	EF469928.1
SNP66	HeAn_C_12028.1	ATGC-Sur	HeAn_C_12028_Contig12028
SNP68	HeAn_R_283.1	GenBank	EF460887.1
SNP69	HeAn_C_9356.1	ATGC-Sur	HeAn_C_9356_Contig9356
SNP7	HeAn_R_PRP2.1	GenBank	JN231455
SNP70	HeAn_C_7016.3	ATGC-Sur	HeAn_C_7016_Contig7016
SNP71	HeAn_C_4498.1	ATGC-Sur	HeAn_C_4498_Contig4498
SNP72	HeAn_C_9709.1	ATGC-Sur	HeAn_C_9709_Contig9709
SNP73	HeAn_C_5628.2	ATGC-Sur	HeAn_C_5628_Contig5628
SNP74	HeAn_R_548.3	GenBank	FJ373527.1
SNP75	HeAn_C_10608.1	ATGC-Sur	HeAn_C_10608_Contig10608
SNP76	HeAn_C_1691.1	ATGC-Sur	HeAn_C_1691_Contig1691
SNP77	HeAn_R_146.1	GenBank	EF461931.1
SNP78	HeAn_R_599.1	GenBank	EU673836

SNP	Locus ID	Base de datos	Número de acceso
SNP79	HeAn_C_10284.1	ATGC-Sur	HeAn_C_10284_Contig10284
SNP80	HeAn_C_12835.1	ATGC-Sur	HeAn_C_12835_Contig12835
SNP81	HeAn_C_11434.1	ATGC-Sur	HeAn_C_11434_Contig11434
SNP82	HeAn_C_4070.3	ATGC-Sur	HeAn_C_4070_Contig4070
SNP84	HeAn_C_12025.1	ATGC-Sur	HeAn_C_12025_Contig12025
SNP85	HeAn_C_8636.1	ATGC-Sur	HeAn_C_8636_Contig8636
SNP86	HeAn_C_12854.1	ATGC-Sur	HeAn_C_12854_Contig12854
SNP87	HeAn_C_11706.1	ATGC-Sur	HeAn_C_11706_Contig11706
SNP88	HeAn_C_7694.1	ATGC-Sur	HeAn_C_7694_Contig7694
SNP89	HeAn_C_5773.2	ATGC-Sur	HeAn_C_5773_Contig5773
SNP9	HeAn_C_2699.1	ATGC-Sur	HeAn_C_2699_Contig2699
SNP90	HeAn_C_6094.1	ATGC-Sur	HeAn_C_6094_Contig6094
SNP92	HeAn_R_126.2	GenBank	EF461978.1
SNP93	HeAn_C_117.1	ATGC-Sur	HeAn_C_117_Contig117
SNP94	HeAn_C_4023.2	ATGC-Sur	HeAn_C_4023_Contig4023
SNP95	HeAn_C_10204.1	ATGC-Sur	HeAn_C_10204_Contig10204
SNP96	HeAn_R_109.1	GenBank	EU112809.1
SNP97	HeAn_C_10383.1	ATGC-Sur	HeAn_C_10383_Contig10383
SNP98	HeAn_C_9002.2	ATGC-Sur	HeAn_C_9002_Contig9002
SNP99	HeAn_R_131.1	GenBank	EF461677.1

ATGC-Sur: http://atgc-sur.inta.gob.ar/init/search/by_name

TGI: DFCI (índice de genes de girasol)

Tablas A2.3 Análisis de la varianza molecular (AMOVA) completos con datos de tipo SSR

PMA vs. OP+CP			
Fuente	GL	SC	% varianza explicada
Entre poblaciones	1	122,078	2%
Entre individuos	370	7990,643	87%
Dentro de individuos	372	463,000	11%
Total	743	8575,720	100%

Estadísticos F	Valor	P-valor
Fst	0,025	0,001
Fis	0,891	0,001
Fit	0,894	0,001

Argentina vs. Otro origen			
Fuente	GL	SC	% varianza explicada
Entre poblaciones	1	51,183	3%
Entre individuos	135	2789,241	95%
Dentro de individuos	137	37,500	3%
Total	273	2877,923	100%

Estadísticos F	Valor	P-valor
Fst	0,028	0,001
Fis	0,974	0,001
Fit	0,975	0,001

HA vs. RHA			
Fuente	GL	SC	% varianza explicada
Entre poblaciones	1	42,517	2%
Entre individuos	132	2725,054	96%
Dentro de individuos	134	36,000	3%
Total	267	2803,571	100%

Estadísticos F	Valor	P-valor
Fst	0,016	0,001
Fis	0,974	0,001
Fit	0,975	0,001

Tablas A2.4 Análisis de la varianza molecular (AMOVA) completos con datos de tipo SNP

Argentina vs. Otro origen			
Fuente	GL	SC	% varianza explicada
Entre poblaciones	1	101,456	2%
Entre individuos	135	6991,515	89%
Dentro de individuos	137	353,500	9%
Total	273	7446,471	100%
Estadísticos F	Valor	P-valor	
Fst	0,018	0,001	
Fis	0,905	0,001	
Fit	0,907	0,001	
HA vs. RHA			
Fuente	GL	SC	% varianza explicada
Entre poblaciones	1	64,286	0,34%
Entre individuos	132	6847,643	90%
Dentro de individuos	134	346,500	9%
Total	267	7258,429	100%
Estadísticos F	Valor	P-valor	
Fst	0,003	0,032	
Fis	0,905	0,001	
Fit	0,905	0,001	

Tablas A2.5 Análisis de la varianza molecular (AMOVA) completos con datos de tipo SSR+SNP

Argentina vs. Otro origen			
Fuente	GL	SC	% varianza explicada
Entre poblaciones	1	152,639	2%
Entre individuos	135	9780,755	90%
Dentro de individuos	137	391,000	7%
Total	273	10324,394	100%
Estadísticos F	Valor	P-valor	
Fst	0,021	0,001	
Fis	0,924	0,001	
Fit	0,926	0,001	
HA vs. RHA			
Fuente	GL	SC	% varianza explicada
Entre poblaciones	1	106,773	1%
Entre individuos	132	9572,951	92%
Dentro de individuos	134	382,500	8%
Total	267	10062,224	100%
Estadísticos F	Valor	P-valor	
Fst	0,007	0,002	
Fis	0,924	0,001	
Fit	0,925	0,001	

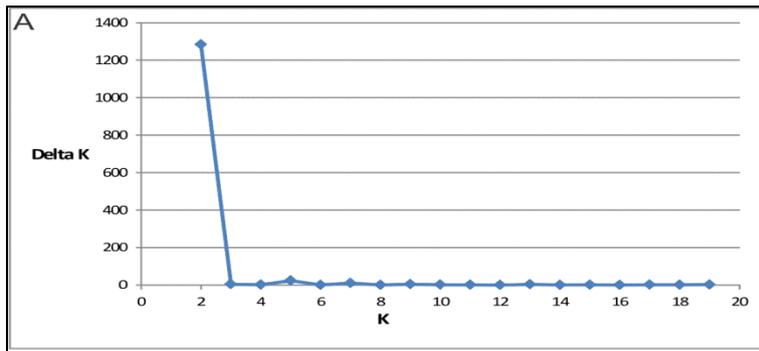


Figura A2.1 Gráficas de los valores de deltaK de STRUCTURE para el conjunto total de materiales (PMA+OP+CP)

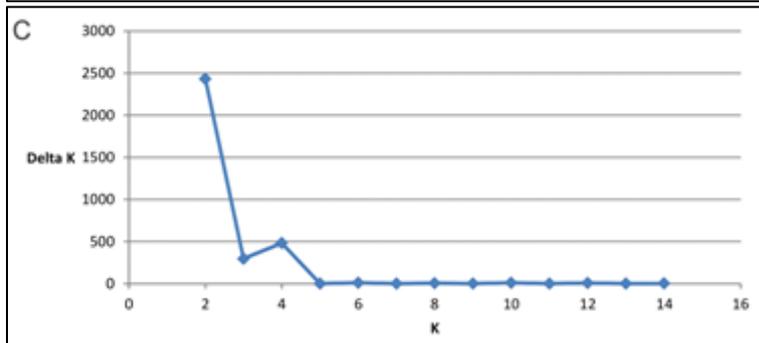
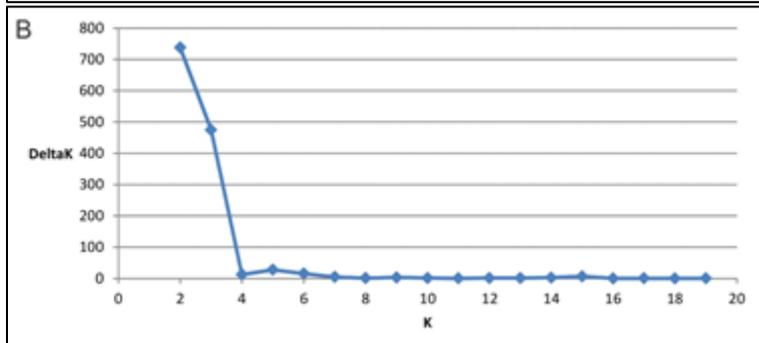
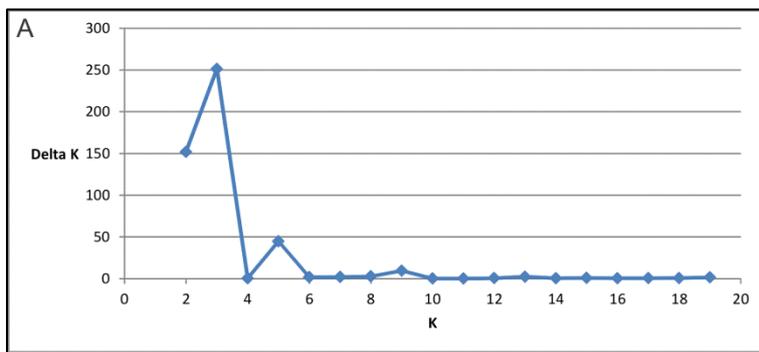


Figura A2.2 Gráficas de los valores de deltaK de STRUCTURE para las líneas que constituyen la PMA A) Información derivada de marcadores SSR; B) Información derivada de marcadores SNP; C) Información derivada de marcadores SSR+SNP.

Tabla A2.6 Líneas endocriadas asignadas a los tres grupos obtenidos por el análisis de DAPC

Línea endocriada	Grupos obtenidos por SSR	Grupos obtenidos por SNP	Grupos obtenidos por SSR + SNP
2071	1	1	1
2125	1	1	1
5075	1	1	1
B481-2	1	1	1
B481-4	1	1	1
B481-5	1	1	1
B483-1	1	1	1
B483-3	1	1	1
B487-4	1	1	1
C438	1	1	1
C454-B	1	1	1
C458B	1	1	1
C816	1	1	1
HA89	1	1	1
5035	1	2	1
B481-3	1	2	1
C813	1	2	1
C818	1	2	1
R413	1	2	1
R415	1	2	1
R422	1	2	1
R425	1	3	1
C977	1	3	3
R073	1	3	3
RHA274	1	3	3
RHA440	1	3	3
RHA299	2	2	1
5023	2	2	2
5307	2	2	2
5313	2	2	2
5393-E	2	2	2
5393-N	2	2	2
ALB2/5261	2	2	2
ALB2/5261/51084	2	2	2
C110M	2	2	2
R417	2	2	2
R419	2	2	2
R420	2	2	2
R423	2	2	2
R424	2	2	2
R426-1	2	2	2
R437	2	2	2
R439	2	2	2
R459-1	2	2	2
R459-2	2	2	2
R459-4	2	2	2
R461-3	2	2	2
R461-4	2	2	2
R465-3	2	2	2
R467-3	2	2	2

Línea endocriada	Grupos obtenidos por SSR	Grupos obtenidos por SNP	Grupos obtenidos por SSR + SNP
R467-5	2	2	2
RHA276	2	2	2
RHA801	2	2	2
5341/51084	2	2	3
5341-1	2	2	3
C610-1	2	2	3
PAC2	2	2	3
R453	2	2	3
RHA275	2	2	3
5381	2	3	3
5383	2	3	3
C208F	2	3	3
C826	2	3	3
GP365	2	3	3
R426-2	2	3	3
R461-2	2	3	3
RHA439	2	3	3
RHA441	2	3	3
2121	3	1	1
2123	3	1	1
51084/5429	3	1	1
C154B	3	1	1
R447	3	1	1
2067	3	2	1
C829B	3	2	1
2021	3	2	2
51084	3	2	2
51084/C820	3	2	2
B10	3	2	2
C820	3	2	2
RHA266	3	2	2
5049	3	2	3
5348	3	2	3
5431	3	2	3
7X1	3	2	3
B475-4	3	2	3
C494-1	3	2	3
C626-M	3	2	3
C627-M1	3	2	3
R432	3	2	3
R433	3	2	3
R463-3	3	2	3
R463-4	3	2	3
R463-5	3	2	3
C112M	3	3	1
7-1-1	3	3	2
2091	3	3	3
2095	3	3	3
2101	3	3	3
5285	3	3	3
5286	3	3	3
5289	3	3	3

Línea endocriada	Grupos obtenidos por SSR	Grupos obtenidos por SNP	Grupos obtenidos por SSR + SNP
B473-2	3	3	3
B473-4	3	3	3
B473-5	3	3	3
B473-6	3	3	3
B475-3	3	3	3
B475-6	3	3	3
B477-1	3	3	3
B481-6	3	3	3
B485-5	3	3	3
B59	3	3	3
B99	3	3	3
C1059	3	3	3
C192-1	3	3	3
C396-2B	3	3	3
C436	3	3	3
C625-U	3	3	3
C811	3	3	3
C812B	3	3	3
C817	3	3	3
c827	3	3	3
C849	3	3	3
C851	3	3	3
C853	3	3	3
F155	3	3	3
HA234	3	3	3
HA302	3	3	3
HA64	3	3	3
HA853	3	3	3
HAR2	3	3	3
HAR4	3	3	3
R429	3	3	3
R435	3	3	3
R443	3	3	3
R449	3	3	3
RK416	3	3	3

Aquellas líneas que fueron asignadas al mismo grupo independientemente del set de marcadores utilizados fueron coloreados de acuerdo al esquema de color utilizado en la Figura 2.7.

