

Tesis Doctoral

Diversidad genética en girasol cultivado: análisis de una colección de germoplasma local para su aplicación en programas de mejoramiento

Moreno, María Valeria

2010

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:

Moreno, María Valeria. (2010). Diversidad genética en girasol cultivado: análisis de una colección de germoplasma local para su aplicación en programas de mejoramiento. Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires.

Cita tipo Chicago:

Moreno, María Valeria. "Diversidad genética en girasol cultivado: análisis de una colección de germoplasma local para su aplicación en programas de mejoramiento". Facultad de Ciencias Exactas y Naturales. Universidad de Buenos Aires. 2010.

EXACTAS UBA

Facultad de Ciencias Exactas y Naturales



UBA

Universidad de Buenos Aires



UNIVERSIDAD DE BUENOS AIRES
Facultad de Ciencias Exactas y Naturales

DIVERSIDAD GENETICA EN GIRASOL CULTIVADO: análisis
de una colección de germoplasma local para su aplicación en
programas de mejoramiento

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

María Valeria Moreno

Director de Tesis: Dra. Verónica V. Lia

Director Asistente: Dra. Norma B. Paniego

Consejero de Estudios: Dr. H. Esteban Hopp

Laboratorio de Biotecnología, INTA EEA Manfredi
Instituto de Biotecnología, CICVyA, INTA Castelar

Buenos Aires, 2010

DIVERSIDAD GENETICA EN GIRASOL CULTIVADO: análisis de una colección de germoplasma local para su aplicación en programas de mejoramiento

Resumen

Los bancos de semillas ofrecen una valiosa fuente de diversidad para el mejoramiento genético de los cultivos. Para un aprovechamiento óptimo de los recursos genéticos, la caracterización fenotípica de los materiales conservados (resistencia a estreses bióticos y abióticos, índices de calidad química, nutricional y rendimiento, entre otros) debe ser acompañada por la caracterización genética de los mismos. De este modo, la diversidad alélica evaluada a través de marcadores moleculares es uno de los mejores indicadores del potencial genético de las entradas (*accessions*) de un banco.

Los objetivos de este trabajo fueron: a) estudiar la diversidad genética en un conjunto de entradas de girasol cultivado preservadas en el Banco Activo de Germoplasma de INTA Manfredi (BAG-IM), b) estudiar la distribución de los polimorfismos detectados y la estructura poblacional, c) evaluar la consistencia entre metodologías de fenotipificación para el carácter de tolerancia a sequía, comparando ensayos de campo con ensayos en condiciones controladas, y d) analizar los patrones de variación nucleotídica en regiones candidatas asociadas al carácter tolerancia a déficit hídrico en líneas endocriadas pertenecientes al programa de mejoramiento de tolerancia a sequía del BAG-IM.

A fin de cuantificar los niveles de variabilidad se genotipificaron 337 individuos pertenecientes a 21 entradas representativas de las distintas categorías conservadas en el banco (líneas, poblaciones y compuestos) utilizando 16 marcadores microsatélites de localización genómica conocida. Esta elección fue respaldada mediante un análisis de diversidad basado en caracteres morfológicos y agronómicos en base a 309 entradas conservadas en el banco. Se obtuvo un número total de 108 alelos, mientras que el número promedio de alelos por locus (A) fue de 6,75. El locus con mayor valor de A para todas las entradas fue HA2077 (3,66), mientras que los que mostraron menor valor fueron HA3582 y HA4239 (1,57 y 1,52, respectivamente). La

heterocigosis esperada promedio fue de 0,29, con valores extremos de 0,476 para la población Prao-Co y de 0,054 para la línea F-164. El número promedio de alelos por locus varió entre 3 y 16. El 100% de las poblaciones (8) y el 83,3% de los compuestos analizados (5 de 6) mostraron desvío a las proporciones de Hardy-Weinberg, con exceso de homocigotas en la mayoría de los loci. La estima global de F_{ST} permitió detectar niveles de diferenciación altos y estadísticamente significativos entre entradas ($F_{ST}=0,413$, $p < 0,05$). La estructura poblacional inferida mediante métodos bayesianos mostró la existencia de 14 entidades panmícticas. Individuos de todas las entradas evidenciaron patrones de mezcla con diferentes proporciones de los acervos génicos detectados contribuyendo a sus constituciones genéticas. Los mayores niveles de mezcla se detectaron para la población Prao-co.

El ensayo bajo condiciones controladas en déficit hídrico moderado resultó eficiente para la fenotipificación rápida de líneas pertenecientes al programa de mejoramiento para tolerancia a sequía. El análisis de los patrones de variación en 7 regiones candidatas para tolerancia a déficit hídrico (Hahb4-p, Hahb4, *Suntip*, FB, HaL1L, HaDhn1 y CK) reveló una frecuencia promedio de SNP de 1/48,3 pb y una diversidad nucleotídica promedio (θ_w) de 0,00501. Se obtuvo un elevado nivel de desequilibrio de ligamiento (DL) ($r^2=0,88$ a las 900 pb), decayendo sólo en *Suntip* ($r^2=0,325$ a las 207 pb) y en HaDhn1 ($r^2=0,249$ a las 78 pb). Las regiones *Suntip* y HaDhn1 serían las más apropiadas como candidatas para mapeo por asociación por la multiplicidad de variantes alélicas en frecuencias similares, mientras que las regiones Hahb4, Hahb4-p, FB y HaL1L resultan interesantes debido a la posible incidencia de procesos selectivos y podrían emplearse en un futuro, complementando el análisis con un mayor número de entradas en la población de mapeo.

Este trabajo constituye el primer aporte en la evaluación de la variabilidad genética contenida en el BAG-IM mediante marcadores moleculares. Los resultados aquí presentados permitirán diseñar estrategias para la caracterización genética exhaustiva del banco y sentar las bases para el desarrollo de estudios de mapeo por asociación para el carácter de tolerancia a déficit hídrico.

Palabras Clave: *Helianthus annuus*, diversidad genética, recursos genéticos, germoplasma, SSR, SNP, diversidad nucleotídica, déficit hídrico.

GENETIC DIVERSITY IN SUNFLOWER: analysis of a local germplasm collection as a tool to assist management and breeding programs

Abstract

Seed banks offer a valuable source of allelic diversity for crop genetic improvement. For genetic resources to reach their maximum utility, the phenotypic characterization of the conserved materials (abiotic and biotic stress resistance, chemical indices, nutritional quality and yield, among others), must be accompanied by their genetic characterization. The genetic composition at the molecular level is the best indicator of the genetic potential of the bank's accessions.

The aims of this study were: a) to assess the genetic diversity of a set of sunflower accessions preserved at the Active Germplasm Bank of INTA Manfredi (AGB-IM), b) to assess the polymorphisms distribution and population structure, c) to study the consistence between two methods for the evaluation of drought tolerance, and d) to study the nucleotide diversity of candidate regions for drought tolerance in a set of lines from the corresponding breeding program.

For the first goal, a total of 337 individuals corresponding to 21 accessions encompassing a wide range of geographic origins and breeding programmes were selected for analysis. These accessions include inbred lines, cultivated populations and compounds used in sunflower breeding programs of INTA and they are representative of the morphological diversity spectrum conserved at the AGB-IM, as confirmed by cluster analysis of a more comprehensive set of accessions. Molecular characterization was conducted using 16 microsatellite markers of known genomic location representing different linkage groups. The 16 loci used in the analysis were polymorphic and revealed a total of 108 alleles. The average gene diversity was 0,29 and the average number of alleles *per locus* was 6,75. Gene diversity indices varied from 0,476 to 0,054, with population Prao-co and line F-164 showing the highest and lowest values, respectively. The mean number of alleles *per locus* ranged from 3 to 16. As expected for breeding materials, departures from Hardy-Weinberg proportions were observed in 100% of populations (8) and 83,3% of composites (5 of 6) studied, with most loci exhibiting an excess of

homozygotes. Global estimates of F_{ST} revealed very high and statistically significant levels of differentiation among accessions ($F_{ST}=0,413$, $p < 0,05$). Inference of population structure was also performed using a bayesian model based approach. The number of gene pools showing the highest likelihood was 14. Individuals from all accessions were admixed, with different proportions of the inferred gene pools contributing to their genomic constitution. Prao-co cultivated population exhibited the highest levels of admixture and showed evidence of several gene pools contributing to their genomes.

The mannitol test proved to be a reliable and potentially useful method for the screening of drought tolerance in sunflower. Analysis of nucleotide diversity in seven candidate genes (Hahb4-p, Hahb4, *Suntip*, FB, HaL1L, HaDhn1 and CK) revealed an average SNP frequency of 1/48,3 bp and an average nucleotide diversity (θ_w) of 0,00501. A high level of LD was obtained for all the regions ($r^2=0,88$ at 900 pb) except for *Suntip* where LD decays to $r^2=0,325$ at 207 pb and HaDhn1 where LD decays to $r^2=0,249$ at 78 pb. *Suntip* and HaDhn1 are the most promising regions for association mapping, while Hahb4, Hahb4-p, FB y HaL1L are interesting for future studies of association mapping with a larger set of lines, because they showed possible incidence of selective process.

This work allowed characterizing a set of AGB-IM accessions at the molecular level based on neutral and functional markers. This information will also be useful for genetic resources management, germplasm selection, association mapping studies and plant breeding programs.

Keywords: *Helianthus annuus*, genetic diversity, genetic resources, germplasm, SSR, SNP, nucleotide diversity, drought.

**A Silvio, por acompañarme y contenerme,
por ser mi fuerza y mi luz.**

**A mis padres Osvaldo y Cristina,
por guiarme y darme lo mejor de ellos.**

**A mis hermanos Rafael y Fernando,
por compartir momentos de complicidad y cariño fraterno.**

AGRADECIMIENTOS

Agradezco al Instituto Nacional de Tecnología Agropecuaria (INTA) por la beca de posgrado otorgada en el año 2006, para completar mi formación profesional. En particular a la Estación Experimental Agropecuaria INTA Manfredi por recibirme y al Instituto de Biotecnología de INTA Castelar por permitir mi estadía temporaria para cumplimentar con los requisitos del posgrado. Agradezco especialmente a la Dra. Ruth Heinz y al Dr. Esteban Hopp por abrirme las puertas del Laboratorio de Marcadores Moleculares.

A la Dra. Verónica Lia por la ayuda, la contención y el apoyo en los momentos más difíciles y por su predisposición a guiarme. ¡Gracias Verito, por la formación en lo profesional, por tu calidez y tu simpleza, por tus valores humanos!.

A la Dra. Norma Paniego, porque desde el primer momento estuvo dispuesta a proponer, guiar, apoyar y estimular. ¡Gracias Norma, por hacerme sentir que siempre *estás* y por tu serenidad!.

Al Dr. Jorge Gieco por recibirme en el Grupo de Biotecnología, por todas las gestiones respecto al posgrado como Director de beca INTA. Gracias Jorge, por la libertad para trabajar y desarrollarme como profesional.

Al Ing. Daniel Álvarez por permitirme trabajar con los recursos genéticos del Banco Activo de Germoplasma de girasol de INTA Manfredi, por el apoyo técnico en cuanto al conocimiento de los materiales objeto de esta tesis.

Al Instituto Nacional de Semillas (INASE), a las Dras. Ana Laura Vicario y Alicia Loray, por el apoyo brindado y por la participación de su Laboratorio, Ali ¡gracias por tu ayuda con las extracciones de ADN!.

A las amigas que el INTA me regaló: Merce, Roxy, Vale, Caro. Por las charlas compartidas, los momentos bellos que pasamos juntas y por aquellos difíciles donde sentimos que vale la pena cultivar amistades sinceras y verdaderas, ¡las

quiero muchísimo!, gracias por soportarme en este largo proceso.....¡sobre todo al final!.

A las chicas del lab: Gaby, Yani y Caro, por los desayunos y almuerzos juntas, por las coincidencias y las charlas de *todo un poco* y, en especial Gaby y Caro, ¡gracias por la ayuda con el ensayo de manitol!. A Juanjo por el apoyo incondicional en los trabajos de campo, a Julio por los matecitos matutinos y las largas charlas. A todas las hermosas personas de los laboratorios “vecinos”, Chachi, Amanda, Jorge M., Seba, Ariel, Patri, Sil.

Al Dr. Cristiano Casini por prestarme la cámara de germinación, a Mauricio y Claudio, por hacerme más gratos los largos días de ensayo, ¡por los matecitos amenizadores del recuento de plántulas de girasol!.

A los chicos de Castelar que me recibieron y me soportaron por 5 años...Vero N., ¡por tu inmensa e invaluable ayuda con los microsátélites! por tu sencillez, calidez y amistad, a Cori por tu acompañamiento con las benditas regiones candidatas y sus SNP, por la buena onda y la complicidad, a Apu por las charlas de genotipificación y ¡de la vida!, por tu calidez y sinceridad, a Pablo por la buena onda, a Luis y Carla M., por nuestras charlas compartidas y las cenas en la residencia del IMYZA!, a todos los chicos del “locutorio” Lula, Carlos, Diego, Laura, Cintia, Guille, Fede, Seba, Jere, Carla F. y si me olvidé de alguien.....¡perdón!.

A las amigas que la vida me regaló: Alita, Laura, Gaby, Luci, fueron, son y serán tan especiales en mi vida que día a día reafirmo el lazo tan profundo y sincero que nos une, ¡las adoro, mil gracias por compartir su camino conmigo!.

A Fer, ¡por su ayuda con el diseño y mejora de los gráficos!, y a Julieta por su aporte con la organización de la bibliografía.

INDICE

Resumen	2
Abreviaturas	12
INTRODUCCION GENERAL	13
11. El girasol cultivado (<i>Helianthus annuus</i> L. var. <i>macrocarpus</i>)	13
12. Importancia del girasol en Argentina y en el mundo	14
13. Recursos genéticos y bancos de germoplasma	16
14. Caracterización genética de cultivos con marcadores moleculares.....	20
15. Mejoramiento vegetal.....	24
16. Programas de mejoramiento de girasol en el país.....	26
HIPOTESIS	29
OBJETIVOS	29

CAPITULO 1: Diversidad genética en entradas de girasol cultivado pertenecientes al Banco Activo de Germoplasma de INTA Manfredi

INTRODUCCION	31
MATERIALES Y METODOS	35
M1. Muestras de material biológico.....	35
M1.1. Material incluido en el análisis de datos morfológicos y agronómicos	35
M1.2. Material para la caracterización genética	35
M2. Análisis de datos morfológicos y agronómicos.....	37
M3. Análisis de variabilidad de loci microsatélites.....	38
M3.1. Métodos experimentales.....	38
M3.2. Análisis de la diversidad y estructura poblacional	47
RESULTADOS	53
R1. Diversidad en base a caracteres morfológicos y agronómicos	53
R2. Diversidad genética en base a variabilidad de loci microsatélites.....	58
R.3. Estructura poblacional.....	70
R.3.1. Equilibrio de Hardy-Weinberg	70
R.3.2. Diferenciación genética.....	72
R.3.3. Análisis de agrupamiento.....	79
DISCUSION	84
D.1. Diversidad genética en el BAG-IM. Implicancias en la conservación de germoplasma	84
D.2. Distribución de la variabilidad genética	92
D.2.1. Estructura genética intrapoblacional.....	92
D.2.2. Diferenciación genética entre poblaciones	93
D.3. Análisis de agrupamiento	97

CAPITULO 2: Caracterización fenotípica y genética de líneas de girasol cultivado del programa de mejoramiento para tolerancia a déficit hídrico del Banco Activo de Germoplasma de INTA Manfredi

INTRODUCCION	101
MATERIALES Y METODOS	112
M1. Caracterización fenotípica para el carácter de tolerancia a sequía... ..	112
M1.1. Material vegetal	112

M1.2. Análisis de datos del ensayo de campo (campaña 2003-2004): evaluación de planta adulta.....	113
M1.3. Ensayo en condiciones controladas: evaluación de plántula.....	114
M2. Caracterización genética de líneas pertenecientes al programa de mejoramiento para tolerancia a sequía.....	116
M2.1. Extracción y cuantificación de ADN.....	116
M2.2. Análisis de variabilidad para marcadores microsatélites	116
M2.3. Selección de regiones candidatas para tolerancia a déficit hídrico	117
M2.4. Diseño de oligonucleótidos iniciadores y amplificación por PCR de las regiones candidatas.....	122
M2.5. Purificación de los productos de amplificación.....	123
M2.6. Secuenciación de los productos de PCR	123
M2.7. Identificación de SNP y análisis de variabilidad en las regiones candidatas.....	124
M2.8. Análisis del desequilibrio de ligamiento en las regiones candidatas	125
RESULTADOS.....	128
R1. Caracterización fenotípica para el carácter de tolerancia a sequía...	128
R1.1. Análisis de datos de ensayo de campo: evaluación de planta adulta	128
R1.2. Ensayo en condiciones controladas: evaluación de plántula.....	131
R2. Caracterización genética de las líneas para tolerancia a sequía	136
R2.1. Análisis de variabilidad microsatélite	136
R2.2. Caracterización de regiones candidatas para el carácter de tolerancia a sequía.....	147
R2.3. Análisis de desequilibrio de ligamiento en las regiones candidatas	159
DISCUSION	162
D.1. Caracterización fenotípica.....	162
D.2. Caracterización genética de las líneas para tolerancia a sequía	166
D.2.1. Diversidad genética y estructura poblacional en base a marcadores SSR	166
D.2.2. Frecuencia de SNP y diversidad nucleotídica	168
D.2.3. Pruebas de neutralidad.....	174
D.2.4. Determinación del desequilibrio de ligamiento	176
DISCUSION GENERAL	179
CONCLUSIONES.....	181
BIBLIOGRAFIA	183
APENDICE	208
Criterios	208
Protocolos.....	210
Soluciones y <i>buffers</i>	213
Tablas.....	214
Figuras.....	236

ABREVIATURAS

°C	grados centígrados
ADN	Ácido desoxirribonucleico
ADNc	ADN copia
ARN	Ácido ribonucleico
ARNm	Ácido ribonucleico mensajero
BAG-IM	Banco Activo de Germoplasma, INTA Manfredi
cm	centímetro
cont.	Continuación
CTAB	Bromuro de cetiltrimetilamonio, <i>Cethyl Trimethyl Ammonium Bromide</i>
DL	Desequilibrio de Ligamiento, <i>Linkage Disequilibrium</i>
Dr.	Doctor/a
EDTA	Ácido etilendiaminotetraacético, <i>Ethylene-diamine-tetraacetic acid</i>
EST	Secuencias que se expresan, <i>Expressed Sequence Tags</i>
<i>et al.</i>	Colaboradores
h	horas
Ing. Agr.	Ingeniero Agrónomo
Kb	Kilobases
MgCl ₂	Cloruro de Magnesio
m	metro
min	minutos
ml	mililitro
mM	mili Molar
µM	micro Molar
Nº	Número
Nm	Nano metro
Pb	pares de bases
PCR	Reacción en cadena de la Polimerasa, <i>Polymerase Chain Reaction</i>
Por ej.	Por ejemplo
QTL	Loci de caracteres cuantitativos, <i>Quantitative Trait Loci</i>
RAPD	ADN Polimórfico Amplificado al Azar, <i>Random Amplified Polymorphic DNA</i>
RFLP	Polimorfismo en el Largo de los Fragmentos de Restricción, <i>Restriction Fragment Length Polymorphism</i>
s	segundos
SNP	Polimorfismo de Simple Nucleótido, <i>Single Nucleotide Polymorphism</i>
SSR	Microsatélites, <i>Simple Sequence Repeats</i>
µg	micro gramo
µl	micro litro
UV	luz ultravioleta
Vs.	<i>Versus</i>

INTRODUCCIÓN GENERAL

11. El girasol cultivado (*Helianthus annuus* L. var. *macrocarpus*)

El género *Helianthus* pertenece a la familia *Asteraceae*, subfamilia *Asteroideae*, tribu *Heliantheae* y contiene 48 especies con distintos niveles de ploidía (diploides, tetraploides y hexaploides) (Seiler and Riesberg, 1997). Uno de sus representantes diploides más conocido por su importancia económica es el girasol cultivado (*Helianthus annuus* L. var. *macrocarpus*). Esta especie posee 34 cromosomas ($2n=2x=34$; Arumuganathan and Earle, 1991) y un tamaño estimado de genoma haploide de 3.000 Mb.

El girasol cultivado es de ciclo anual y se encuentra dentro del grupo de plantas autógamas. El tallo es el órgano de sostén de las hojas y de la inflorescencia. En la etapa de floración la planta puede medir de 1 a 3 m y el diámetro del tallo puede variar entre 1 y 5 cm, dependiendo de las condiciones del cultivo (Figura 1-A). El sistema radical del girasol se encuentra constituido por un eje principal y un conjunto de raíces laterales secundarias y terciarias. Bajo condiciones favorables, la raíz principal puede alcanzar profundidades mayores a 2 m. Sus ramificaciones (raíces secundarias y terciarias) son numerosas en la cercanía al cuello de la planta, disminuyendo su densidad significativamente a 15 cm del mismo (Figura 1-B). La raíz es el primer órgano que atraviesa las cubiertas de la semilla durante la germinación. Las flores se reúnen en una inflorescencia denominada cabeza o torta, que botánicamente se define como capítulo (Figura 1-C). Lo que vulgarmente se conoce como “semilla” de girasol (denominada botánicamente cipsela) es un fruto seco, uniseminado, con pericarpio separado de la verdadera semilla o pepita. El nombre de cipsela es más adecuado que el de aquenio dado que en las plantas compuestas los ovarios son bicarpelares e ínferos, mientras que el aquenio corresponde a un fruto derivado de ovarios unicarpelares y súperos. La semilla es la que contiene la mayor parte de la materia grasa o aceite (Aguirrezábal *et al.*, 2001) (Figura 1-D).

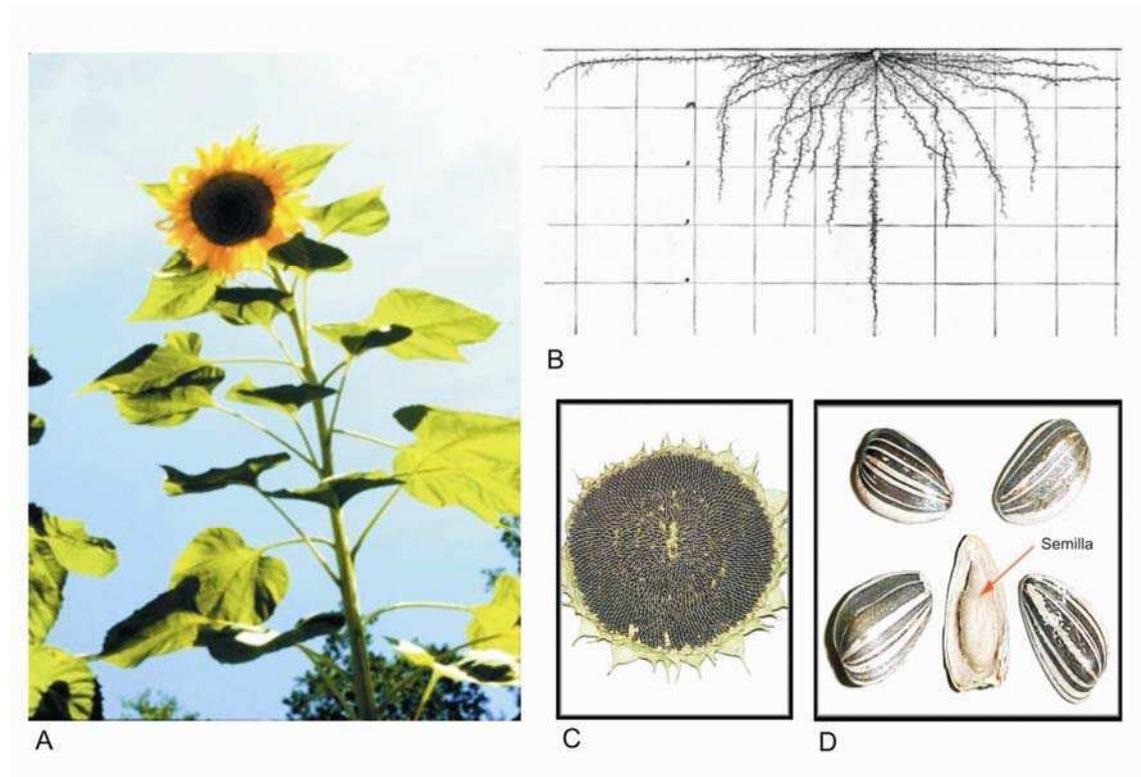


Figura 1.A: Estructura de la parte aérea de la planta adulta de girasol. B: Sistema radical principal y secundario. C: Capítulo de girasol. D: Semillas de girasol.

I2. Importancia del girasol en Argentina y en el mundo

Los principales centros de producción de girasol a nivel internacional son la Argentina, el Este Europeo y la Unión Soviética. En los últimos años, el cultivo se está expandiendo en las regiones áridas del Mediterráneo y Norte de África, fundamentalmente debido a que es moderadamente tolerante a condiciones de estrés por sequía y salinidad, y a que tiene elevada importancia agronómica (Connor and Hall, 1997).

El principal destino del girasol es su industrialización para la elaboración de aceite, siendo además empleada la semilla en confitería y productos panificados. Se obtienen productos diferenciados como semillas ricas en ácido oleico y antioxidantes (tocoferoles), semillas con alta relación de ácidos esteárico/palmitico o semillas ricas en lecitinas. En cuanto a la molienda, se produce aceite crudo y refinado, además de los subproductos de la misma como alimentos balanceados para la producción animal y granulados, siendo Argentina el mayor exportador de éste último (Casini, 2010). Aunque de menor

importancia, el girasol cultivado es conocido también por su uso ornamental (Tang *et al.*, 2002).

El total de la producción mundial de semillas de girasol es de 25,8 millones de toneladas y son utilizadas casi en su totalidad para la extracción de aceite, proporcionando 8,2% del total de aceites elaborados mundialmente, estimado en 107 millones de toneladas (Cantamutto and Poverene, 2007). El aceite de girasol, por su volumen de producción, es el cuarto en importancia después de los de soja, palma y colza. Nuestro país es uno de los cuatro grandes productores mundiales de girasol y el segundo exportador de aceite, generando el 3% del total de grano, el 34% de harina y el 32% de aceite que se consume globalmente. Sin embargo, el consumo local en relación al mundo es bajo, dado que sólo es del 5% para el caso de la harina y del 4% para el aceite (http://www.asagir.org.ar/asagir2008/news_detail.asp?id=94). La Argentina cuenta con ambientes agroecológicos sumamente favorables para su cultivo, pero puesto que el girasol posee rusticidad y adaptación a diversidad de ambientes, es sembrado en áreas donde el suelo y/o el clima condicionan severamente la productividad de otros cultivos.

El incremento en la producción de aceite de palma y de soja y la mayor demanda mundial de proteína vegetal tuvieron como consecuencia la expansión del cultivo de soja y el desplazamiento del resto de los cultivos estivales, entre ellos el girasol, hacia ambientes menos favorables. Esta situación se agudizó al producirse un desplazamiento de los productores tradicionalmente “girasoleros” hacia el cultivo de soja.

Actualmente, el cultivo de girasol en Argentina se extiende desde el sudeste de la Región Pampeana (39°S) hasta la Región Chaqueña (26°S). La mayor superficie sembrada se halla principalmente en la provincia de Buenos Aires, seguida por La Pampa y Córdoba (Figura 2), cubriendo una superficie de 2 millones de hectáreas (SAGPyA, 2009). Entre las campañas 2007-2008 y 2008-2009, se observó una disminución del rendimiento entre 10 al 19% en la región Norte-Centro y la zona Sur de la Provincia de Córdoba, cuya causa principal fue el impacto de la sequía. En la campaña 2007-2008 se destinaron a molienda 3,060 millones de toneladas de girasol, de los cuales 1,223 millones de toneladas se emplearon para producción de aceite (70% para exportación y

30% para consumo interno), mientras que 1,315 millones de toneladas se destinaron para producción de harinas (60% para exportación y 40% para consumo interno). Las reducciones en los niveles de exportación con respecto a campañas anteriores se adjudicaron al precio final del girasol, que se vio afectado por derechos de exportación superiores a los de trigo y maíz (CREA, 2009, www.redcrea.org.ar).

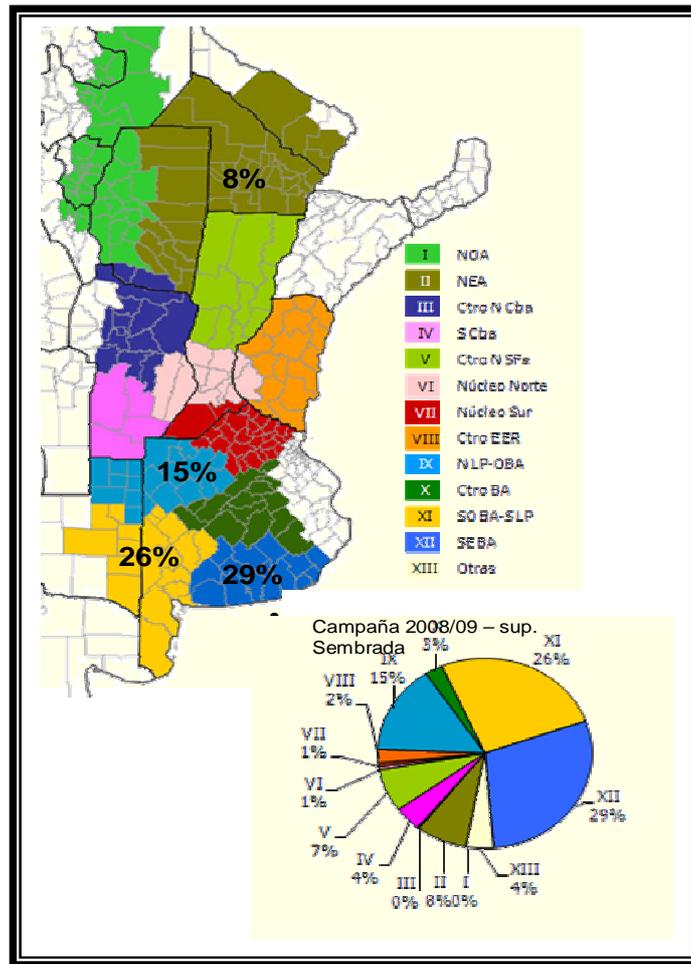


Figura 2. Distribución de las áreas productivas de girasol en Argentina. (Modificado de: Movimiento Red CREA, 2009).