Tabla A3.1 Identificación utilizada (ID) para reconocer el género y la especie de las secuencias de GenBank utilizadas en la construcción de los árboles MP

Género-especie	ID	Género-especie	ID
<i>Actinidia chinensis</i>	Ac	<i>Elaeis guineensis</i>	EI
<i>Actinidia deliciosa</i>	Ad	<i>Eragrostis curvula</i>	Ec
<i>Actinidia eriantha</i>	An	<i>Eschscholzia californica</i>	Es
<i>Actinidia hemsleyana</i>	Ah	<i>Eucalyptus globulus</i>	Eg
<i>Adonis aestivalis</i>	Ae	<i>Euphorbia esula</i>	Ee
<i>Agrostis alba</i>	Al	<i>Euphorbia tirucalli</i>	Et
<i>Agrostis capillaris</i>	Ag	<i>Festuca arundinacea</i>	Fa
<i>Agrostis stolonifera</i>	As	<i>Fragaria vesca</i>	Fv
<i>Allium cepa</i>	Ap	<i>Ginkgo biloba</i>	Gi
<i>Amborella trichopoda</i>	Ab	<i>Glycine max</i>	Gm
<i>Antirrhinum majus</i>	Am	<i>Gnetum gnemon</i>	Gg
<i>Arabidopsis thaliana</i>	At	<i>Gossypium arboreum</i>	Ga
<i>Arachis hypogaea</i>	Ai	<i>Gossypium barbadense</i>	Gb
<i>Aristolochia fimbriata</i>	Af	<i>Gossypium hirsutum</i>	Gh
<i>Artemisia annua</i>	Aa	<i>Gossypium raimondii</i>	Gr
<i>Barnadesia spinosa</i>	Be	<i>Guizotia abyssinica</i>	Gy
<i>Beta vulgaris</i>	Bv	<i>Helianthus agrestis</i>	HA
<i>Brachypodium distachyon</i>	Bd	<i>Helianthus angustifolius</i>	HG
<i>Brassica napus</i>	Bn	<i>Helianthus annuus</i>	Ha
<i>Brassica rapa</i>	Br	<i>Helianthus anomalus</i>	Ho
<i>Cannabis sativa</i>	Cv	<i>Helianthus argophyllus</i>	Hg
<i>Capsicum annuum</i>	Ca	<i>Helianthus arizonensis</i>	HR
<i>Carica papaya</i>	Cp	<i>Helianthus atrorubens</i>	HZ
<i>Carthamus tinctorius</i>	Ct	<i>Helianthus bolanderi</i>	HB
<i>Catharanthus roseus</i>	Cq	<i>Helianthus californicus</i>	HC
<i>Cenchrus ciliaris</i>	Cx	<i>Helianthus carnosus</i>	HN
<i>Centaurea maculosa</i>	Cm	<i>Helianthus ciliaris</i>	Hc
<i>Centaurea solstitialis</i>	Co	<i>Helianthus cusickii</i>	Hk
<i>Ceratopteris richardii</i>	Ch	<i>Helianthus debilis</i>	Hd
<i>Chamaecyparis formosensis</i>	Cz	<i>Helianthus decapetalus</i>	HE
<i>Chlamydomonas reinhardtii</i>	CH	<i>Helianthus deserticola</i>	HS
<i>Cichorium endivia</i>	Ce	<i>Helianthus divaricatus</i>	HV
<i>Cichorium intybus</i>	Ci	<i>Helianthus eggertii</i>	HK
<i>Citrullus lanatus</i>	Ck	<i>Helianthus exilis</i>	He
<i>Citrus aurantiifolia</i>	cA	<i>Helianthus floridanus</i>	FI
<i>Citrus aurantium</i>	CA	<i>Helianthus giganteus</i>	Gu
<i>Citrus clementina</i>	Cc	<i>Helianthus gracilentus</i>	Gn
<i>Citrus latifolia</i>	Cf	<i>Helianthus grandiflorus</i>	Gd
<i>Citrus reshni</i>	CE	<i>Helianthus grosseserratus</i>	Gr
<i>Citrus reticulata</i>	Cr	<i>Helianthus heterophyllus</i>	He
<i>Citrus sinensis</i>	Cs	<i>Helianthus hirsutus</i>	HZ
<i>Citrus uinshu</i>	CU	<i>Helianthus laciniatus</i>	Hq
<i>Coffea canephora</i>	Cn	<i>Helianthus laevigatus</i>	HQ
<i>Cryptomeria japonica</i>	Cj	<i>Helianthus longifolius</i>	HW
<i>Cucumis melo</i>	Cl	<i>Helianthus maximiliani</i>	HX
<i>Curcuma longa</i>	Cg	<i>Helianthus microcephalus</i>	HM
<i>Cyamopsis tetragonoloba</i>	Cb	<i>Helianthus mollis</i>	Mo
<i>Cycas rumphii</i>	Cu	<i>Helianthus neglectus</i>	Ne
<i>Cynara scolymus</i>	Cy	<i>Helianthus niveus</i>	Hi
<i>Cynodon dactylon</i>	Cd	<i>Helianthus nuttallii</i>	Hn
<i>Dunaliella salina</i>	Ds	<i>Helianthus occidentalis</i>	HO
<i>Helianthus paradoxus</i>	Hx	<i>Panax ginseng</i>	Px

Género-especie	ID	Género-especie	ID
<i>Helianthus petiolaris</i>	Hp	<i>Panax quinquefolius</i>	Pq
<i>Helianthus porteri</i>	HP	<i>Panicum virgatum</i>	Pv
<i>Helianthus praecox</i>	Hr	<i>Papaver somniferum</i>	Pn
<i>Helianthus pumilus</i>	Hu	<i>Phaseolus vulgaris</i>	Pl
<i>Helianthus radula</i>	HD	<i>Physcomitrella patens</i>	Py
<i>Helianthus resinosus</i>	Hs	<i>Picea glauca</i>	Pg
<i>Helianthus salicifolius</i>	Hf	<i>Picea sitchensis</i>	Ps
<i>Helianthus schweinitzii</i>	Hw	<i>Pinus taeda</i>	Pt
<i>Helianthus silphioides</i>	HU	<i>Poncirus trifoliata</i>	Pf
<i>Helianthus simulans</i>	Hm	<i>Populus alba</i>	Pa
<i>Helianthus smithii</i>	Hh	<i>Populus nigra</i>	Po
<i>Helianthus strumosus</i>	HT	<i>Populus tomentosa</i>	Pm
<i>Helianthus tuberosus</i>	Ht	<i>Populus tremula</i>	Pe
<i>Helianthus verticillatus</i>	HL	<i>Populus tremuloides</i>	Pi
<i>Hevea brasiliensis</i>	Hb	<i>Populus trichocarpa</i>	Pr
<i>Hordeum vulgare</i>	Hv	<i>Prunus armeniaca</i>	PA
<i>Humulus lupulus</i>	HI	<i>Prunus persica</i>	Pp
<i>Ipomea batata</i>	Ib	<i>Pseudoroegneria spicata</i>	Pu
<i>Lactuca aculeata</i>	Lc	<i>Raphanus sativus</i>	Rs
<i>Lactuca altaica</i>	LI	<i>Ricinus communis</i>	Rc
<i>Lactuca canadensis</i>	Ln	<i>Saccharum officinarum</i>	So
<i>Lactuca dissecta</i>	Ld	<i>Salvia fruticosa</i>	Sf
<i>Lactuca dregeana</i>	Le	<i>Saruma henryi</i>	Sh
<i>Lactuca indica</i>	Li	<i>Selaginella moellendorffii</i>	Sm
<i>Lactuca lessertiana</i>	LL	<i>Senecio cambrensis</i>	SB
<i>Lactuca perennis</i>	Lp	<i>Solanum chacoense</i>	Sc
<i>Lactuca saligna</i>	Lg	<i>Solanum lycopersicum</i>	Sl
<i>Lactuca sativa</i>	Ls	<i>Solanum pennellii</i>	Sp
<i>Lactuca serriola</i>	Lr	<i>Solanum purheja</i>	Sj
<i>Lactuca sibirica</i>	Lb	<i>Solanum tuberosum</i>	St
<i>Lactuca tatarica</i>	LT	<i>Sorghum bicolor</i>	Sb
<i>Lactuca tenerrima</i>	Lm	<i>Stevia rebaudiana</i>	Sr
<i>Lactuca viminea</i>	LV	<i>Suaeda salsa</i>	Ss
<i>Lactuca virosa</i>	Lv	<i>Syntrichia ruralis</i>	SR
<i>Lathyrus odoratus</i>	Lo	<i>Taraxacum officinale</i>	To
<i>Lemna gibba</i>	LB	<i>Thellungiella halophila</i>	Th
<i>Limnanthes alba</i>	La	<i>Theobroma cacao</i>	Tc
<i>Liriodendron tulipifera</i>	Lt	<i>Trifolium pratense</i>	Tr
<i>Lotus japonicus</i>	Lj	<i>Triphysaria pusilla</i>	Tp
<i>Lupinus albus</i>	LA	<i>Triphysaria versicolor</i>	Tv
<i>Malus sieboldii</i>	Mi	<i>Triticum aestivum</i>	Ta
<i>Manihot esculenta</i>	Me	<i>Triticum monococcum</i>	Tm
<i>Medicago sativa</i>	Ms	<i>Triticum turgidum</i>	Tt
<i>Medicago truncatula</i>	Mt	<i>Tropaeolum majus</i>	Tj
<i>Mesembryanthemum crystallinum</i>	Mc	<i>Vaccinium corymbosum</i>	Vc
<i>Mimulus guttatus</i>	Mu	<i>Vigna unguiculata</i>	Vu
<i>Musa acuminata</i>	Ma	<i>Vitis shuttleworthii</i>	Vs
<i>Nicotiana benthamiana</i>	Nb	<i>Vitis vinifera</i>	Vv
<i>Nicotiana tabacum</i>	Nt	<i>Zantedeschia aethiopica</i>	Za
<i>Nuphar advena</i>	Na	<i>Zinnia violacea</i>	Zv
<i>Ocimum basilicum</i>	Ob	<i>Zea mays</i>	Zm
<i>Oryza sativa</i>	Os	<i>Zingiber officinale</i>	Zo

Tabla A3.2 Iniciadores utilizados para amplificar, secuenciar y genotipificar los genes candidato en la PMA

Gen candidato	Iniciadores utilizados para amplificar la región genómica en girasol	Iniciadores utilizados para secuenciar 10 líneas de girasol	Iniciadores utilizados para genotipificar las 137 LE de girasol
G21 Familia de proteínas inhibidoras Invertasa/pectin metilesterasa	FA: 5'CTCACCATTTCTCGACTCA 3' RA: 5'CACAAACACCGGAATCATTG 3' FB: 5'TCACTATGAAGGTCCATCTCCTC 3' RB: 5'GTCGCTAAAGCTCCGTTTAC 3'	FB: 5'TGAAGGTCCATCTCCTCTTATTTTC 3' RB: 5'GTCGCTAAAGCTCCGTTTAC 3'	FB: 5'TGAAGGTCCATCTCCTCTTATTTTC 3' RB: 5'GTCGCTAAAGCTCCGTTTAC 3'
G22 Familia de proteínas GH3 de respuesta a auxinas	FA: 5'AGGAGTTGTTCCCGAATGTG 3' RA: 5'GCAAATCTTTCTCGGTGTCC 3' FB: 5'TCCGACATAAAAACCGGAAC 3' RB: 5'GGAACCCGACAACCTTAACA 3'	F: 5'GGAACCTTATCCACCCGAAT 3' R: 5'GCAAATCTTTCTCGGTGTCC 3'	F: 5'TGTTTGAGTTTGAGCAACTGG 3' R: 5'TGAAGCATGTCACACTGAACAA 3'
G26 Calmodulina	F: 5'CGCCGGTAAAATACACCCTA 3' R: 5'AACGCCTCAAACAGTTTCGTC 3'	F: 5'CTGATGAGTGGTCCGAGGTT 3' R: 5'CCATCATTGTCGGCGTCTA 3'	F: 5'GAGCTGCCTTTGGAAAAGT 3' R: 5'CAGAGCTTATGGCTCCGAAC 3'
G28 Región de unión a ATP, dominio ATP-asa	F: 5'GCTATGACAGGATAGTAGTCC 3' R: 5'GCTCCATATTACGTCATCATC 3'	F: 5'CAGGATAGTAGTCCCGATGGTG 3' R: 5'TCAAAGGTTCCAGTCGCAAA 3'	
G30 Proteína tipo traumatina	F: 5'CCCGACACTTATTAAGACTCG 3' R: 5'GCTGATCCTAGTCAACAACCTGC 3'	F: 5'CCCGACACTTATTAAGACTCG 3' R: 5'GCTGATCCTAGTCAACAACCTGC 3'	F: 5'CCCGACACTTATTAAGACTCG 3' R: 5'CTGCGTGTGAGGTTCGATTTA 3'
G33 Proteína con dominio radical SAM	F: 5'TGTTTCGATGCACCTTCATA 3' R: 5'TATGGGTGACAAACGGGAAG 3'	F: 5'TTGATTATCCTGATGTTATTATCTTG 3' R: 5'TTGATGAATTATACCAACCTACCAAA 3'	F: 5'FAM-TTGCATTATCCTGATGTTATTATCTTG 3' R: 5'TTGATGAATTATACCAACCTACCAAA 3'
G34 Glicoproteína P	F: 5'ATGGTGCATCTTTGGTAGG 3' R: 5'GCAGTTGATTTCCAGCTTC 3'	F: 5'TAGGCGATCGATGCTCACTT 3' R: 5'TCAAGGCGAGTAGGGTAAGC 3'	F: 5'TGCCTTCTCATCTCAAACACA 3' R: 5'TGTAATGATCTCGGGCTTCA 3'
HeAn_315 Chaperona HSP20-like	F: 5'CCACTCAATCTCCCCAAAAA 3' R: 5'CACCTTCGCCACTTTAATCC 3'	F: 5'CCACTCAATCTCCCCAAAAA 3' R: 5'CACCTTCGCCACTTTAATCC 3'	F: 5'CCACTCAATCTCCCCAAAAA 3' R: 5'CACCTTCGCCACTTTAATCC 3'
HeAn_41021 Gamma tionina	F: 5'CGTCTTCGTTGTTGCAATCT 3' R: 5'AGGAGGAGGATGAGGAGCAT 3'	F: 5'CGTCTTCGTTGTTGCAATCT 3' R: 5'AGGAGGAGGATGAGGAGCAT 3'	R: 5'AGGAGGAGGATGAGGAGCAT 3'
HeAn_41022 Glicoproteína 1- Unión a polen	F: 5'CAGGCCAAAGTCGAGAATGT 3' R: 5'CACAAGGGAAACACAAACTG 3'	F: 5'CAGGCCAAAGTCGAGAATGT 3' R: 5'CACAAGGGAAACACAAACTG 3'	
HeAn_36589 Chaperona HSP20-like	F: 5'CTCCGCAACCTCCACTACAT 3' R: 5'GAGGCAGCTTCTCCACAGTC 3'	F: 5'CTCCGCAACCTCCACTACAT 3' R: 5'GAGGCAGCTTCTCCACAGTC 3'	R: 5'GAGGCAGCTTCTCCACAGTC 3'
HeAn_12761 Factor de virulencia	F: 5'ATGCAAGCATGTGAAGCAGT 3' R: 5'CCCCAAGCAACATGACTTTT 3'	F: 5'ATGCAAGCATGTGAAGCAGT 3' R: 5'CCCCAAGCAACATGACTTTT 3'	
HeAn_6452 ATPasa	F: 5'AGGGCTTGGGATAATCTGCT3' R: 5'GGCGCTTCAAGAACTAGCAC3'	F: 5'AGGGCTTGGGATAATCTGCT3' R: 5'GGCGCTTCAAGAACTAGCAC3'	
HeAn_35791 Desconocida	F: 5'CCGTCACCATTTTACATTG3' R: 5'TAAATTTTGGGAACCGGACA3'	F: 5'CCGTCACCATTTTACATTG3' R: 5'TAAATTTTGGGAACCGGACA3'	

Gen candidato	Iniciadores utilizados para amplificar la región genómica en girasol	Iniciadores utilizados para secuenciar 10 líneas de girasol	Iniciadores utilizados para genotipificar las 137 LE de girasol
HeAn_12562 Gen mayoritario de respuesta a Vd	F: 5' AGCCAGTGTCAAATCAACC3' R: 5' TCCATACTTACGGTTCCTCA 3'	F: 5' AGCCAGTGTCAAATCAACC3' R: 5' TCCATACTTACGGTTCCTCA 3'	F: 5' AGCCAGTGTCAAATCAACC3'
HeAn_28738 <i>Heat shock protein Hsp70</i>	F: 5' GGTGGACCATTCATCAAGG 3' R: 5' CGTGGGACAGCAACTAAA 3'	F: 5' GGTGGACCATTCATCAAGG 3' R: 5' CGTGGGACAGCAACTAAA 3'	
HeAn_36825 Desconocida	F1: 5' GCGACTAACCAAAGGCTGTT 3' F2: 5' AATTCTGGTGACGGAGCATC 3' R: 5' ACGGACGTTAATTCGAGCTG 3'		
HeAn_1756 Guanilil ciclasa	F: 5' AGGAGCAATTGGCTAACGAT 3' R: 5' CCATCACGAATTAACCATTGC 3'		
HeAn_18494 Dominio EF-hand-like	F: 5' GTGGCATCTGCCGATATGAT 3' R: 5' CCGAAACCCAAAGAACAAAA 3'		
HeAn_38637 Dominio EF-hand-like	F: 5' AATTCTGGTGACGGAGCATC 3' R: 5' ACCGCAATTCACCTGAGAAC 3'		
HeAn_10191 Lipidasa	F: 5' ATCCGGTAAACCACAAGTCC 3' F2: 5' GGTAACCACAAGTCCCATGTT 3' R: 5' CAAAGCCTTTAGTTGCCGATA 3'		
HaGLP3 Proteína tipo germina		F: 5' TCACATTCTTTCATCCTATGCTC 3' R: 5' CCAGTTCACCAAGAACAGC 3'	F: 5' GACTTTTGTGTTGCGGCCT 3' R: 5' TAGCCTCCACCAATTTTCGAC 3'
HaGLP4 Proteína tipo germina		1F: 5' TGCTTGTAACCTCCTCCTCCT 3' 1R: 5' TGTGGGAAAAGCATGATGTC 3' 2F: 5' GCTCCGAACCTTTTGGTGGT 3' 2R: 5' AGCTAAGCTTCAATCCAGAAACA 3'	F: 5' FAM-TGGCTGCAACAACCTTTCCTT 3' R: 5' TTCAATCCAGAAACAACTTcTAA 3'
HaGLP5 Proteína tipo germina		1F: 5' TCAAGTGCTATGGCTACTAATGATTT 3' 1R: 5' TCGATGTATAAATGTATGGTAAAAGAA 3' 2F: 5' TTTTCTCCAATTTTAGTTCCAA 3' 2R: 5' CACCTTCTTGTATGGTCTTGG 3'	F: 5' TGGGTTGATAATTGGCATGAT 3' R: 5' TAAACGGCAACATTTTTCATGT 3'
HaWRKY5 Factor de transcripción		F2: 5' ACACCTCTCAACTTGCACCAAA 3' R2: 5' GCCATGGATCTGAAGACAAA 3'	F: 5' FAM- CCGATCAAAGGCTCAATCTA 3' R: 5' CACATCCGCTAGTTCACACC 3'
HaWRKY7 Factor de transcripción		F3: 5' CAAGCCGCTAGAGATGGAAC 3' R3: 5' TCAACCCTGTGGTGTTTTGA 3' F4: 5' TGATTCTTCGGCATCATTTG 3' R4: 5' GGCCAAAACCTGATTCATGGT 3'	F: 5' FAM- CATTGTTGGTCAACCCTGTG 3' R: 5' AGGGAAGCATAACCATGACG 3'
HaPAL Fenilalanina amonio-liasa		F: 5' TGTGGTCTTCAAATTCATTAATAACC 3' R: 5' GGCCATTCCTAACAGGATCA 3'	F: 5' TGTGGTCTTCAAATTCATTAATAACC 3' R: 5' GGCCATTCCTAACAGGATCA 3'

Gen candidato	Iniciadores utilizados para amplificar la región genómica en girasol	Iniciadores utilizados para secuenciar 10 líneas de girasol	Iniciadores utilizados para genotipificar las 137 LE de girasol
RhoBp Proteína con dominio <i>p21-rho-binding</i>		FB1: 5' TTGAGGGATTCTAATTGTTATAGTTGA3'	F: 5' FAM- GGCTTGCGTTACATCTCTGA 3'
		RB1: 5' TTCGGGTGTTCTGCCTTTT 3'	R: 5' CCCAACTAGGAGCATTGGAA 3'
HaDRP Proteína RD22 con respuesta a deshidratación	F: 5' CAAAGGAAGCAACAACACCA 3'	F: 5' CAAAGGAAGCAACAACACCA 3'	F: 5' TCTTCAGGTAAGAAATGGCAA3'
	R: 5' TCGCTGATTGATACCCCTT 3'	R: 5' TCGCTGATTGATACCCCTT 3'	R: 5' GCATACGCGTCTTTTATTG 3'
HaCP Cistein-proteinasa	F: 5' TGCTATTGATGCTGGCAGTT 3'	F: 5' TGCTATTGATGCTGGCAGTT 3'	
	R: 5' AATCATGTTTCACATACCAAATCTT 3'	R: 5' AATCATGTTTCACATACCAAATCTT 3'	
HATRP Proteína relacionada con la formación de tumores		F: 5' TTCTTTAGGCCAACCCCTCAC 3'	F: 5' FAM- TCCAACAGTGTGTGACCTTTG 3'
		R: 5' CCCTTAATCATAAATCACGAATGTC 3'	R: 5' CATTAGTTACGTTACAAAGCTAT 3'
HaWP Proteína de pared		F: 5' CCAAACCCGATGATGATAAA 3'	F: 5' FAM- CAGGAATCACGGTCCCTAGT 3'
		R: 5' AACAAACAAAACAAGCCACATT 3'	R: 5' TGAACATGAGGGATGAGCA 3'
HaCOI1 Reguladoras de la inducción de Jasmonato	F: 5' CCGATTTGCCACTGGATAAC 3'	F: 5' CCGATTTGCCACTGGATAAC 3'	R: 5' ACACGCTGGATAGTCGTTCC 3'
	R: 5' ACACGCTGGATAGTCGTTCC 3'	R: 5' ACACGCTGGATAGTCGTTCC 3'	
HaCOI2 Reguladoras de la inducción de Jasmonato	F: 5' ATTTCTTTAGCGCGAGGTTG 3'		
	R: 5' ATTCGCAGGTGGGTGTAGAG 3'		
HA77 QTL severidad		F: 5' HEX-TGTAATCTGTATCACTTCCACC 3'	F: 5' HEX-TGTAATCTGTATCACTTCCACC 3'
		R: 5' GTTGTTCTGTTAGGTCGTTCCG 3'	R: 5' GTTGTTCTGTTAGGTCGTTCCG 3'
HA1848 QTL IEMAX		F: 5' HEX-CATCCCCTTCTGAATAGAAA 3'	F: 5' HEX-CATCCCCTTCTGAATAGAAA 3'
		R: 5' TAGGCTCGTTAAACTTACGG 3'	R: 5' TAGGCTCGTTAAACTTACGG 3'
HA2989 QTL IEMAX		F: 5' HEX –GCCTGTGTCCTCCTCAGAGTC 3'	F: 5' HEX –GCCTGTGTCCTCCTCAGAGTC 3'
		R: 5' GCCTGTGTCCTCCTCAGAGTC 3'	R: 5' GCCTGTGTCCTCCTCAGAGTC 3'
HA3039 QTL IEMAX		F: 5' FAM –GGTTGATTAAGTGTGGGTC 3'	
		R: 5' GGTTGATTAAGTGTGGGTC 3'	
HA3103 QTL IEMAX		F: 5' NED –CCTTAAAAGGCCAAACGATTCTCC 3'	F: 5' NED –CCTTAAAAGGCCAAACGATTCTCC 3'
		R: 5' CCTTAAAAGGCCAAACGATTCTCC 3'	R: 5' CCTTAAAAGGCCAAACGATTCTCC 3'
HA3847 QTL IEMAX		F: 5' NED –CATCACTTCAACATGCCTCC 3'	F: 5' NED –CATCACTTCAACATGCCTCC 3'
		R: 5' CATCACTTCAACATGCCTCC 3'	R: 5' CATCACTTCAACATGCCTCC 3'
HA4264 QTL IEMAX		F: 5' HEX –TCCCAAACCTACACCTCTAA 3'	F: 5' HEX –TCCCAAACCTACACCTCTAA 3'
		R: 5' TCCCAAACCTACACCTCTAA 3'	R: 5' TCCCAAACCTACACCTCTAA 3'

Gen candidato	Iniciadores utilizados para amplificar la región genómica en girasol	Iniciadores utilizados para secuenciar 10 líneas de girasol	Iniciadores utilizados para genotipificar las 137 LE de girasol
HA3950 QTL IEMAX		F: 5' NED –CATTCTCACCAACATAACA 3' R: 5' CATTCTCACCAACATAACA 3'	
ORS331 QTL IEMAX		F: 5' FAM –TGAAGAAGGGTTGTTGATTACAAG 3' R: 5' TGAAGAAGGGTTGTTGATTACAAG 3'	F: 5' FAM –TGAAGAAGGGTTGTTGATTACAAG 3' R: 5' TGAAGAAGGGTTGTTGATTACAAG 3'
ORS460 QTL IEMAX		F: 5' FAM –ACTCGGCTACCACCTCACAC 3' R: 5' ACTCGGCTACCACCTCACAC 3'	F: 5' FAM –ACTCGGCTACCACCTCACAC 3' 5' R: ACTCGGCTACCACCTCACAC 3'
ORS1041 QTL IEMAX		F: 5' FAM-AAACAAACCTTAATGGGGTCGTA 3' R: 5' ATATTGGCTGGTTGATGCTGAT 3'	F: 5' FAM-AAACAAACCTTAATGGGGTCGTA 3' R: 5' ATATTGGCTGGTTGATGCTGAT 3'