13. Recursos genéticos y bancos de germoplasma

Los bancos de semillas ofrecen una valiosa fuente de diversidad alélica para el futuro mejoramiento genético de los cultivos. Para un aprovechamiento óptimo de los recursos genéticos, la caracterización fenotípica de los materiales

conservados (resistencia a patógenos, índices de calidad química, nutricional y rendimiento, entre otros), debe ser acompañada por la caracterización genética de los mismos (Jansson and Buckler, 2007). La constitución genética a nivel molecular es el mejor indicador del potencial genético de las entradas de un banco. Los genotipos cultivados son el resultado directo de la acumulación de alelos benéficos en genes que controlan caracteres de interés agronómico. Sin embargo, considerando que se espera que un alto número de loci participen en la expresión fenotípica de tales caracteres, es bastante improbable que los genotipos modernos retengan variantes alélicas benéficas para todos los genes involucrados en el control de caracteres agronómicamente importantes (Tanksley and McCouch, 1997). Por otro lado, muchos de los caracteres agronómicos tienen una base genética compleja que incluye loci de caracteres cuantitativos (QTL, *Quantitative Trait Loci*) cuya influencia relativa puede variar de acuerdo con la especificidad de cada uno de los agroecosistemas por lo que, es también esperable, la conservación de diversidad genética en características ecológicamente adaptativas. Finalmente, para completar la compleja constitución del germoplasma cultivado, la dinámica de coevolución con patógenos implicó que los mejoradores apelaran frecuentemente a cruzamientos amplios buscando diversidad genética, incluyendo especies silvestres emparentadas para incorporar genes de resistencia, los cuales suelen venir acompañados de amplios segmentos cromosómicos en desequilibrio de ligamiento. De este modo, el nivel de diversidad existente en el germoplasma de un cultivo dependerá de la magnitud de los cuellos de botella que hayan sufrido los materiales durante la domesticación y de los subsiguientes ciclos de selección desarrollados por los mejoradores (Smith *et al.*, 2009). El método tradicional para la identificación de cultivares, basado en características morfológicas de las plantas, depende de las condiciones ambientales, del estado nutricional y de salud de los individuos analizados (Hvarleva *et al.*, 2007). Debido a ello, la evaluación precisa de la diversidad genética y el control de los genotipos usados en los programas de mejoramiento depende del desarrollo y aplicación de marcadores moleculares confiables.

Hoy en día la conservación de recursos genéticos es aceptada de forma generalizada como una responsabilidad social, dentro de un contexto mucho

más amplio para la preservación de la biodiversidad. En este escenario, a la pérdida de recursos genéticos dentro de las plantas cultivadas como consecuencia de la sustitución de variedades tradicionales por cultivares modernos, se suma la ocasionada en especies vegetales silvestres a causa del deterioro de los ecosistemas naturales por la creciente actividad humana (Irrondo Alegría, 2001). Se espera que para el año 2030 la población humana alcance la cifra de 8,9 billones; lo cual corresponde a un aumento de casi el 50% con respecto a la cifra actual (5,7 billones). Por otro lado, la disminución de las reservas de petróleo (base de los fertilizantes y pesticidas) y de los reservorios naturales de agua, en conjunto con el aumento en la contaminación agrícola y ambiental, tendría como consecuencia un aumento escaso, o sólo el mantenimiento, de los materiales cultivables en el mundo (Lobell *et al.*, 2008; Stocking, 2003).

La pérdida de diversidad genética ocasionada por el mejoramiento origina una vulnerabilidad importante de los cultivos frente a las Enfermedades Infecciosas Emergentes (EIEs), por ello la conservación de recursos genéticos representa la estrategia que potencialmente podría solucionar estas limitaciones. La causa fundamental de la mayoría de las EIEs en plantas es de origen antropogénico, ya sea por introducción de patógenos y sus vectores, por cambios en las técnicas agrícolas, o por disturbios en el hábitat inducidos por la globalización, el desarrollo socio-económico y los cambios de paradigma en el uso de las tierras. Otra causa importante de emergencia de las EIEs son los eventos climáticos severos, cuya frecuencia y severidad muestra una tendencia creciente a causa del cambio climático global (Daszak *et al.*, 2000). En muchos países en desarrollo, la disminución del precio de mercado de los granos básicos para la alimentación (trigo, arroz, maíz, papa) y la disponibilidad de irrigación todo el año ha promovido la intensidad de siembra de cultivos no tradicionales, condición que favorece la contaminación con patógenos e induce la pérdida de diversidad genética en la producción de cultivos y el aumento de la susceptibilidad a enfermedades (Anderson *et al.*, 2004; Harlan, 1995; Van der Wouw *et al.*, 2010).

Otro factor que impacta negativamente sobre la diversidad genética de los cultivos es la selección continua a través de los cruzamientos entre cultivares genéticamente relacionados. Esto ha producido un estrechamiento

de la base genética de aquellos cultivos en los cuales se basa la agricultura moderna, contribuyendo a la erosión genética de los programas de mejoramiento (Plucknett *et al.*, 1987; Van de Wouw *et al.*, 2009; Van Hintum *et al.*, 2007). Asimismo, el aumento del dominio de las grandes compañías multinacionales en mejoramiento de semillas podría generar una importante reducción de diversidad a escala global, como resultado de la diseminación masiva de semillas, aspecto que puede ser fuertemente limitante en la diversidad si se eligen sistemáticamente las mismas variedades. Si bien, existen en el mercado pocos cultivares de alto rendimiento muy cercanos entre sí, los avances que se han logrado en mejoramiento tendrán su impacto en el campo en los próximos años (Van de Wouw *et al.*, 2010).

En este contexto, resulta evidente que el potencial genético contenido en los bancos de semillas es enorme y sólo será posible aprovechar este recurso si se alcanza un cambio de paradigma desde la búsqueda de fenotipos adaptados hacia la búsqueda de genes superiores con el auxilio de los mapas moleculares (Tanksley and McCouch, 1997). La conservación de germoplasma tiene entre sus objetivos preservar la diversidad de determinada especie y proveer de insumos para los programas de mejoramiento de la misma. Estos programas están orientados a optimizar diversos caracteres agronómicos como el rendimiento, la productividad, la calidad y la sustentabilidad, entre otros. Por ello, la disponibilidad de germoplasma constituye la base de los programas de mejoramiento, y su caracterización fenotípica y molecular es fundamental para optimizar la preservación de los recursos genéticos, evitando duplicaciones, y para la conformación de una fuente de genes que pueda ser utilizada en el mejoramiento agrícola de los cultivos (Koornneef and Stam, 2001; Takeda and Matsuoka, 2008).

La conservación de germoplasma implica tanto el mantenimiento de colecciones vegetales como la creación de bancos de germoplasma. La primera de estas aproximaciones incluye colecciones *ex situ* tales como jardines botánicos, o colecciones a campo, e involucra la recolección de muestras representativas de la variabilidad genética de una especie y su mantenimiento fuera de las condiciones naturales en las que la especie ha evolucionado. Las ventajas de este tipo de colecciones son el control directo sobre el material, la fácil accesibilidad y la disponibilidad (Reid and Miller,

1989). El segundo enfoque involucra a los bancos de semillas, de cultivo *in vitro*, de polen y de ADN. Estos métodos permiten la conservación de un gran número de entradas de una misma especie en espacios pequeños, un mayor control del estado fitosanitario de las mismas, la reducción de los tiempos de multiplicación y la facilidad de intercambio de material genético (Iriando Alegría, 2001). Tradicionalmente, el enfoque para la utilización de materiales mantenidos en bancos de semillas ha sido el estudio de las entradas por su apariencia fenotípica. El análisis comienza con las observaciones fenotípicas y luego se profundiza a nivel genético y molecular, enfoque conocido como “*top-down*”. En contraposición con esta estrategia, existe la alternativa de abordar inicialmente un estudio a nivel de secuencias genómicas y luego el análisis del fenotipo, lo cual se conoce como genética reversa o “*bottom-up*” (Prada, 2009). El desafío actual consiste en articular ambos enfoques, dado que el estudio fenotípico debe complementarse con el genotípico, y lograr un equilibrio entre el número de entradas manipuladas en el campo, el costo de la fenotipificación, la accesibilidad de las técnicas moleculares, la eficiencia de las mismas, el costo de su implementación, etc.

14. Caracterización genética de cultivos con marcadores moleculares

En las últimas dos décadas el uso de marcadores de ADN para el estudio de diversidad genética de cultivos comenzó a ser rutinario. Las técnicas que se han desarrollado otorgan mayor precisión y rapidez, y son más económicas para la evaluación de la variación genética. Los distintos marcadores (RFLP: *Restriction Fragment Length Polymorphism* o Polimorfismo de Longitud del Fragmento Digerido, RAPD: *Random Amplified Polymorphic DNA* o Polimorfismo en la Amplificación Aleatoria de ADN, AFLP: *Amplified Fragment Length Polymorphism* o Polimorfismo en la Longitud del Fragmento Amplificado, SSR: *Simple Sequence Repeat* o Secuencias Repetidas Simples, SNP: *Single Nucleotide Polymorphisms* o Polimorfismos de Nucleótido Simple e Indel: *Insertion/Deletion* o pequeñas inserciones o deleciones) difieren con respecto a características tales como abundancia en el genoma, nivel de polimorfismo detectado, especificidad de los loci analizados, reproducibilidad, requerimientos técnicos e inversión económica. La identificación del marcador

genético más apropiado dependerá de la aplicación específica, del nivel de polimorfismo requerido, de la presencia de facilidades técnicas suficientes y del conocimiento previo o *know-how*, de la disponibilidad de tiempo y de las limitaciones económicas (Spooner *et al.*, 2005).

Los marcadores de ADN tienen aplicaciones diversas como la construcción de mapas de ligamiento genético, la identificación de caracteres genéticos simples, el análisis de loci de caracteres cuantitativos y la aplicación de estrategias de mejoramiento molecular a través de la selección e introgresión asistida por marcadores moleculares. En cuanto al germoplasma, también han sido ampliamente utilizados en análisis de diversidad, estudios de relaciones filogenéticas entre especies y taxones superiores, identificación de materiales duplicados y análisis molecular de procesos de regeneración de germoplasma. Dentro de los primeros ejemplos de aplicación de SSR en la caracterización de germoplasma, se encuentran la detección de entradas redundantes en una colección de sorgo (Dean *et al.*, 1999), la identificación de diferentes cultivares de vid (Thomas *et al.*, 1994), la comparación de razas locales y cultivares de arroz (Yang *et al.*, 1994) y el desarrollo de perfiles de ADN únicos en genotipos de soja (Rongwen *et al.*, 1995). Además, se emplearon SSR en arroz para discriminar cultivares relacionados (Akagi *et al.*, 1996) y para diferenciar entradas de una colección núcleo de genotipos de manzana (Hokanson *et al.*, 1998). Actualmente, la aplicación de marcadores moleculares se ha transformado en una actividad de rutina para la caracterización del germoplasma de diversas especies (Bracco *et al.*, 2009; Cai *et al.*, 2005; Furini and Wunder, 2004; Fusari *et al.*, 2008; Fusari, 2010; Giancola *et al.*, 2006; Kumar *et al.*, 2009; Laurentin, 2009; Mago *et al.*, 2005; Yan *et al.*, 2009; Yu *et al.*, 2009; entre otros).

Otra aplicación de los marcadores moleculares es el mapeo genético. Los mapas basados en marcadores moleculares han sido desarrollados para la mayoría de las principales especies cultivadas. Estos mapas permiten la localización de loci que están relacionados con caracteres complejos (QTL, *Quantitative Trait Loci* o Loci de Caracteres Cuantitativos) agronómicamente importantes, permiten hacer clonado genético posicional, mapeo comparativo y selección asistida por marcadores. Existen actualmente numerosos ejemplos de mapeo de QTL en diversas especies (Allinne *et al.*, 2009; Baxter *et al.*,

2005; Causse *et al.*, 2004; Doebley *et al.*, 2006; Ebrahimi *et al.*, 2008; Haddadi *et al.*, 2010; Hyten *et al.*, 2004; Li *et al.*, 2006; Poormohammad Kiani *et al.*, 2007; Poormohammad Kiani *et al.*, 2009; entre otros). Recientemente, se ha implementado una estrategia nueva conocida como mapeo por asociación, con el objeto de detectar regiones genómicas asociadas a caracteres agronómicos de interés en materiales de germoplasma de diversas especies (Breseghello and Sorrells, 2006; Malosetti *et al.*, 2007; Yu and Buckler, 2006; entre otros). El mapeo por asociación tiene por objetivo establecer una correlación entre marcadores genéticos y el fenotipo de interés en poblaciones no estructuradas, que en el caso de las plantas generalmente consisten en colecciones de germoplasma. La ventaja de esta aproximación es que puede examinarse simultáneamente un elevado número de alelos, mientras que en el mapeo de QTL tradicional (poblaciones biparentales) sólo se evalúan los alelos presentes en los parentales.

En girasol, los marcadores moleculares han sido empleados para la caracterización genética de líneas, cultivares y especies silvestres (Brunel, 1994; Burke *et al.*, 2002; Dehmer and Friedt, 1998; Ellis *et al.*, 2006; Liu and Burke, 2006; Paniago *et al.*, 2002; Solodenko and Sivolap, 2005; Solodenko *et al.*, 2004; Tang and Knapp, 2003; Yu *et al.*, 2002). Por otro lado, se han empleado marcadores SSR en el desarrollo de mapas genéticos (Al-Chaarani *et al.*, 2004; Lai *et al.*, 2005; Paniago *et al.*, 2002; Poormohammad Kiani *et al.*, 2007; Talia *et al.*, 2010; Tang *et al.* 2003, Tang *et al.*, 2002; Yu *et al.*, 2003).

Asimismo, se han descrito QTL para varios caracteres agronómicos como resistencia a *Sclerotinia sclerotiorum*, *Plasmopara halstedii* y *Phoma macdonaldii* (Al-Chaarani *et al.*, 2004; Bert *et al.*, 2004; Maringolo, 2007; Micic *et al.*, 2004; Micic *et al.*, 2005; Poormohammad Kiani *et al.*, 2007; Rönicke *et al.*, 2005), para caracteres de germinación (Al-Chaarani *et al.*, 2005) y para caracteres fisiológicos, entre ellos la variación del nivel de agua en la planta y el ajuste osmótico bajo condiciones de estrés de agua (Hervé *et al.*, 2001; Poormohammad Kiani *et al.*, 2007). Por último, recientemente se han desarrollado estudios genéticos en base a marcadores funcionales de tipo SNP, tendientes a establecer los niveles de diversidad nucleotídica y desequilibrio de ligamiento en girasol, como paso inicial de los estudios de

mapeo por asociación (Fusari *et al.*, 2008; Fusari, 2010; Kolkman *et al.*, 2007; Lai *et al.*, 2005; Liu and Burke, 2006).

Los marcadores microsatélites son precisos para identificación de cultivares, para evaluación de diversidad genética y para protección de la propiedad intelectual, pero se ha sugerido que deberían ser complementados con marcadores funcionales como los SNP (Hvarleva *et al.*, 2007). Estos últimos son más estables (baja frecuencia de mutación) que los SSR y están asociados no sólo a regiones neutras sino a regiones codificantes, por lo tanto serían más adecuados para la prueba de estabilidad de variedades y para los análisis poblacionales de las mismas. La distancia genética estimada en base a marcadores moleculares es un parámetro importante para identificar variedades y determinar las relaciones de parentesco entre las mismas. De este modo, la distancia genética estimada se utiliza como un indicador de la verdadera distancia genética entre dos genotipos (Heckenberger *et al.*, 2002; Heckenberger *et al.*, 2003; Heckenberger *et al.*, 2005). En resumen, para generar resultados confiables que puedan ser empleados con fines legales, se requieren estimas de diversidad precisas y reproducibles que involucren el empleo de marcadores moleculares capaces de alcanzar altos niveles de discriminación, utilizando técnicas automatizables y con gran capacidad de procesamiento.

La utilización de marcadores moleculares es también un aspecto relevante desde el punto de vista de la regulación de obtenciones vegetales. La selección de descriptores reviste un problema, dado que los caracteres agronómicos suelen ser complejos de medir y/o estandarizar y no existen suficientes antecedentes que avalen el uso de marcadores moleculares como descriptores. Actualmente, la caracterización de variedades vegetales está basada en una serie de caracteres fenotípicos (fisiológicos, morfológicos, fenológicos y de comportamiento sanitario), pero se evalúa la incorporación de técnicas moleculares para la prueba de Distinguibilidad, Homogeneidad y Estabilidad de las variedades vegetales (prueba DHE) según la normativa propuesta por la Unión Internacional para la Protección de Nuevas Variedades de Plantas (UPOV: *International Union for the Protection of New Varieties of Plants*) y para la identificación con fines de certificación según la Asociación Internacional de Ensayos de Semillas (ISTA: *International Seed Testing*

Association) y OECD (*Organization for Economic Co-operation and Development Trade and Agriculture Directorate*). Los marcadores moleculares proveen un mayor poder discriminante que los marcadores morfológicos, además de una mejor consistencia con la información de *pedigree* de los materiales evaluados. Ello indica que es posible integrarlos en los protocolos de análisis para evaluación de variedades descriptos por la UPOV (Gunjaca *et al.*, 2008).

15. Mejoramiento vegetal

La agricultura moderna es dependiente del cultivo de unas pocas especies altamente productivas. Tal como lo mencionan Tanksley y McCouch (1997) y Koornneef y Stam (2001), los cultivos alimenticios fueron inicialmente domesticados desde especies silvestres hace aproximadamente 10.000 años atrás, durante la transición del hombre de la vida nómada, con hábitos recolectores y cazadores, hacia la vida en “sociedades agrarias”. Los agricultores recolectaban semillas y órganos reproductivos vegetativos (tubérculos y bulbos) y almacenaban parte de ellos para plantar en la próxima estación. Durante el proceso fueron seleccionando las variantes que mostraban mayor rendimiento y mejor adaptación para el cultivo. Esas variantes produjeron semillas para propagarse en mayor cantidad y con buena sanidad, dando origen al “síndrome de la domesticación”. La hibridación no controlada en los campos de cultivo, en ambiente natural, pudo haber aumentado la variación en el germoplasma disponible para los agricultores. Luego de muchos siglos de selección consciente o inconsciente por parte de los agricultores, con la presión de selección impuesta por condiciones de crecimiento heterogéneas a nivel espacial y temporal, se originaron las razas locales que son poblaciones genéticamente heterogéneas que están adaptadas a condiciones impuestas por el hombre y el ambiente físico. De este modo, las razas locales, representan el primer paso en el proceso de domesticación y, junto con las especies silvestres, proveen una amplia representación de la variación natural que ocurre en el total de las especies (Ferne *et al.*, 2006; Tanksley and McCouch, 1997).

El primer cambio en el mejoramiento de plantas a nivel más especializado ocurrió con pequeños productores que comenzaron deliberadamente a seleccionar cultivares y a comercializar las semillas de calidad mejorada. Se fueron seleccionando variantes superiores mediante la selección en masa y ocasionalmente por selección de familias, aumentando así la uniformidad del cultivo. Hasta comienzos del siglo 20, la hibridación controlada para crear por selección nuevas variedades fue poco desarrollada.

En síntesis, la domesticación y el mejoramiento de plantas han sido indudablemente exitosos para el aumento de las frecuencias de alelos beneficiosos para caracteres agronómicamente importantes, pero muchos de estos alelos beneficiosos se han perdido en el cuello de botella generado como consecuencia del proceso de domesticación y a través de los años de mejoramiento moderno y selección (Figura 3).

En las últimas décadas del siglo 20 se han producido grandes avances tecnológicos que han permitido un desarrollo creciente del mejoramiento, entre ellos el cultivo de células y de tejidos, la fusión de protoplastos, la transformación genética y el mejoramiento asistido por marcadores moleculares (Koornneef and Stam, 2001).

En particular, la aplicación de marcadores moleculares ha sido una de las herramientas de mayor impacto sobre el mejoramiento. Estas técnicas han permitido asistir a la introgresión de genes desde germoplasma exótico por monitoreo de la fracción de genoma receptor y dador a través de generaciones sucesivas de autofecundaciones. La principal ventaja de esta aplicación es la reducción a la mitad del número de generaciones necesarias para alcanzar una línea isogénica de introgresión, comparado con el proceso clásico o tradicional. Por otro lado, el enfoque de QTL con la combinación de las autofecundaciones y retrocruzas, resulta útil para el control de caracteres cuantitativos que han sido identificados en materiales de germoplasma no adaptados a la agricultura, pero que son aptos para la transferencia a líneas de mejoramiento avanzadas (Koornneef and Stam, 2001). Cabe destacar que todas las herramientas tecnológicas descritas y sus respectivas aplicaciones requieren, necesariamente, de la materia prima para implementarlas. En este sentido, cobran vital importancia los recursos genéticos. El establecimiento y mantenimiento de bancos de semillas debe complementarse con la habilidad

para utilizar activamente los materiales de estas colecciones. La mayoría de las entradas de los bancos no contribuyen sustantivamente a la formación de variedades modernas, especialmente con respecto a los caracteres complejos, tales como rendimiento y calidad nutricional. Los análisis moleculares realizados hasta la fecha sugieren que la diversidad genética almacenada en los bancos de semillas puede ser empleada más eficientemente. Es posible que muchos alelos favorables estén ocultos entre las numerosas entradas mantenidas en los bancos de semillas, por ende la integración de las bases de datos a nivel fenotípico y genómico es una estrategia sumamente útil para ser aplicada en el mejoramiento (Ferne *et al.*, 2006).

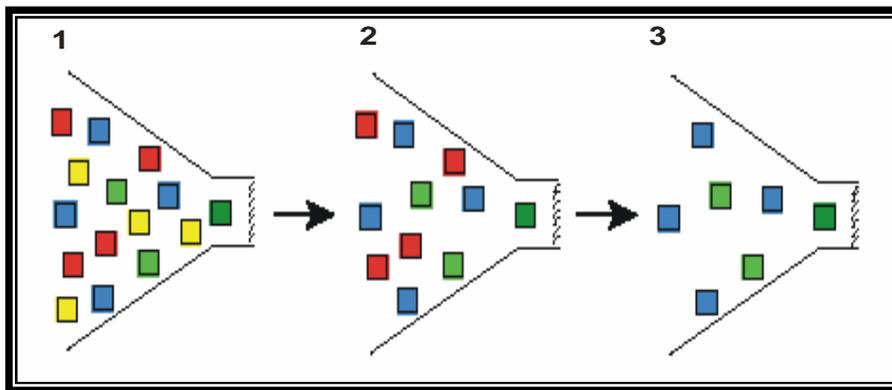


Figura 3. Efecto genético o cuello de botella impuesto durante el proceso de domesticación de plantas y la aplicación de programas de mejoramiento modernos. Los cuadros de colores representan las variantes alélicas halladas en las especies silvestres. Se muestra la pérdida gradual de las variantes a través de los procesos de domesticación y mejora. 1. Especies silvestres. 2. Especies recientemente domesticadas. 3. Variedades modernas. (Modificado de: Tanksley and McCouch, 1997).

16. Programas de mejoramiento de girasol en el país

Los programas de mejoramiento para girasol, conducidos tanto por instituciones oficiales como privadas, trabajan en la generación de nuevas combinaciones para el posterior desarrollo de líneas y la creación y evaluación de nuevos híbridos. El objetivo principal de cualquier programa es lograr mayor rendimiento, que para el caso de girasol, en general, se expresa en kilos de aceite por hectárea. Sin embargo, 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).

Con respecto al control eficiente de malezas, los programas se focalizan en la incorporación de resistencia a herbicidas de la familia de las imidazolinonas (IMI) en híbridos comerciales, aspecto que muestra avances concretos tras varios años de trabajo (Sala *et al.*, 2008; Vega, 2008).

Para el manejo de las enfermedades se tienen en cuenta cuatro patógenos que son los más importantes en Argentina: *Sclerotinia sclerotiorum* que causa la Podredumbre Húmeda de Capítulo en Girasol (PHCG), *Verticillium dahliae* que produce el Secado Anticipado y Quebrado del Tallo (SAyQT), *Plasmophara halstedii* conocido como *downy mildew* que origina enanismo y *Puccinia helianthii* o roya negra.

En cuanto a la nutrición mineral, se encuentran en desarrollo métodos de fertilización, estudios de deficiencias de nutrientes esenciales y micronutrientes, formas de aplicación y su interacción con sistemas de labranza (Bruniard, 2002).

Por otro parte, las tendencias del mercado apuntan a productos de calidades diferenciadas y esta necesidad se traslada al desarrollo de los nuevos materiales. En el mejoramiento para calidad de aceite se apunta a aumentar el contenido de ácido oleico como principal carácter, dado que así se incrementa la estabilidad oxidativa, la longevidad de las frituras y se agiliza la cocción logrando menor absorción de aceite. Dentro de esta línea de mejoramiento, también se intenta desarrollar materiales que muestren bajo porcentaje de ácido palmítico y esteárico, lo cual disminuiría el contenido de ácidos grasos saturados. A estas características se suma el mejoramiento para aumentar el contenido de tocoferoles, precursores de la vitamina E.

En relación con la línea de mejoramiento focalizado a fisiología, se estudian mecanismos que permitan aumentar el rendimiento y mantenerlo estabilizado a lo largo de las campañas. Los esfuerzos biotecnológicos en este campo están centrados en la modificación de la estructura de la planta, la mejora de la resistencia a las condiciones adversas y la adaptación a nuevos ambientes (Bruniard, 2002; López Pereira *et al.*, 2008; Radonic *et al.*, 2008; Zavallo *et al.*, 2010).

Recientemente, se estimó que el 24,2% del área geográfica mundial es potencialmente arable, mientras que sólo el 10,6% de la misma está actualmente bajo cultivo, la porción restante no está disponible debido a la existencia de uno o más factores abióticos que ocasionan estrés (Singh, 2000). De acuerdo con este autor, la sequía es el principal factor abiótico, seguido en orden de importancia por la deficiencia y toxicidad por minerales y, en tercer lugar, el frío. La sequía es el factor más limitante dado que afecta por encima de un tercio de los suelos del mundo. Por otro lado, no sólo causa diferencias entre el rendimiento promedio y el rendimiento potencial, sino que además origina variaciones de rendimiento de año a año y consecuentemente inestabilidad de este carácter (Monte, 1986). En concordancia con estas observaciones, uno de los caracteres más importantes para el mejoramiento de girasol es el desarrollo de materiales con tolerancia al déficit hídrico (*Ing. Álvarez, comunicación personal*) (Pereyra-Irujo *et al.*, 2008).

El Banco Activo de Germoplasma de INTA Manfredi (BAG-IM) es el único banco activo de germoplasma de girasol a nivel nacional y funciona desde hace más de 30 años, albergando cerca de 1200 entradas que provienen de diversos orígenes geográficos y que están generadas a partir de diferentes metodologías de cruzamiento. Por lo tanto, el material disponible dentro del BAG-IM incluye diversas categorías tales como poblaciones, compuestos y líneas. Los orígenes geográficos de las entradas coleccionadas 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, Yugoslavia. Las colecciones fueron realizadas directamente en los lugares de cultivo y posteriormente recibidas, conservadas y multiplicadas por el BAG-IM. Como se mencionó anteriormente, los materiales allí conservados constituyen una invaluable fuente de materia prima para el desarrollo de los programas de mejoramiento en curso. De este modo, el programa de mejoramiento para tolerancia al déficit hídrico que lleva adelante el BAG-IM, es uno de los más importantes del banco, coincidentemente con los antecedentes bibliográficos expuestos en esta sección.

HIPÓTESIS

La hipótesis de este trabajo es que existe diversidad genética cuantificable entre los materiales de girasol cultivado almacenados en el Banco Activo de Germoplasma de girasol de INTA Manfredi (BAG-IM), capaz de ser caracterizada mediante la utilización de marcadores moleculares, y que su conocimiento aportará valiosos insumos para los programas de mejoramiento del cultivo.

OBJETIVOS

Objetivo general

El objetivo general de este trabajo es generar conocimiento sobre el nivel de diversidad genética en las entradas del BAG-IM para orientar la caracterización exhaustiva del mismo y asistir a los programas de mejoramiento de girasol de INTA.

Objetivos específicos

- I. Estudiar los niveles de diversidad genética en entradas de girasol cultivado de diversas categorías (poblaciones, compuestos y líneas) pertenecientes al BAG-IM, mediante el empleo de marcadores microsatélites.
- II. Estudiar la distribución de los polimorfismos detectados y la estructura poblacional en las entradas analizadas.
- III. Evaluar la consistencia entre los resultados obtenidos en el ensayo de campo para tolerancia a sequía con respecto al ensayo bajo condiciones controladas de déficit hídrico para la fenotipificación de un conjunto de líneas de girasol pertenecientes al programa de mejoramiento del BAG-IM para tolerancia a este estrés abiótico.
- IV. Estudiar la diversidad nucleotídica, el desequilibrio de ligamiento y la presencia de desvíos en los patrones de neutralidad en regiones candidatas asociadas al carácter tolerancia a déficit hídrico, utilizando el conjunto de líneas de girasol del correspondiente programa empleadas en la caracterización fenotípica.

CAPÍTULO 1

Diversidad genética en entradas de girasol cultivado pertenecientes al Banco Activo de Germoplasma de INTA Manfredi

INTRODUCCIÓN

El objetivo final del mejoramiento vegetal es encontrar aquellos materiales que presenten mayor variabilidad genética, facilitando así la detección de alelos favorables asociados a caracteres de interés. De este modo, se seleccionan materiales que se cruzan para reunir en uno solo las características de los progenitores elegidos. En función del modo de obtención de las distintas entradas, se distinguen diversas categorías tales como las poblaciones, los compuestos, los cultivares y las líneas endocriadas.

Las poblaciones se generan partiendo desde materiales que poseen intrínsecamente un vigor híbrido asociado, es decir un grado de heterosis alto. Pueden ser de dos tipos, el primero son las poblaciones heterogéneas homocigotas representadas por un conjunto de líneas puras y razas locales domesticadas. Estas poblaciones son fuentes de adaptación a clima, suelo y resistencia a enfermedades. El segundo tipo son las poblaciones heterogéneas heterocigotas construidas desde individuos segregantes pertenecientes a una generación F₂-F₃ (Conrero, 2007; Cubero, 1999). Una forma alternativa de generar poblaciones es a partir de la polinización abierta de un conjunto grande de individuos provenientes de distintos híbridos y/o especies silvestres, donde cada uno aporta su genoma, combinando tanto caracteres deseables como no deseables (*Ing. Álvarez, comunicación personal*). Luego, mediante el proceso de selección se actuará sobre caracteres deseables, heredables y genéticamente variables.

Para la construcción de los compuestos de plantas autógamias, se eligen individuos de líneas puras diferentes que muestran adaptación a diversos ambientes (Conrero, 2007). Las líneas elegidas para conformar el compuesto deben mostrar caracteres deseables para el mejorador y se cruzan entre sí, intentando crear un material en el que confluyan todas estas buenas características. Teóricamente, los valores de heterocigosis para esta categoría podrán ser iguales o inferiores a los de las poblaciones, dado que el conjunto de materiales de partida es más homogéneo.

Por su parte, los cultivares se obtienen a partir de materiales con altos niveles de heterosis (Brickell *et al.*, 2004) y pueden denominarse también variedades cultivadas, haciendo referencia al cultivo de las mismas por el

productor. Es decir que, los cultivares representan la etapa final del proceso de mejoramiento puesto que se comercializan en el mercado, mientras que las poblaciones y los compuestos son categorías empleadas sólo en etapas iniciales del mejoramiento y por ende únicamente el mejorador dispone de las mismas.

En las especies vegetales autóгамas, como el girasol, la descendencia de cada planta se produce como consecuencia de la fecundación de las ovocélulas por su propio polen. Esto da lugar a un aumento de la homocigosis o disminución de la heterocigosis hasta una “eliminación práctica” en el material en estudio. De este modo se obtienen las líneas puras endocriadas. El término “eliminación práctica” hace referencia a que la determinación de la homocigosis se realiza por evaluación del fenotipo en ensayos de campo. Lograr la homocigosis de todos los loci es un proceso lento, dado que en cada generación de autofecundación se reduce a la mitad la proporción de heterocigotas. En promedio, se requiere de 8 generaciones de autofecundación para alcanzar un nivel de homocigosis superior al 99% en un único locus (Meksem and Kahl, 2005), mientras que para 5 loci deberán realizarse 11 ciclos de autofecundación (Cubero, 1999). Las líneas de girasol que se designan como “B” corresponden a líneas mantenedoras (por ej.: B67), en relación con su comportamiento frente al carácter de fertilidad. Todas aquellas líneas cuya designación tiene una letra “A”, son líneas recurrentes o recesivas porque no poseen polen viable y por ende son andro-estériles (por ej.: A59), mientras que si se designan con una letra “R” en su nomenclatura, son líneas donantes o dominantes porque restauran la fertilidad (por ej.: R432, R307, entre otras).

Banco de germoplasma de INTA

Como ya se mencionó en la introducción general, el BAG-IM es el único banco activo de girasol a nivel nacional y funciona desde hace más de 30 años, albergando cerca de 1200 entradas o introducciones de *Helianthus annuus* que provienen de diversos orígenes geográficos, y que están generadas a partir de diferentes metodologías de cruzamiento. Dado que la caracterización de la totalidad de los recursos genéticos mantenidos en un banco de semillas es prácticamente inalcanzable por el costo efectivo que esto implica, Frankel

(1984) propuso el desarrollo de colecciones núcleo o *core collections*. Las colecciones núcleo, que comprenden aproximadamente un 10% de la colección original, están conformadas por un subgrupo de materiales elegidos para representar la mayor proporción posible de la diversidad genética hallada en la colección original. Las colecciones núcleo resultan sumamente eficaces para mantener germoplasma adecuadamente caracterizado y se utilizan como fuente primaria de semilla en el fitomejoramiento de los cultivos (Brown, 1995). De este modo, la estrategia de identificación de colecciones núcleo resulta una valiosa herramienta para el estudio del espectro de diversidad genética dentro de un banco. La riqueza alélica determinada mediante marcadores moleculares constituye un complemento imprescindible de la información fenotípica para caracterizar materiales y constituir colecciones núcleo (Kresovich *et al.*, 1995).

Marcadores microsatélites

Los marcadores microsatélites o SSR (*Simple Sequence Repeat*) consisten en repeticiones en tándem de secuencias cortas, entre 1 a 6 pares de bases. Cuando la secuencia de nucleótidos en las regiones flanqueantes del microsatélite es conocida, se pueden diseñar iniciadores o cebadores específicos de 20 a 25 pb para amplificar el microsatélite por reacción en cadena de la polimerasa (PCR: *Polymerase Chain Reaction*). El deslizamiento (*slippage*) de la enzima polimerasa durante la replicación del ADN es la principal causa de variación en el número de unidades repetidas de un microsatélite, generando polimorfismos de distinta longitud que pueden detectarse por electroforesis en geles de agarosa de alta pureza al 3%, en geles de poliacrilamida al 6% y por electroforesis capilar, en orden creciente de precisión. La detección de la variación microsatélite puede ser automatizada si se emplea un secuenciador automático (Spooner *et al.*, 2005).

Las ventajas de este tipo de marcadores son la codominancia de alelos, la alta abundancia genómica en eucariotas y la distribución al azar a través del genoma con asociación preferencial a regiones de bajo número de copias (Morgante *et al.*, 2002). No requiere alta cantidad de ADN porque recurre al uso de la amplificación por PCR. Múltiples microsatélites pueden ser amplificados durante una misma reacción de PCR y detectados por electroforesis capilar

automatizada, siempre que no exista superposición en el rango de tamaño de los alelos de diferentes loci (Ghislain *et al.*, 2004). Para ello se requiere también que los fluoróforos empleados como marcadores en los cebadores a combinar en la mezcla de PCR sean distintos.

Las desventajas de este tipo de marcadores son el alto costo para el desarrollo de iniciadores si las secuencias específicas de interés son desconocidas, la presencia de mutaciones en el sitio de hibridación del iniciador que conducirá a la aparición de alelos nulos (ausencia de amplificación), lo cual generará errores en el genotipado (Brownstein *et al.*, 1996).

Los microsatélites son frecuentemente utilizados para mapeo genético y en estudios genético-poblacionales, desde el nivel de individuos hasta el de especies estrechamente relacionadas (Spooner *et al.*, 2005). En girasol existen antecedentes previos que muestran niveles altos de variabilidad SSR entre poblaciones naturales de *Helianthus annuus*, comparables a la diversidad observada en otras especies diploides del género, como *H. deserticola* y *H. petiolaris* (Gross *et al.*, 2003), o las más raras como *H. verticillatus* y *H. angustifolius* (Ellis *et al.*, 2006; Zhang *et al.*, 2005). Del mismo modo, los marcadores microsatélites han sido también de gran utilidad para la caracterización de líneas endocriadas y razas nativas de girasol cultivado (Tang and Knapp, 2003).

En base a los antecedentes presentados en los párrafos precedentes, en el presente Capítulo se abordarán como objetivos específicos el estudio de los niveles de diversidad genética mediante marcadores microsatélites en entradas de girasol cultivado de diversas categorías (poblaciones, compuestos y líneas) pertenecientes al BAG-IM, y el estudio de la distribución de los polimorfismos detectados y la estructura poblacional en las entradas analizadas. Previo al desarrollo de estos objetivos, se presentará el análisis de diversidad para caracteres morfológicos y agronómicos, el cual fue realizado con el fin de corroborar que los materiales seleccionados para el estudio de variabilidad microsatelite fueran representativos del espectro de diversidad fenotípica conservada en el BAG-IM.

MATERIALES Y MÉTODOS

M1. Muestras de material biológico

M1.1. Material incluido en el análisis de datos morfológicos y agronómicos

Los materiales empleados para el análisis morfo-agronómico pertenecen a la colección de entradas almacenadas en el BAG-IM. Se eligieron 309 introducciones de orígenes geográficos diversos, entre ellos Alemania Oriental, Argentina, Armenia, Australia, Bolivia, Brasil, Bulgaria, Canadá, Chile, China, España, Estados Unidos, Francia, Grecia, Hungría, Israel, Italia. Los materiales utilizados fueron asignados a diversas categorías de acuerdo con el método empleado para su generación. Esta selección de materiales correspondió al criterio del mejorador en el marco de las campañas anuales de multiplicación del BAG-IM (*Ing. Álvarez, comunicación personal*).

M1.2. Material para la caracterización genética

Para la caracterización genética se eligieron 21 materiales del BAG-IM, considerando su origen geográfico, los diferentes programas de mejoramiento a los que pertenecen y los resultados del análisis de agrupamiento realizado sobre el conjunto de las 309 entradas descriptas en la sección M1.1 en base a caracteres morfológicos y agronómicos.

Las 21 entradas (337 individuos en total) incluyen representantes de cuatro categorías, 8 poblaciones, 5 compuestos, 1 cultivar y 7 líneas (Tabla 1). Las poblaciones empleadas en este trabajo provienen de orígenes geográficos diversos y fueron obtenidas por el método de obtención de polinización abierta descrito anteriormente (ver Introducción-Capítulo 1).

Las 8 poblaciones se seleccionaron con el objeto de encontrar alelos exclusivos propios de cada programa de mejoramiento, de acuerdo con los diversos orígenes geográficos. En la categoría de los compuestos, todos ellos de origen argentino, HAR-1, HAR-2, HAR-3 y HAR-4 fueron elegidos por

mostrar resistencia a distintas razas de *Puccinia helianthi*, agente causal de la roya negra y el primer patógeno conocido para girasol en Argentina (Pereyra and Escande, 1994). Por otro lado, el hecho de provenir de variedades originales de Argentina, constituye una ventaja al momento de adaptar los materiales a la región. El quinto compuesto seleccionado fue Comangir, que también se formó a partir de materiales locales y que mostró una amplia variación en diversos caracteres de importancia agronómica. Dentro de la categoría de los compuestos se incluyó además un cultivar (Colliguay), dado que fue el único representante de su categoría y que se asume un grado similar de heterocigosis. Finalmente, la categoría de las líneas incluyó un total de 7, de las cuales 4 son provenientes de Argentina (L-156, P-94, F-164 y MP-557), situación que resulta óptima por la adaptabilidad de las mismas a la zona de cultivo y a las condiciones existentes en nuestro país. Por su parte, PAC-1 y CM-307-1, provienen de Francia y HA336 de Estados Unidos. La línea PAC-1 se eligió por haber demostrado tolerancia frente al patógeno *Sclerotinia sclerotiorum* que produce ataque severo a nivel del capítulo de la planta de girasol, dando origen a la PHCG (Escande *et. al.*, 2002; Escande *et al.*, 2006). La línea CM-307-1 posee un ciclo precoz y por ende resulta un material interesante para emplear al momento de cruzarlo con otras entradas de ciclo más largo, logrando reducir el mismo. Por último, HA336, se eligió porque mostró resistencia frente al ataque del patógeno *Plasmopara halstedii* que genera la enfermedad conocida como enanismo o *downy mildew* (Pereyra and Escande, 1994).

Tabla 1. Entradas del BAG-IM incluidas en el estudio de diversidad genética. KLM 209: KLM, IAC-Anhandy: IAC, Puntano x Smena: PxS, VNIIMK 8931-1: VNIIMK, Forestal cambá: F. cambá.

<i>Nombre del material</i>	<i>Categoría</i>	<i>Nº individuos evaluados</i>	<i>Localidad mejoramiento o re-selección</i>
KLM 209	Población	24	Argentina
IAC-Anhandy	Población	24	Brasil
Sunrise	Población	19	Canadá
Júpiter	Población	24	Alemania
Puntano x Smena	Población	20	Argentina
VNIIMK 8931-1	Población	24	Rusia
Prao-co	Población	24	Italia
Forestal cambá	Población	15	Argentina
HAR-1	Compuesto	24	Argentina
HAR-2	Compuesto	20	Argentina
HAR-3	Compuesto	14	Argentina
HAR-4	Compuesto	32	Argentina
Comangir	Compuesto	24	Argentina
Colliguay	Cultivar	23	Chile
L-156	Línea	4	Argentina
P-94	Línea	2	Argentina
CM-307-1	Línea	4	Francia
F-164	Línea	4	Argentina
PAC-1	Línea	4	Francia
MP-557	Línea	4	Argentina
HA336	Línea	4	Estados Unidos

M2. Análisis de datos morfológicos y agronómicos

Con el fin de realizar un análisis parcial de las aproximadamente 1200 entradas del BAG-IM, se eligieron 309 que fueron analizadas en campañas sucesivas de cultivo desde la temporada 2002-2003 hasta 2007-2008. Los resultados fueron obtenidos y gentilmente cedidos por el Ing. Álvarez (Grupo de Mejoramiento Vegetal de Girasol, INTA EEA Manfredi). Los caracteres evaluados fueron tanto cualitativos como cuantitativos. Dentro de los primeros se encontraron: tipo de superficie de la hoja (SH), forma de la hoja (FH) y margen de la hoja (MH), grado de inclinación del capítulo (IC), presencia de antocianinas en la pálea (AP), tipo de ramificación (RAM), coloración del grano (CG) y presencia de estrías en la superficie del mismo (EG). Dentro de los caracteres cuantitativos se analizaron ancho y largo de hoja (AH y LH), área foliar (AF), altura a floración (altura F), diámetro de capítulo (DC), diámetro de

tallos (DT), días de emergencia a madurez (SAM), peso de cien semillas (PCS) y porcentaje de materia grasa (MG).

Las referencias para evaluar cada carácter mencionado se tomaron en base a directivas propuestas por el Instituto Nacional de Semillas (INASE). La unificación de los criterios empleados en la evaluación, a fin de lograr que todos los caracteres tengan el mismo peso en la matriz final de distancias (Crisci, 1983), se describe en el Apéndice.

A partir de los datos correspondientes a las 309 entradas seleccionadas se construyó una matriz empleando la distancia taxonómica (Sneath and Sokal, 1973; Sokal and Sneath, 1963) y se construyó un dendrograma mediante el método de agrupamiento de encadenamiento promedio (UPGMA: *Unweighted Pair-Group Methods using Arithmetic Average*). Se calculó el coeficiente de correlación cofenética para determinar la bondad de ajuste del agrupamiento obtenido en base al conjunto de datos (Rohlf and Sokal, 1981) empleando la prueba de Mantel para normalizar los datos utilizando 1000 permutaciones. El coeficiente de correlación cofenética mide la correlación entre las interdistancias en el dendrograma y las interdistancias en la matriz de distancia sobre la que se aplicó el procedimiento de cálculo. Por ello se emplea como criterio de selección del algoritmo, dado que mientras mayor sea el valor del coeficiente, mejor ajusta el algoritmo elegido a los datos de partida. Para ello se empleó el programa NTSYS 4.0.

M3. Análisis de variabilidad de loci microsatélites

M3.1. Métodos experimentales

M3.1.1. Selección de loci SSR

La elección de los loci SSR se llevó a cabo en base a la información disponible en la bibliografía publicada para la especie en estudio y desarrollados en INTA Castelar (Fernández *et al.*, 2003; Paniego *et al.*, 2002; Poormohammad Kiani *et al.*, 2007). Los 16 loci analizados fueron seleccionados a partir de la evaluación experimental de un total de 35

marcadores, tomando como criterio su reproducibilidad, su facilidad de interpretación y la baja frecuencia de alelos nulos (menos al 10%) (Tabla 2). Todos los iniciadores empleados son públicos y pertenecen a la serie HA (Paniego *et al.*, 2002). Se eligió un iniciador microsatélite por cada Grupo de Ligamiento (GL) de la especie (17), excepto para los grupos 11 y 13. En el GL 11 existen dos marcadores HA disponibles, uno de ellos (HA3640) no generó producto de amplificación y el otro (HA3448) resultó multi-locus para todas las muestras ensayadas. Para el GL 13, ninguno de los marcadores estudiados generó productos de amplificación, mientras que para el GL 14 se genotipificaron dos marcadores. La numeración de los GL se corresponde con la del mapa de referencia de girasol (Tang *et al.*, 2002).

Tabla 2. Descripción de los loci microsatélites ensayados. Los marcadores seleccionados para el estudio poblacional se encuentran resaltados en gris.

a. Acceso del GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>)

Locus	Localización cromosómica	Motivo de repetición	Iniciador Forward 5'-3'	Iniciador Reverse 5'-3'	Nº de acceso ^a	Fluoróforo utilizado
HA 77	1	GA	HEX-TGTAATCTGTATCACTTCCACC	GTTGTTCTGTTAGGTCGTTCCG	BV727857	HEX
HA 95	5	GA	HEX-ACGCTTGATAGACAAATGCT	TAGGCAGCAGGTTTTACTCT	G67405	HEX
HA 293	14	GA	HEX-GGGGACATCTCCCGTCCACC	CCTCATCCATCTCCATCCAATC	G67519	HEX
HA 432	4	GT	HEX-GGGTTTAGTGGCCAGTAGTTGTC	CTTTATCCCCACCCCCTCC	G67407	HEX
HA 876	15	TGG	HEX-CGTTATTTATAAGACCCACCCC	CAACTTGGATGTGGCCGGAG	BV727953	HEX
HA 928	10	GT ATT	HEX-CCTTTGTAGTCCCTTACTGG	GGTCGATCATGTATGCGTGTTG	BV728013	HEX
HA 991	4	GA	FAM-GCCCCCTTGATGCCCTTTTC	GAATCGCCATTTGAATCGCCAG	G67520	FAM
HA 1402	17	GT	FAM-GTGTGAATGTGAGTGTGAATCC	GTCACAGTTGCATATCCATCC	BV727899	FAM
HA 1837	15	ATT	FAM-GTTATAAGCGCCAAATGGG	CACACACGTGCACTTTAAC	BV728003	FAM
HA 1848	7	ATT	HEX-CATCCCCTTCTGAATAGAAA	TAGGCTCGTTAAACTTACGG	BV728005	HEX
HA 2063	9	GA GT	FAM-CACCGCAGATGGCCAATTCAAC	GGTTCTGCTAACCTGCAATGCG	BV728113	FAM
HA 2077	14	GA	HEX-GATAAGGTTTTCTCTCTCCC	GAGAAAAATGAGCTGATACCG	BV727907	HEX
HA 2193	16	GA	FAM-CAATACACGCAAACCCCGAAC	GCACTGCCAAAACAGTTGAC	BV727902	FAM
HA 2920	3	ATG	FAM-AACGTGTTTCATTCCGGATGC	GAAGATTGGTATGATGGGGC	BV728085	FAM
HA 3059	12	GA	NED-GAGTGAGAGTAATCAATG	CTCCCTGTCATCACATCAC	BV728163	NED
HA 3073	12	GA	NED-GAGTCAATTGCTTGTATG	GCACTGAAACATTAGCATC	BV728166	NED
HA 3102	15	GA	NED-CATGATGAAGTACCACCTCCATC	CCTCTCTCATCATGTAAGCC	BV728366	NED
HA 3103	7	GA	NED-CCTTAAAAGGCAAACGATTCTCC	GTTACAGTACGGCGAGAATGTAGG	BV728174	NED
HA 3239	2	ATG	NED-CTTAATTCGGACATAAGCAGG	CTTTCATTATTCACAACCCCC	BV728208	NED
HA 3272	7	TA GA	FAM-TGTTCTCACACTTTTCACCA	TAGGCCCTTAAATTCAGATG	BV728357	FAM
HA 3278	8	GA TA	FAM-TAGCCAACCTTCTCTTTCCA	CTTTCCCCTGCACCCATA	BV728212	FAM
HA 3325	2	GA	FAM-GTCATTGCTTGAGTGTGAGA	CGACGCTGAACATATACTTG	BV728106	FAM
HA 3396	12	GT ATT	FAM-CACTGGAACACCTGAAAGAT	ATTCTTTGTTGCCTTTTGGAG	BV728228	FAM
HA 3448	11	TA GT	FAM-TCAAGATAGAACCTGATGTGCG	TTTATTAAGTTTCCGGGTTT	BV728225	FAM
HA 3555	12	CAT	FAM-GATATCTCTCATAAGTGCCG	GGTCTTGTGATGACGAAGC	BV728375	FAM
HA 3581	8	ATG	NED-GTAGAGGAGTATCGCCAATAGC	GTTCCAATGGCCAACGATTGTG	BV728253	NED
HA 3582	16	ATG	NED-TGACCTTAATCCTTGGAGAG	CCTCTAATTCGAAAACCTAAC	BV728254	NED
HA 3627	5	GT	HEX-GATGAAATGCCTCACAGACT	CATTGAAGATTGATGGATGA	BV728237	HEX
HA 3640	11	GA	NED-GCCATATCAATAACCAATCCC	CAAAGCATGTGAGGACCAGAATG	BV728259	NED
HA 3700	5	GA	FAM-GAAAAGATTGGGAGAGACCT	TTCTTCTTCCACCCACCTAT	BV728302	FAM
HA 3878	7	ATC ATT	FAM-TTTGTTTAGCATCATCATCATC	GAGACCCTAACCATACATGA	BV728314	FAM

<i>Locus</i>	<i>Localización cromosómica</i>	<i>Motivo de repetición</i>	<i>Iniciador Forward 5´-3´</i>	<i>Iniciador Reverse 5´-3´</i>	<i>Nº de acceso ^a</i>	<i>Fluoróforo utilizado</i>
HA 4011	13	GA	FAM-AC TTCTCCCTCCCCTTCTT	CTGTACACGTGCTGCTTTAG	BV728360	FAM
HA 4090	1	GT	NED-GCCATGATTGGCTAAGGTTCG	TGCGTTACCGACAACAAGG	BV728336	NED
HA 4103	6	GA	FAM-TCACTTATCACCAATCTCATCCA	CTGTCTCAAATCGGGTGGTT	BV728338	FAM
HA 4239	15	ATT	FAM- GAATGATAGTGAATTGAGACAGG	CTGGCATCTATATCCATGGATAG	BV728202	FAM

M3.1.2. Extracción de ADN genómico total

El ADN genómico total de los 337 individuos analizados fue extraído de hojas jóvenes no lesionadas por agentes patógenos ni insectos, provenientes de plantas crecidas en invernáculo por 5 semanas. Las hojas fueron liofilizadas luego del muestreo y se almacenaron en cámara a 4°C. La extracción se llevó a cabo con el equipo (*kit*) de extracción *NucleoSpin Plant II* (Macherey-Nagel, Alemania), empleando como agente extractivo el bromuro de cetil-trimetil amonio (CTAB) (ver Apéndice). Se empleó esta metodología dado que con los protocolos de extracción con CTAB (Saghai-Maroo *et al.*, 1984) y de Dellaporta *et al.* (1983) no fue posible obtener productos de amplificación a partir de cada una de las muestras de ADN. Por otro lado, en muchos casos los individuos mostraban signos de impurezas y contaminantes asociados al ADN, lo cual se solucionó mediante el empleo de columnas de purificación.

M3.1.3. Cuantificación del ADN

El ADN genómico se cuantificó empleando un fluorómetro Versaflúor (Biorad) mediante el uso del reactivo Hoechst N° 33258, estableciendo una curva patrón con estándares de ADN de timo de ternero (Biorad) de concentraciones conocidas (ver Apéndice).

Para determinar la calidad del ADN genómico obtenido se empleó electroforesis horizontal en geles de agarosa al 0,8% (p/v) y tinción con bromuro de etidio (0,5 µg/ml), utilizando como referencia concentraciones conocidas de ADN de timo de ternero (Biorad). Se sembró un volumen de muestra de 5 µl, incluyendo 2 µl de buffer de siembra (ver Apéndice). La electroforesis se llevó a cabo en buffer TAE 1X a un voltaje constante de 5 Volts/cm durante 30 min. La visualización del ADN se realizó por transiluminación con luz UV, en digitalizador de imágenes acoplado a PC con el programa Gene Quantitive (Biorad).

M3.1.4. Amplificación por PCR

La mezcla de reacción para la amplificación por PCR de todos los individuos estudiados se preparó como se indica en la Tabla 3.

Tabla 3. Mezcla de reacción para amplificación por PCR

Reactivo	Volumen
Buffer con (NH ₄) ₂ SO ₄ [10X] (Invitrogen)	1,20 µl
MgCl ₂ [50 mM]	0,36 µl
Desoxirribonucleótidos trifosfato (dNTPs) [10 mM]	0,24 µl
Iniciadores <i>Forward</i> + <i>Reverse</i> [5 µM]	0,30 µl
<i>Taq</i> Polimerasa Platinum (Invitrogen) [5 U/µl]	0,10 µl
H ₂ O Mili-Q o tridestilada estéril	8,80 µl
ADN [15 ng/µl]	1,00 µl

Las amplificaciones se realizaron en placa de 96 pocillos empleando un termociclador GeneAmp system 9700 (Applied Biosystems) con el programa de amplificación descrito en la Tabla 4.

Tabla 4. Programa de amplificación empleado en este estudio

Etapas del programa	Tiempo
Desnaturalización inicial a 94°C	4 min
<i>Touchdown</i> (12 ciclos)	
Desnaturalización a 94°C	45 s
Hibridación a 64°C (-1°C/ciclo)	45 s
Elongación a 72 °C	45 s
Amplificación (35 ciclos)	
Desnaturalización a 94°C	45 s
Hibridación a 52°C	45 s
Elongación a 72 °C	45 s
Extensión final a 72 °C	10 min
Mantenimiento a 15°C	

Todas las placas preparadas para la amplificación incluyeron controles realizados con ADN de la línea pura HA89 de genotipo conocido para algunos de los marcadores utilizados y un control negativo (sin adición de ADN de ningún tipo).

Las reacciones de PCR se llevaron adelante usando el método de disminución progresiva de la temperatura de hibridación (*touchdown*). Cuando los iniciadores no lograron producir amplificación empleando el *touchdown* descrito anteriormente, se probaron programas con temperaturas de

hibridación específicas (T_m : *melting temperature*) para cada par de iniciadores en base a su contenido de CG y de AT. Los valores de T_m básica (en solución acuosa) y de T_m ajustada en solución salina fueron obtenidos desde sitios públicos on-line (http://www.promega.com/biomatch/calc11.htm#melt_results).

Todos los iniciadores empleados en este trabajo se marcaron en el extremo 5' del iniciador *Forward* con alguno de los siguientes fluoróforos: 6-FAM (azul), HEX (verde) o NED (amarillo). La mezcla de reacción se preparó con dos marcadores SSR juntos para cada una de las muestras (*multiplex*) (Tang *et al.*, 2003), mientras que sólo en los casos negativos o problemáticos en cuanto a la obtención de bandas, se repitió la amplificación con un sólo marcador.

M3.1.5. Visualización de los productos de amplificación

M3.1.5.1. Empleo de geles de agarosa

La presencia de producto de amplificación luego de la reacción de PCR fue verificada mediante electroforesis horizontal en geles de agarosa al 2% (p/v) y tinción con bromuro de etidio (0,5 µg/ml). Se sembró un volumen de muestra de 5 µl, incluyendo 2 µl de buffer de siembra (ver Apéndice). La electroforesis se llevó a cabo en buffer TAE 1X a un voltaje constante de 5 Volts/cm durante 20 min. La visualización del ADN se realizó por transiluminación con luz UV, en digitalizador de imágenes acoplado a PC con el programa Gene Quantitive (Biorad). Se utilizaron dos marcadores de peso molecular conocido (10 y 25 pb, Invitrogen). Para la observación de los mismos fragmentos de PCR con la fluorescencia correspondiente según el fluoróforo marcador, se empleó el equipo *Typhoon* (Amersham).

M3.1.5.2. Empleo de electroforesis capilar

La separación de los productos de amplificación se efectuó mediante electroforesis capilar en secuenciador ABI 3130xl que mide la fluorescencia inducida por láser usando un detector de múltiples longitudes de onda. Se empleó como estándar de peso molecular el compuesto *GeneScan 500 (-250)*

ROX (6-carboxil-X-rodamina) que está constituido por un conjunto de fragmentos de tamaño conocido marcados con un fluoróforo de color rojo (Applied Biosystems). Las muestras se prepararon con 10 µl de formamida (*HiDi Formamide*) como agente desnaturalizante y 0,1 µl de *GeneScan 500 (-250) ROX* por cada placa de 96 pocillos. Se adicionaron 2 µl de las muestras (en una dilución adecuada para favorecer la visualización de los picos fluorescentes), se centrifugó, se desnaturalizó a 94°C 5 min y se centrifugó nuevamente previo a la inyección de las muestras en el equipo.

Los datos colectados en forma de electroferogramas fueron analizados automáticamente con el programa Genemapper 3.4 y 4.0 (Applied Biosystems) con el objeto de clasificar los alelos detectados según sus pesos moleculares.

Los criterios empleados para definir los alelos en base a los picos detectados por el programa se detallan a continuación:

1. Intensidad de fluorescencia: el valor mínimo de la misma deberá ser de 100, correspondiendo así a un posible alelo del marcador analizado. Cuando existen dos alelos de la misma forma, para determinar si el genotipo del individuo es realmente heterocigota, la fluorescencia del pico principal del segundo alelo deberá ser como mínimo 1/3 de la fluorescencia del pico principal del primer alelo (el de mayor intensidad) (Figura 1). Cuando un alelo excede la fluorescencia (desde aproximadamente 8000 o más), se suele quebrar el pico principal originando dos picos solapados que podrían generar perfiles falsos.
2. Forma del alelo: deberá ser constante en todas las muestras analizadas, lo cual es una medida de la certeza en cuanto a la especificidad del alelo detectado. El parámetro con respecto a la forma del alelo para el marcador en estudio será la línea control HA89. El alelo tomará el número de pares de bases del pico más alto, llamado pico principal. En algunos casos la forma del alelo intercambia el orden del pico principal, mostrando el pico adyacente más alto. En estos casos se tomará como criterio de alelo el primer pico (originalmente el principal). Estas situaciones suelen originarse en los casos donde hay agregado de una

adenina (adenilación) por parte de la polimerasa (Brownstein *et al.*, 1996).

3. Selección del rango (*bin*) del alelo: si el tamaño molecular del alelo corresponde a 148 pb, su rango será de 147,5-148,5; es decir que todos los picos principales que se ubiquen dentro de este rango serán adjudicados como “alelo 148”. Si el *bin* de ese alelo está mal definido, posiblemente la adjudicación de alelos en las muestras también se vea desplazada. Si luego de verificar el *bin* sigue estando corrido el alelo, posiblemente sea un nuevo alelo de ese marcador, se deberá controlar la forma del mismo y si se repite en algún otro individuo. La inserción de una sola base más al alelo se puede justificar como una situación de deslizamiento de la polimerasa. Una solución posible para esta situación será repetir la PCR de esa muestra para corroborar la existencia del alelo “raro”.
4. Requerimiento de amplificación unilocus: los marcadores que muestren dos o más sitios de amplificación serán descartados, dado que no será posible asignar los alelos que segregan juntos ni determinar el genotipo real del individuo heterocigota.