Tabla A3.3 Genotipos de los genes candidato

Línea	HeAn_12562	HeAn_315	HeAn_36589	HeAn_41021	G21	G22	G26	G30	G33	G34	HaPRD	HaCP	HaGLP5	HaGLP3	HaGLP4	HaWRKY7
2021	AA	AA	AA	NA	AA	AA	AA	BB	AA	AA	BB	BB	AA	AA	DD	BB
2067	AA	AA	AA	NA	AA	BB	AA	BB	AA	AA	BB	AA	BB	BB	DD	AA
2071	AA	AA	AA	BB	AA	AA	AA	AA	BB	AA	AA	BB	AA	BB	NA	AA
2091	AA	AA	BB	AA	AA	BB	AA	BB	AA	AA	AA	BB	BB	AA	DD	AA
2095	AA	AA	AA	AA	AA	AA	AA	BB	BB	AA	AA	BB	AA	AA	DD	AA
2101	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	BB	AA	BB	CC	AA
2121	AA	AA	AA	AA	AA	BB	AA	BB	BB	AA	AA	CC	AA	AA	NA	AA
2123	AA	AA	AA	BB	AA	AA	AA	BB	NA	AA	BB	BB	AA	BB	NA	AA
2125	AA	AA	AA	AA	AA	BB	AA	BB	BB	AA	BB	AA	AA	AA	CC	NA
5023	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA	NA	BB
5035	AA	BB	AA	AA	AA	AA	AA	BB	NA	AA	AA	BB	BB	AA	CC	BB
5049	BB	BB	AA	BB	AA	BB	BB	AA	AA	AA	BB	AA	BB	AA	AA	BB
5075	AA	AA	AA	CC	AA	BB	AA	AA	AA	AA	BB	BB	AA	AA	NA	BB
5285	AA	AA	AA	BB	AA	AA	AA	BB	AA	AA	AA	BB	AA	BB	NA	BB
5286	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	AA	BB	BB	CC	BB
5289	AA	AA	AA	NA	AA	BB	AA	BB	BB	AA	AA	DD	BB	BB	CC	AA
5307	AA	BB	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	BB	DD	BB
5313	AA	AA	AA	AA	AA	AA	AA	BB	AA	BB	BB	AA	BB	BB	DD	BB
5341,1	AA	BB	AA	CC	AA	AA	AA	CC	DD	AA	CC	AA	AA	AA	NA	BB
5348	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	BB	BB	AA	AA	BB
5381	AA	AA	AA	EE	AA	BB	AA	BB	AA	AA	AA	AA	BB	BB	DD	BB
5383	AA	AA	AA	BB	AA	AA	AA	AA	BB	AA	AA	AA	BB	BB	AA	AA
5431	DD	AA	AA	AA	AA	AA	BB	AA	AA	AA	BB	AA	BB	AA	AA	BB
51084	BB	AA	AA	BB	AA	BB	BB	BB	BB	BB	AA	AA	BB	AA	DD	AA
51084/5429	BB	BB	AA	AA	AA	AA	BB	AA	AA	AA	BB	AA	AA	AA	CC	BB
51084/C820	BB	AA	AA	BB	AA	BB	AA	AA	BB	BB	BB	AA	AA	AA	DD	AA
5341/51084	BB	AA	AA	AA	AA	BB	BB	AA	AA	AA	AA	AA	AA	AA	CC	BB
5393-E	AA	AA	AA	BB	AA	AA	AA	BB	AA	AA	AA	BB	BB	BB	CC	BB
5393-N	AA	AA	AA	BB	AA	AA	AA	BB	CC	BB						
7-1-1	AA	AA	AA	BB	BB	AA	AA	AA	AA	CC	AA	AA	AA	AA	CC	BB
7X1	NA	AA	AA	AA	BB	AA	AA	AA	AA	CC	AA	AA	AA	AA	DD	NA
ALB2/5261	NA	AA	AA	AA	AA	BB	BB	AA	BB	AA	AA	AA	AA	AA	CC	AA
ALB2/5261/51084	BB	BB	AA	BB	AA	BB	BB	BB	BB	AA	AA	AA	AA	AA	CC	AA
B10	AA	BB	AA	BB	AA	BB	AA	AA	AA	CC	BB	AA	AA	BB	AA	BB

Línea	HeAn_12562	HeAn_315	HeAn_36589	HeAn_41021	G21	G22	G26	G30	G33	G34	HaPRD	HaCP	HaGLP5	HaGLP3	HaGLP4	HaWRKY7
B473-2	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA	CC	BB
B473-4	AA	AA	AA	BB	AA	AA	AA	BB	AA	AA	AA	AA	BB	AA	CC	BB
B473-5	AA	AA	AA	BB	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA	CC	BB
B473-6	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	BB	AA	AA	CC	AA
B475-3	AA	BB	AA	CC	AA	BB	AA	AA	AA	AA	AA	AA	AA	AA	CC	BB
B475-4	AA	BB	AA	NA	AA	BB	AA	BB	AA	AA	AA	AA	BB	AA	NA	BB
B475-6	BB	AA	AA	CC	AA	AA	BB	AA	AA	AA	AA	AA	AA	AA	CC	BB
B477-1	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	AA	AA	AA	DD	AA
B481-2	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	BB	AA	AA	CC	AA
B481-3	BB	AA	AA	BB	AA	AA	BB	AA	AA	AA	AA	BB	AA	AA	CC	AA
B481-4	AA	AA	AA	AA	AA	BB	BB	AA	BB	AA	BB	BB	AA	AA	AA	AA
B481-5	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	BB	AA	AA	CC	AA
B481-6	AA	BB	AA	NA	AA	BB	AA	BB	NA	AA	AA	AA	AA	AA	AA	AA
B483-1	NA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	BB	AA	AA	DD	AA
B483-3	NA	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	CC	AA
B485-5	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	CC	BB
B487-4	AA	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	CC	AA
B59	AA	BB	AA	AA	AA	AA	AA	AA	AA	CC	AA	AA	BB	AA	CC	BB
B99	AA	AA	AA	AA	AA	AA	AA	AA	AA	CC	AA	BB	BB	BB	CC	AA
C1059	AA	AA	AA	CC	AA	AA	AA	AA	NA	AA	BB	BB	AA	AA	CC	BB
C110M	BB	AA	AA	AA	AA	AA	BB	AA	BB	AA	CC	AA	AA	AA	CC	AA
C112M	AA	AA	AA	BB	AA	AA	AA	BB	NA	AA	AA	BB	AA	BB	DD	BB
C154B	AA	AA	AA	BB	AA	AA	AA	AA	AA	CC	BB	BB	AA	AA	CC	BB
C192-1	AA	AA	CC	AA	AA	AA	AA	BB	AA	AA	AA	AA	AA	BB	NA	NA
C208F	NA	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	AA	BB	CC	BB
C396-2B	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA	BB	BB	AA	AA	DD	BB
C436	AA	AA	AA	NA	AA	AA	AA	AA	AA	AA	BB	AA	BB	BB	CC	BB
C438	AA	BB	AA	AA	AA	AA	AA	BB	NA	BB	AA	BB	AA	AA	CC	AA
C454-B	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	AA	BB	AA	AA	NA	AA
C458B	AA	AA	AA	AA	AA	BB	AA	AA	BB	AA	BB	BB	AA	BB	NA	AA
C494-1	AA	AA	AA	BB	AA	BB	AA	BB	AA	AA	BB	AA	BB	AA	NA	BB
C610	EE	AA	AA	BB	AA	AA	BB	AA	AA	BB	BB	EE	BB	BB	AA	BB
C625-U	AA	AA	AA	EE	AA	BB	AA	BB	AA	AA	BB	AA	AA	CC	DD	BB
C626-M	AA	AA	AA	BB	AA	BB	AA	AA	AA	AA	AA	AA	BB	AA	DD	BB
C627-1M	BB	AA	AA	NA	AA	AA	BB	BB	AA	AA	BB	AA	AA	BB	DD	BB

Línea	HeAn_12562	HeAn_315	HeAn_36589	HeAn_41021	G21	G22	G26	G30	G33	G34	HaPRD	HaCP	HaGLP5	HaGLP3	HaGLP4	HaWRKY7
R426-2	CC	BB	AA	CC	AA	AA	AA	BB	AA	AA	AA	AA	BB	AA	CC	BB
R429	AA	BB	AA	BB	AA	AA	AA	BB	CC	BB						
R432	AA	BB	AA	CC	AA	AA	BB	AA	CC	BB						
R433	AA	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	AA	AA	AA	NA	BB
R435	NA	BB	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	CC	BB
R437	AA	AA	AA	FF	AA	AA	AA	AA	AA	CC	AA	AA	AA	AA	CC	BB
R439	AA	AA	AA	BB	AA	AA	AA	AA	AA	CC	AA	AA	BB	AA	DD	BB
R443	AA	AA	AA	BB	AA	BB	BB	AA	AA	BB						
R447	BB	AA	AA	AA	AA	BB	BB	BB	BB	CC	AA	AA	AA	AA	CC	AA
R449	BB	BB	AA	BB	AA	AA	BB	DD	AA	AA	AA	AA	BB	BB	BB	BB
R453	AA	AA	AA	BB	AA	AA	AA	AA	AA	CC	AA	AA	BB	AA	DD	BB
R459-1	AA	BB	AA	BB	AA	AA	BB	AA	NA	BB						
R459-2	AA	BB	AA	BB	AA	AA	AA	BB	AA	AA	AA	AA	BB	AA	CC	AA
R459-4	AA	BB	AA	BB	AA	AA	AA	AA	CC	AA						
R461-2	AA	BB	AA	AA	AA	AA	AA	AA	BB	AA	AA	AA	AA	BB	CC	AA
R461-3	BB	BB	AA	BB	AA	AA	BB	AA	AA	AA	AA	AA	BB	BB	CC	AA
R461-4	BB	AA	AA	BB	AA	AA	BB	AA	AA	AA	AA	AA	BB	BB	CC	AA
R463-3	BB	BB	AA	CC	AA	AA	BB	AA	AA	AA	AA	AA	AA	AA	CC	BB
R463-4	AA	BB	AA	CC	AA	AA	AA	AA	CC	BB						
R463-5	AA	BB	AA	BB	AA	AA	BB	BB	NA	BB						
R465-3	AA	BB	AA	AA	AA	BB	AA	BB	AA	AA	AA	AA	BB	AA	CC	BB
R467-3	AA	BB	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	AA	CC	AA
R467-5	AA	AA	AA	CC	AA	AA	AA	AA	CC	AA						
RHA266	AA	BB	AA	CC	AA	AA	AA	BB	DD	AA						
RHA274	AA	AA	AA	DD	AA	AA	AA	BB	AA	BB	AA	AA	BB	BB	AA	BB
RHA275	AA	BB	AA	CC	AA	BB	AA	BB	NA	BB	AA	AA	BB	BB	EE	AA
RHA276	BB	BB	AA	BB	AA	AA	BB	BB	AA	AA	CC	AA	BB	AA	EE	AA
RHA299	NA	AA	AA	BB	AA	BB	BB	BB	AA	AA	BB	AA	AA	AA	EE	BB
RHA439	NA	AA	AA	AA	AA	BB	BB	BB	AA	AA	AA	AA	AA	AA	DD	BB
RHA440	AA	BB	NA	NA	AA	AA	AA	BB	AA	AA	AA	BB	AA	AA	AA	BB
RHA441	AA	AA	AA	NA	AA	BB	AA	BB	AA	BB	AA	AA	BB	AA	DD	BB
RHA801	BB	BB	NA	CC	AA	AA	BB	AA	AA	AA	AA	AA	BB	BB	EE	BB
RK416	AA	AA	AA	DD	AA	AA	AA	BB	CC	AA	BB	BB	AA	BB	DD	BB

Tabla A3.3 Genotipos de los genes candidato (continuación)

Línea	HaCP	HaWRKY5	HaTRP	RHOBP	HaPAL	HACOI_1	HA77	HA1848	HA2989	HA3103	HA3847	HA4264	ORS331	ORS460	ORS1041
2021	AA	BB	AA	AA	AA	AA	AA	EE	AA	AA	DD	AA	AA	BB	CC
2067	AA	AA	BB	AA	BB	AA	AA	JJ	AA	DD	AA	BB	AA	DD	BB
2071	AA	BB	AA	AA	AA	BB	CC	AA	AA	BB	DD	FF	AA	AA	EE
2091	AA	AA	AA	AA	BB	AA	CC	DD	AA	BB	AA	AA	AA	AA	AA
2095	AA	AA	AA	AA	AA	BB	CC	GG	AA	BB	AA	AA	AA	AA	DD
2101	NA	BB	NA	AA	AA	AA	CC	GG	AA	BB	DD	AA	AA	DD	CC
2121	AA	AA	AA	AA	AA	AA	CC	KK	AA	AA	DD	AA	AA	BB	BB
2123	AA	NA	AA	AA	AA	AA	CC	NA	AA	AA	AA	AA	NA	BB	BB
2125	AA	AA	AA	AA	AA	NA	CC	II	NA	AA	AA	AA	AA	BB	BB
5023	AA	DD	AA	AA	AA	AA	CC	JJ	AA	AA	DD	DD	AA	DD	BB
5035	AA	AA	AA	AA	AA	BB	CC	KK	AA	AA	BB	FF	AA	BB	BB
5049	BB	BB	AA	AA	AA	AA	BB	EE	AA	BB	BB	FF	AA	AA	DD
5075	AA	AA	AA	AA	AA	AA	CC	GG	AA	AA	BB	AA	AA	BB	BB
5285	AA	BB	AA	AA	AA	AA	CC	EE	AA	NA	NA	AA	NA	BB	BB
5286	AA	BB	AA	AA	AA	BB	CC	DD	AA	BB	BB	AA	AA	AA	AA
5289	AA	NA	AA	AA	AA	NA	CC	BB	AA	DD	AA	BB	AA	BB	EE
5307	AA	BB	AA	AA	AA	AA	CC	FF	AA	AA	DD	DD	AA	AA	BB
5313	AA	DD	BB	AA	BB	AA	CC	JJ	AA	DD	BB	CC	DD	AA	BB
5341,1	AA	AA	AA	AA	AA	AA	CC	II	AA	AA	DD	AA	AA	BB	BB
5348	AA	DD	AA	AA	AA	DD	AA	EE	NA	DD	NA	BB	BB	NA	CC
5381	AA	AA	AA	CC	AA	AA	CC	DD	NA	AA	NA	EE	AA	NA	AA
5383	AA	BB	AA	NA	BB	NA	CC	CC	AA	BB	DD	AA	AA	AA	AA
5431	BB	DD	AA	AA	AA	AA	BB	II	AA	BB	BB	FF	AA	AA	BB
51084	BB	BB	AA	AA	AA	BB	CC	II	DD	AA	BB	AA	AA	BB	BB
51084/5429	AA	BB	AA	AA	AA	BB	CC	BB	AA	NA	DD	BB	DD	DD	EE
51084/C820	BB	BB	AA	AA	AA	EE	CC	HH	AA	BB	BB	FF	AA	BB	BB
5341/51084	BB	DD	AA	AA	AA	BB	CC	EE	DD	BB	BB	FF	NA	AA	DD
5393-E	BB	BB	AA	CC	AA	AA	CC	EE	AA	BB	BB	FF	AA	AA	BB
5393-N	AA	BB	AA	AA	AA	AA	CC	EE	AA	AA	DD	AA	AA	BB	BB
7-1-1	AA	BB	AA	AA	AA	AA	CC	HH	AA	AA	DD	AA	AA	BB	CC
7X1	AA	BB	AA	AA	AA	AA	CC	HH	BB	DD	DD	BB	AA	DD	CC
ALB2/5261	BB	BB	AA	AA	AA	EE	CC	II	AA	NA	BB	FF	AA	AA	BB
ALB2/5261/51084	BB	AA	AA	BB	AA	EE	CC	BB	DD	BB	BB	FF	DD	AA	EE
B10	AA	BB	AA	AA	AA	BB	CC	HH	AA	NA	DD	BB	BB	AA	CC

Línea	HaCP	HaWRKY5	HaTRP	RHOBP	HaPAL	HACOI_1	HA77	HA1848	HA2989	HA3103	HA3847	HA4264	ORS331	ORS460	ORS1041
B473-2	AA	BB	BB	AA	BB	AA	CC	BB	DD	CC	AA	AA	BB	DD	EE
B473-4	AA	DD	AA	AA	AA	AA	CC	GG	AA	BB	AA	FF	BB	AA	BB
B473-5	AA	BB	AA	AA	AA	AA	AA	GG	DD	BB	AA	FF	BB	DD	BB
B473-6	AA	BB	AA	AA	AA	BB	AA	GG	AA	BB	DD	FF	AA	AA	BB
B475-3	AA	BB	AA	AA	AA	AA	CC	HH	DD	CC	AA	AA	AA	DD	BB
B475-4	AA	BB	AA	AA	AA	AA	CC	HH	AA	BB	AA	FF	AA	AA	BB
B475-6	BB	BB	AA	AA	AA	NA	CC	HH	DD	CC	BB	AA	BB	AA	BB
B477-1	AA	BB	AA	AA	AA	BB	CC	HH	AA	CC	AA	AA	AA	DD	BB
B481-2	AA	BB	AA	AA	AA	BB	CC	AA	AA	BB	DD	FF	AA	AA	EE
B481-3	BB	BB	AA	BB	AA	AA	CC	JJ	AA	AA	BB	DD	AA	DD	BB
B481-4	NA	BB	AA	BB	AA	BB	CC	II	AA	AA	DD	AA	BB	BB	BB
B481-5	AA	BB	AA	AA	AA	BB	CC	AA	AA	BB	DD	FF	AA	AA	EE
B481-6	AA	BB	AA	BB	AA	AA	AA	FF	AA	BB	DD	FF	AA	AA	BB
B483-1	AA	BB	AA	BB	AA	BB	CC	AA	AA	BB	DD	FF	AA	AA	EE
B483-3	AA	BB	AA	AA	AA	BB	CC	AA	AA	BB	DD	FF	AA	AA	EE
B485-5	AA	AA	AA	AA	AA	BB	CC	DD	DD	BB	AA	FF	AA	AA	AA
B487-4	AA	BB	AA	AA	AA	NA	CC	AA	AA	BB	DD	FF	AA	AA	EE
B59	AA	AA	AA	AA	BB	BB	CC	EE	AA	NA	DD	DD	DD	AA	DD
B99	AA	DD	AA	NA	AA	CC	AA	BB	AA	NA	AA	DD	AA	DD	EE
C1059	AA	CC	AA	BB	BB	DD	BB	KK	AA	DD	BB	BB	AA	DD	BB
C110M	BB	AA	AA	AA	AA	BB	CC	II	AA	AA	BB	AA	AA	AA	BB
C112M	BB	AA	AA	AA	AA	BB	AA	EE	AA	AA	BB	BB	AA	BB	BB
C154B	AA	BB	AA	AA	AA	AA	CC	DD	AA	AA	DD	AA	AA	BB	AA
C192-1	AA	DD	AA	CC	NA	GG	CC	HH	NA	CC	NA	AA	AA	NA	CC
C208F	BB	AA	AA	BB	AA	BB	CC	JJ	AA	BB	BB	FF	AA	AA	BB
C396-2B	AA	NA	AA	AA	AA	AA	AA	FF	AA	AA	DD	AA	AA	BB	CC
C436	AA	BB	AA	BB	BB	BB	CC	KK	AA	CC	DD	AA	AA	DD	BB
C438	AA	AA	AA	AA	AA	BB	CC	II	AA	NA	AA	EE	AA	BB	BB
C454-B	AA	BB	AA	AA	AA	BB	CC	AA	AA	BB	NA	FF	AA	AA	EE
C458B	AA	NA	AA	AA	AA	AA	CC	II	AA	AA	DD	AA	AA	BB	BB
C494-1	AA	BB	AA	AA	AA	BB	CC	EE	AA	CC	DD	AA	BB	DD	CC
C610	AA	EE	AA	AA	AA	HH	CC	II	CC	DD	CC	BB	CC	NA	CC
C625-U	AA	NA	AA	BB	AA	BB	CC	EE	DD	BB	BB	FF	DD	AA	AA
C626-M	AA	AA	NA	AA	BB	CC	CC	HH	AA	BB	AA	GG	AA	DD	CC
C627-1M	AA	CC	AA	AA	AA	BB	AA	NA	AA	AA	BB	AA	AA	BB	CC

Línea	HaCP	HaWRKY5	HaTRP	RHOBP	HaPAL	HACOI_1	HA77	HA1848	HA2989	HA3103	HA3847	HA4264	ORS331	ORS460	ORS1041
C811	AA	AA	AA	AA	AA	BB	AA	EE	AA	DD	BB	BB	DD	BB	AA
C812B	BB	BB	AA	BB	BB	BB	AA	II	AA	AA	DD	BB	AA	DD	CC
C813	AA	AA	AA	AA	AA	AA	AA	II	AA	AA	DD	AA	AA	BB	BB
C816	AA	BB	NA	AA	AA	AA	CC	BB	AA	BB	DD	FF	DD	AA	EE
C817	AA	BB	AA	AA	AA	BB	BB	EE	AA	DD	DD	BB	AA	NA	AA
C818	AA	DD	AA	AA	AA	FF	CC	BB	AA	AA	BB	AA	DD	BB	EE
C820	AA	NA	BB	AA	BB	BB	CC	BB	DD	DD	DD	BB	AA	AA	EE
C826-M	BB	BB	AA	AA	AA	AA	AA	HH	AA	AA	NA	DD	AA	NA	BB
c827	AA	BB	AA	AA	AA	CC	AA	DD	AA	BB	BB	AA	AA	AA	AA
C829B	AA	BB	BB	AA	BB	BB	CC	EE	DD	DD	DD	CC	AA	BB	DD
C849	AA	BB	BB	AA	BB	BB	AA	HH	AA	DD	DD	BB	AA	DD	CC
C851B	AA	AA	AA	AA	AA	DD	BB	CC	AA	CC	BB	AA	AA	DD	EE
C853B	AA	BB	AA	AA	AA	CC	CC	II	AA	AA	AA	DD	AA	BB	BB
C977B	AA	BB	AA	AA	AA	GG	AA	JJ	DD	AA	DD	AA	AA	BB	BB
F155B	AA	BB	AA	AA	AA	AA	CC	DD	AA	NA	BB	EE	AA	BB	AA
GP365	AA	BB	AA	CC	AA	EE	CC	II	DD	BB	BB	FF	AA	AA	BB
HA234	AA	AA	AA	AA	AA	DD	BB	CC	AA	CC	BB	AA	AA	DD	EE
HA302	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
HA64	BB	BB	AA	AA	AA	BB	CC	HH	DD	CC	AA	AA	BB	DD	BB
HA853	AA	BB	BB	AA	BB	BB	AA	HH	AA	DD	DD	BB	AA	DD	CC
HA89	AA	AA	AA	AA	AA	AA	CC	HH	AA	AA	DD	GG	AA	BB	BB
HAR2	AA	AA	AA	AA	BB	DD	NA	LL	AA	CC	AA	AA	BB	DD	CC
HAR4	AA	AA	AA	AA	AA	NA	CC	DD	AA	BB	BB	FF	AA	AA	NA
PAC2	AA	AA	AA	AA	BB	CC	CC	EE	DD	CC	BB	AA	DD	DD	CC
R073	AA	BB	NA	AA	AA	AA	CC	GG	AA	AA	BB	FF	AA	DD	CC
R413	AA	BB	BB	AA	BB	BB	CC	BB	AA	DD	BB	BB	AA	AA	EE
R415	AA	BB	BB	AA	BB	BB	CC	BB	AA	DD	BB	BB	AA	AA	EE
R417	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
R419	AA	AA	AA	AA	AA	AA	CC	JJ	DD	AA	BB	AA	AA	BB	BB
R420	AA	NA	AA	BB	AA	AA	CC	JJ	AA	AA	DD	DD	AA	AA	BB
R422	AA	BB	BB	AA	BB	BB	CC	BB	AA	DD	BB	BB	AA	BB	EE
R423	AA	CC	AA	AA	AA	AA	CC	II	AA	AA	DD	DD	AA	BB	BB
R424	AA	DD	AA	NA	AA	AA	CC	JJ	AA	DD	BB	BB	AA	DD	BB
R425	AA	BB	AA	AA	BB	AA	CC	II	AA	BB	BB	FF	AA	AA	BB
R426-1	AA	BB	AA	CC	AA	AA	CC	JJ	AA	AA	NA	DD	AA	DD	BB

Línea	HaCP	HaWRKY5	HaTRP	RHOBP	HaPAL	HACOI_1	HA77	HA1848	HA2989	HA3103	HA3847	HA4264	ORS331	ORS460	ORS1041
R426-2	AA	BB	AA	CC	AA	AA	CC	JJ	AA	AA	NA	DD	AA	DD	BB
R429	AA	AA	BB	AA	BB	BB	CC	BB	DD	DD	AA	BB	AA	AA	EE
R432	AA	BB	AA	BB	AA	CC	CC	EE	AA	BB	AA	FF	DD	AA	DD
R433	AA	BB	AA	AA	AA	CC	CC	EE	DD	AA	AA	DD	DD	BB	CC
R435	AA	BB	AA	AA	AA	AA	CC	EE	AA	AA	AA	DD	DD	BB	CC
R437	AA	BB	AA	AA	AA	BB	CC	JJ	AA	AA	DD	DD	AA	DD	BB
R439	AA	BB	AA	BB	BB	AA	CC	II	AA	DD	DD	BB	AA	DD	BB
R443	AA	BB	AA	BB	AA	AA	AA	NA	AA	AA	DD	AA	AA	BB	CC
R447	BB	AA	AA	BB	AA	AA	BB	EE	AA	BB	BB	FF	DD	AA	DD
R449	BB	AA	AA	AA	AA	DD	CC	NA	AA	DD	BB	BB	BB	DD	DD
R453	AA	BB	AA	AA	BB	AA	CC	II	AA	DD	DD	BB	AA	DD	BB
R459-1	AA	CC	AA	AA	AA	AA	CC	KK	AA	AA	DD	DD	BB	DD	BB
R459-2	AA	NA	BB	AA	BB	AA	CC	BB	AA	DD	DD	BB	AA	DD	EE
R459-4	AA	CC	BB	AA	BB	AA	CC	BB	AA	DD	DD	BB	AA	DD	EE
R461-2	AA	CC	AA	AA	AA	AA	AA	JJ	AA	BB	DD	FF	AA	AA	BB
R461-3	BB	CC	AA	AA	AA	AA	CC	GG	AA	BB	BB	FF	BB	AA	BB
R461-4	BB	CC	AA	AA	AA	AA	AA	JJ	AA	AA	BB	DD	AA	DD	BB
R463-3	BB	DD	AA	AA	AA	AA	CC	EE	AA	BB	BB	FF	BB	AA	DD
R463-4	AA	BB	BB	AA	BB	AA	CC	BB	NA	DD	NA	BB	AA	NA	EE
R463-5	AA	BB	BB	AA	BB	AA	CC	BB	AA	BB	DD	FF	BB	AA	EE
R465-3	AA	AA	AA	AA	AA	AA	CC	JJ	AA	AA	DD	DD	AA	DD	BB
R467-3	AA	BB	AA	AA	AA	AA	CC	JJ	DD	CC	DD	AA	AA	DD	BB
R467-5	AA	BB	AA	AA	AA	AA	CC	JJ	AA	AA	DD	DD	AA	DD	BB
RHA266	AA	BB	AA	AA	AA	BB	AA	BB	AA	BB	DD	FF	AA	AA	EE
RHA274	NA	BB	AA	AA	AA	BB	CC	EE	AA	DD	BB	BB	DD	AA	DD
RHA275	AA	DD	AA	AA	AA	BB	CC	JJ	AA	AA	BB	AA	DD	BB	BB
RHA276	AA	DD	AA	AA	AA	FF	CC	CC	AA	DD	BB	BB	AA	AA	EE
RHA299	BB	AA	AA	AA	AA	FF	CC	KK	AA	AA	DD	DD	AA	CC	BB
RHA439	AA	BB	AA	AA	AA	BB	AA	HH	AA	AA	BB	DD	AA	NA	BB
RHA440	AA	BB	NA	AA	AA	CC	CC	DD	AA	BB	BB	FF	AA	AA	BB
RHA441	NA	DD	AA	AA	BB	BB	CC	JJ	AA	AA	BB	AA	DD	BB	BB
RHA801	AA	BB	AA	AA	AA	BB	CC	JJ	AA	AA	BB	DD	AA	DD	BB
RK416	AA	AA	BB	AA	AA	AA	CC	HH	AA	BB	DD	FF	AA	BB	CC