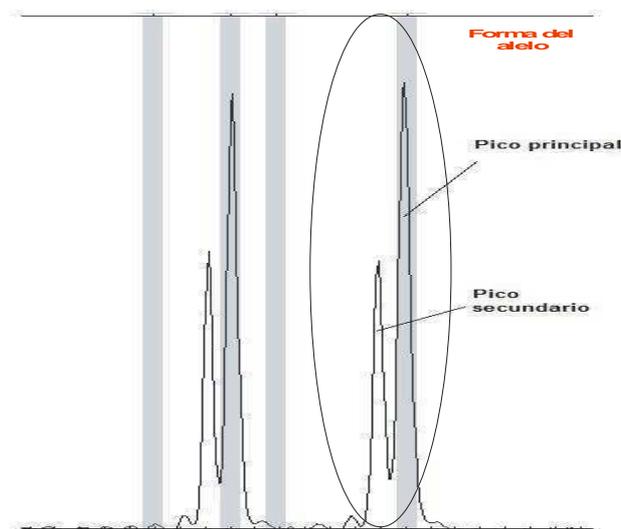


Figura 1. Electroferogramas visualizados con Genemapper. Las líneas grises verticales corresponden a cada uno de los *bines* detectados en el conjunto de muestras de germoplasma analizadas. La forma del alelo se indica con una elipse. En este caso el genotipo corresponde a un individuo heterocigota.

M3.2. Análisis de la diversidad y estructura poblacional

M3.2.1. Cálculo de frecuencias alélicas

Las frecuencias alélicas fueron calculadas mediante el método de conteo directo a partir de los genotipos individuales encontrados en cada una de las entradas analizadas. La frecuencia de alelos nulos se calculó desde la matriz de datos en función de la ausencia de producto de amplificación, luego de repetir la reacción de PCR por triplicado para el individuo y el marcador en estudio.

M3.2.2. Índices de diversidad genética

El contenido de información polimórfica (PIC) se calculó por cada locus SSR y en promedio de acuerdo con Anderson *et al.* (1993) utilizando el programa PowerMarker versión 3.25 (Liu and Muse, 2005).

Como estimadores de la variabilidad genética de las distintas entradas se emplearon los siguientes parámetros:

- ✓ La heterocigosis promedio esperada por locus (H_e) o diversidad genética: fue calculada a partir de las frecuencias alélicas utilizando el método descrito por Nei (1972, 1987).

$$H_e = (1/L) \sum 1 - \sum x_i^2$$

Donde: L=número de loci analizados; x_i = frecuencia de los i-alelos

Siendo: $0 < H_e < 1$

- ✓ El número promedio de alelos por locus (A): se obtuvo calculando la media aritmética del número de alelos presentes en cada uno de los loci analizados.
- ✓ La riqueza alélica (R_s) (El Mousadik and Petit, 1996): constituye una medida del número de alelos y es independiente del tamaño de la

muestra. Estima el número de alelos diferentes que se espera encontrar en una submuestra de tamaño $2n$ tomada a partir de una población compuesta por $2N$ genes ($n \leq N$).

Los estadísticos mencionados fueron calculados utilizando el programa F-Stat (Goudet, 2001). La estimación del número de heterocigotas observados por población se calculó por recuento directo de la matriz de datos. Se determinó la presencia de alelos exclusivos presentes en un material en particular y ausente en el resto de ellos.

La comparación de las estimas de diversidad H_e , H_o (heterocigosis observada) y R_s obtenidas para las diferentes categorías (poblaciones, compuestos y líneas) se realizó según lo propuesto por Goudet (2001) utilizando el programa F-Stat.

La idoneidad de los tamaños muestrales utilizados para obtener las estimas de diversidad fue evaluada generando submuestras de individuos al azar mediante el método de *bootstrapping* incorporado en el programa Infogen (Balzarini *et al.*, 2005). Se realizaron 250 réplicas para diversos tamaños de muestra por categorías (poblaciones y compuestos) y por entrada.

La identificación de grupos núcleo (*core sets*) se llevó a cabo según el método propuesto por Liu *et al.* (2003). Este método se basa en un algoritmo que permite, a partir del total de individuos analizados, identificar subconjuntos de individuos que maximicen la riqueza alélica. De este modo, los materiales que contengan mayor número de alelos exclusivos mostrarán una mayor probabilidad de estar representados en el grupo núcleo. Este cálculo se realizó con el programa PowerMarker versión 3.25 (Liu and Muse, 2005) empleando los parámetros descritos en el manual. Se probaron tamaños de grupo núcleo de entre 10 y 50 individuos para las poblaciones y entre 10 y 40 para los compuestos. Se realizaron 5 repeticiones para cada tamaño de muestra, con 1000 réplicas cada una.

Para distinguir las diferentes causas de variación dentro de las líneas endocriadas se emplearon los criterios citados por Heckenberger *et al.* (2002), los cuales incluyen 10 casos posibles para las comparaciones entre individuos de una misma línea (Tabla 5). Se considera un caso normal cuando, al comparar dos individuos de una misma línea, éstos poseen el mismo genotipo

homocigota (por ej: individuo 1: 150-150; individuo 2: 150-150). Otras alternativas posibles se producen cuando uno de los individuos o ambos presentan genotipos heterogéneos, es decir heterocigotas para los mismos alelos (caso 2), heterocigotas para alelos diferentes (casos 5 y 6) o un individuo homocigota y el otro heterocigota con un alelo común o diferente (casos 3 y 4). Las restantes posibilidades incluyen situaciones donde uno o ambos individuos poseen alelo nulo (casos 7, 8 y 9), o ambos individuos son homocigotas para distintos alelos (caso 10).

Tabla 5. Casos posibles de heterogeneidad en líneas. A, B, C y D representan diferentes alelos para un locus marcador dado. M representa un dato perdido o alelo nulo.

Individuo	Caso									
	1	2	3	4	5	6	7	8	9	10
1	A	A+B	A	A	A+B	A+B	A+B	A	M	A
2	A	A+B	A+B	B+C	A+C	C+D	M	M	M	B

Los resultados de este análisis se expresan como la ocurrencia en porcentaje de cada uno de los casos descritos en la Tabla 5. Se efectúa el recuento del número de veces que aparece cada caso en todas las comparaciones de una línea dada y para todos los marcadores evaluados. Luego se divide este resultado por el número total de comparaciones por línea (en base al número de individuos que pertenecen a la misma y al total de marcadores analizados), y finalmente este valor se multiplica por 100.

M3.2.3. Equilibrio de Hardy-Weinberg

Para analizar si los datos muestran ajuste a las proporciones de Hardy-Weinberg (HW) se empleó la prueba de los scores (Rousset and Raymond, 1995) para loci individuales en cada entrada. En cuanto a las hipótesis alternativas de exceso y defecto de heterocigotas, se analizaron separadamente según lo descrito por el programa GENEPOP versión 3.4 (Raymond and Rousset, 1995). Los límites de significación fueron establecidos según la corrección de Bonferroni ($\alpha = 5\%$) ajustada en base al número total de loci analizados ($0,05/16 = 0,003125$).

M3.2.4. Estructura poblacional

El grado de diferenciación entre las entradas ensayadas se analizó según el modelo de mutación de alelos infinitos (IAM: *Infinite Allele Model*), donde la estructura genética de las poblaciones fue evaluada por medio de los índices de fijación F_{IS} , F_{IT} y F_{ST} (Wright, 1965; Wright, 1978). Los mismos pueden expresarse en función de los valores de heterocigosis esperada y observada a través de las reformulaciones propuestas por Nei (1977):

$$F_{IT} = (H_T - H_I)/H_T \quad F_{IS} = (H_S - H_I)/H_S \quad F_{ST} = (H_T - H_S)/H_T$$

Donde:

H_T = heterocigosis esperada en la población total

H_I = heterocigosis observada promedio por población

H_S = heterocigosis esperada promedio por población

La interpretación de los resultados obtenidos se basó en que F_{IT} es una medida del exceso o defecto de homocigotas en la población total, F_{IS} es una medida del efecto del apareamiento no aleatorio dentro de las subpoblaciones, es decir que representa la endogamia local. Por último, F_{ST} es una medida que indica subdivisión o estructuración de la población total.

Se calculó el índice de fijación de Wright F_{IS} (Wright, 1978) de acuerdo con el análisis de la varianza de Weir y Cockerham (1984) usando el programa F-Stat. Se calculó también, del mismo modo, el índice F_{ST} para detectar estructura poblacional. Los niveles de significación fueron calculados a través del método de permutaciones incluido en el programa F-Stat y modificados de acuerdo con la corrección de Bonferroni con un α global de 0,05 (Rice, 1989) ajustado en base al número total de loci analizados.

Método bayesiano para la inferencia de estructura poblacional

La estructura poblacional fue también examinada usando el modelo bayesiano según Pritchard *et al.* (2000), implementado en el programa Structure 2.2. Este enfoque permite identificar grupos de individuos cuya

distribución de genotipos se ajusta a lo esperado para una población panmíctica en ausencia de desequilibrio de ligamiento, y permite obtener las frecuencias alélicas en cada grupo. El programa Structure 2.2 utiliza el marco teórico de la estadística bayesiana y los métodos de Monte Carlo basados en cadenas de Markov (MCMC: *Markov Chain Monte Carlo*) para la estimación de determinados parámetros (por ej. frecuencias alélicas) que maximicen la probabilidad de obtener los datos observados. Se considera que el conjunto de individuos analizados representa una mezcla proveniente de un número desconocido de poblaciones (K), por ende se determinará un valor de K que maximice la probabilidad de los datos, o dicho de otro modo, el número de “unidades genéticas ideales” de la muestra. Para este modelo, no se requiere que las poblaciones sean delimitadas *a priori*. Se obtendrá un valor de K que podrá ser menor, igual o mayor al número de entradas originalmente definidas como poblaciones o compuestos en base a criterios geográficos, morfológicos o agronómicos. De este modo se logra detectar el patrón de estructuración que muestran los materiales analizados. Se realizaron 10 repeticiones para cada valor de K, efectuándose corridas con un largo de 500.000 iteraciones tras un período de *burn-in* de 100.000 iteraciones. Se asumió que el genoma de cada individuo podía estar constituido por la contribución de una o más de las poblaciones inferidas (modelo de mezclas o *admixture model*) con frecuencias correlacionadas. La elección del valor óptimo de K se basó en las recomendaciones provistas por la documentación del programa. Se considera que a igual probabilidad para dos valores de K diferentes, el más adecuado será el menor, lo cual evitará patrones de estructuración aleatorios o irreales surgidos por un incremento ficticio del número de poblaciones. La evaluación de los resultados obtenidos por medio de este programa fue realizada empleando dos criterios *ad hoc*, uno basado en la estabilidad de las corridas individuales y el otro analizando el cambio logarítmico de las probabilidades para cada corrida y calculando el cambio de ΔK vs K (Evanno *et al.*, 2005).

M3.2.5. Análisis de agrupamiento

Para las poblaciones, compuestos y el cultivar Colliguay, las distancias genéticas fueron calculadas en base a dos algoritmos diferentes. Uno de ellos

fue la distancia entre pares de poblaciones y compuestos empleando la distancia de Nei (1972, 1987). El segundo fue la distancia entre pares de individuos utilizando la proporción de alelos compartidos entre ellos (Dps). Esta última distancia resulta de una transformación al logaritmo de la proporción de alelos compartidos entre los individuos de a pares, se eligió esta estima dado que posee baja varianza (Matsuoka *et al.*, 2002). La estima de distancias de Nei se realizó con el programa PowerMarker versión 3.25 (Liu and Muse, 2005), mientras que la de Dps se efectuó con el programa Microsat (Minch *et al.*, 1995-1996). Los agrupamientos fueron realizados según el método de Neighbor-Joining (Saitou and Nei, 1987) implementado en el paquete Phylip 3.6 (Felsenstein, 1991-2004). Las medidas de apoyo de ramas fueron obtenidas mediante el método de *bootstrapping*, utilizando la rutina incluida en cada uno de los programas utilizados para el cálculo de distancias genéticas entre poblaciones y entre individuos, respectivamente.

En el caso de las líneas, para poder determinar el grado de asociación entre los individuos pertenecientes a cada una de ellas y la cohesión entre los mismos como miembros de un grupo, se realizó un dendrograma a partir de las distancias calculadas con el índice de Jaccard, según lo descrito por Heckenberger *et al.* (2002). Se empleó el programa NTSYS con el método de agrupamiento UPGMA.

RESULTADOS

Los resultados expuestos en este Capítulo corresponden a la diversidad obtenida en base a caracteres morfo-agronómicos y, en un enfoque más profundo, a la diversidad obtenida en base a la caracterización molecular mediante marcadores microsatélites.

R1. Diversidad en base a caracteres morfológicos y agronómicos

El objetivo del análisis de diversidad en base a caracteres morfológicos y agronómicos fue validar el diseño de muestreo de materiales del BAG-IM para el posterior análisis molecular.

El análisis de agrupamiento del total de entradas evaluadas se muestra en un dendrograma adjunto en el Apéndice (Figura A.1.). A modo de simplificación se incorporó a continuación un esquema del mismo, que representa la distribución en conglomerados (*clusters*) de los materiales elegidos para el análisis molecular (Figura 2).

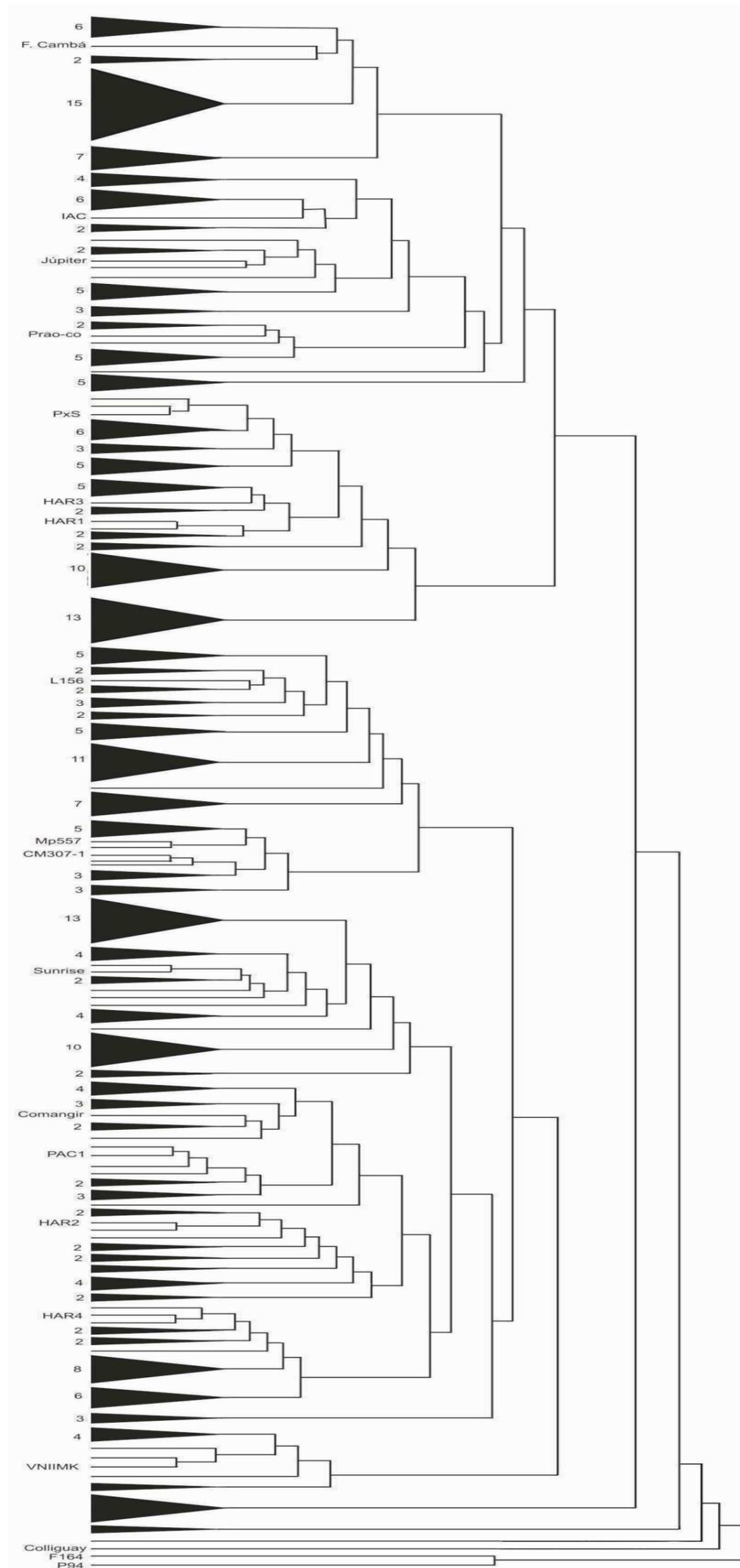


Figura 2. Esquema del árbol obtenido en base a caracteres morfológicos y agronómicos. Como índice se empleó la distancia taxonómica (*taxonomic distance*). Sólo se identificaron con su nombre aquellas entradas que fueron elegidas para el análisis molecular (19 de 309 materiales). A la izquierda de cada triángulo se incluye el número de entradas que representa cada uno (1 entrada = 2 mm de altura).

El coeficiente de correlación cofenética obtenido de los datos en estudio fue de 0,983.

En el árbol obtenido se observan dos conglomerados principales, uno de ellos con sólo dos líneas (F-164 y P-94), mientras que el otro incluyó a los restantes materiales (Figura 2, Figura A.1.). Dentro de éste último, a su vez se determinó un patrón similar porque sólo una entrada (Colliguay) quedó separada del resto. En un tercer y cuarto nivel de complejidad se repitió la misma situación, dado que GMS801-1 formó un grupo y los restantes materiales se agruparon en otro, del cual se separó Gigante Ruso e Iregikorai, mientras que los demás conformaron otro conglomerado. Las entradas HA336 y KLM 209 (muestreadas para el análisis molecular), no se encuentran en ninguna de las ramas del árbol (Figura A.1.) debido a que éstas no cuentan con datos morfológicos y agronómicos, ya que no han sido evaluadas en campo en ninguna de las campañas analizadas en el presente trabajo. Para el caso del material KLM 209, se dispone de la caracterización de las dos líneas de esta entrada (KLM 209-2 y KLM 209-10) que derivaron de la población original, pero su evaluación en campo no está registrada aún en los archivos del BAG-IM.

La distribución de las entradas elegidas para el análisis molecular en los distintos grupos del árbol obtenido en base a caracteres morfológicos y agronómicos, permitió confirmar que éstas abarcan el espectro de diversidad de los materiales conservados en el BAG-IM.

Dado el volumen de los datos, en esta sección sólo se muestran los resultados de las evaluaciones fenotípicas correspondientes a las entradas seleccionadas para el análisis molecular (19/309). Las Tablas 6 y 7 muestran los resultados obtenidos para cada categoría y entrada en cuanto a los caracteres cuantitativos y cualitativos, respectivamente. La información correspondiente a las 309 entradas puede consultarse en los registros del BAG-IM.

Tabla 6. Caracteres cuantitativos por categoría en las entradas seleccionadas para el análisis molecular. AH: ancho de la hoja. LH: largo de la hoja. AF: área foliar. Altura F: altura a floración. DC: diámetro del capítulo. DT: diámetro del tallo. SAM: días de emergencia a madurez. PCS: peso de cien semillas. MG: materia grasa. Los valores destacados en color rojo corresponden a los promedios totales de las 309 entradas analizadas agrónomicamente. sd: sin dato.

Material	AH	LH	AF	Altura F	DC	DT	SAM	PCS	MG
Poblaciones									
Prao-co	27	26	702	120	18	250	94	2,8	56,8
PxS	31	32	992	190	21,5	303	86	6,41	37,57
Sunrise	23,2	25,3	586,96	128	12,7	196	95	5,8	37,5
VNIIMK	17,3	19,7	340,81	135	11,5	116	101	3,6	38,83
F. cambá	36	32	1152	194	18,3	271	92	8,17	41,62
IAC	23,5	23,2	545,2	190	18,5	251	86	9,46	41,6
Júpiter	22,5	23	517,5	148	15	244	101	6,2	41,11
Promedio	25,786	25,886	690,924	157,857	16,5	233	93,571	6,063	40,132
Desvío	6,160	4,636	285,208	32,447	3,563	60,740	6,188	2,342	1,867
Compuestos									
Colliguay	30	29	870	215	21	310	sd	sd	38,56
Comangir	22,1	22,9	506,09	153	10,3	215	111	7,4	39,72
HAR-1	25	25	625	131	19,75	283	99	10,3625	37,55
HAR-4	24	22,3	535,2	150	19	230	84	8,74	41,7
HAR-2	19,6	22,3	437,08	130	11,1	230	118	5,2	33,72
HAR-3	38,9	34,4	994,16	180	26	284	102	5,93	41,38
Promedio	26,6	25,98	661,255	159,833	17,858	258,667	102,8	7,5265	38,25
Desvío	6,95	4,85	221,78	32,603	6,064	38,521	12,911	2,091	2,965
Líneas									
F-164	Sd	sd	Sd	125	15	220	sd	sd	32,73
P-94	Sd	sd	sd	140	9	200	sd	sd	27,95
CM-307-1	12,3	16	196,8	118	9,3	134	95	3,2	30,8
L-156	18,2	17,8	323,96	100	10,3	174	118	4,9	42,5
MP-557	18,5	18,5	342,25	84	7,1	153	104	4,6	38,74
PAC-1	23	22	506	125	10	210	90	3,5	48,20
Promedio	18	18,575	342,253	115,333	10,117	181,833	101,75	4,05	34,998
Desvío	4,389	2,514	126,891	20,096	2,642	33,944	12,285	0,827	6,772
Prom. global	27,81	27,31	771,26	149,5	21,63	262	92,5	9	39,52
Desvío global	4,51	5,22	268,33	7,78	0,88	1,41	9,19	0,36	2,27

Tabla 7. Caracteres cualitativos por categoría en las entradas seleccionadas para el análisis molecular. IC: grados de inclinación del capítulo. AP: presencia de antocianinas en la pálea. RAM: tipo de ramificación. CG: color de los granos. EG: presencia de estrías en los granos. SH: superficie de la hoja. FH: forma de la hoja. MH: margen de la hoja. sd: sin dato.

Material	IC	AP	RAM	CG	EG	SH	FH	MH
Poblaciones								
Prao-co	4	1	total	1	1	2	4	1
PxS	4	1	sin	3	1	2	4	1
Sunrise	4	1	sin	1	1	2	2	0
VNIIMK	3	1	sin	1	1	3	4	1
F. cambá	5	2	sin	1	0	2	4	1
IAC	4	1	sin	1	1	2	4	1
Júpiter	3	1	sin	1	1	2	4	1
Compuestos								
Colliguay	5	2	sd	1	1	sd	4	1
Comangir	2	1	sin	1	1	2	4	1
HAR-1	3 y 4	2	sin y basal	2	0	2	4	1
HAR-4	4	3	sin	1	0	2	4	1
HAR-2	4	1	sin	1	1	2	4	1
HAR-3	4	1	sin	1	1	1	4	1
Líneas								
F-164	4	sd	sd	sd	1	sd	sd	sd
P-94	5	sd	total	sd	1	sd	sd	sd
CM-307-1	3	3	sin	1	0	2	2	1
L-156	3 y 4	1	total	1	1	3	4	0
MP-557	3	1	sin	1	1	2	4	0
PAC-1	3 y 4	1	total	1	1	2	4	1

Las tendencias promedio para cada categoría con respecto a caracteres cuantitativos se muestran en la Figura 3. Las mayores diferencias entre categorías se observaron para el carácter de AF, donde las líneas mostraron un menor valor promedio que las poblaciones y los compuestos. Un patrón similar se observó para los caracteres de DT y alt F en cuanto a las líneas. Las poblaciones mostraron mayor AF promedio que los compuestos, pero menores valores de DT y alt F.

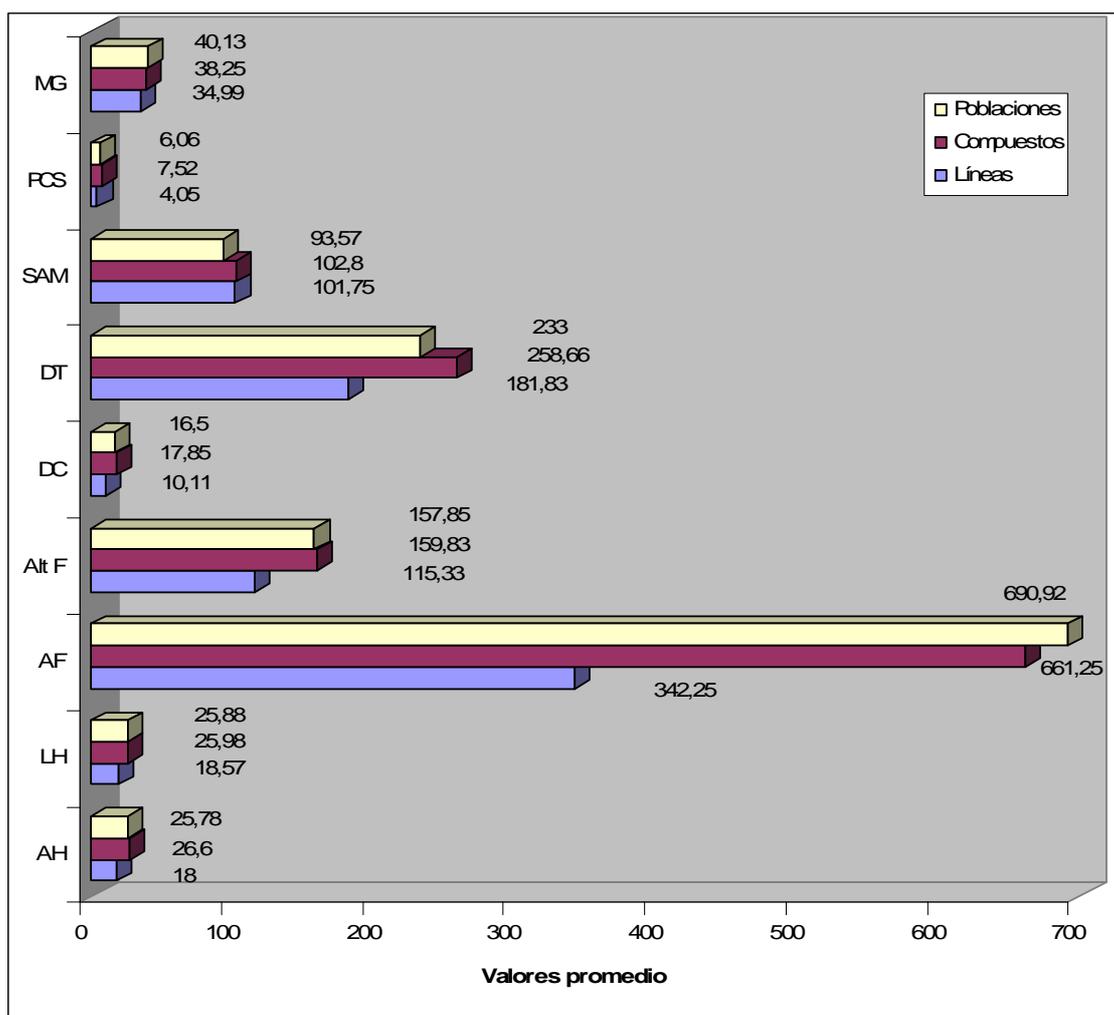


Figura 3. Tendencias promedio por categoría para caracteres cuantitativos. AH: ancho de la hoja. LH: largo de la hoja. AF: área foliar. Alt F: altura a floración. DC: diámetro del capítulo. DT: diámetro del tallo. SAM: días de emergencia a madurez. PCS: peso de cien semillas. MG: materia grasa.

R2. Diversidad genética en base a variabilidad de loci microsatélites

Las extracciones de ADN de los distintos materiales del BAG-IM se realizaron inicialmente con el protocolo que emplea CTAB (Saghai-Marooft *et al.*, 1984) y con el de Dellaporta *et al.* (1983). Para algunas entradas no fue posible obtener ADN, mientras que para otras la cantidad de ADN extraído fue escasa y, en muchos casos, la calidad del mismo fue baja por la presencia de polifenoles y polisacáridos. Por lo tanto, se decidió emplear el equipo (*kit*) de extracción *NucleoSpin II* de la firma Macherey-Nagel. La Figura 4 incluye dos

fotografías tomadas de geles de agarosa al 0,8% con ADN genómico. La imagen de la izquierda muestra corridas de individuos de los cuales no se logró extraer ADN o donde el mismo está degradado, correspondiendo a extracciones con los métodos tradicionales. La fotografía de la derecha muestra el ADN de individuos extraídos con el *kit* citado anteriormente, donde se observa que tanto la calidad como la cantidad de ADN fueron satisfactorias para los fines del presente trabajo.

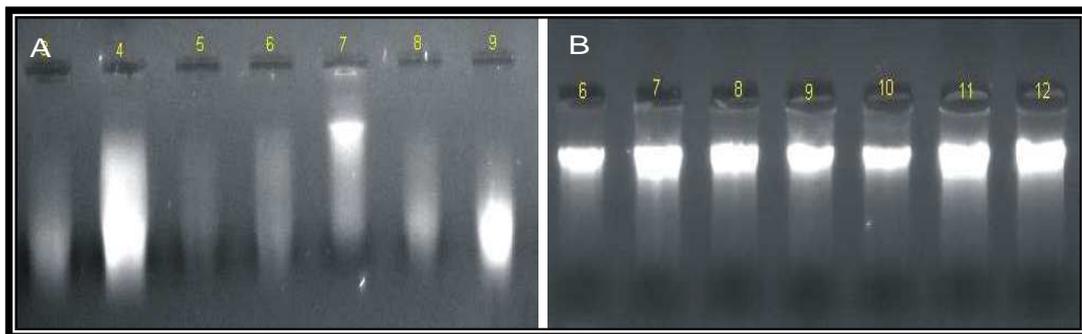


Figura 4. Imágenes de geles de agarosa al 0,8% con ADN genómico. Foto A: ADN extraído con métodos tradicionales. Foto B: ADN extraído con *kit*. Los números corresponden a la identificación de cada individuo dentro de una misma entrada.

Se examinaron 35 loci SSR, pero fueron seleccionados para este trabajo 16 de ellos, teniendo en cuenta como criterio la reproducibilidad, la facilidad de interpretación (ausencia de bandas producto del deslizamiento enzimático y sitio único de amplificación), la baja frecuencia de alelos nulos (menor al 10%) y la codominancia. De los 17 GL, están representados 16 porque en uno de ellos (GL 11) hay sólo dos loci SSR de la serie HA (Paniego *et al*, 2002). Ambos fueron probados, pero se obtuvo un patrón de amplificación multilocus para el loci HA3448 y ausencia de amplificación para el loci HA3640. Para el GL 13 ninguno de los marcadores estudiados generó productos de amplificación, mientras que para el GL 14 se genotificaron dos marcadores.