Tabla A4.1 Medias ajustadas de los cuatro caracteres fenotípicos

Línea	AUDPCinc	IEmax	Pinc	SEV	Línea	AUDPCinc	IEmax	Pinc	SEV
2021	3,90	43,54	20,44	50,35	B485-5	9,32	77,20	15,96	78,73
2067	9,78	85,52	16,94	64,76	B487-4	5,71	57,32	18,13	75,58
2071	5,25	46,23	18,37	54,82	B59	4,31	52,30	21,06	40,03
2091	9,45	70,45	17,47	61,12	B99	5,36	39,17	18,97	44,42
2095	7,38	65,57	17,12	60,38	C1059	5,25	52,47	17,37	55,35
2101	6,09	62,18	19,14	46,34	C110M	8,02	75,55	18,46	69,10
2121	6,37	68,93	19,60	49,09	C112M	5,12	50,03	18,88	58,02
2123	6,01	59,88	18,41	55,88	C154B	7,49	60,28	17,32	65,88
2125	5,47	58,54	19,06	54,38	C192-1	3,49	41,18	19,65	42,73
5023	7,47	76,17	17,49	72,44	C208F	4,48	30,78	19,25	57,23
5035	5,31	34,27	17,63	73,34	C396-2B	5,60	43,48	19,02	58,15
5049	4,92	42,62	17,76	73,68	C436	3,92	37,24	20,53	41,13
5075	4,45	39,78	19,07	58,66	C438	2,20	23,95	20,84	31,50
5285	5,11	42,72	18,83	63,38	C454-B	5,49	55,04	19,18	55,83
5286	6,60	49,22	15,53	74,14	C458B	6,54	63,39	18,11	63,33
5289	11,23	87,04	14,65	78,67	C494-1	5,96	51,32	17,93	65,50
5307	7,51	58,40	17,73	62,50	C610	4,24	23,24	17,54	73,29
5313	4,86	40,41	19,42	44,77	C625-U	5,24	50,62	18,51	49,59
5341,1	3,24	26,98	18,53	38,16	C626-M	6,63	47,52	19,51	80,69
5348	4,37	37,84	20,04	27,15	C627-1M	4,57	36,84	19,64	67,29
5381	4,13	26,76	19,42	43,94	C811	6,25	55,18	19,13	49,49
5383	2,05	18,48	24,55	23,52	C812	6,64	49,18	17,76	64,28
5431	9,68	82,34	17,82	73,68	C813	5,63	50,54	18,63	58,61
7-1-1	1,41	21,67	14,24	24,35	C816	4,38	31,37	19,95	48,12
51084	3,81	32,13	19,88	45,57	C817	7,01	80,51	20,13	49,90
51084/5429	2,06	39,41	22,47	31,35	C818	6,29	50,68	18,83	63,46
51084/C820	4,03	46,21	21,20	49,44	C820	5,19	53,08	19,87	59,31
5341/51084	7,82	73,87	17,68	56,21	C826-M	7,58	62,09	17,08	73,84
5393-E	7,48	68,16	16,75	63,99	C827	4,83	40,26	19,50	57,74
5393-N	7,31	66,80	18,41	70,57	C829B	3,00	25,57	20,89	37,15
7X1	5,58	52,03	19,81	45,44	C849	4,87	52,15	20,78	36,22
ALB2/5261	0,86	22,57	24,79	22,07	C851B	7,94	61,91	17,49	66,91
ALB2/5261/51084	4,02	29,00	20,07	39,04	C853B	8,35	78,27	16,21	63,28
B10	2,90	23,34	22,02	41,76	C977B	7,16	60,33	17,96	77,35
B473-2	7,24	73,57	18,57	55,81	F155B	3,60	33,76	22,88	41,91
B473-4	6,52	68,58	18,84	66,15	GP365	6,00	57,54	17,92	57,75
B473-5	4,15	48,54	20,98	51,22	HA234	8,22	61,99	15,44	81,93
B473-6	4,60	42,49	19,31	70,73	HA64	5,00	45,53	19,08	54,96
B475-3	8,29	71,20	18,00	83,94	HA853	3,71	27,44	21,66	42,86
B475-4	6,74	73,80	20,33	53,97	HA89	4,53	42,67	18,41	58,10
B475-6	6,83	69,60	19,35	51,15	HAR2	7,26	73,62	19,34	50,27
B477-1	4,18	69,01	22,86	30,27	HAR4	6,75	62,04	16,93	67,54
B481-2	4,88	65,50	21,35	41,39	PAC2	2,05	29,48	22,34	37,88
B481-3	9,35	86,20	17,07	76,53	R073	7,84	61,76	16,19	73,11
B481-4	6,17	44,82	16,69	66,81	R413	6,35	55,09	20,08	48,05
B481-5	8,79	76,70	16,58	77,88	R415	6,24	58,61	19,38	54,01
B481-6	6,72	77,47	19,24	58,15	R419	7,58	65,75	17,35	69,32
B483-1	4,57	54,11	20,75	55,34	R420	8,58	82,96	18,59	67,13
B483-3	5,58	50,65	18,85	56,02	R422	4,75	51,35	19,50	50,51

Línea	AUDPCinc	IEmax	Pinc	SEV
R423	5,85	42,68	17,42	64,09
R424	8,13	72,39	16,92	77,22
R425	5,02	38,66	18,29	70,29
R426-1	5,82	51,15	17,58	79,81
R426-2	8,46	56,98	16,52	74,09
R429	5,27	54,87	19,94	56,10
R432	8,92	76,23	15,39	76,19
R433	6,10	75,24	20,13	54,08
R435	9,36	87,89	17,37	69,95
R437	7,43	85,47	19,95	47,31
R439	9,91	80,25	16,84	78,58
R443	4,49	46,84	21,72	63,57
R447	3,06	24,86	20,50	27,79
R449	3,47	43,86	21,56	39,91
R453	9,20	82,82	17,31	66,51
R459-1	6,08	64,60	21,25	46,50
R459-2	2,03	20,87	19,27	34,59
R459-4	10,54	89,02	15,08	84,38
R461-2	5,88	51,92	18,32	65,10
R461-3	6,15	67,76	20,69	55,15
R461-4	8,04	73,34	18,10	73,18
R463-3	9,32	67,20	14,80	86,02
R463-4	7,41	81,03	19,36	58,92
R463-5	5,26	52,04	19,52	65,27
R465-3	8,10	69,00	16,50	76,71
R467-3	10,68	84,60	15,22	82,05
R467-5	9,31	78,86	16,27	68,59
RHA266	3,57	31,95	19,35	51,83
RHA274	3,04	27,32	21,06	36,16
RHA275	6,35	60,95	19,33	52,25
RHA276	6,11	70,35	19,55	58,95
RHA299	3,70	28,27	21,61	50,63
RHA439	3,85	37,81	19,06	57,15
RHA440	6,21	63,55	19,12	44,55
RHA441	2,52	21,84	21,53	32,94
RHA801	3,75	31,88	20,79	50,57
RK416	2,37	23,30	21,83	35,27

Tabla A4.2 Resultados de los análisis de asociación

Marcador	Incidencia				AUDPC incidencia				Severidad				Período de incubación				MAF
	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	
G22	0,2026	0,40	0,00	0,18	0,1274	0,30	1,05	0,23	0,0895	0,31	1,51	0,31	0,0323	0,52	5,30	1,21	0,27
G26	0,0965	0,29	4,44	0,82	0,2136	0,38	0,95	0,84	0,3904	0,46	0,00	0,08	0,5307	0,73	0,00	1,47	0,19
G30	0,5783	0,56	0,00	0,13	0,2737	0,41	0,00	0,28	0,2265	0,37	0,00	0,17	0,3503	0,63	0,00	0,12	0,47
G33	0,0436	0,19	5,03	0,76	0,0201	0,20	8,84	1,21	0,0090	0,19	9,87	1,73	0,0523	0,52	4,93	1,28	0,26
G34	0,0102	0,11	12,17		0,1574	0,33	3,02		0,1338	0,32	2,13		0,5962	0,73	0,00		
HaGLP3	0,0403	0,18	4,89	1,81	0,1025	0,29	1,52	0,41	0,4807	0,50	0,00	0,19	0,5094	0,73	0,00	0,01	0,34
HaGLP4	0,1495	0,35	2,73		0,2104	0,38	1,40		0,4736	0,50	0,00		0,2136	0,57	2,95		
HaGLP5	0,5062	0,53	0,00	0,04	0,7518	0,61	0,00	0,10	0,8571	0,63	0,00	0,01	0,7513	0,73	0,00	0,02	0,34
HA1848	0,0030	0,11	6,79		0,0003	0,02	10,16		0,0065	0,19	10,32		0,0252	0,52	5,91		
HA2989	0,9636	0,72	0,00	0,17	0,7716	0,62	0,00	0,09	0,6793	0,59	0,00	0,01	0,8569	0,74	0,00	0,11	0,15
HA3103	0,4762	0,53	0,00		0,2722	0,41	0,00		0,1834	0,35	1,45		0,1225	0,57	1,66		
HA3847	0,0254	0,16	11,01	0,21	0,1678	0,35	3,01	0,06	0,5613	0,54	0,00	0,09	0,8150	0,73	0,00	0,08	0,38
HA4264	0,0747	0,24	1,30		0,1786	0,35	0,00		0,5137	0,51	0,00		0,6896	0,73	0,00		
HA77	0,7716	0,66	0,00		0,9710	0,68	0,00		0,1502	0,32	4,07		0,7399	0,73	0,00		
HaCOI_1	0,0142	0,12	8,76		0,0022	0,08	11,08		0,0036	0,19	9,15		0,0039	0,17	9,11		
HaCP	0,6016	0,58	0,00	0,30	0,8424	0,65	0,00	0,02	0,6654	0,59	0,00	0,02	0,8225	0,73	0,00	0,33	0,26
HaPAL	0,4787	0,53	7,05	0,18	0,5925	0,58	1,08	0,18	0,3552	0,46	0,00	0,59	0,4649	0,70	0,00	1,67	0,21
HaPRD	0,0121	0,11	0,00	1,56	0,1420	0,31	0,00	0,81	0,7686	0,62	0,00	0,00	0,8893	0,75	0,00	0,13	0,27
HaTRP	0,9175	0,71	0,00	0,04	0,6253	0,59	0,00	0,13	0,3187	0,44	0,00	0,32	0,5712	0,73	0,00	0,38	0,12
HaWP	0,2581	0,42	0,00	0,03	0,4243	0,52	0,00	0,22	0,2283	0,37	0,00	0,12	0,1973	0,57	0,35	0,00	0,17
HaWrky5	0,6270	0,59	0,00		0,5305	0,56	0,00		0,3389	0,45	0,00		0,1992	0,57	1,13		
HaWrky7	0,8136	0,68	0,00	0,61	0,5835	0,58	0,00	0,50	0,0467	0,31	4,44	1,78	0,4812	0,70	0,00	0,00	0,39
HeAn12562	0,5727	0,56	0,00	0,43	0,6926	0,60	0,00	0,17	0,5040	0,51	0,00	0,21	0,9366	0,75	0,00	0,01	0,17
HeAn315	0,2239	0,42	0,00	0,79	0,0792	0,27	1,96	0,63	0,0378	0,30	4,17	0,91	0,0359	0,52	3,61	1,02	0,33
HeAn41021	0,1132	0,32	2,86		0,0704	0,26	5,72		0,1690	0,33	3,10		0,1741	0,57	1,75		
ORS1041	0,0105	0,11	4,74		0,0811	0,27	3,71		0,3762	0,46	1,26		0,8343	0,73	0,00		
ORS331	0,1654	0,36	2,68		0,2672	0,41	1,02		0,0554	0,31	5,82		0,4586	0,69	0,29		
ORS460	0,0573	0,21	4,74		0,0441	0,20	5,80		0,2410	0,37	1,35		0,3383	0,63	0,23		
RhoBp	0,8371	0,68	0,00	0,02	0,8974	0,67	0,00	0,11	0,6929	0,59	0,00	0,02	0,7211	0,73	0,00	0,10	0,12
SNP100	0,3124	0,47	0,00	0,06	0,3073	0,43	0,00	0,00	0,8393	0,63	0,00	0,05	0,7412	0,73	0,00	1,72	0,27
SNP101	0,0647	0,23	14,22	0,47	0,1340	0,30	8,49	0,35	0,0714	0,31	8,72	0,21	0,1897	0,57	3,80	0,18	0,07
SNP102	0,0484	0,19	16,57	0,03	0,0359	0,20	15,84	0,01	0,2403	0,37	0,00	0,09	0,2992	0,63	2,92	0,07	0,07

Marcador	Incidencia				AUDPC incidencia				Severidad				Período de incubación				MAF
	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	
SNP104	0,8662	0,70	0,00	0,01	0,7210	0,61	0,00	0,19	0,1479	0,32	2,27	0,69	0,2299	0,57	0,06	1,79	0,33
SNP105	0,8263	0,68	0,00	0,00	0,9706	0,68	0,00	0,02	0,7099	0,59	0,00	0,11	0,8114	0,73	0,00	0,07	0,13
SNP106	0,7692	0,66	0,00	0,32	0,5396	0,56	0,00	0,37	0,9952	0,67	0,00	0,03	0,4051	0,67	0,00	0,44	0,16
SNP107	0,3385	0,49	0,00	0,04	0,4415	0,53	0,00	0,00	0,3857	0,46	2,79	0,00	0,2670	0,60	1,40	0,01	0,11
SNP109	0,4298	0,52	0,00	2,61	0,9653	0,68	0,00	1,26	0,4338	0,48	0,00	0,87	0,0787	0,54	3,89	0,40	0,47
SNP11	0,0354	0,18	12,87	0,51	0,0225	0,20	18,88	0,41	0,0781	0,31	12,73	0,31	0,0784	0,54	8,64	0,17	0,07
SNP111	0,1963	0,39	2,08	3,99	0,6602	0,60	0,00	1,91	0,9980	0,67	0,00	1,81	0,7595	0,73	0,00	0,61	0,20
SNP112	0,9298	0,71	0,00	0,10	0,5867	0,58	0,00	1,17	0,1283	0,32	2,29	2,61	0,9376	0,75	0,00	0,65	0,29
SNP113	0,0380	0,18	4,65	0,99	0,0479	0,21	4,91	1,13	0,0172	0,19	7,41	2,24	0,8195	0,73	0,00	0,02	0,32
SNP114	0,6847	0,62	0,00	0,11	0,6306	0,59	0,00	0,01	0,9372	0,66	0,00	1,00	0,8025	0,73	0,00	0,00	0,38
SNP115	0,0362	0,18	11,95	0,70	0,1317	0,30	7,21	1,55	0,2113	0,37	6,71	0,64	0,8680	0,74	0,00	0,10	0,10
SNP116	0,7262	0,64	0,00	0,01	0,7482	0,61	0,00	0,03	0,6598	0,59	0,00	0,00	0,1340	0,57	6,27	0,34	0,07
SNP117	0,0007	0,08	23,89	1,86	0,0004	0,02	25,00	2,88	0,0138	0,19	12,74	0,82	0,0032	0,17	20,51	3,28	0,18
SNP118	0,4030	0,52	0,00	0,01	0,7582	0,61	0,00	0,00	0,8521	0,63	0,00	0,04	0,8347	0,73	0,00	0,01	0,07
SNP119	0,3747	0,51	0,00	0,22	0,3364	0,44	0,00	0,19	0,0726	0,31	7,63	0,80	0,9992	0,78	0,00	0,01	0,11
SNP121	0,3602	0,50	0,00	0,02	0,2693	0,41	0,19	0,11	0,8116	0,63	0,00	0,00	0,7917	0,73	0,00	0,36	0,31
SNP122	0,0329	0,18	17,42	0,50	0,0373	0,20	20,60	0,57	0,0957	0,31	12,88	0,22	0,1434	0,57	7,62	0,90	0,07
SNP123	0,0365	0,18	5,55	0,79	0,1153	0,29	1,84	0,53	0,6332	0,59	0,00	0,00	0,3448	0,63	0,00	0,62	0,26
SNP124	0,0153	0,12	12,49	2,15	0,0304	0,20	7,71	1,29	0,0726	0,31	0,43	0,33	0,2732	0,60	0,00	0,32	0,19
SNP125	0,0098	0,11	10,19	1,91	0,0281	0,20	6,09	0,58	0,1464	0,32	0,00	0,33	0,1522	0,57	1,66	0,22	0,43
SNP127	0,2212	0,42	0,18	1,30	0,6136	0,59	0,00	0,24	0,4907	0,50	0,00	0,19	0,3157	0,63	0,00	0,06	0,49
SNP128	0,0069	0,11	13,94	1,63	0,0093	0,18	12,31	1,12	0,0136	0,19	10,51	1,61	0,1936	0,57	3,00	0,29	0,20
SNP129	0,5060	0,53	0,00	0,65	0,4767	0,54	0,00	0,41	0,1577	0,32	0,00	0,58	0,3221	0,63	0,00	0,33	0,14
SNP130	0,3570	0,50	0,00	0,05	0,4656	0,53	0,00	0,00	0,4189	0,48	0,00	0,00	0,8204	0,73	0,00	0,02	0,26
SNP131	0,3224	0,47	0,00	0,04	0,2013	0,38	0,79	0,08	0,2367	0,37	0,90	0,00	0,1480	0,57	2,14	0,03	0,22
SNP133	0,9275	0,71	0,00	0,04	0,6674	0,60	0,00	0,01	0,3371	0,45	1,51	0,06	0,2522	0,60	2,12	0,01	0,10
SNP134	0,5144	0,53	0,00	0,06	0,8497	0,65	0,00	0,00	0,2989	0,42	0,00	0,10	0,1914	0,57	1,51	0,45	0,13
SNP136	0,0032	0,11	11,39	1,80	0,0038	0,11	12,24	2,20	0,0051	0,19	9,87	1,40	0,1086	0,57	1,91	0,82	0,27
SNP139	0,0115	0,11	16,55	0,80	0,0248	0,20	17,49	0,67	0,1336	0,32	9,65	0,25	0,7177	0,73	0,00	0,03	0,09
SNP140	0,0118	0,11	19,01	1,67	0,0397	0,20	11,98	1,38	0,7487	0,62	0,00	0,59	0,3585	0,63	0,00	0,57	0,10
SNP141	0,1231	0,34	14,70	0,39	0,1718	0,35	5,77	0,19	0,6450	0,59	0,00	0,14	0,3084	0,63	0,92	0,31	0,11
SNP142	0,9728	0,72	0,00	0,04	0,3450	0,44	0,35	0,52	0,1143	0,31	0,48	1,15	0,6024	0,73	0,00	0,06	0,43
SNP143	0,3768	0,51	0,00	1,74	0,2559	0,41	0,00	2,89	0,1064	0,31	3,49	2,40	0,7461	0,73	0,00	1,96	0,39
SNP144	0,0041	0,11	27,72	0,18	0,0089	0,18	22,42	0,28	0,0346	0,30	13,11	0,11	0,2104	0,57	2,12	0,42	0,09

Marcador	Incidencia				AUDPC incidencia				Severidad				Período de incubación				MAF
	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	
SNP146	0,5998	0,58	0,00	0,74	0,3201	0,44	0,00	0,22	0,3821	0,46	0,00	0,53	0,0664	0,52	5,47	0,55	0,31
SNP147	0,1955	0,39	1,09	0,27	0,0293	0,20	8,05	3,74	0,4463	0,49	0,00	0,36	0,2482	0,60	1,05	1,18	0,22
SNP149	0,5586	0,56	0,00	0,06	0,4764	0,54	0,00	0,25	0,2532	0,39	0,00	2,67	0,6902	0,73	0,00	0,29	0,36
SNP150	0,5072	0,53	0,00	0,01	0,5524	0,56	0,00	0,36	0,7905	0,62	0,00	3,41	0,4741	0,70	0,00	1,75	0,38
SNP152	0,5538	0,56	0,00	0,07	0,3067	0,43	0,00	0,02	0,2209	0,37	1,18	0,09	0,0600	0,52	3,69	1,62	0,49
SNP154	0,4664	0,53	0,00	0,04	0,1971	0,37	2,90	0,00	0,1125	0,31	2,83	0,02	0,6262	0,73	0,00	0,05	0,11
SNP155	0,2431	0,42	1,82	0,11	0,1803	0,35	3,10	0,02	0,1394	0,32	3,63	0,06	0,1842	0,57	2,50	0,03	0,13
SNP156	0,1311	0,34	0,64	1,68	0,4542	0,53	0,00	1,05	0,2989	0,42	0,00	0,85	0,8335	0,73	0,00	0,00	0,22
SNP157	0,5015	0,53	0,00	0,16	0,0738	0,26	2,40	0,13	0,0488	0,31	2,22	0,28	0,0079	0,26	8,19	3,34	0,46
SNP158	0,7631	0,66	0,00	0,28	0,8853	0,66	0,00	0,87	0,8738	0,63	0,00	0,08	0,6810	0,73	0,00	1,08	0,36
SNP159	0,8771	0,70	0,00	0,32	0,9445	0,68	0,00	0,15	0,8206	0,63	0,00	0,36	0,9618	0,77	0,00	0,13	0,32
SNP160	0,5742	0,56	0,00	1,13	0,8053	0,64	0,00	0,03	0,6970	0,59	0,00	0,08	0,5294	0,73	0,00	0,78	0,44
SNP161	0,2591	0,42	8,55	0,02	0,4629	0,53	0,49	0,01	0,2405	0,37	0,00	0,08	0,8356	0,73	0,00	0,01	0,07
SNP163	0,1531	0,35	3,44	1,08	0,2514	0,41	1,02	1,08	0,9227	0,65	0,00	0,12	0,2040	0,57	2,23	1,54	0,15
SNP164	0,7119	0,64	0,00	0,00	0,9673	0,68	0,00	0,03	0,3552	0,46	1,62	0,04	0,3115	0,63	0,01	0,01	0,10
SNP165	0,0953	0,29	2,33	0,34	0,1081	0,29	3,15	0,73	0,1068	0,31	3,36	1,95	0,6787	0,73	0,00	0,47	0,38
SNP166	0,4057	0,52	0,00	0,60	0,4611	0,53	0,00	0,82	0,8880	0,63	0,00	0,79	0,5773	0,73	0,00	0,05	0,27
SNP167	0,0480	0,19	13,17	0,66	0,2434	0,41	4,10	0,74	0,7772	0,62	0,00	0,24	0,3655	0,63	0,94	0,90	0,08
SNP168	0,6476	0,60	0,00	0,00	0,2555	0,41	1,52	0,70	0,1169	0,31	0,12	1,51	0,1903	0,57	0,83	1,32	0,35
SNP169	0,1011	0,30	2,81	0,26	0,0351	0,20	9,03	0,46	0,4459	0,49	0,00	0,00	0,0906	0,57	5,31	0,81	0,13
SNP17	0,2959	0,45	0,00	0,64	0,1102	0,29	2,36	2,01	0,2187	0,37	0,75	0,92	0,0477	0,52	4,86	3,77	0,33
SNP170	0,2544	0,42	0,00	0,11	0,1302	0,30	0,00	0,07	0,0876	0,31	0,72	0,53	0,6874	0,73	0,00	0,00	0,40
SNP171	0,0399	0,18	7,21	0,78	0,1153	0,29	0,00	0,19	0,9983	0,67	0,00	0,03	0,8861	0,75	0,00	0,22	0,15
SNP172	0,5244	0,53	0,00	0,00	0,6793	0,60	0,00	0,32	0,8128	0,63	0,00	0,13	0,5825	0,73	0,00	1,78	0,36
SNP174	0,8374	0,68	0,00	0,44	0,6568	0,60	0,00	0,45	0,4529	0,49	0,00	0,63	0,2159	0,57	0,87	0,02	0,23
SNP175	0,1891	0,39	0,67	1,16	0,2051	0,38	1,00	2,16	0,3917	0,46	0,00	1,24	0,2109	0,57	0,31	3,31	0,22
SNP178	0,9650	0,72	0,00	0,13	0,7016	0,60	0,00	0,32	0,4282	0,48	0,03	0,01	0,1781	0,57	2,54	0,26	0,28
SNP180	0,7170	0,64	0,00	0,02	0,9757	0,68	0,00	0,48	0,2736	0,40	0,00	0,00	0,6467	0,73	0,00	0,22	0,31
SNP181	0,3149	0,47	0,00	0,01	0,6187	0,59	0,00	0,02	0,9939	0,67	0,00	0,05	0,7410	0,73	0,00	0,05	0,26
SNP182	0,0974	0,29	3,94	0,22	0,1456	0,31	3,69	0,47	0,6678	0,59	0,00	0,19	0,7147	0,73	0,00	0,87	0,13
SNP184	0,6514	0,60	0,00	0,02	0,5167	0,56	0,00	0,14	0,4925	0,50	0,00	0,10	0,2703	0,60	0,00	0,28	0,15
SNP185	0,5241	0,53	0,00	0,25	0,6104	0,59	0,00	0,00	0,7424	0,62	0,00	0,15	0,3982	0,67	0,00	0,34	0,42
SNP186	0,3844	0,51	0,00	0,02	0,5696	0,57	0,00	0,50	0,7724	0,62	0,00	0,96	0,7785	0,73	0,00	0,88	0,40
SNP187	0,9968	0,73	0,00	0,38	0,6980	0,60	0,00	1,04	0,2711	0,40	0,00	0,51	0,0671	0,52	1,61	0,04	0,46