En la Tabla 8 se muestra una descripción de los niveles de polimorfismo detectados para los marcadores analizados en este trabajo, considerando el conjunto total de individuos. Los marcadores que mostraron mayor PIC fueron HA1848, HA2063, HA2077 y HA3239, con valores de 0,80, 0,69, 0,80 y 0,68;

respectivamente. Mientras que los marcadores que arrojaron menores valores de PIC fueron HA3581 y HA3582 con 0,20 y 0,24; respectivamente. El valor promedio de PIC para los 16 loci SSR fue de 0,50. La frecuencia promedio del alelo más frecuente fue de 0,554, la heterocigosis esperada promedio (He) fue de 0,56 y la heterocigosis observada promedio (Ho) fue de 0,22.

Tabla 8. Información de los marcadores analizados. El número de observaciones se registra a partir del total de individuos estudiados (337). AMF: alelo más frecuente. He: heterocigosis esperada promedio. Ho: heterocigosis observada promedio. PIC: contenido de información polimórfica. Los marcadores más informativos están destacados con fondo gris.

ID del marcador HA	Frecuencia alélica (AMF)	Número de genotipos	Número de observaciones	Número de alelos	He	Ho	PIC
77	0,4984	8	325	5	0,56	0,06	0,47
95	0,6065	21	305	8	0,60	0,25	0,57
293	0,4706	11	324	5	0,65	0,28	0,58
928	0,7439	10	330	6	0,41	0,03	0,38
991	0,4630	12	311	6	0,60	0,20	0,52
1402	0,6000	10	330	7	0,52	0,26	0,44
1848	0,3723	48	329	14	0,82	0,40	0,80
2063	0,4697	22	314	10	0,72	0,19	0,69
2077	0,3528	37	333	16	0,82	0,45	0,80
2920	0,4617	9	314	5	0,65	0,23	0,58
3239	0,3381	10	312	5	0,73	0,17	0,68
3555	0,5632	4	316	3	0,49	0,40	0,37
3581	0,8881	12	322	6	0,21	0,11	0,20
3582	0,8535	8	321	5	0,26	0,11	0,24
4103	0,5254	5	314	4	0,50	0,18	0,38
4239	0,6701	6	337	3	0,45	0,14	0,35
Promedio	0,5548	14,56	321,13	6,75	0,56	0,22	0,50

Los electroferogramas correspondientes a los loci con mayores niveles de variación se presentan en la Figura 5.

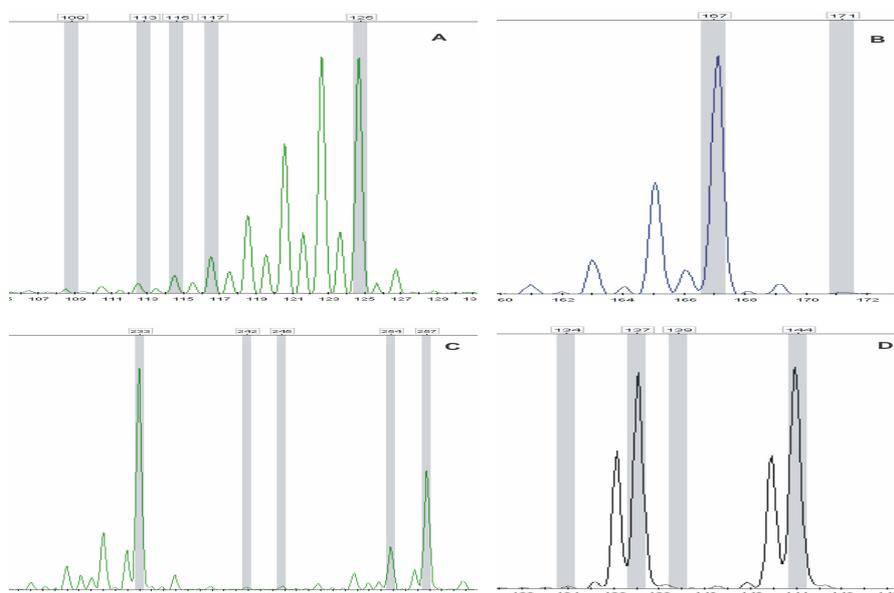


Figura 5. Electroferogramas de los loci microsatélites con mayor nivel de variación (página anterior). A. marcador HA2077 (individuo homocigota 125-125). B. marcador HA2063 (individuo homocigota 167-167). C. marcador HA1848 (individuo heterocigota 233-257). D. marcador HA3239 (individuo heterocigota 137-144).

El análisis de 16 loci microsatélites en 337 individuos provenientes de 21 entradas permitió detectar un total de 108 alelos, siendo el número promedio de alelos por locus (A) de 6,75.

En la Tabla 9 se presenta un resumen de los resultados obtenidos en el análisis de diversidad genética. En el Apéndice se expone la tabla completa con los resultados en cuanto a las estimas de A, Rs, He y número de alelos exclusivos por material y por locus (Tabla A.1.). Para el conjunto de las 8 poblaciones (174 individuos) se detectaron 91 alelos (valor promedio de 5,68), mientras que para el conjunto de los 6 compuestos (137 individuos) se hallaron 74 alelos (valor promedio de 4,62). Se encontraron 60 alelos en 7 líneas que fueron representadas por 26 individuos en total. Los alelos identificados presentaron frecuencias variables, desde valores bajos tales como 0,023 hasta 1,000; sin mostrar un patrón de acuerdo con la categoría del material. La frecuencia de alelos nulos para las poblaciones fue de 0,027, para los compuestos fue de 0,020, mientras que para las líneas fue de 0,0026. Los valores de A por entrada presentaron un rango general de 1,1875 a 3,3125 (Tabla 9). Dentro de las poblaciones analizadas (8), el menor valor de A correspondió a F. cambá con 1,75 y el mayor valor fue 3,3125 para IAC. Con respecto a los compuestos y el cultivar, el menor número de alelos promedio fue de 1,75 para el material HAR-1 y el mayor fue de 3,0625 para Colliguay. Todas las líneas estudiadas arrojaron valores de A mayores a 1, evidenciando diferentes grados de heterocigosis dentro de las mismas.

Las estimas de riqueza alélica (Rs) mostraron un valor mínimo de 1,475 para la población KLM y un máximo de 1,946 para la población Prao-co; mientras que para los compuestos la mínima Rs fue de 1,385 para HAR-1 y la máxima fue de 1,884 para Colliguay (Tabla 9).

Tabla 9. Estimaciones de diversidad genética para las entradas del BAG-IM analizadas en este trabajo. NAE: número de alelos exclusivos.

Material por categorías	A promedio	Rs promedio	He promedio	NAE
Poblaciones				
Júpiter	3,25	1,923	0,456	3
F. cambá	1,75	1,491	0,277	3
VNIIMK	2,0625	1,481	0,265	2
Sunrise	2,3125	1,558	0,298	1
Prao-co	3,0625	1,946	0,476	1
PxS	2,8125	1,759	0,39	1
IAC	3,3125	1,841	0,409	1
KLM	2,1875	1,475	0,26	0
Promedio	2,5937	1,684	0,354	
Compuestos				
HAR-2	2,5625	1,661	0,353	3
HAR-4	2,6875	1,651	0,336	1
Comangir	2,75	1,760	0,38	1
HAR-1	1,75	1,385	0,21	0
HAR-3	1,9375	1,630	0,339	0
Colliguay	3,0625	1,884	0,433	0
Promedio	2,4583	1,662	0,342	
Líneas				
F-164	1,1875	1,112	0,054	2
P-94	1,1875	1,188	0,093	2
HA 336	1,8125	1,581	0,304	2
CM-307-1	1,6875	1,432	0,239	1
MP-557	1,5625	1,344	0,187	0
PAC-1	1,5	1,332	0,177	0
L-156	1,5	1,4	0,265	0
Promedio	1,4911	1,341	0,188	

Del total de alelos encontrados en el conjunto de entradas analizadas, 24 (22,2%) constituyen alelos únicos o exclusivos de alguna introducción en particular, no estando presentes en ninguna de las demás (Tablas 9 y 10). Las entradas con mayor número de alelos exclusivos fueron F. cambá, Júpiter y HAR-2; siendo las dos primeras poblaciones y el tercero un compuesto. Para cada una de ellas se detectaron 3 alelos exclusivos. Seguidos en orden decreciente de abundancia por VNIIMK, F-164, P-94 y HA336 que mostraron 2 alelos exclusivos y Sunrise, CM-307-1, HAR-4, IAC, Comangir, PxS y Prao-co, donde se observó 1 alelo exclusivo. Para los loci más variables, HA1848, HA2063 y HA2077, se encontró el mayor número de alelos exclusivos, 4, 5 y 4 respectivamente (Tabla 10). La cantidad y variedad de alelos exclusivos detectados en las poblaciones se corresponde con los diversos orígenes geográficos de cada material.

Tabla 10. Alelos exclusivos detectados en las entradas del BAG-IM analizadas en este trabajo. Se destacan en gris los valores de frecuencia más bajo y más alto del total de datos obtenidos.

Material	Categoría	Locus	Alelo exclusivo (pb)	Frecuencia
F. cambá (N=15)	Población	HA77	244	0,067
		HA928	203	0,036
		HA1848	230	0,767
Sunrise (N=19)	Población	HA77	262	0,026
VNIIMK (N=24)	Población	HA95	149	0,682
		HA1402	159	0,023
PxS (N=20)	Población	HA2077	145	0,100
Prao-co (N=24)	Población	HA3555	106	0,024
IAC (N=24)	Población	HA2063	168	0,063
Júpiter (N=24)	Población	HA1848	239	0,476
		HA1848	266	0,048
		HA2063	167	0,325
HAR-2 (N=20)	Compuesto	HA1848	272	0,050
		HA2063	161	0,056
		HA3582	143	0,036
HAR-4 (N=32)	Compuesto	HA2063	166	0,203
Comangir (N=24)	Compuesto	HA2063	174	0,091
F-164 (N=4)	Línea	HA2077	121	0,125
		HA2077	125	0,625
P-94 (N=2)	Línea	HA2077	127	1,000
		HA4103	152	0,250
CM-307 (N=4)	Línea	HA928	205	0,125
HA336 (N=4)	Línea	HA991	151	0,500
		HA3581	137	0,125

En la Figura 6 se observa la cantidad de loci que presentaron los diferentes números de alelos detectados, discriminados por categoría. Para las líneas se observa que en la mayoría de los loci se identificaron 2 alelos, variando entre 1 y 4. Mientras que para las poblaciones y compuestos la mayoría de los loci poseen entre 4 y 5 alelos, con un rango que osciló entre 1 y 9.

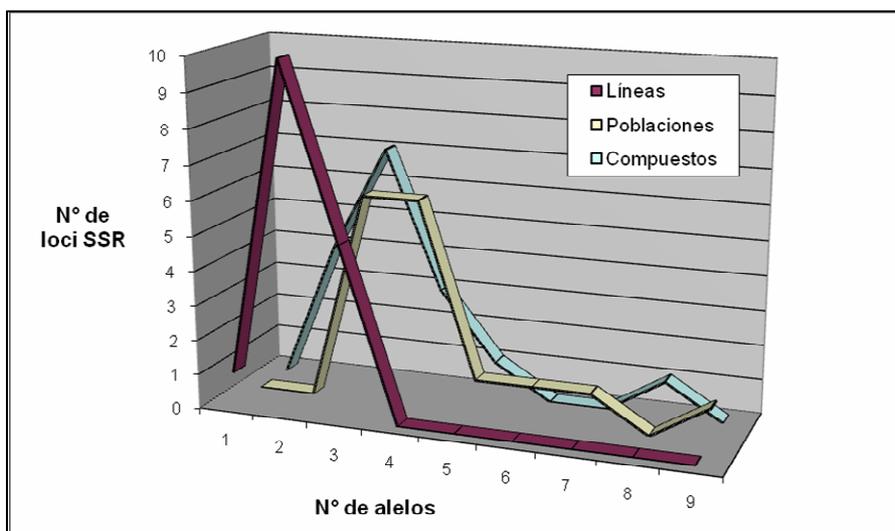


Figura 6. Número de alelos por locus para 16 SSR genotificados en las distintas categorías de germoplasma evaluadas

Los valores de heterocigosis promedio esperada (H_e) oscilaron entre 0,26 para KLM y el valor máximo fue de 0,476 para Prao-co. En cuanto a los compuestos, la H_e mínima fue de 0,21 para el compuesto HAR-1 y el valor máximo fue de 0,433 para Colliguay. Los valores de H_e promedio para las líneas mostraron un rango que varió entre 0,054 para F-164 y 0,304 para HA336.

Para completar el análisis de diversidad, se evaluó la existencia de diferencias entre los materiales organizados por grupos de acuerdo con la categoría a la que pertenecen. Se construyeron tres grupos, el primero conteniendo todas las poblaciones, el segundo constituido por los compuestos y el tercero conformado por las líneas. Se calcularon los estimadores R_s , H_e y H_o para cada uno de los grupos. Se realizaron las comparaciones de los valores obtenidos de a pares de grupos y se estableció la significación según el método de permutaciones detallado en la sección M3.1 (Materiales y Métodos). Al analizar la R_s promedio por categoría, el valor de las poblaciones fue de 1,684 y de los compuestos fue de 1,662; mientras que los valores promedio de H_e fueron de 0,354 para las poblaciones y de 0,342 para los compuestos (Tablas 9 y 11).

Tabla 11. Estimaciones de diversidad calculadas por categoría. Letras diferentes indican diferencias significativas entre grupos ($p < 0.05$).

Estimador	Poblaciones	Compuestos	Líneas
Rs	1,684 ^a	1,662 ^a	1,341 ^b
Ho	0,215 ^a	0,235 ^a	0,128 ^a
He	0,354 ^a	0,342 ^a	0,188 ^b

Ninguno de los estimadores mostró diferencias significativas entre poblaciones y compuestos. La comparación entre poblaciones y líneas mostró diferencias significativas tanto para Rs ($p=0,0045$) como para He ($p=0,0010$); obteniéndose un patrón similar para la comparación de Rs ($p=0,0140$) y He ($p=0,0085$) entre compuestos y líneas. No se observaron diferencias significativas para los valores de Ho, probablemente debido a que hubo una elevada proporción de homocigotas en las poblaciones y compuestos y un alto número de heterocigotas en las líneas.

Con el objeto de establecer los tamaños muestrales mínimos requeridos para efectuar una caracterización molecular más exhaustiva del BAG-IM, se generaron submuestras aleatorias de tamaño creciente dentro de cada categoría y entrada, y se computó el número de alelos en cada una de ellas. El número de alelos obtenido para cada tamaño muestral en el análisis global de poblaciones y compuestos se presenta en la Figura 7.

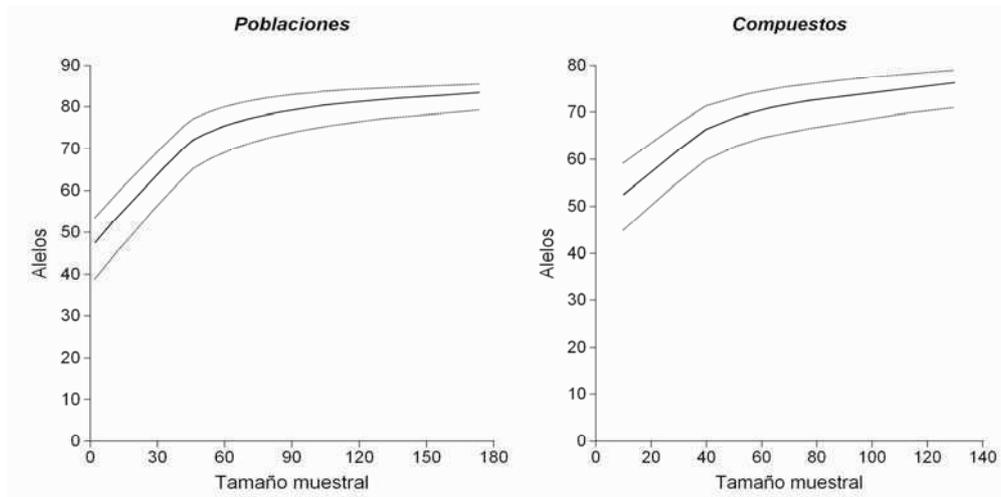


Figura 7. Análisis de muestra mínima para diversidad en poblaciones y compuestos. Las líneas enteras corresponden al número promedio de alelos en 250 réplicas generadas por cada tamaño muestral. Las líneas punteadas indican los intervalos de confianza al 95%.

Se observó un punto de inflexión entre 40 y 50 individuos muestreados para la categoría de poblaciones, mientras que para los compuestos fue cercano a 40 individuos. El cambio de pendiente resultó más abrupto para las poblaciones que para los compuestos. El valor promedio de alelos (línea entera, Figura 7) mostró una tendencia al incremento del número de alelos con el aumento del tamaño muestral para ambas categorías. Sin embargo, dado que las pendientes a mayores tamaños muestrales fueron leves y se alcanzó un *plateau*, se podría asumir que los tamaños utilizados resultaron adecuados para capturar la mayor parte de la variación presente en el conjunto de materiales estudiados.

El número de alelos obtenido para cada tamaño muestral al considerar cada entrada en forma independiente se presenta en las Figuras 8 y 9. De acuerdo con estas inferencias, el 50% de las poblaciones analizadas alcanzaron un *plateau*, es decir que sus pendientes se aproximaron a cero antes de alcanzar el total de individuos analizados en cada caso. Para F. cambá este punto se alcanzó con 12 individuos, para IAC con 18, para PxS con 17 y para Sunrise con 14; mientras que Júpiter, KLM, Prao-co y VNIIMK no alcanzaron el *plateau* para el número de individuos estudiados (Figura 8).

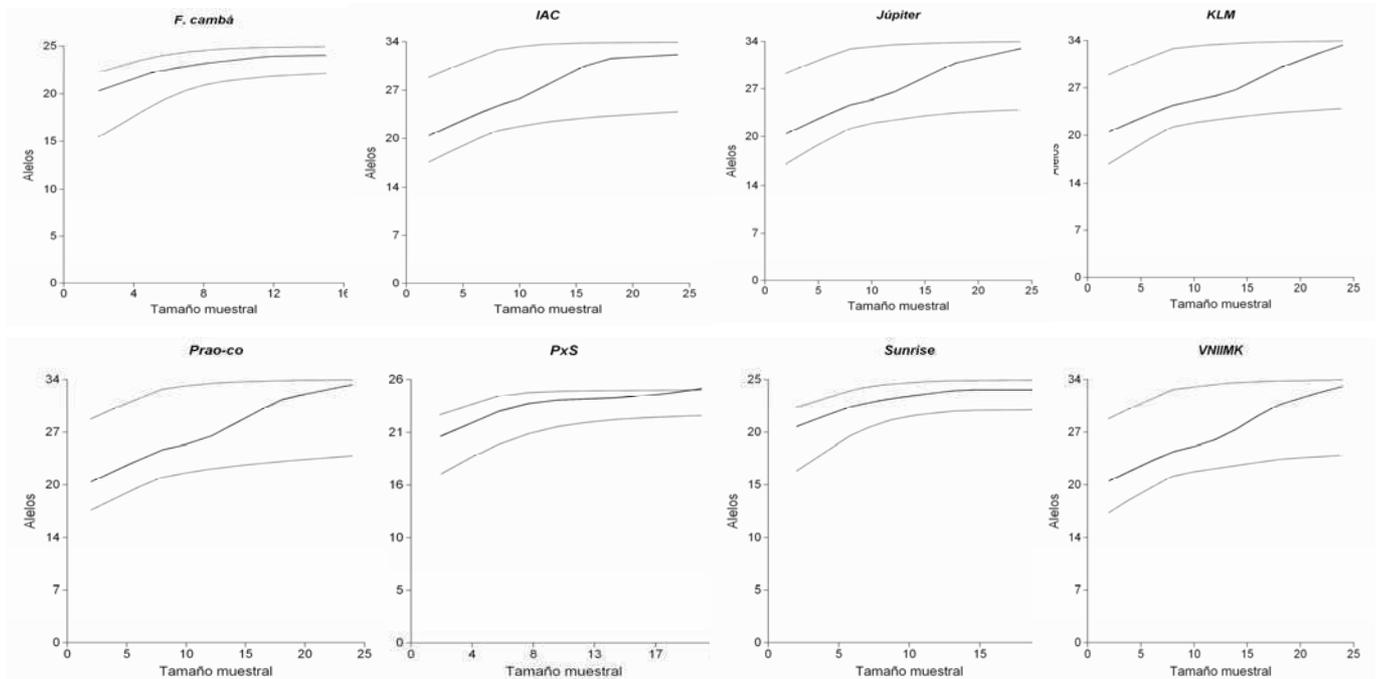


Figura 8. Análisis de muestra mínima para diversidad en las 8 poblaciones estudiadas.

Las líneas enteras corresponden al número promedio de alelos en las 250 réplicas generadas por cada tamaño muestral. Las líneas punteadas indican los intervalos de confianza al 95%.

Tres de los seis compuestos (Colliguay, HAR-2 y HAR-3) alcanzaron el *plateau* con un rango de individuos entre 10 y 20, HAR-4 mostró un cambio importante de pendiente, mientras que Comangir y HAR-1 mostraron un aumento continuo del número de alelos con el tamaño muestral (Figura 9).

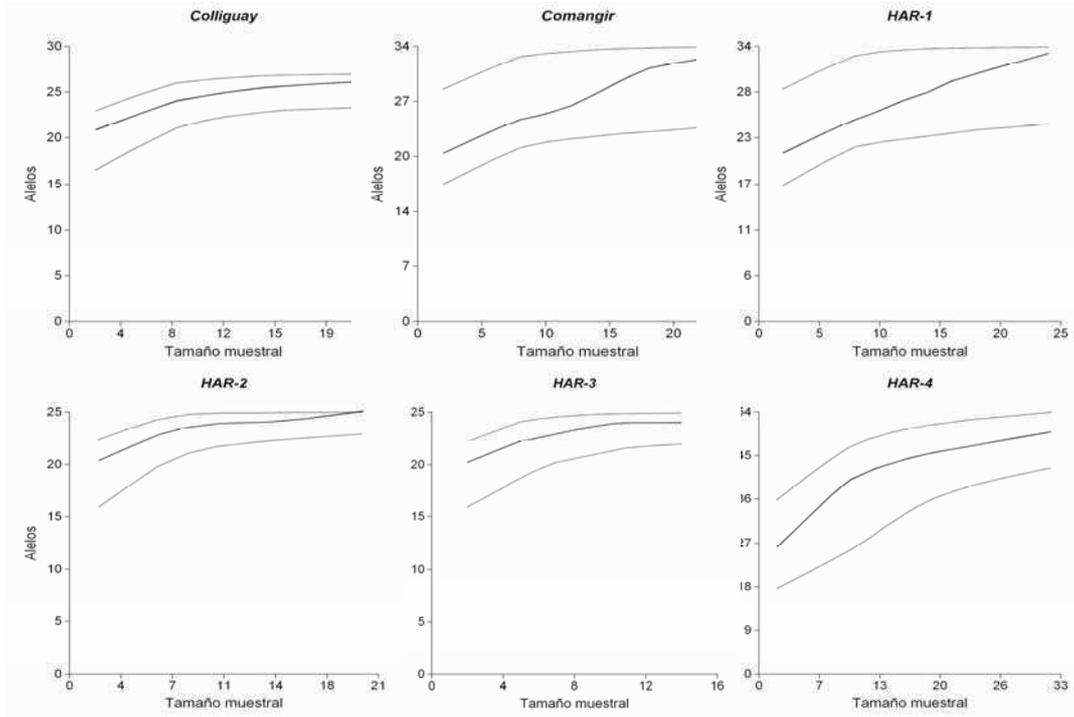


Figura 9. Análisis de muestra mínima para diversidad en los 6 compuestos estudiados. Las líneas enteras corresponden al número promedio de alelos en las 250 réplicas generadas por cada tamaño muestral. Las líneas punteadas indican los intervalos de confianza al 95%.

Para encontrar el número mínimo de individuos que maximicen el número de alelos dentro de cada categoría (poblaciones y compuestos), se analizó la cantidad de alelos capturados en submuestras de tamaño creciente obtenidas mediante el método de identificación de grupos núcleo (*core sets*) de Liu *et al.* (2003) (ver sección M3.2. en Materiales y Métodos). Los resultados de este análisis se muestran en la Tabla 12.

Tabla 12. Grupos núcleo (core sets) para poblaciones y compuestos del BAG-IM. a. Tamaño de la submuestra. b. Número de alelos detectados para cada tamaño de submuestra analizado.

Categoría	Tamaño ^a	Alelos ^b	Entradas (número de individuos de cada una)
Poblaciones (N=174)	10	71	PxS (2), Praoco (2), VNIIMK (2), Sunrise (3), KLM (1)
	20	80	PxS (2), Praoco (3), VNIIMK (3), Sunrise (2), KLM (1), Júpiter (3), IAC (3), F. cambá (3)
	30	85	PxS (5), Praoco (4), VNIIMK (4), Sunrise (2), KLM (2), Jupiter (5), IAC (4), F. cambá (4)
	40	87	PxS (7), Praoco (6), VNIIMK (6), Sunrise (2), KLM (4), Jupiter (6), IAC (4), F. cambá (5)
	50	91	PxS (8), Praoco (7), VNIIMK (8), Sunrise (4), KLM (6), Jupiter (7), IAC (5), F. cambá (5)
Compuestos (N=137)	10	65	Colliguay (2), Comangir (1), HAR-2 (1), HAR-3 (2), HAR-4 (4)
	20	69	Colliguay (5), Comangir (2), HAR-2 (3), HAR-3 (3), HAR-4 (3), HAR-1 (4)
	30	74	Colliguay (7), Comangir (5), HAR-2 (3), HAR-3 (2), HAR-4 (9), HAR-1 (4)
	40	74	Colliguay (9), Comangir (8), HAR-2 (3), HAR-3 (2), HAR-4 (14), HAR-1 (4)

Para las poblaciones se observó que las muestras de 10 individuos que maximizaron el número de alelos capturaron el 78,02% (71) de los 91 alelos hallados en el total de individuos, las muestras de 20 capturaron el 87,91% (80), las de 30 capturaron el 93,4% (85), las de 40 captaron el 95,6% (87) y las de 50 individuos lograron capturar el 100% de los 91 alelos detectados (Figura 10). Se observó que con el 29% (50/174) del total de individuos muestreados para el caso de las poblaciones se arribó a cubrir la variabilidad total detectada en esta categoría. Para el caso de los compuestos, 10 individuos capturaron el 87,84% (65) de 74 alelos totales, 20 capturaron el 93,2% (69), mientras que 30 capturaron el 100% (74 alelos) (Figura 10). Se observó que con el 22% (30/137) del total de individuos muestreados para el caso de los compuestos se alcanzó a cubrir la variabilidad total detectada en esta categoría.

Este análisis no se realizó para la categoría de líneas a causa de la heterogeneidad detectada dentro de las mismas.

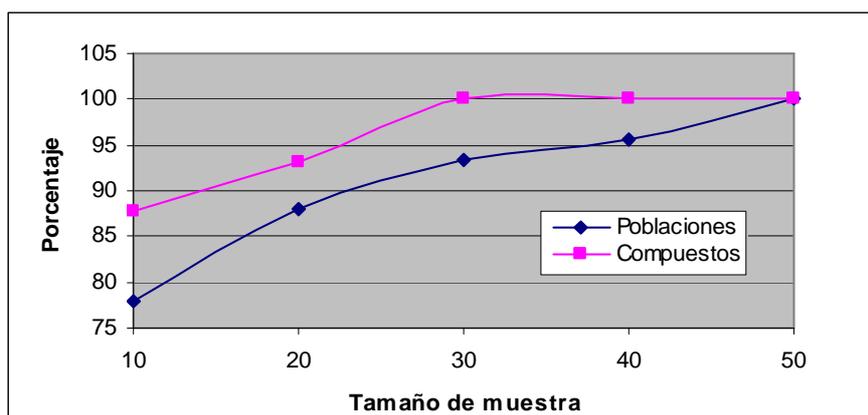


Figura 10. Porcentaje del número de alelos capturados en diferentes tamaños de muestras generadas por el método de identificación de grupos núcleo (Liu *et al.*, 2003) (página anterior). Los tamaños de submuestra fluctúan entre 10 y 50 individuos. El porcentaje se calculó en base al total de alelos detectados por categoría.

Debido a la heterogeneidad observada en las líneas, se realizó el análisis propuesto por Heckenberger *et al.* (2002) a fin de identificar las fuentes de variación dentro de cada entrada. Este análisis se efectuó para todos los loci dentro de cada una de las líneas. Se observaron tanto casos normales como casos de heterogeneidad y ocurrencia de alelos nulos o datos perdidos, según las categorías propuestas por Heckenberger *et al.* (2002) (ver sección M3.2. en Materiales y Métodos). Los casos normales se definen como aquellas situaciones donde ambos individuos comparados son homocigotas para el mismo alelo (igual genotipo). Las situaciones de heterogeneidad entre dos individuos de la misma línea corresponden a situaciones donde ambos son heterocigotas, uno de ellos es heterocigota y el otro homocigota o ambos homocigotas con alelos diferentes (casos 2 al 6 y caso 10). La ocurrencia de alelos nulos se contempla en los casos 7 al 9. Los resultados obtenidos mostraron que la ocurrencia de casos normales varió desde 52,1 para PAC-1 a 87,5 para P-94, con un valor promedio de 67,6 para todas las líneas analizadas (Tabla 13). El caso de heterogeneidad más frecuente en todas las líneas analizadas fue un individuo con un alelo en homocigosis, y otro individuo heterocigota que comparte el alelo del individuo homocigota (caso 3; 13,8). En cuanto a la ocurrencia de alelos nulos o datos perdidos, el mayor valor promedio fue obtenido para el caso 8 (4,8) donde un individuo es homocigota y el otro individuo no posee dato. El surgimiento de alelos nuevos se reflejó en la ocurrencia del caso 10, cuyo promedio de aparición fue de 6,3.

Tabla 13. Ocurrencia de diferentes casos de heterogeneidad entre los individuos de una misma línea evaluada con 16 SSR. Ncp: número de comparaciones de a pares.