Marcador	Incidencia				AUDPC incidencia				Severidad				Período de incubación				MAF
	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	
SNP188	0,0207	0,15	10,55	2,14	0,0227	0,20	8,61	1,73	0,0596	0,31	2,30	0,42	0,1860	0,57	1,10	0,79	0,19
SNP19	0,4543	0,53	0,00	0,57	0,3254	0,44	0,00	1,05	0,9719	0,67	0,00	0,01	0,8252	0,73	0,00	0,08	0,17
SNP190	0,1446	0,35	2,88	0,23	0,3348	0,44	0,00	0,40	0,1608	0,32	3,38	0,65	0,4369	0,68	0,00	0,95	0,13
SNP191	0,8523	0,69	0,00	0,04	0,9482	0,68	0,00	0,39	0,8476	0,63	0,00	1,75	0,5368	0,73	0,00	0,00	0,34
SNP192	0,6177	0,58	0,00	0,16	0,9415	0,68	0,00	0,05	0,9867	0,67	0,00	0,08	0,8120	0,73	0,00	0,08	0,16
SNP193	0,8377	0,68	0,00	0,01	0,7447	0,61	0,00	0,13	0,4890	0,50	0,00	0,22	0,6349	0,73	0,00	0,51	0,21
SNP194	0,0133	0,12	6,93	1,03	0,0247	0,20	5,79	1,18	0,0201	0,19	5,94	1,54	0,3406	0,63	0,00	0,34	0,27
SNP195	0,4546	0,53	0,00	0,00	0,3500	0,44	0,00	0,00	0,0948	0,31	10,01	0,03	0,0168	0,44	22,51	0,01	0,06
SNP196	0,9208	0,71	0,00	0,02	0,7338	0,61	0,00	0,07	0,3006	0,42	0,21	0,44	0,5560	0,73	0,00	0,01	0,17
SNP21	0,2636	0,42	0,40	0,09	0,3255	0,44	0,00	0,00	0,3898	0,46	0,00	0,14	0,4369	0,68	0,00	0,22	0,19
SNP23	0,1493	0,35	1,86	0,01	0,0480	0,21	12,62	0,11	0,0205	0,19	17,53	0,22	0,0025	0,17	25,31	0,60	0,10
SNP25	0,9565	0,72	0,00	0,53	0,7254	0,61	0,00	0,22	0,7837	0,62	0,00	0,09	0,8912	0,75	0,00	0,46	0,26
SNP27	0,6079	0,58	0,00	0,41	0,4927	0,54	0,00	0,03	0,8501	0,63	0,00	0,02	0,7199	0,73	0,00	0,03	0,47
SNP29	0,4252	0,52	0,00	0,17	0,3076	0,43	0,00	0,02	0,6670	0,59	0,00	0,00	0,5928	0,73	0,00	0,16	0,46
SNP30	0,7318	0,64	0,00	0,00	0,8687	0,66	0,00	0,01	0,3436	0,45	0,00	1,65	0,6657	0,73	0,00	0,13	0,31
SNP32	0,2185	0,42	2,23	0,47	0,1005	0,29	3,73	0,48	0,1342	0,32	0,00	0,09	0,1865	0,57	0,65	0,36	0,14
SNP34	0,1306	0,34	0,69	0,61	0,0410	0,20	5,32	2,31	0,1415	0,32	1,85	1,06	0,2255	0,57	0,00	1,27	0,35
SNP36	0,5078	0,53	0,00	0,13	0,8496	0,65	0,00	0,01	0,4651	0,49	0,00	0,03	0,8145	0,73	0,00	0,07	0,28
SNP38	0,3597	0,50	0,00	8,32	0,8109	0,64	0,00	3,34	0,6466	0,59	0,00	1,92	0,7027	0,73	0,00	0,88	0,39
SNP40	0,2631	0,42	0,00	0,00	0,2969	0,43	0,00	0,01	0,1847	0,35	0,31	0,18	0,5891	0,73	0,00	0,00	0,24
SNP43	0,4301	0,52	0,00	0,20	0,4210	0,52	0,00	0,00	0,9652	0,67	0,00	0,01	0,7725	0,73	0,00	0,04	0,47
SNP44	0,0673	0,23	4,18	6,98	0,0341	0,20	5,11	11,52	0,0105	0,19	7,35	7,07	0,0669	0,52	3,07	5,06	0,35
SNP45	0,0578	0,21	7,09	2,51	0,2824	0,42	0,78	0,83	0,8333	0,63	0,00	0,43	0,8091	0,73	0,00	0,07	0,30
SNP46	0,2672	0,42	2,24	0,02	0,5246	0,56	0,00	0,03	0,3981	0,46	0,00	0,04	1,0000	0,78	0,00	0,01	0,10
SNP47	0,1514	0,35	0,45	0,07	0,0730	0,26	4,11	0,00	0,1839	0,35	0,21	0,07	0,0675	0,52	4,58	0,24	0,16
SNP48	0,1655	0,36	0,31	0,71	0,1126	0,29	1,63	2,70	0,0494	0,31	2,71	1,98	0,1196	0,57	0,87	0,68	0,44
SNP50	0,1496	0,35	2,06	0,01	0,0976	0,29	5,25	0,00	0,1449	0,32	4,47	0,03	0,4169	0,68	0,00	0,25	0,24
SNP51	0,9902	0,73	0,00	0,85	0,6924	0,60	0,00	0,72	0,8702	0,63	0,00	0,26	0,4176	0,68	0,00	0,15	0,21
SNP53	0,4864	0,53	0,00	0,21	0,4324	0,52	0,00	0,39	0,4642	0,49	0,00	0,09	0,3613	0,63	0,00	0,03	0,27
SNP55	0,1617	0,36	0,86	0,42	0,0916	0,29	2,82	0,46	0,0806	0,31	2,71	0,29	0,1381	0,57	2,11	0,45	0,30
SNP56	0,8962	0,71	0,00	0,08	0,6854	0,60	0,00	0,01	0,2136	0,37	5,31	0,00	0,2542	0,60	0,00	0,03	0,07
SNP58	0,2502	0,42	0,75	0,49	0,7502	0,61	0,00	0,00	0,6108	0,58	0,00	0,28	0,2227	0,57	0,13	3,27	0,43
SNP59	0,0484	0,19	16,57	0,02	0,0359	0,20	15,84	0,01	0,2403	0,37	0,00	0,11	0,2992	0,63	2,92	0,06	0,07
SNP60	0,0095	0,11	8,30	3,58	0,0443	0,20	4,53	4,52	0,1156	0,31	1,22	1,92	0,2096	0,57	0,30	9,15	0,38

Marcador	Incidencia				AUDPC incidencia				Severidad				Período de incubación				MAF
	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	p-valor	q-valor Storey	Contrib, varianza	Efecto (BLR)	
SNP62	0,9830	0,73	0,00	0,03	0,9379	0,68	0,00	0,18	0,6979	0,59	0,00	0,19	0,0413	0,52	8,50	0,12	0,16
SNP63	0,0225	0,15	6,67	0,05	0,0372	0,20	4,70	0,24	0,2784	0,40	0,00	0,01	0,0648	0,52	3,85	1,43	0,34
SNP64	0,5119	0,53	0,00	0,79	0,5876	0,58	0,00	1,29	0,1165	0,31	2,49	4,54	0,6669	0,73	0,00	2,15	0,39
SNP65	0,2417	0,42	0,00	0,05	0,3297	0,44	0,00	0,03	0,1326	0,32	4,02	0,33	0,9035	0,75	0,00	0,01	0,10
SNP66	0,5069	0,53	0,00	2,95	0,5165	0,56	0,00	1,39	0,4266	0,48	0,00	0,15	0,7807	0,73	0,00	0,61	0,36
SNP68	0,8258	0,68	0,00	0,03	0,4820	0,54	0,00	0,27	0,0154	0,19	8,34	3,49	0,0814	0,54	4,14	0,60	0,36
SNP69	0,4000	0,52	0,00	0,01	0,2635	0,41	0,00	0,11	0,0194	0,19	6,04	0,52	0,1141	0,57	1,84	0,32	0,40
SNP7	0,0254	0,16	4,63	2,12	0,0275	0,20	5,54	1,72	0,7530	0,62	0,00	0,00	0,9266	0,75	0,00	0,11	0,25
SNP70	0,0499	0,19	8,39	1,21	0,0694	0,26	6,99	0,92	0,6719	0,59	0,00	0,07	0,2833	0,61	1,04	1,09	0,19
SNP71	0,4937	0,53	0,00	0,06	0,9069	0,67	0,00	0,02	0,6244	0,59	0,00	0,47	0,8082	0,73	0,00	0,42	0,28
SNP72	0,6777	0,62	0,00	0,24	0,9499	0,68	0,00	0,33	0,8583	0,63	0,00	1,99	0,3895	0,67	0,00	0,04	0,37
SNP73	0,1723	0,37	0,74	0,35	0,1202	0,29	1,89	0,21	0,1100	0,31	1,84	0,16	0,1933	0,57	1,14	0,14	0,29
SNP74	0,2327	0,42	0,00	0,16	0,2323	0,40	0,00	0,26	0,6721	0,59	0,00	0,02	0,6127	0,73	0,00	0,65	0,17
SNP75	0,0685	0,23	2,87	0,23	0,0634	0,26	3,33	0,26	0,6549	0,59	0,00	0,10	0,3381	0,63	0,00	0,06	0,35
SNP76	0,2629	0,42	0,00	0,18	0,1218	0,29	1,06	0,15	0,0761	0,31	0,55	0,87	0,5092	0,73	0,00	0,46	0,38
SNP77	0,8099	0,68	0,00	0,87	0,8466	0,65	0,00	0,03	0,3951	0,46	0,00	0,78	0,9026	0,75	0,00	1,42	0,45
SNP79	0,9340	0,71	0,00	0,03	0,5528	0,56	0,00	0,25	0,2730	0,40	0,64	0,56	0,6396	0,73	0,00	0,34	0,10
SNP80	0,0094	0,11	12,48	2,70	0,1018	0,29	3,96	2,60	0,1536	0,32	1,84	1,23	0,9271	0,75	0,00	0,11	0,24
SNP81	0,5069	0,53	0,00	1,50	0,6980	0,60	0,00	0,93	0,9265	0,65	0,00	0,14	0,8183	0,73	0,00	0,12	0,39
SNP82	0,3924	0,52	0,00	2,23	0,7733	0,62	0,00	1,82	0,3777	0,46	0,00	3,77	0,6729	0,73	0,00	3,17	0,46
SNP84	0,7147	0,64	0,00	0,58	0,9044	0,67	0,00	0,33	0,7624	0,62	0,00	0,03	0,6155	0,73	0,00	0,02	0,39
SNP85	0,1060	0,30	1,59	0,00	0,3025	0,43	0,00	0,00	0,2052	0,37	0,68	1,06	0,3293	0,63	0,26	1,05	0,19
SNP87	0,2395	0,42	1,17	0,90	0,1784	0,35	1,50	0,60	0,2121	0,37	0,00	0,39	0,6113	0,73	0,00	0,01	0,30
SNP88	0,2365	0,42	0,00	0,68	0,8487	0,65	0,00	0,24	0,5327	0,52	0,00	0,00	0,9807	0,78	0,00	0,00	0,29
SNP89	0,2971	0,45	0,49	1,09	0,5565	0,56	0,00	1,00	0,1106	0,31	0,79	3,79	0,8508	0,74	0,00	2,52	0,35
SNP9	0,9195	0,71	0,00	0,33	0,6769	0,60	0,00	0,05	0,8865	0,63	0,00	0,01	0,1189	0,57	3,29	0,33	0,15
SNP90	0,1925	0,39	0,05	0,06	0,0748	0,26	2,67	0,45	0,0599	0,31	2,97	0,70	0,4358	0,68	0,00	0,25	0,47
SNP92	0,2907	0,45	0,69	0,28	0,1115	0,29	5,59	0,36	0,0740	0,31	4,49	0,36	0,2636	0,60	0,30	0,10	0,14
SNP94	0,9059	0,71	0,00	0,82	0,9882	0,68	0,00	0,44	0,7105	0,59	0,00	0,01	0,9276	0,75	0,00	0,07	0,44
SNP95	0,4298	0,52	0,00	0,15	0,5313	0,56	0,00	0,03	0,5305	0,52	0,00	0,03	0,6202	0,73	0,00	0,01	0,16
SNP96	0,3419	0,49	0,00	0,07	0,2292	0,40	0,45	0,03	0,0517	0,31	5,78	0,01	0,4284	0,68	0,00	0,09	0,38
SNP97	0,4703	0,53	0,00	0,22	0,3718	0,46	0,00	0,46	0,0759	0,31	4,76	3,20	0,5509	0,73	0,00	1,78	0,43
SNP98	0,4112	0,52	0,00	0,02	0,3546	0,45	0,00	0,04	0,3266	0,44	0,00	0,00	0,1683	0,57	4,09	0,15	0,13
SNP99	0,4338	0,52	0,00	0,04	0,2190	0,38	1,30	0,05	0,5369	0,52	0,00	0,55	0,4577	0,69	0,00	0,30	0,20