Línea	Ncp	Casos										
		1	2	3	4	5	6	7	8	9	10	
L-156	6	68,8	6,3	3,1	0,0	0,0	0,0	0,0	0,0	4,2	1,0	16,7
F-164	6	83,3	0,0	5,2	2,1	1,0	0,0	0,0	0,0	7,3	1,0	0,0
CM-307-1	6	64,6	0,0	22,9	2,1	0,0	0,0	0,0	0,0	0,0	0,0	10,4
PAC-1	3	52,1	8,3	22,9	0,0	0,0	0,0	2,1	10,4	0,0	4,2	
MP-557	6	62,5	1,0	21,9	1,0	0,0	0,0	0,0	9,4	0,0	4,2	
HA336	6	54,2	4,2	20,8	2,1	4,2	0,0	4,2	2,1	0,0	8,3	
P-94	1	87,5	6,3	0,0	6,3	0,0	0,0	0,0	0,0	0,0	0,0	
Media		67,6	3,7	13,8	1,9	0,7	0,0	0,9	4,8	0,3	6,3	

R.3. Estructura poblacional

R.3.1. Equilibrio de Hardy-Weinberg

El compuesto HAR-1 mostró un patrón general de ajuste a las proporciones de Hardy-Weinberg. Por el contrario, las restantes poblaciones y compuestos analizados (KLM, IAC, HAR-4, Colliguay, HAR-3, PxS, Prao-co, VNIIMK, Júpiter, Sunrise, Comangir, HAR-2 y F. cambá) mostraron deficiencia de heterocigotas (exceso de homocigotas) respecto a lo esperado bajo la hipótesis de unión de gametas al azar, presentando F_{IS} globales positivos y significativos, a pesar de que en general los loci no mostraron desvíos tan marcados (Tabla 14).

Del análisis de locus individuales, la mayoría de las desviaciones fueron halladas al considerar la hipótesis alternativa de defecto de heterocigotas (Tabla 14, Tabla A.2 del Apéndice), mientras que sólo se observó un exceso de heterocigotas significativo para el locus HA2077 en el compuesto HAR-1 ($F_{IS} = -0,7388$; $p=0$) (Tabla 14). Para el resto de las combinaciones entrada/locus no se observaron desviaciones significativas en cuanto a esta prueba.

Las entradas que mostraron mayor número de loci con desvíos al equilibrio fueron KLM y Júpiter con 8 loci, mientras que HAR-4 presentó 6 loci con desvíos (Tabla 14). Los loci que presentaron mayor número de entradas con desvíos fueron HA77 con 7 y HA928 y HA2063 con 5 introducciones cada uno.

Tabla 14. Ajuste a las proporciones de Hardy-Weinberg. F_{IS} por locus y entrada. M: locus monomórfico. Casillas grises: ajuste al equilibrio de Hardy-Weinberg (unión de gametas al azar). Casillas blancas: F_{IS} significativos bajo la hipótesis alternativa de defecto de heterocigotas. Casilla naranja: F_{IS} significativo bajo la hipótesis alternativa de exceso de heterocigotas. $F_{IS(G)}$: F_{IS} globales. Nivel de significación corregido según Bonferroni $p < 0,003125$.

Material	Locus																$F_{IS(G)}$
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
KLM		*			*			*		*	*	*	M	*	M	*	0,740*
IAC	*			*		*											0,206*
HAR-4	*			*	*		*						M	M	*	*	0,512*
Colliguay																M	0,226*
HAR-1	*	M	M	M	*			M	*		M		M	M		M	0,158
HAR-3				M		*		*					M	M	M	M	0,230*
PxS			*	*										M			0,425*
PraoCo				*		*								M			0,277*
VNIIMK	*		*	M				*					M	M	M	M	0,484*
Júpiter	*	*				*	*	*	*	*					*	M	0,463*
Sunrise	*			M									M			M	0,364*
Comangir				M													0,174*
HAR-2	*			*	*		*				M		M		M		0,443*
F. camba				M	M			*		M		M	M			M	0,455*

R.3.2. Diferenciación genética

Se estimaron los niveles de diferenciación mediante el índice de fijación F_{ST} de Wright (1978) para las poblaciones y los compuestos, dado que no se hallaron diferencias significativas entre F_{ST} y R_{ST} en la prueba de permutación de tamaños alélicos de Hardy *et al.* (2003). La estima global de F_{ST} reveló un grado de diferenciación considerable y estadísticamente significativo entre las entradas del BAG-IM aquí estudiadas ($F_{ST}= 0,413$). Los valores de diferenciación más altos correspondieron a los pares de materiales conformados por el compuesto HAR-1 con VNIIMK ($F_{ST} 0,6394$), con Sunrise ($F_{ST} 0,6151$) y con F. cambá ($F_{ST} 0,6305$) (Tabla 15). Esta marcada diferenciación podría derivar de la diversidad de orígenes geográficos entre estos materiales. Los valores más bajos de diferenciación y también significativos se detectaron para los pares IAC con Colliguay ($F_{ST} 0,1899$) y Colliguay con Júpiter ($F_{ST} 0,1795$). Dentro del conjunto de poblaciones y compuestos, se encontró que para 7 pares de entradas, los valores de F_{ST} no resultaron estadísticamente significativos. Los mismos corresponden a los siguientes pares: Colliguay con HAR-3, Colliguay con F. cambá, HAR-2 con HAR-3, HAR-3 con F. cambá, PxS con F. cambá, VNIIMK con F. cambá y HAR-2 con F. cambá.

Tabla 15. Diferenciación genética entre entradas del BAG-IM. Cuadrante superior: valores obtenidos para distancias calculadas según Nei (1972). Cuadrante inferior: valores de F_{ST} entre pares de entradas (poblaciones y compuestos). * F_{ST} : significativos ($p < 0,003125$). Las casillas verdes corresponden a los valores de F_{ST} no significativos, las casillas celestes a los valores de F_{ST} más altos y significativos, las casillas fucsias a los valores de F_{ST} más bajos y significativos.

	KLM	IAC	HAR-4	Colliguay	HAR-1	HAR-3	PxS	Prao-co	VNIIMK	Júpiter	Sunrise	Comangir	HAR-2	F.cambá
Distancias de Nei														
KLM	0.0000	0.3517	0.2337	0.3880	0.3599	0.4552	0.3825	0.3061	0.4097	0.4822	0.3902	0.3280	0.2843	0.4590
IAC	0.4783*	0.0000	0.2944	0.1233	0.3467	0.2453	0.2753	0.1662	0.2731	0.2591	0.2182	0.2607	0.3460	0.3825
HAR-4	0.3949*	0.3663*	0.0000	0.3142	0.2818	0.3671	0.2845	0.2398	0.4263	0.4006	0.2581	0.2834	0.1903	0.4731
Colliguay	0.4959*	0.1899*	0.3831*	0.0000	0.3603	0.1855	0.2096	0.1603	0.2648	0.1908	0.2478	0.2175	0.3337	0.3277
HAR-1	0.5947*	0.4903*	0.4692*	0.4859*	0.0000	0.4372	0.3341	0.3079	0.5033	0.4851	0.4507	0.3705	0.2640	0.4831
HAR-3	0.5483*	0.2255*	0.4414*	0.1841	0.5611*	0.0000	0.2901	0.2518	0.2593	0.2703	0.2777	0.3234	0.3750	0.4032
PxS	0.4862*	0.3198*	0.3393*	0.2180*	0.4693*	0.3195*	0.0000	0.2059	0.3284	0.2830	0.2467	0.2251	0.3151	0.4252
Prao-co	0.3986*	0.2140*	0.2375*	0.2018*	0.4226*	0.2873*	0.2067*	0.0000	0.2792	0.2216	0.1896	0.1912	0.2477	0.3058
VNIIMK	0.5952*	0.3407*	0.5431*	0.3367*	0.6394*	0.3447*	0.4212*	0.3640*	0.0000	0.2855	0.2894	0.3060	0.4278	0.4543
Júpiter	0.4865*	0.2524*	0.4040*	0.1795*	0.5233*	0.2459*	0.2713*	0.2022*	0.3034*	0.0000	0.2914	0.2376	0.3679	0.3732
Sunrise	0.5314*	0.2965*	0.3526*	0.3035*	0.6151*	0.3798*	0.3254*	0.2181*	0.4471*	0.3006*	0.0000	0.2294	0.3165	0.4302
Comangir	0.4485*	0.3477*	0.3387*	0.2868*	0.4942*	0.3751*	0.2725*	0.2212*	0.4096*	0.2284*	0.3121*	0.0000	0.2573	0.3938
HAR-2	0.4239*	0.4083*	0.3350*	0.4091*	0.4418*	0.4533	0.4101*	0.2864*	0.5472*	0.4029*	0.4441*	0.3459*	0.0000	0.4027
F.cambá	0.5592*	0.4347*	0.5299*	0.3816	0.6305*	0.4811	0.4839	0.3556*	0.5857	0.4147*	0.5453*	0.4894*	0.4783	0.0000

En cuanto al análisis de estructura poblacional mediante el empleo del modelo basado en el enfoque de Pritchard *et al.* (2000), si bien no se cumplen todos los supuestos del modelo dado que se hallaron desvíos de la panmixia en las entradas analizadas, se empleó esta metodología de manera exploratoria con el objeto de detectar grandes grupos genéticos. En base al conjunto de datos obtenidos, se analizaron los valores logarítmicos de las probabilidades para cada posible K [$\ln P(D)$] y el cambio de ΔK en función de K (Evanno *et al.*, 2005) para poder inferir el número de poblaciones ideales o *pooles* génicos que conforman la muestra total. Según la documentación del programa Structure, el valor de K más apropiado es aquel más bajo capaz de capturar la mayor estructuración en el conjunto total de datos. Dicho de otro modo, se deberá seleccionar aquel valor de K que presente la más alta probabilidad o mayor $\ln P(D)$, sin incorporar estructura ficticia. El análisis incluyó la evaluación de K=1 hasta K=20, obteniéndose como valor más probable K=14 (Tabla 16). Se observó un incremento en la probabilidad entre sucesivos valores de K a medida que éste aumentó, alcanzándose una estabilización de las probabilidades o una meseta en valores de K superiores a 14 (Figura 11-A, Tabla 16).

Tabla 16. Estimación del número de poblaciones mediante el método bayesiano para las entradas estudiadas. $\ln P(D)$ promedio: promedio del logaritmo natural de la probabilidad calculada para 10 réplicas de cada valor de K.

Valor de K	$\ln P(D)$ promedio	Desvío estándar	Valor de K	$\ln P(D)$ promedio	Desvío estándar
1	-10406,52	1,24	11	-6370,14	120,36
2	-9144,62	24,67	12	-6255,1	198,32
3	-8672,69	48,67	13	-6110,68	171,30
4	-8252,69	76,56	14	-5940,03	80,82
5	-7920,28	100,27	15	-5894,86	86,20
6	-7632,21	104,38	16	-5889,61	107,66
7	-7290,01	107,24	17	-5920,44	147,30
8	-6984,1	140,15	18	-5904,83	173,87
9	-6797,69	103,48	19	-5805,72	162,97
10	-6520,98	99,71	20	-5863,17	175,01

Los resultados de las 10 repeticiones realizadas para cada valor de K se adjuntan en la Tabla A.3 del Apéndice.

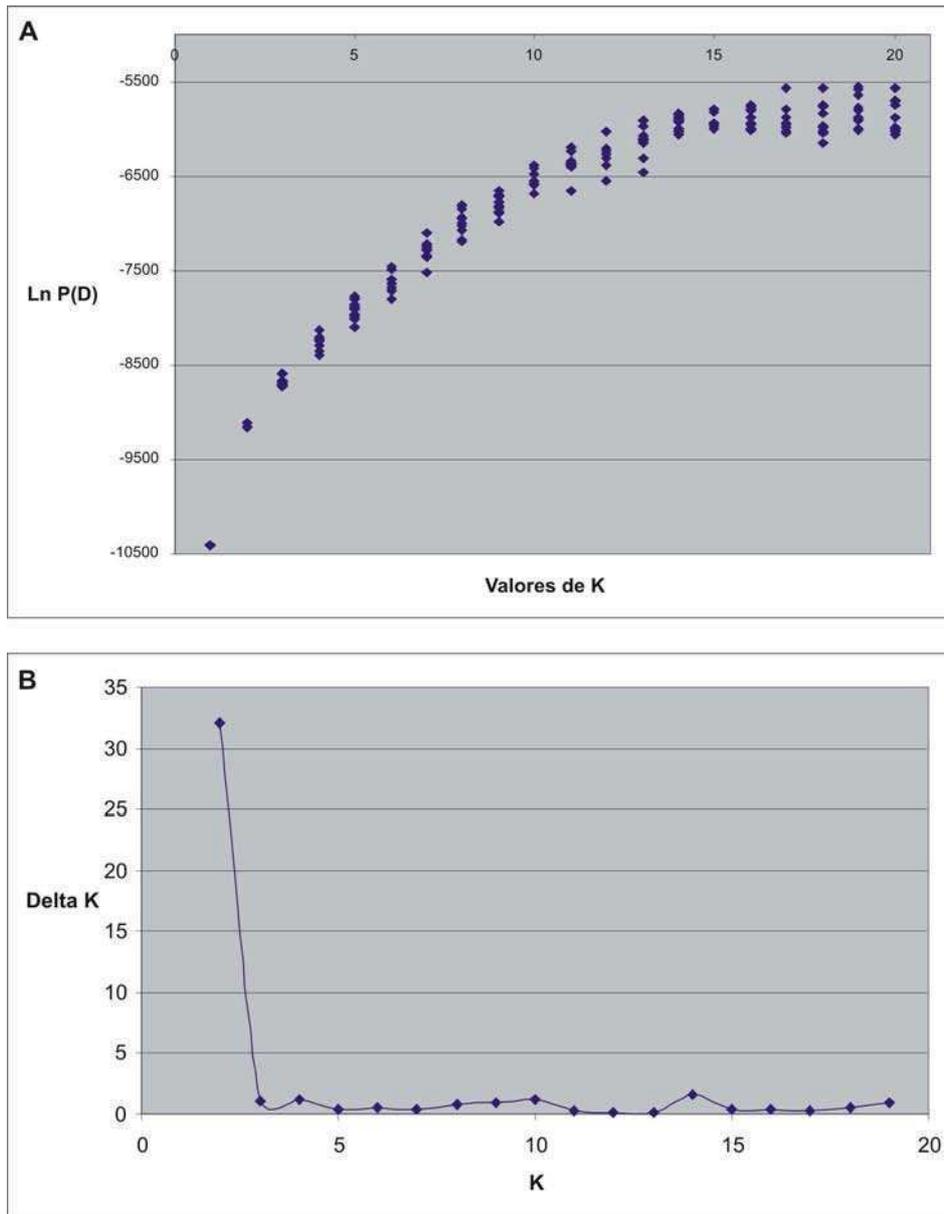


Figura 11. Análisis de estructura poblacional con método bayesiano. A. Logaritmo de las probabilidades para cada K. B. Gráfico de ΔK vs. K.

En base al método de Evanno *et al.* (2005) se obtuvo el máximo ΔK en $K=2$ (32,017), con tres picos de menor valor en $K=4$ (1,144); $K=10$ (1,26) y $K=14$ (1,55) (Figura 11-B). Se analizó la distribución de los materiales en los diversos grupos para cada uno de dichos valores de K. Un nivel superior de estructuración se observó en $K=2$, con un grupo rojo que incluyó a KLM, HAR4, HAR1 y HAR2 con más del 80% del genoma de cada individuo asignado a este grupo, mientras que el conglomerado verde agrupó a los restantes materiales (Figura 12).

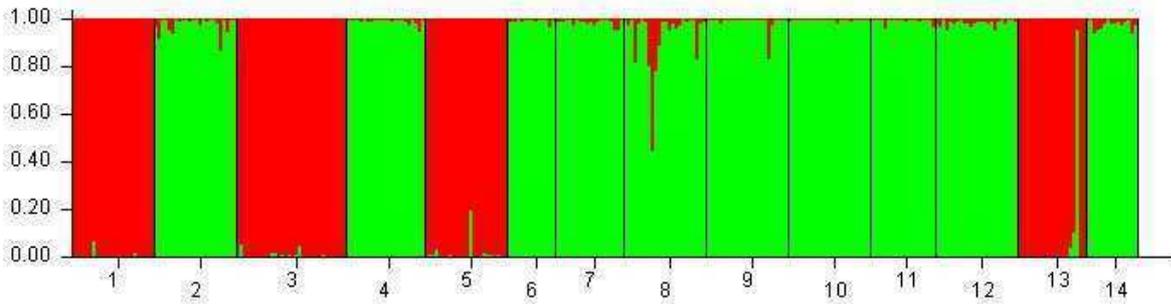


Figura 12. Estructura poblacional estimada para K=2 en el conjunto de poblaciones y compuestos analizados. Cada individuo está representado por un segmento vertical el cual es dividido en K segmentos de colores que representan el aporte de cada *pool* génico. Los números dispuestos horizontalmente corresponden a la identificación de cada material del BAG-IM. KLM (1), IAC (2), HAR-4 (3), Colliguay (4), HAR-1 (5), HAR-3 (6), PxS (7), Prao-co (8), VNIIMK (9), Júpiter (10), Sunrise (11), Comangir (12), HAR-2 (13), F. cambá (14).

En un nivel de estructuración más profundo se observó que las entradas se dividieron en 4 conglomerados (K=4), donde cada uno de los dos grupos obtenidos en K=2 se dividió en dos (Figura 13).

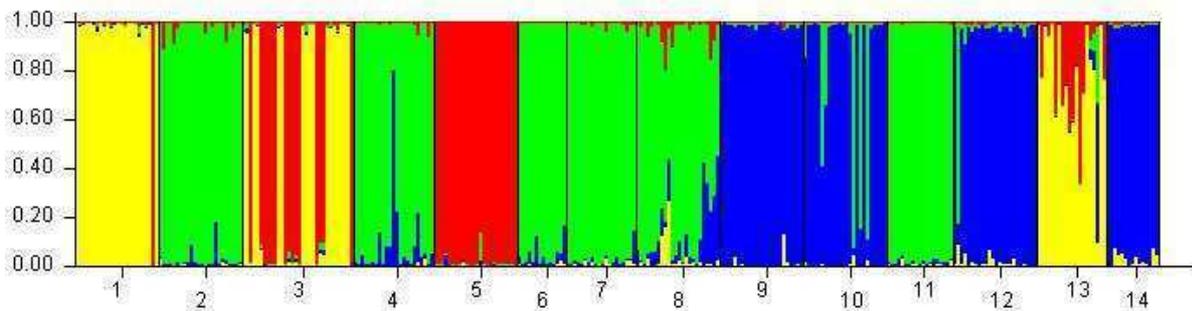


Figura 13. Estructura poblacional estimada para K=4 en el conjunto de poblaciones y compuestos analizados. Los números dispuestos horizontalmente corresponden a la identificación de cada material del BAG-IM. KLM (1), IAC (2), HAR-4 (3), Colliguay (4), HAR-1 (5), HAR-3 (6), PxS (7), Prao-co (8), VNIIMK (9), Júpiter (10), Sunrise (11), Comangir (12), HAR-2 (13), F. cambá (14).

Todas las entradas, excepto HAR-4, mostraron más del 75% del genoma de cada individuo asignado a un determinado grupo. De este modo, el grupo amarillo incluyó a KLM (94,4% del genoma de cada individuo dentro de ese grupo), HAR-2 (75,6%) y 17/32 individuos de HAR-4 (93,3% del genoma de los

17 individuos amarillos), mientras que el rojo se formó con HAR-1 (98,3%) y 15/32 individuos de HAR-4 (90,6% del genoma de los 15 individuos rojos). Los materiales IAC (95,6%), Colliguay (90,4%), HAR-3 (94,8%), PxS (96,2%), Prao-co (81,9%) y Sunrise (97,5%) se agruparon por un lado (*cluster* verde), mientras que VNIIMK (97,5), Júpiter (82,8), Comangir (92,3) y F. cambá (95,6) conformaron el grupo azul.

El valor de $K=10$ también mostró un máximo de ΔK (1,26), lo cual indicaría un nivel aún mayor de sub-estructuración en los materiales estudiados. En este análisis se corroboró la propuesta de que HAR-4 está formado por dos acervos genéticos en casi igual proporción porque se separaron sus individuos en dos grupos distintos (16 individuos celeste claro: 92,5%, 14 individuos naranja: 89,5%) (Figura 14). KLM, HAR-1, PxS, VNIIMK, HAR-2 y F. cambá formaron un único cluster con más del 87% del genoma de sus individuos adjudicados a un grupo excepto IAC (84,9%), Colliguay (87,4%) y HAR-3 (93,3%) que constituyeron el grupo verde. Las introducciones Prao-co (66,1%) y Sunrise (91,3%) quedaron en el grupo rojo y Júpiter (69,6%) con Comangir (88,9%) en el grupo morado oscuro (Figura 14). Cabe destacar que Prao-co fue la entrada que mostró contribuciones de mayor número de acervos genéticos (K) en el genoma de cada uno de sus individuos.

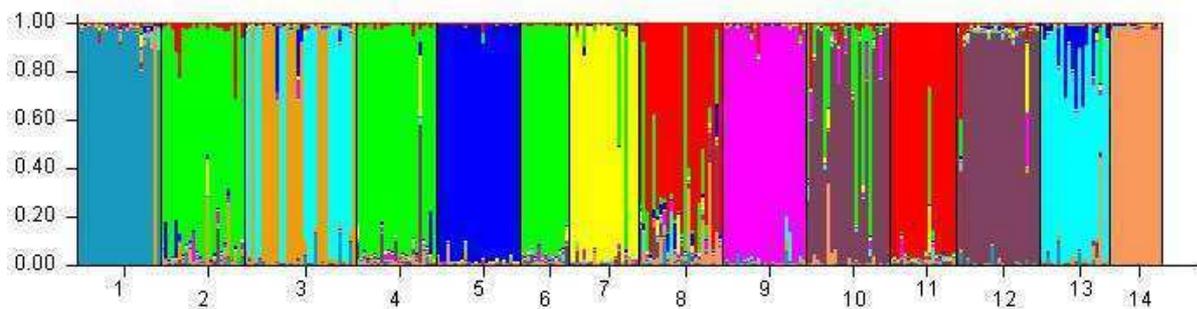


Figura 14. Estructura poblacional estimada para $K=10$ en el conjunto de poblaciones y compuestos analizados. Los números dispuestos horizontalmente corresponden a la identificación de cada material del BAG-IM. KLM (1), IAC (2), HAR-4 (3), Colliguay (4), HAR-1 (5), HAR-3 (6), PxS (7), Prao-co (8), VNIIMK (9), Júpiter (10), Sunrise (11), Comangir (12), HAR-2 (13), F. cambá (14).

Finalmente se eligió el valor de $K=14$ como el más probable para la asignación de la estructura en los materiales estudiados. Esta decisión se tomó en base a que $K=14$, no sólo mostró un máximo para ΔK (1,55), sino que

además este resultado determinó el inicio de la meseta en los valores de probabilidades para cada K . Dada la elevada diferenciación observada en base a los F_{ST} y a los diversos orígenes de los programas de mejoramiento, resulta lógico esperar que el valor de K más probable sea alto.

De las 14 entradas incluidas en este análisis, 13 mostraron pertenecer a un grupo genético diferente, excepto HAR-4 que evidenció la contribución de un acervo genético común a HAR-1 y otro común con HAR-2. Por su parte, Júpiter mostró estar compuesto por la contribución de dos *pooles* genéticos distintos y no comunes con los restantes materiales (Figura 15). Si bien la mayoría de los individuos mostraron un alto porcentaje de pertenencia a un grupo determinado, muchos de ellos presentaron, además, contribuciones menores de uno o varios grupos adicionales. Al igual que lo observado para los valores de K descritos previamente, Prao-co fue la entrada que mostró contribuciones de mayor número de acervos genéticos (K) en el genoma de cada uno de sus individuos.

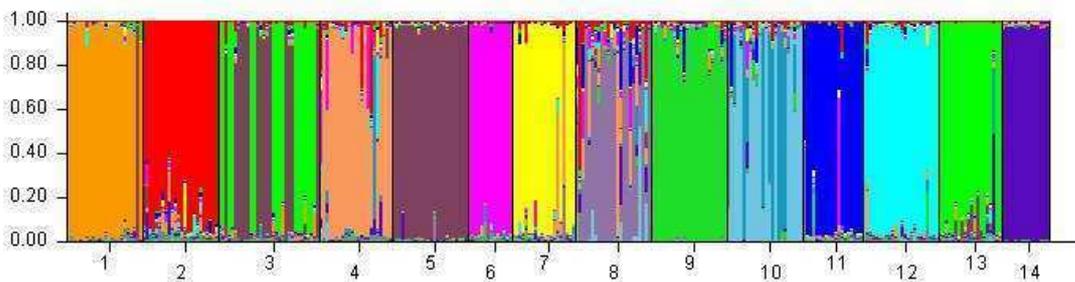


Figura 15. Estructura poblacional estimada para $K=14$ en el conjunto de poblaciones y compuestos analizados. Los números dispuestos horizontalmente corresponden a la identificación de cada material del BAG-IM. KLM (1), IAC (2), HAR-4 (3), Colliguay (4), HAR-1 (5), HAR-3 (6), PxS (7), Prao-co (8), VNIIMK (9), Júpiter (10), Sunrise (11), Comangir (12), HAR-2 (13), F. cambá (14).

Se muestra a continuación la distribución de cada material en el grupo al cual fue asignado y la proporción mayoritaria de miembros de cada material que pertenecen a ese grupo (Tabla 17). Esta proporción fue expresada como porcentaje.

Tabla 17. Distribución de las entradas analizadas por grupos

Grupo	Entrada	Porcentaje
1: naranja	KLM	90,1
2: rojo	IAC	86,3
3: morado y verde intenso	HAR-4	90,5 y 94,3
4: durazno	Colliguay	73,7
5: morado	HAR-1	95,2
6: fucsia	HAR-3	93,5
7: amarillo	PxS	83,7
8: lila	Prao-co	57,7
9: verde oscuro	VNIIMK	91,5
10: celeste claro y oscuro	Júpiter	85,2 y 90,3
11: azul	Sunrise	87,1
12: celeste intenso	Comangir	89,7
13: verde intenso	HAR-2	86,1
14: azul oscuro	F. cambá	96,5

R.3.3. Análisis de agrupamiento

Se construyó un árbol de Neighbor-Joining en base a las distancias genéticas de Nei (1972) entre las poblaciones y los compuestos analizados. El retículo resultante se muestra en la Figura 16.

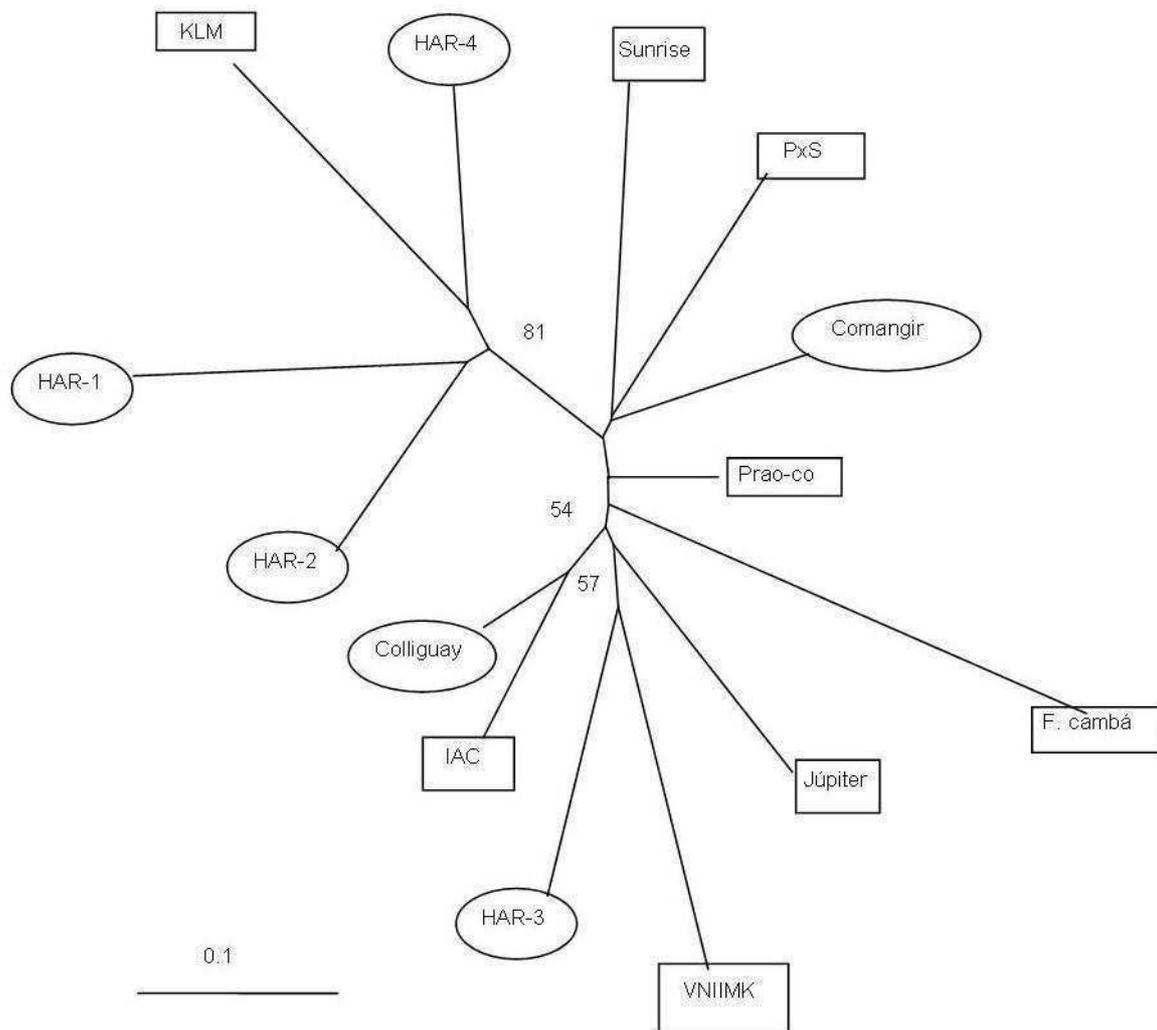


Figura 16. Retículo de Neighbor-Joining obtenido para las poblaciones y compuestos en base a la distancia genética de Nei (1972). Se obtuvieron valores de soporte de *bootstrap* entre 15 y 81%, en la Figura sólo se agregaron los valores superiores a 50. Las entradas dentro de cuadrados corresponden a las poblaciones, mientras que las entradas dentro de óvalos corresponden a la categoría de compuestos.

La rama que mostró soporte de *bootstrap* de 81% dividió a los materiales de manera coincidente con los grupos obtenidos con el método bayesiano en $K=2$. Por otro lado, la rama que sostiene a HAR-1 es más larga, lo cual indicaría que acumuló mayor cantidad de cambios, concordando con el resultado de $K=4$ donde la sub-estructuración separó claramente a HAR-1 de KLM, HAR-2 y HAR-4. La población Prao-co se unió al árbol por medio de una rama más corta que las demás y en una posición intermedia, lo cual resulta

coincidente con los altos niveles de mezcla detectados para estos individuos mediante el método bayesiano.

Se realizó un análisis de agrupamiento para el conjunto total de individuos que conforman las poblaciones y los compuestos. El árbol de distancias genéticas generado en base a la proporción de alelos compartidos entre pares de individuos mostró una clara delimitación para las entradas HAR-1, F. cambá y VNIIMK, agrupándose el 100% de sus individuos en un único conglomerado. Los individuos de los restantes materiales se agruparon en dos o más *clusters*. Aquellos que formaron dos grupos (KLM, Comangir, IAC) mostraron uno mayoritario que incluyó entre el 91,67% y el 95,83% de sus individuos. Los individuos de los materiales HAR-2, Júpiter, PxS, Sunrise y HAR-3 formaron tres grupos, donde el mayoritario incluyó entre el 50% y el 89,5%. Un individuo de KLM se intercaló entre un *cluster* de HAR-4 y éstos se mezclaron con individuos de HAR-2, coincidentemente con el patrón observado en el árbol de distancias de Nei y con los agrupamientos obtenidos mediante el método bayesiano, resultando consistente con el origen de estos compuestos. Los individuos de los materiales Júpiter, HAR-3 y VNIIMK también siguieron el mismo patrón de comportamiento que su disposición en la Figura 16. Los materiales cuyos individuos formaron más de tres grupos (HAR-4 y Colliguay), mostraron uno mayoritario que incluyó entre el 39,13% y el 43,75% de sus individuos. Cabe destacar que, en concordancia con los análisis previos, la población Prao-co fue la única que se diseminó de forma notoria, exhibiendo un patrón disperso con individuos dentro de 10 grupos distintos a lo largo de todo el árbol (Figura 17).

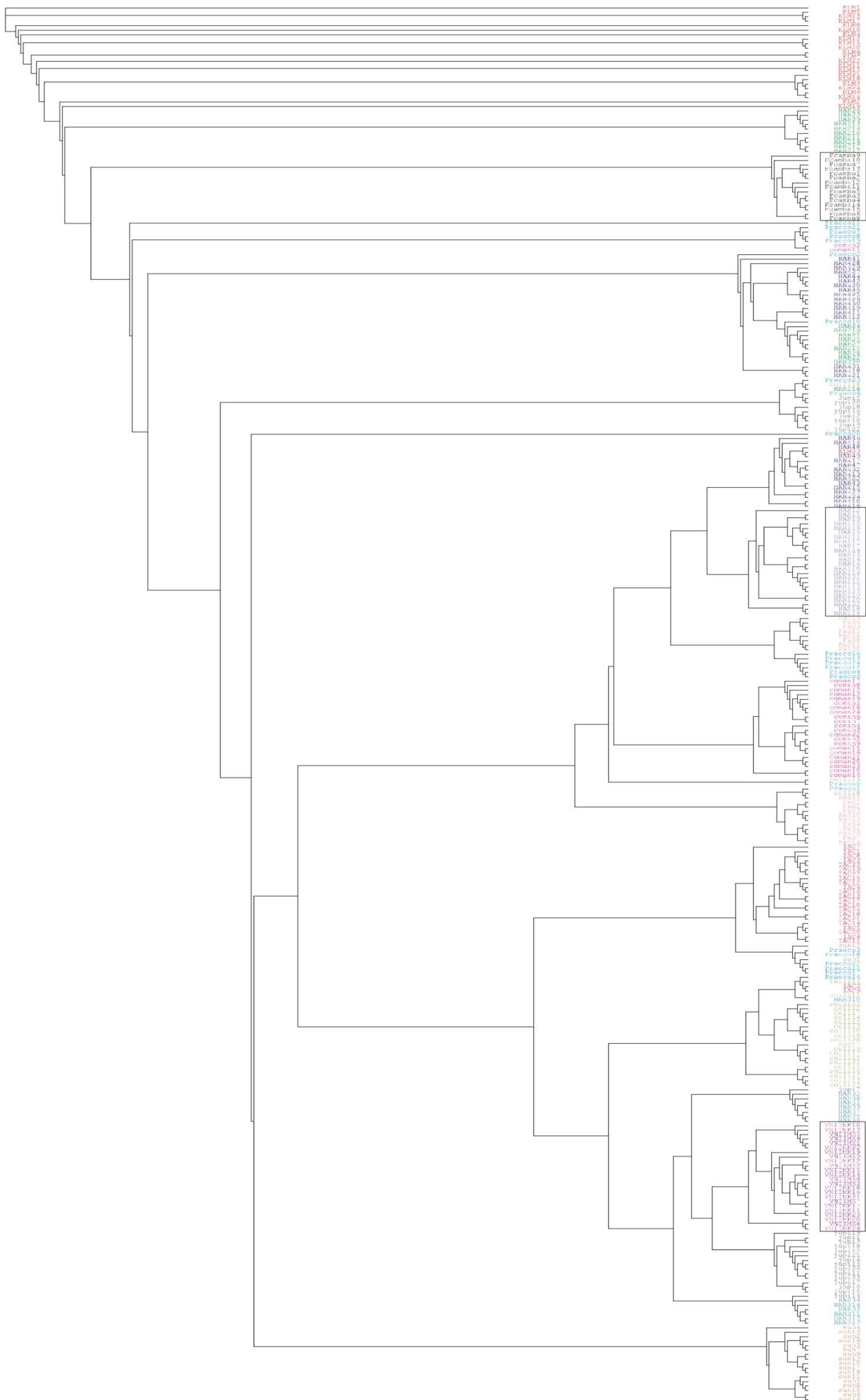


Figura 17. Árbol del total de individuos en base a la distancia de alelos compartidos (Dps) para poblaciones y compuestos. Los rectángulos indican aquellas entradas donde el 100% de sus individuos se agruparon juntos, constituyendo un único conglomerado (F. cambá, HAR-1 y VNIIMK).

El análisis de agrupamiento correspondiente a las líneas también fue realizado en forma separada, retomando los procedimientos de evaluación propuestos por Heckenberger *et al.* (2002). En el árbol obtenido en base al índice de Jaccard, todos los individuos de una misma línea se unen al mismo grupo con diversos valores de similitud, los cuales varían entre 0,50 y 1 (Figura 18). A pesar de la falta de homocigosis dentro de las líneas, el dendrograma permitió corroborar que existe cohesión entre individuos pertenecientes a la misma línea, confirmando lo que fue observado en el campo por el mejorador. Sólo el individuo 3 de la línea CM 307-1 quedó excluido de su grupo.

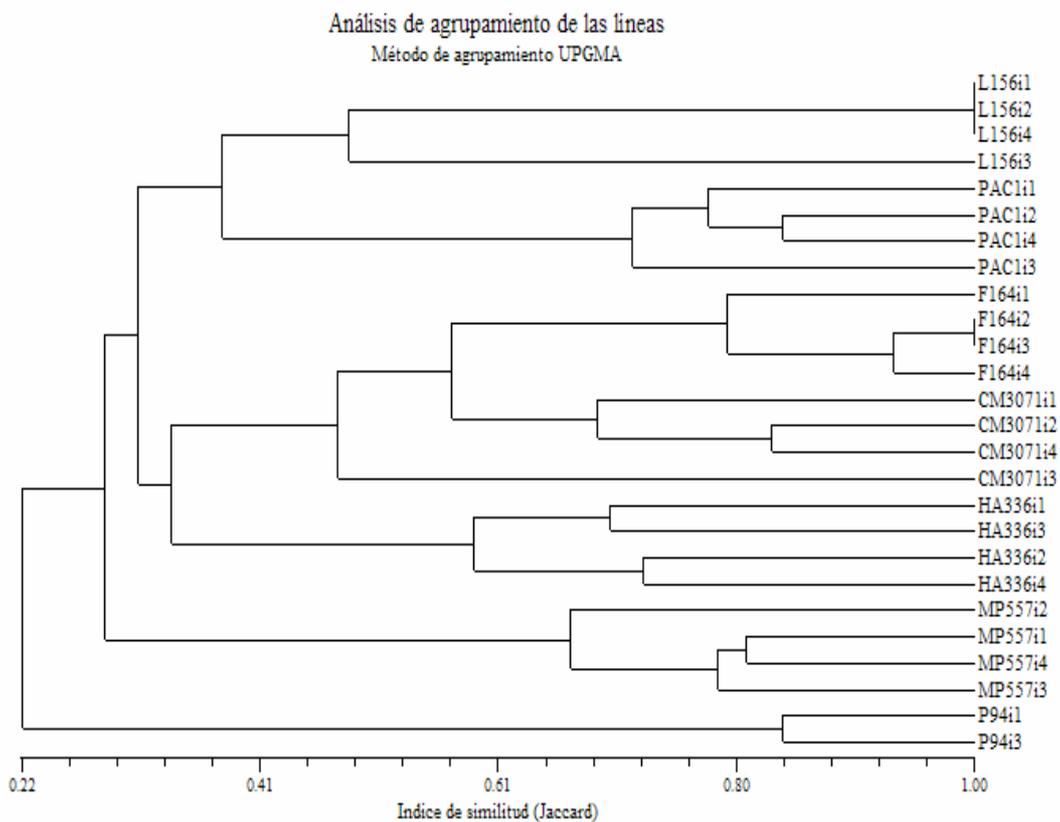


Figura 18. Agrupamiento entre los individuos pertenecientes a cada una de las líneas estudiadas en base al índice de Jaccard con 16 loci SSR. Se identificó a cada individuo con el nombre de la línea a la que pertenece, seguido de la letra “i” junto al respectivo número de individuo.

DISCUSIÓN

D.1. Diversidad genética en el BAG-IM. Implicancias en la conservación de germoplasma

El conocimiento de los materiales coleccionados en el BAG-IM, tanto a nivel morfológico como agronómico y molecular, resulta de vital importancia como herramienta para el desarrollo eficiente del programa de mejoramiento de girasol. El estudio de la diversidad molecular del germoplasma establece una base sobre la cual se construirá conocimiento acerca de procesos selectivos, consecuencias de la domesticación, fenómenos poblacionales, perfiles funcionales, correlación entre los mismos y caracteres de interés agronómico; todo ello con un enfoque hacia el aporte directo para el mejorador. La estrategia propuesta es recurrir a las fuentes de germoplasma como dadoras de alelos que potencien estos caracteres deseables, los cuales deberán incorporarse en fondos genéticos apropiados y en ambientes favorables (Takeda and Matsuoka, 2008). El estudio de los patrones de diversidad y de las relaciones genéticas entre las entradas de girasol del BAG-IM, es una herramienta invaluable en las estrategias de mejoramiento del cultivo. El análisis de diversidad genética en colecciones de germoplasma facilita la clasificación confiable de las entradas y la identificación de un conjunto de materiales núcleo o *core sets accessions* con posible utilidad para propósitos específicos de mejoramiento.

El análisis de conglomerados (*clusters*) de los datos obtenidos a partir de la caracterización morfológica y agronómica de 309 entradas del BAG-IM durante las campañas 2002-2003 a 2007-2008 corroboró el criterio de selección *a priori* de los materiales analizados en este trabajo. El dendrograma obtenido separó a las 21 entradas elegidas para este estudio en distintos grupos dispersos a lo largo de todo el dendrograma, validando así el criterio de cubrir el espectro de diversidad del banco. La diferenciación detectada a nivel morfológico sugiere la existencia de diferenciación a nivel del ADN. Sin embargo, dado que la varianza fenotípica es la sumatoria de la

varianza genética más la varianza ambiental, las relaciones obtenidas a partir del árbol fenotípico y del molecular pueden no ser comparables.

Este Capítulo del presente trabajo de tesis reviste suma importancia dado que constituye el primer antecedente de análisis de diversidad con marcadores moleculares en un conjunto de entradas pertenecientes al BAG-IM. De los 35 SSR ensayados, 16 se eligieron por su reproducibilidad y facilidad de interpretación, mientras que 19 (54,28%) se descartaron por mostrar alto número de alelos nulos en los individuos analizados, o por amplificar en dos o más regiones del genoma como consecuencia de la presencia de dos o más loci SSR duplicados. Estos resultados concuerdan con Tang *et al.* (2002) que detectaron un 53,6% de SSR con los mencionados problemas. La mayoría de los loci dinucleotídicos elegidos para este trabajo correspondieron a la generación de productos $(GA)_n$ porque los motivos $(GT)_n$ producen perfiles de menor calidad al igual que los productos de motivos $(ATC)_n$ (Paniego *et al.*, 2002).

Se observó que algunos de los alelos detectados experimentalmente no fueron coincidentes con los tamaños esperados de acuerdo a la teoría de surgimiento de la variabilidad en loci SSR. Una de las causas podría ser que estos loci mostraran patrones mutacionales complejos como los descritos en otros cultivos tales como maíz (Lia *et al.*, 2007; Mogg *et al.*, 2002).

El contenido de información polimórfica (PIC) promedio fue de 0,50; similar al detectado en 124 líneas con 78 loci SSR (0,51) (Zhang *et al.*, 2005) y levemente menor al encontrado para 16 líneas endocriadas de germoplasma con 170 loci SSR (0,55) (Paniego *et al.*, 2002).

Los marcadores microsatélites que mostraron mayor variabilidad en el conjunto de entradas estudiadas (HA1848, HA2063, HA2077 y HA3239) podrían implementarse para la caracterización de líneas e híbridos comerciales inscriptos en el INASE con objeto de asistir a la protección de los materiales y de garantizar la calidad de las semillas que el productor adquiere. Este organismo público se ocupa de controlar el comercio, la fiscalización de semillas, la inscripción de nuevos cultivares y el otorgamiento de títulos de propiedad intelectual para las obtenciones vegetales. Estos títulos se conceden según los lineamientos internacionales fijados por la UPOV (organismo del cual la Argentina es miembro) y, como

ya se mencionó, se basan en la descripción de las variedades vegetales en función de características fenotípicas para el cumplimiento de las pruebas DHE. Debido a las limitaciones de los marcadores morfológicos, desde hace varios años se estudia la posibilidad de incorporar a los descriptores los marcadores de ADN. En este sentido, los marcadores SSR polimórficos detectados en este trabajo constituyen un aporte para la caracterización e identificación de materiales de girasol que lleva adelante el INASE.

El número promedio de alelos por locus (A) obtenido en este trabajo fue de 6,75 (Tabla 9), valor que resultó intermedio en comparación con los obtenidos por Tang y Knapp (2003) para poblaciones silvestres (9,7), para un conjunto de materiales domesticados exóticos (5,2) y para líneas endocriadas (3,5). Este último valor fue coincidente con los informados por Paniego *et al.* (2002) y Zhang *et al.* (2005). Mientras que para el conjunto de las poblaciones estudiadas se obtuvo un valor promedio de A por población de 2,59, para los compuestos de 2,46 y 1,49 para las líneas (Tabla 9). Los valores de A obtenidos en este trabajo para las poblaciones y compuestos de girasol cultivado fueron inferiores a los reportados por Gross *et al.* (2003) y por Tang y Knapp (2003) para poblaciones silvestres y razas nativas de *Helianthus annuus*. La bibliografía menciona valores de A mayores a los obtenidos en este trabajo para otras especies de *Helianthus*, entre ellas *H. deserticola* y *H. petiolaris* (Gross *et al.*, 2003) además de *H. verticillatus* y *H. angustifolius* (Ellis *et al.*, 2006). Estas diferencias probablemente se deban a que las mencionadas especies están constituídas por poblaciones naturales que no han sufrido un proceso de domesticación.

El máximo número de alelos por locus para las poblaciones fue de 9 en el locus HA1848, mientras que para los compuestos fue de 8 en el locus HA2077. Para el conjunto de las líneas, 5 loci mostraron 3 alelos, 10 loci SSR detectaron 2 alelos, mientras que solamente en un locus SSR se detectó 1 alelo (Figura 6). En base a estos resultados, que ponen de manifiesto distintos niveles de heterogeneidad dentro de las líneas, se analizaron las fuentes de variación según el criterio propuesto por Heckenberger *et al.* (2002) (discutido posteriormente).

Tomando el criterio de Tang *et al.* (2002) que detectó una frecuencia de alelos nulos de 0,065, para el presente estudio se determinó una

frecuencia global de alelos nulos de 0,050. Según los resultados expuestos por Tang y Knapp (2003) la frecuencia de alelos nulos fue aumentando según se tratara de líneas (0,0142), de entradas exóticas domesticadas (0,0170) o de poblaciones silvestres (0,0508). En concordancia con lo descrito por los mencionados autores, en este trabajo de tesis la frecuencia de alelos nulos para las líneas fue de 0,0026, para los compuestos fue de 0,020 y para las poblaciones fue de 0,027. Las diferencias observadas entre las poblaciones aquí estudiadas y las poblaciones silvestres empleadas por los mencionados autores podrían atribuirse a la base genética más restringida de las primeras.

Los materiales F. cambá y Júpiter, que mostraron mayor número de alelos exclusivos, también presentaron un patrón de elevada diferenciación en los análisis de estructura (método bayesiano y retículo de Neighbor-Joining). En general, las frecuencias detectadas para los alelos exclusivos fueron altas (0,023 a 1, Tabla 10); coincidentemente con el elevado grado de diferenciación hallada, comparado con trabajos previos en maíz (Lia, 2004). Estos resultados podrían ser consecuencia de los orígenes geográficos diversos de los materiales y de los efectos de la deriva génica durante el proceso de mejoramiento.

La heterocigosis (H_e) promedio obtenida en el presente estudio fue de 0,29 para el conjunto de marcadores SSR (tanto dinucleótidos como trinucleótidos) y considerando todas las entradas (Tabla 9), menor que los resultados reportados por Tang *et al.* (2002) para 4 líneas endocriadas, con una H_e promedio de 0,44 para SSR trinucleótidos y 0,52 para SSR dinucleótidos. Los resultados de H_e promedio por categorías obtenidos en este trabajo fueron de 0,354 para el conjunto de poblaciones; 0,342 para los compuestos y 0,188 para las líneas (Tablas 9 y 11). Estos valores fueron menores a los hallados por Tang y Knapp (2003), Ellis *et al.* (2006) y Gross *et al.* (2003) para poblaciones naturales de *Helianthus annuus* (entre 0,57 y 0,67) y materiales domesticados (0,64). En cuanto a las comparaciones con otras especies de *Helianthus*, las estimas obtenidas en el presente trabajo también fueron inferiores a las de poblaciones de *H. verticillatus* y de *H. angustifolius* (Ellis *et al.*, 2006). El mismo patrón se observó al comparar los valores aquí obtenidos con los de otras especies de *Helianthus* analizadas

por Gross *et al.* (2003) (*H. deserticola* y *H. petiolaris*). En comparación con materiales silvestres o provenientes de otros programas, los valores de H_e obtenidos en este trabajo fueron inferiores, pero mostraron similitud con los valores hallados para el caso de *H. angustifolius* ($H_e=0,34$).

Resulta importante destacar que las especies de girasol cultivado han sido intensamente mejoradas generando así una disminución marcada de la heterocigosis del germoplasma progresivamente más domesticado (Tang and Knapp, 2003), reteniendo sólo entre el 40 y 50% de la diversidad presente en las especies silvestres (Liu and Burke, 2006). Este patrón es coincidente con lo observado en nuestro trabajo, donde todas las estimas de diversidad fueron menores a las de materiales silvestres o de especies de *Helianthus* no cultivadas. Las entradas de germoplasma correspondientes a poblaciones y compuestos analizadas en este trabajo poseen ciclos sucesivos de cruzamientos dirigidos dado que provienen de líneas cultivadas o son materiales con años de uso en los programas de mejoramiento.

Se halló una gradual disminución en las estimas de diversidad dentro del espectro de germoplasma estudiado (poblaciones, compuestos y líneas), concordante con lo detectado por Tang y Knapp (2003). Las poblaciones muestran los valores más altos de cada estima de diversidad (A , R_s , H_e y alelos exclusivos). Esto podría explicarse por el modo de obtención de estos materiales (polinización abierta) y por los distintos orígenes geográficos de sus programas. Dado que las poblaciones se construyen partiendo de materiales que poseen intrínsecamente un vigor híbrido asociado (alto grado de heterosis), muestran mayor H_e que los compuestos. Para el programa de mejoramiento del cultivo es sumamente importante determinar las categorías de materiales que muestran mayores estimas de diversidad, dado que son las más promisorias para la búsqueda de alelos favorables enfocados a caracteres de interés agronómico.

En base a los resultados obtenidos en el análisis de muestras mínimas para evaluar diversidad, tanto para poblaciones como para compuestos, se alcanzó el *plateau* en torno a un número semejante de individuos que osciló entre 40 y 50 (Figura 7). Dentro de cada categoría se detectó el mismo patrón de comportamiento, puesto que el 50% de las entradas alcanzaron el *plateau* con un número de individuos que para las

poblaciones varió entre 12 y 18, mientras que para los compuestos osciló entre 10 y 20 (Figuras 8 y 9). Por lo tanto, no se observó un comportamiento diferencial entre ambas categorías estudiadas. Si bien, potencialmente habría mayor cantidad de alelos que sólo podrían capturarse con un aumento en el número de individuos analizados, el hecho de haber alcanzado el *plateau* permite definir un número estimado de individuos (alrededor de 20) a muestrear como estrategia para la caracterización exhaustiva del banco.

De acuerdo con los resultados obtenidos en el análisis de grupos núcleo (*core sets*), el conjunto de las poblaciones mostró conservar el 100% de la diversidad detectada en este trabajo con un muestreo de 50 individuos (29% del total genotipificado), mientras que para los compuestos se alcanzó con 30 individuos (22% del total genotipificado). Por lo tanto, la proporción de individuos requeridos para alcanzar el total de la diversidad resultó ligeramente mayor para las poblaciones que para los compuestos. Si bien no son sustanciales, estas diferencias podrían explicarse por una heterocigosis observada levemente mayor para los compuestos (Tabla 11). Es decir que, cuando se analizan entradas que no corresponden a la categoría de líneas, para definir el tamaño de los *core sets* no sólo se requiere tener en cuenta la variabilidad en cuanto al número de alelos, sino también los niveles de heterocigosis observada. El porcentaje de individuos que logró conservar el total de la diversidad detectada para cada categoría resultó similar al reportado en arroz (Kojima *et al.*, 2005). Estos autores conformaron un *core set* de 67 entradas de arroz cultivado (*Oryza sativa* L.) (20% del total genotipificado) que retuvo el 91% de la diversidad total hallada en 332 materiales analizados con 179 marcadores RFLP. Mientras que nuestros resultados fueron superiores al porcentaje de individuos reportado para trigo (alrededor de 10% del total genotipificado) (Balfourier *et al.*, 2007). Estos autores analizaron 3942 entradas de trigo pan (*Triticum aestivum*) con 38 SSR y conformaron un grupo núcleo de 372 entradas que retuvieron el 98% de la diversidad alélica total.

Los individuos que portan alelos exclusivos poseen una mayor probabilidad de estar representados en el grupo núcleo (*core set*). En los materiales aquí estudiados, sólo Júpiter y F. cambá (ambos con 3 alelos

exclusivos), no fueron incluidos en el primer grupo núcleo (*core set*) de tamaño 10. En los restantes tamaños de muestra evaluados, todas las entradas estuvieron igualmente representadas (Tabla 10).

En el contexto de las líneas, identificar grupos núcleo (*core sets*) sirve para rescatar un bajo número de individuos que retengan el 100% de la variabilidad detectada en el conjunto global. El hecho de reducir el número de individuos a muestrear disminuye el costo y agiliza el análisis, sin afectar la precisión de los resultados obtenidos. Considerando el modo de obtención de poblaciones y compuestos, los resultados obtenidos por medio de este análisis no se emplearán como criterio para definir un grupo núcleo (*core set*) específico dado que la constitución genética de cada entrada está definida por un conjunto de individuos. Por lo tanto, este análisis se utilizará sólo como un indicador indirecto para el diseño de muestreos de materiales del BAG-IM con el objeto de caracterizarlos molecularmente de manera más exhaustiva y ajustando el número de individuos. En base a lo hallado en este trabajo, sería mejor orientar los esfuerzos a caracterizar más entradas del BAG-IM con un menor número de individuos muestreados (entre 10 y 20), lo cual se apoya también en los valores de F_{ST} obtenidos y la aparición de alelos exclusivos. Esta propuesta también está avalada por los resultados obtenidos en el análisis de grupos núcleo, donde todas las entradas están representadas con al menos un individuo (Tabla 12). Probablemente en muestreos iniciales los números deberían ser iguales para ambas categorías y, en base a cálculos posteriores de las estimas de diversidad, ajustar el tamaño de muestra. Cabe destacar que el análisis de *core set* propuesto por Liu *et al.* (2003) por categorías, incluye el número total de individuos analizados dentro de ellas, sin información previa acerca de la asignación de los mismos a un material determinado. Con lo cual, el comportamiento o patrón de cada categoría fue consecuencia del global del genotipo de los individuos dentro de la misma.

En cuanto a las líneas, no se realizó el análisis de muestra mínima para diversidad ni el de grupos núcleo (*core sets*) a causa de la elevada heterocigosis detectada, sumado a que el número de individuos resultaría deficiente para calcular las estimas de individuos a muestrear.

El estudio de la diversidad genética en líneas endocriadas resulta sumamente importante debido a la potencialidad de su aplicación en el análisis de relaciones de *pedigree* y empleo de criterios legales que permitan discriminar líneas independientes y perfiles específicos de un genotipo dado. El análisis con marcadores moleculares es una herramienta ideal para realizar la pre-selección de líneas nuevas y contribuye a los estudios de pureza varietal y a la identificación obtenciones vegetales con fines legales. Dado que en este trabajo se observó heterocigosis y situaciones de heterogeneidad entre los individuos de una misma línea, se empleó el criterio propuesto por Heckenberger *et al.* (2002) para describir la variación observada. La ocurrencia promedio de casos normales para todas las líneas (67,6) fue menor que el reportado por estos autores (86,6), lo cual podría deberse a que evaluaron un mayor número de loci SSR. Además, se obtuvieron valores de ocurrencia superiores para los casos 2, 3, 4, 5, 7, 8 y 10; mientras que para el caso 9 fueron menores y para el caso 6 coincidente con lo reportado por los mismos autores ya que no se registraron comparaciones entre individuos con este comportamiento (Tabla 13). Asimismo, en el presente trabajo se observó una considerable variación de la distancia genética entre los diversos pares de individuos de una misma línea. Estas observaciones pueden ser explicadas por motivos genéticos y/o técnicos. Dentro de las razones genéticas se incluyen la segregación de ADN parental en las generaciones estudiadas, dado que las líneas empleadas en este trabajo se encuentran en la cuarta o quinta generación (*Ing. Álvarez, comunicación personal*). Podría suceder que las líneas aún no estén estabilizadas, es decir que hayan alcanzado la homocigosis práctica, no así a nivel genético. La homocigosis práctica se refiere a la homogeneidad entre los individuos de una misma línea en cuanto a su fenotipo, es decir sus características morfológicas y agronómicas observadas en el campo, pero que conservan a nivel genético loci heterocigotas que no pueden ser detectados fenotípicamente. Otro motivo podrían ser las mutaciones *de novo*, tanto en la región del iniciador que generarían alelos nulos, como las mutaciones en las regiones fuera del mismo que resultarían en alelos nuevos. La presencia de heterocigosis residual también se ha reportado en la bibliografía como una posible causa

de la aparición de variación dentro de las líneas (Cooke *et al.*, 2003; Gunjaca *et al.*, 2008). Dentro de las causas técnicas, la contaminación con polen foráneo es muy poco probable dado que los fenotipos se mostraron homogéneos en las evaluaciones de campo. Otra causa técnica aplicable a nuestro estudio es el efecto inesperado de deslizamiento de la ADN polimerasa durante la reacción de PCR que podría originar surgimiento de alelos nuevos. En base a antecedentes previos (Gunjaca *et al.*, 2008; Heckenberger *et al.* 2002; Heckenberger *et al.* 2003; Heckenberger *et al.*, 2005), dada la heterocigosis observada en las líneas analizadas en este trabajo, sería conveniente realizar un seguimiento de los genotipos a lo largo de varias generaciones de autofecundación subsiguientes para una misma línea. Heckenberger *et al.* (2002, 2003, 2005) detectaron variación en las estimas de distancia genética basada en SSR de más de 0,12 (escala de 0 a 1) entre diferentes generaciones de mejoramiento de una misma línea de maíz. Gunjaca *et al.* (2008) hallaron 6 de 41 líneas de maíz fuera de tipo en las pruebas DHE llevadas a cabo en 3 años consecutivos, lo cual podría indicar presencia de heterocigosis residual.

D.2. Distribución de la variabilidad genética

D.2.1. Estructura genética intrapoblacional

El girasol es una especie autógena, es decir de polinización cerrada o que puede autofecundarse, pero puede polinizarse de manera cruzada. En base al modo de obtención de las entradas analizadas (poblaciones y compuestos) se esperaría panmixia, pero dado que derivan de unos pocos individuos con ciclos sucesivos de multiplicación, pueden no ajustar a las proporciones propuestas por Hardy-Weinberg (HW). De acuerdo con esta situación, el único material que mostró un ajuste global al equilibrio de HW fue el compuesto HAR-1 (Tabla 14). Todos los loci polimórficos analizados mostraron desvíos significativos al equilibrio de HW, en al menos una de las poblaciones estudiadas, coincidente con lo observado por Ellis *et al.* (2006) para *H. verticillatus* y *H. angustifolius*. Si bien la tendencia general (F_{IS} global) fue un exceso de homocigotas, el análisis locus por locus no mostró

desviaciones tan marcadas (Tabla 14). Esta situación puede ser consecuencia de las correcciones empleadas dado que las pruebas están diseñadas para estudiar sólo un locus y, por ende, las aproximaciones multilocus pueden generar distorsiones (Goudet, 2001).

El hecho de que los materiales KLM, Júpiter y HAR-4 hayan presentado mayor número de loci con desvíos puede ser consecuencia de las mezclas génicas que conforman cada muestra, sobre todo en las dos últimas entradas (Figura 15), lo cual originaría la observación de un exceso de homocigotas en la población global como consecuencia de la subestructuración (Efecto Wahlund). Esto podría atribuirse al sesgo propio de los métodos tradicionales para estudiar estructura poblacional (como los índices de Wright), donde la entidad es la población y no el individuo. Otras posibles causas del exceso de homocigotas global serían los errores de muestreo al coleccionar el material y la deriva génica. En particular, Júpiter podría haber sufrido un estrechamiento de la base genética puesto que se conoce previamente que muchos materiales de origen alemán son el resultado de autofecundaciones de variedades consanguíneas que luego fueron cruzadas entre sí (*Ing. Álvarez, comunicación personal*).

D.2.2. Diferenciación genética entre poblaciones

El valor de F_{ST} promedio hallado en nuestro trabajo (0,413) (Tabla 15) fue notoriamente mayor al reportado por Gross *et al.* (2003) para poblaciones naturales de *Helianthus annuus* (0,159). El mismo patrón se observó al comparar con resultados reportados para poblaciones de otras especies de *Helianthus* (Ellis *et al.*, 2006; Petit *et al.*, 2005; Gross *et al.*, 2003). Ellis *et al.* (2006) reportaron en poblaciones de *H. verticillatus* un valor de F_{ST} de 0,118 y en *H. angustifolius* de 0,207. Por otro lado, Gross *et al.* (2003) y Petit *et al.* (2005) detectaron para 6 poblaciones de *H. petiolaris* un valor de F_{ST} de 0,194 y para 8 poblaciones de *H. deserticola* un valor de 0,249.

El valor máximo de F_{ST} de a pares (0,6394) se registró para la población VNIIMK con el compuesto HAR-1, mientras que el F_{ST} más bajo

obtenido en el presente trabajo fue de 0,1795 entre Colliguay y Júpiter. Este trabajo permitió detectar un elevado nivel de diferenciación genética entre los materiales analizados. Las discrepancias observadas al comparar con trabajos previos pueden ser interpretadas como consecuencia de la intervención del mejorador en cuanto al modo de manejo del banco de semillas, y en menor medida a la diversidad de orígenes geográficos de los materiales analizados. El modo de manejo del banco incluye varios puntos, entre ellos la elección del tamaño práctico de las entradas de la colección, lo cual origina cambios en la distribución de las frecuencias de los alelos presentes. Por otro lado, dado que en la mayor parte de los muestreos para un banco las zonas muestreadas representan sólo una pequeña porción de la distribución, se acentúa la estratificación de los materiales (Sharbel *et al.*, 2000). Asimismo, existe una reducción de la viabilidad de las semillas asociada al almacenamiento *ex situ* de las mismas, además las estrategias de regeneración empleadas contribuyen a cambios en la composición genética de las entradas del banco (Prada, 2009). La marcada diferenciación entre los materiales analizados podría ser consecuencia de un proceso de deriva genética dada por bajo número (N) de individuos en las entradas o por fluctuaciones del N, que generarían cambios significativos en las frecuencias alélicas. La restricción del flujo génico con materiales silvestres es otro factor que puede aumentar la diferenciación poblacional (Sharbel *et al.*, 2000).

Dentro del conjunto de poblaciones y compuestos, se encontró que para 7 pares de materiales, los valores de F_{ST} resultaron no significativos (Tabla 15), por ende no existiría diferenciación genética entre ellos (Colliguay-HAR-3, Colliguay-F. cambá, HAR-2-HAR-3, HAR-3-F. cambá, PxS-F. cambá, VNIIMK-F. cambá, HAR-2-F. cambá). Estas observaciones sugerirían que Colliguay, que deriva de un programa de mejoramiento chileno, tendría un origen similar al de HAR-3 y F. cambá. Por otro lado, F. cambá deriva de Pehuén INTA, material que fue obtenido a partir del cruzamiento de entradas canadienses (Sunrise y 953-102-1-1-22-4) con un grupo de introducciones rusas, entre ellas VNIIMK (Figura 19) (Bertero de Romano and Vázquez, 2003). De todos modos, a nivel global la

diferenciación es alta, lo cual resulta acorde con sus orígenes geográficos diversos.

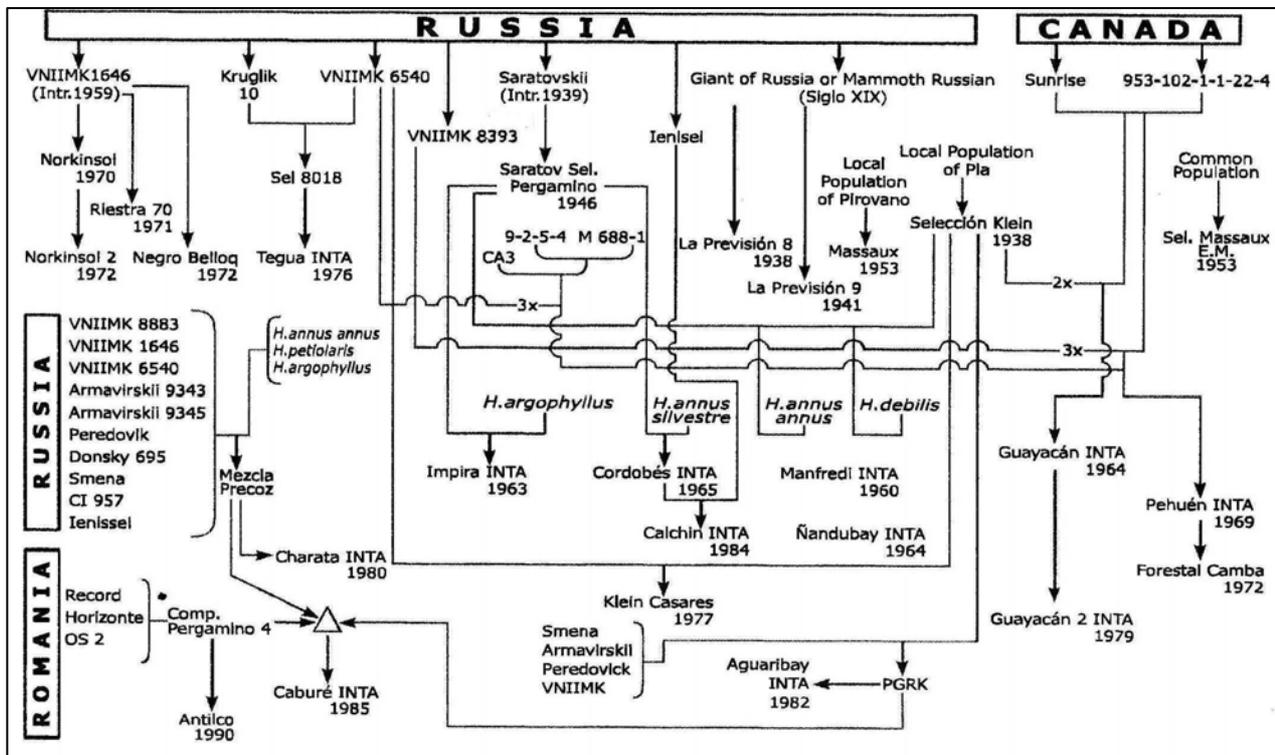


Figura 19. Origen de las variedades argentinas. (Extraído: Bertero de Romano and Vázquez, 2003).

El análisis bayesiano logró discriminar más claramente los grupos que el método de conglomerados (*clusters*) (Figuras 12, 13, 14 y 15). Se detectó una elevada diferenciación entre las poblaciones y los compuestos, coincidente con sus diversos orígenes geográficos y con la información agronómica provista por el mejorador (*Ing. Álvarez, comunicación personal*). La naturaleza intrínseca de los materiales analizados hace que la gráfica de $\ln P(D)$ vs. K no se ajuste exactamente a las predicciones teóricas, puesto que los valores logarítmicos de las probabilidades deberían disminuir posteriores al K más probable. Se observó cierta dificultad para determinar el valor de K , probablemente como consecuencia de los desvíos de la panmixia en la mayoría de las poblaciones y compuestos estudiados (Ehrenreich *et al.*, 2009). No se obtuvo un pico de probabilidad de K , sino una meseta de valores y un incremento de la varianza entre corridas, que fueron

coincidentes con lo reportado por el manual del programa (Pritchard and Wen, 2003) y con trabajos previos (Evanno *et al.*, 2005). Los agrupamientos obtenidos con valores de $K=2$, $K=4$ y $K=10$ (que correspondieron a picos de ΔK , Figura 11-B), mostraron coherencia biológica y representarían niveles anidados de estratificación de los materiales hasta alcanzar la máxima estructuración con sentido biológico en $K=14$ (Figura 11). El patrón de estructuración observado en este trabajo se ajustaría al modelo jerárquico de islas (*hierarchical island model*) (Slatkin and Voelm, 1991) propuesto por Evanno *et al.* (2005). Se detectaron subgrupos adjudicados a subniveles de estructuración, desde un nivel jerárquico superior de partición genética entre las entradas ($K=2$) hasta niveles más inferiores ($K=14$). Esta situación es coincidente con lo reportado por los mencionados autores en situaciones reales, donde lo más frecuente es detectar estratos o capas de estructura poblacional entre los individuos estudiados. Según antecedentes previos (Vigouroux *et al.*, 2008), el método de Evanno favorece a $K=2$ como resultado posible, dado que siempre la diferencia entre las probabilidades de $K=2$ y $K=1$ son las más elevadas. Cabe destacar que el cálculo de ΔK implica una diferencia entre los logaritmos de las probabilidades para $K=2$ y $K=1$, como éste último posee valores sumamente bajos, siempre esa diferencia resulta favorecida. Por esto es que el ΔK entre $K=1$ y $K=2$ fue el valor más alto obtenido en este trabajo. La estructuración en $K=2$ generó dos grupos, coincidentes con la partición de mayor soporte obtenida en el retículo de Neighbor-Joining para compuestos y poblaciones (Figura 16). Este resultado ($K=2$) sería el estrato superior de diferenciación en las entradas analizadas. Asimismo, se encontró consistencia de los grupos obtenidos en $K=4$ con grandes orígenes geográficos, puesto que los acervos génicos amarillo y rojo (Figura 13) incluyeron introducciones originadas a partir de materiales argentinos provenientes de los diversos programas de mejoramiento de INTA (Impira INTA, Cordobés INTA, Manfredi INTA, Guayacán INTA, Charata INTA). El grupo verde reflejó un componente geográfico principalmente americano con entradas de Brasil, Chile y Canadá; mientras que el grupo azul mostró mayor aporte de materiales europeos (rusos y alemanes). Por lo tanto, habría dos situaciones geográficas definidas, genomas similares entre algunos materiales

argentinos y americanos por un lado, y genomas similares entre materiales argentinos y rusos-alemanes por otro lado. Estos resultados concuerdan con Bertero de Romano y Vázquez (2003) que postulan como principales orígenes del material argentino los germoplasmas rusos y canadienses. La identificación de grandes grupos genéticos correlacionados con el origen de los materiales resultó coincidente con lo reportado en la bibliografía para trigo, *Arabidopsis* y maíz, respectivamente (Chao *et al.*, 2010; Ehrenreich *et al.*, 2009; Liu *et al.*, 2003). Para niveles más estratificados ($K=10$ y $K=14$), todos los materiales mostraron la contribución de varios acervos génicos al genoma de sus individuos, algunos con mayor mezcla (Prao-co y Júpiter) que otros, lo cual resulta lógico puesto que no constituyen poblaciones naturales. Los elevados niveles de mezcla de los individuos de Prao-co fueron detectados por todas las metodologías analíticas empleadas (método bayesiano, árbol de Nei, árbol en base a Dps), pero no se dispone de información adicional que pueda dar cuenta de este patrón. El compuesto HAR-4 mostró contribución de dos acervos genéticos, uno coincidente con HAR-1 y el otro con HAR-2, lo cual es concordante con sus orígenes geográficos comunes de acuerdo a los registros del BAG-IM. Por otro lado, Júpiter mostró aporte de dos grupos génicos diferentes de los del resto de los materiales (Figura 15). No fue posible realizar la comparación de los resultados obtenidos con los *pedigrees* de todas las entradas, porque no se dispuso de la información completa sobre este aspecto.

En base a los resultados discutidos en esta sección se infiere que dada las marcadas diferencias genéticas entre la mayor parte de las entradas, el mejorador deberá evaluarlas en el campo puesto que de no ser así, podrían perderse alelos favorables para algún carácter agronómico deseable.

D.3. Análisis de agrupamiento

La información de *pedigree* provee una guía útil para el estudio de la diversidad genética en germoplasma. Sin embargo la selección y deriva genética durante el mejoramiento pueden causar discrepancias entre

pedigree y constitución genética (Liu *et al.*, 2003). Basándose en el registro de *pedigree* podría inferirse si las distancias genéticas halladas en el árbol se relacionan con el mismo, es decir podrían detectarse grupos de germoplasma estrechamente relacionados (Tang and Knapp, 2003). Cuando estos registros están incompletos, algunos materiales no logran ser agrupados, o algunos de los grupos obtenidos pueden ser espúreos (Cox *et al.*, 1985; Sneller, 1994). A partir del árbol obtenido por medio de las distancias genéticas de Nei (1972), la partición que separa al grupo formado por HAR-1, HAR-2, HAR-4 y KLM del resto de las poblaciones y compuestos mostró el soporte de *bootstrap* más alto (81%) (Figura 16). Esto concuerda con los datos de *pedigree* de KLM y los compuestos HAR, dado que todos ellos son de origen argentino y tienen constitución genética similar porque provienen de programas de mejoramiento de girasol que INTA mantiene en sus estaciones experimentales. La población KLM es producto de un cruzamiento múltiple entre el cultivar Klein x un local de Pergamino (*pool* de variedades del programa de mejoramiento de INTA Pergamino como Guayacán INTA, Charata INTA, etc.) x un local de Manfredi (*pool* de variedades del programa de mejoramiento de INTA Manfredi como Impira INTA, Cordobés INTA, Manfredi INTA, etc.). Comangir, PxS y Sunrise forman un grupo diferenciado del resto de los materiales. El grupo formado por Colliguay e IAC está apoyado por un valor de *bootstrap* del 54%. El material HAR-3 se unió a VNIIMK con un soporte de 57% y ellos a Júpiter con valor inferior (Figura 16). Esto concuerda con reportes previos (Paniego *et al.*, 2002), dado que HAR-3 deriva de una variedad rusa (Peredovick 301) y VNIIMK es también de origen ruso, con lo cual podrían compartir alelos, pero en menor medida que con Júpiter cuyo origen es alemán. El material F. cambá es el más diferenciado de todos, dado que acumuló el mayor número de cambios (mayor longitud de ramas) (Figura 16). Las distancias de Nei de esta entrada con respecto al resto son las más altas (Tabla 15). La población Prao-co se unió a la zona intermedia del árbol, con una rama corta, confirmando que este material no está muy diferenciado del resto. Lo mismo se observa en el árbol de individuos, donde los representantes de esta entrada, se intercalan entre individuos de las restantes a lo largo de todo el árbol (Figura 17). La mezcla genética entre sus individuos también se

evidenció en el análisis bayesiano (Figura 15). La única información disponible en cuanto a su *pedigree* es que se formó a partir de híbridos de origen europeo con alto contenido de ácido oleico. Dado que, se desconoce la información de *pedigree* de algunas entradas empleadas en este trabajo, se dificulta la asociación entre los resultados moleculares, la ascendencia y los orígenes geográficos de las mismas.

CAPÍTULO 2

Caracterización fenotípica y genética de líneas de girasol cultivado del programa de mejoramiento para tolerancia a déficit hídrico del Banco Activo de Germoplasma de INTA Manfredi

INTRODUCCIÓN

Programa de mejoramiento para tolerancia a sequía

Una definición agronómicamente aceptada de la sequía es la inadecuada disponibilidad de agua, incluyendo las precipitaciones y la capacidad de almacenamiento de humedad en el suelo, en cantidad y distribución, durante el ciclo de vida del cultivo, que puede disminuir la expresión del máximo potencial genético de rendimiento (Skoric, 2009). El rendimiento del girasol (peso de frutos por unidad de superficie) puede ser dividido en tres componentes: el número de capítulos por unidad de superficie, el número de frutos llenos por capítulo y el peso individual de esos frutos. El número de capítulos por unidad de superficie resulta del número de plantas por unidad de superficie capaces de desarrollar una inflorescencia. Este componente del rendimiento se define principalmente durante la germinación y la emergencia de la planta, ya que las pérdidas posteriores de plantas o capítulos son menos frecuentes y sólo se producen en casos de ataques por patógenos o por vuelco de las plantas (Pedraza *et al.*, 2000). El número de frutos llenos por capítulo y el peso individual de esos frutos son características que dependen principalmente de la información genética de la planta y de las condiciones del ambiente. Los procesos fisiológicos modificados en situaciones de déficit hídrico en girasol producen una reducción del rendimiento a campo por disminución del período de transferencia de asimilados a la semilla (Hall *et al.*, 1985; Hall *et al.*, 1990) y por modificaciones en la composición bioquímica de las semillas, tales como el aumento de la relación entre ácido oleico/linoleico y proteína/aceite (Roche *et al.*, 2004). En cualquier estadio de desarrollo de la planta, un déficit de humedad en el suelo puede tener un impacto negativo sobre el rendimiento del girasol, pero la severidad del efecto de la sequía depende del estadio de desarrollo en el que se encuentren las plantas. Las mayores reducciones del rendimiento se observan cuando la sequía ocurre desde el período de floración hasta la maduración de las semillas. Los efectos de la sequía sobre la calidad de las semillas de girasol son mayores después de la floración (Robel *In*, 1967). La sequía en este estadio causa

una disminución del 7-8% del porcentaje de aceite de la semilla, mientras que en la organogénesis produce una reducción significativa de la viabilidad y funcionalidad del polen (Sitalov and Ikonikov, 1969).

El carácter de resistencia a sequía puede definirse como el o los mecanismos que causan una pérdida mínima de rendimiento en un ambiente de sequía, relativo al máximo rendimiento en un ambiente óptimo para el cultivo (Skoric, 2009). Se han mencionado en la literatura dos mecanismos que muestran los cultivos para minimizar la pérdida de rendimiento por efecto de la sequía, escapar de la misma o evitar la deshidratación. La madurez temprana es un vehículo importante para alcanzar el escape a la sequía, principalmente en ambientes sujetos a estrés hídrico en estaciones tardías (Singh, 2000). Los híbridos tempranos de girasol generalmente tienen índices de área foliar más bajos, menor evapotranspiración y potencial de rendimiento más bajos que los tardíos. El segundo mecanismo es la capacidad de una planta para retener un nivel relativamente alto de hidratación bajo condiciones de déficit hídrico tanto en suelo como atmosférico. El estatus de agua en los tejidos se alcanza por reducción de la transpiración (plantas que economizan agua) o aumento de los canales de agua (plantas consumidoras). Existen varios parámetros empleados para medir el impacto de la sequía en las plantas que emplean estos mecanismos, tales como la reducción de la transpiración, el ajuste osmótico, el contenido de ácido absísico (ABA), la cera de la cutícula y las características de las hojas (Skoric, 2009).

El déficit hídrico produce una importante reducción del poder germinativo en la mayoría de los cultivos que son sensibles a esta condición (Priano and Pilatti, 1989). Por ello, la etapa de implantación, que comprende la siembra, la germinación, la emergencia y la aparición del primer par de hojas verdaderas, es sumamente útil al momento de estudiar los efectos de este tipo de estrés abiótico, dado que la plántula se nutre únicamente de sus reservas seminales. La etapa de germinación y el posterior crecimiento de la plántula son estadios críticos de su historia de vida que a menudo se encuentran sujetos a altas tasas de mortalidad. La germinación y la emergencia de las plántulas deben producirse en forma rápida y uniforme, para poder así, obtener un conjunto de plantas similares en tamaño y con

una distribución por unidad de superficie que permita explorar óptimamente los recursos ambientales (Aguirrezábal *et al.*, 2001). En comparación con la planta adulta, las semillas y plántulas toleran en menor medida el estrés, por lo tanto la habilidad de las semillas para convertirse en una planta joven está fuertemente condicionada por diferentes factores (Bigo, 2005).

La sequía, el estrés por salinidad del suelo y el estrés por altas y bajas temperaturas comparten pasos y patrones fisiológicos de respuesta en la planta. El estudio molecular de plantas sometidas a estos estreses ha demostrado la acumulación de osmoprotectores y de proteínas de choque térmico (*heat shock*), un metabolismo mediado por la acción de fitohormonas como el ABA y los jasmonatos (Vigliocco *et al.*, 2007). El aumento de estas respuestas por manipulación genética resulta en un aumento de la tolerancia al estrés abiótico, pero el funcionamiento de muchos de los genes identificados aún es desconocido en términos de la disponibilidad para el mejoramiento (Takeda and Matsuoka, 2008).

En líneas generales, se ha desarrollado mejoramiento de girasol orientado a resistencia a sequía en distintas fuentes de germoplasma incluyendo líneas, variedades e híbridos cultivados, especies silvestres del género *Helianthus* y poblaciones mutageneizadas (Skoric, 2009). El uso de los híbridos *stay-green* en los programas de mejoramiento para resistencia a sequía permitió obtener una mejor producción en girasoles comerciales (Vranceanu, 2000). Aumentos más importantes en la tolerancia a sequía para girasol cultivado se han logrado usando especies silvestres principalmente *H. argophyllus*, y en el futuro se espera incrementarla aún más incluyendo otras especies silvestres como *H. deserticola*, *H. hirsutus*, *H. maximiliani* y *H. tuberosus*, entre otros (Skoric, 2009).

El objetivo del programa de mejoramiento para déficit hídrico del BAG-IM es seleccionar materiales con tolerancia frente a este carácter combinando la capacidad de adaptación a varios ambientes. Por ello, los ensayos de evaluación de genotipos se llevan adelante en varias localidades de diferentes provincias. Los materiales sobre los cuales se centra el programa son líneas endocriadas. Éstas tienen la ventaja de que su genoma alcanzó la homocigosis práctica y resulta más fácil poder manejarlas para

selección de los genes deseados y posterior cruzamiento con otras entradas de mayor rendimiento (*Ing. Álvarez, comunicación personal*). El programa incluye alrededor de 100 líneas que han sido pre-seleccionadas por su comportamiento frente al déficit hídrico, y cada año se evalúan nuevos materiales para ser empleados activamente en el programa. La evaluación fenotípica consiste en ensayos de campo tanto en condición de déficit hídrico como de irrigación. Los caracteres que se miden son la altura de las plantas, el peso de cien semillas, el número de semillas por capítulo, el rendimiento de las mismas en Kg/ha y el porcentaje de materia grasa. Hasta el momento, existe un grupo de 30 líneas pertenecientes al programa que han sido caracterizadas en ensayo de campo, y de ellas 7 fueron evaluadas también en estudios de respuesta frente al déficit hídrico inducido por manitol *in vitro* (Alemano *et al.*, 2005). Dado que, ninguna de estas líneas ha sido caracterizada molecularmente, resulta importante conocer el nivel de diversidad genética presente en ellas y explorar si existe alguna asociación entre los genotipos y los fenotipos observados.

El mapeo de asociación como herramienta para el mapeo de QTL en caracteres complejos

La búsqueda de asociaciones entre genotipo y fenotipo a escala poblacional o “mapeo por asociación” es un enfoque muy poderoso y relativamente reciente para identificar genes o alelos que contribuyan a la variación en caracteres adaptativos o agronómicamente interesantes (Long and Langley, 1999). A diferencia de las estrategias convencionales, el mapeo por asociación emplea conjuntos de individuos que pueden o no estar relacionados, como por ejemplo, poblaciones naturales, colecciones de germoplasma, líneas o variedades *elite* de un programa de mejoramiento. Esto permite emplear de forma más eficiente y económica los recursos genéticos, ya que varios caracteres fenotípicos pueden ser estudiados en un único conjunto de individuos utilizando la misma información genotípica. El mapeo de asociación consiste en utilizar eventos de recombinación históricos a lo largo de un linaje, y no solamente aquellos ocurridos en una

determinada población de mapeo, para establecer posibles relaciones entre genotipo y fenotipo (Oraguzie *et al.*, 2007).

La existencia de subestructuración poblacional es la más común de las fuentes sistemáticas de desvío en estudios de asociación (Buckler and Thornsberry, 2002; Hirschhorn and Daly, 2005). Para detectar la estructuración y otros procesos demográficos que podrían causar falsos positivos en este tipo de estudios, se emplean marcadores moleculares considerados potencialmente neutros, tales como los SSR nucleares, en forma concomitante con los marcadores funcionales para los que se pretende establecer asociaciones con el carácter de interés (Rosenberg *et al.*, 2002). Mediante el empleo de marcadores neutros se obtienen medidas de diversidad generales que permiten hacer inferencias en cuanto a los eventos demográficos que han sufrido las entradas en análisis. Los estudios de asociación requieren del conocimiento de los patrones de diversidad nucleotídica y de desequilibrio de ligamiento (DL) para cada especie en particular y para cada gen candidato. El DL determina la densidad de marcadores a emplear y el diseño experimental que se implementará, dado que refleja la historia poblacional de recombinación. Para especies de autofecundación como el girasol, el DL suele tener gran alcance, lo cual disminuye el número de marcadores a evaluar en la estrategia de asociación de genoma completo. Sumado a ello, la selección artificial del cultivo y los factores geográficos pueden originar estructuración poblacional compleja, aumentando el riesgo de falsas asociaciones (Takeda and Matsuoka, 2008). La asociación no aleatoria entre un polimorfismo dado y la variación fenotípica para cierto rasgo de interés, se obtiene a través de correlaciones estadísticas, es decir que se logra identificar el desequilibrio de ligamiento (DL) entre las variantes genéticas causales de la variación del carácter y alguno/s del o de los polimorfismo/s analizado/s. Los alelos altamente penetrantes con frecuencias intermedias tienen mayor probabilidad de ser detectados por mapeo por asociación, mientras que los alelos en frecuencias más bajas ($\leq 0,05$) y con modesta contribución al fenotipo son prácticamente imposibles de detectar (Risch, 2000).

El alcance del DL para girasol resulta similar a los valores reportados para especies vegetales típicamente autógamas (Fusari, 2010). En especies

silvestres el alcance del DL se extiende tan solo hasta las 200 pb (Liu and Burke, 2006); mientras que para materiales cultivados como líneas endocriadas este valor es superior (0,32 a las 5500 pb) (Kolkman *et al.*, 2007) como consecuencia de una historia sucesiva de autofecundaciones (Fusari, 2010; Kolkman *et al.*, 2007).

El desarrollo de un estudio de mapeo por asociación involucra la evaluación fenotípica del carácter agronómico de interés (por ej.: peso del grano, contenido de materia grasa, altura de la planta, días a floración, etc.) y la genotipificación de un conjunto apropiado de marcadores para cada una de las entidades a comparar (líneas endocriadas, variedades, etc.). El número y naturaleza de los polimorfismos a caracterizar dependerá de la estrategia de análisis que se haya seleccionado.

Debido a su gran abundancia y ubicuidad en el genoma, los marcadores moleculares más frecuentemente utilizados para el mapeo por asociación son los SNP. Un SNP se define como una diferencia a nivel de base nucleotídica entre dos secuencias de ADN. Los casos de inserciones o deleciones (InDel) no se consideran formalmente SNP, pero a los fines de evaluación del polimorfismo de un material genético, son sumamente importantes. De acuerdo con la teoría, las variantes de tipo SNP podrían involucrar las cuatro bases nucleotídicas (A, T, G o C), pero en general sólo se observan dos posibilidades para un determinado locus. Por ello, este tipo de marcadores se consideran bialélicos (Brookes, 1999). Esta característica podría ser considerada una desventaja con respecto a los marcadores SSR que son polialélicos, pero como ventaja frente a éstos se observa que los loci SNP son más abundantes y estables (Oraguzie *et al.*, 2007).

Los marcadores SNP han sido empleados con éxito para diversas aplicaciones como el mapeo genético, el clonado posicional, la detección de caracteres asociados a genes a través del mapeo por asociación y la determinación de relaciones genéticas entre individuos (Syvanen, 2001). En los últimos años, se desarrollaron numerosos proyectos de descubrimiento de SNP a gran escala en especies cultivadas con el fin de realizar análisis de asociación que permitan el mejoramiento genético de los cultivos (Feltus *et al.*, 2006; Giancola *et al.*, 2006; Monna *et al.*, 2006; Rostoks *et al.*, 2005; Schmid *et al.*, 2003).

Existen dos aproximaciones para desarrollar un estudio de mapeo por asociación, una a través de la genotipificación de marcadores SNP distribuidos a lo largo de todo el genoma (GWA: *genome-wide association*) y la otra por medio del estudio de los polimorfismos presentes en genes candidatos. Ambos enfoques requieren de la identificación de la estructura poblacional y de las relaciones de parentesco entre los individuos de la población de mapeo, las cuales son determinadas a través del empleo de marcadores neutros. La aproximación GWA involucra una exploración exhaustiva del genoma para la localización de variación genética causativa. Se analizan un número elevado de marcadores uniformemente distribuidos en el genoma (10.000 a 100.000), para encontrar asociaciones con varios caracteres complejos, sin requerir ningún tipo de información previa acerca del control genético del carácter. La estrategia de genes candidatos se basa en la selección de genes potencialmente involucrados en el control de la variación fenotípica del carácter en estudio. Por lo tanto, en esta estrategia, el costo de genotipificación se reduce sustancialmente respecto de GWA. De este modo, la estrategia de genes candidatos es la elegida por la mayor parte de los grupos de investigación académicos (Buckler and Thornsberry, 2002).

Para realizar un estudio de mapeo por asociación se requiere cumplimentar con precisión cinco aspectos relacionados: 1) la selección del germoplasma que constituirá la población de mapeo por asociación (PMA), 2) la fenotipificación de la PMA para el carácter a evaluar, 3) la genotipificación de la PMA a través de GWA o genes candidatos, 4) la comprobación de la existencia (o no) de estructura poblacional en la muestra y, 5) el análisis estadístico y los modelos de comparación de los genotipos y fenotipos para detectar asociaciones (Ersoz *et al.*, 2007). La correcta selección del germoplasma para conformar la PMA es crítica para asegurar el éxito del mapeo por asociación (Zhu *et al.*, 2008). Lo ideal es trabajar con individuos no emparentados, pero como esto en la mayoría de los casos no es factible, la población experimental debe ser lo más representativa posible de la población para la cual se realizará la inferencia de asociación (Ersoz *et al.*, 2007). Las PMA pueden clasificarse en colecciones núcleo de los bancos de germoplasma, poblaciones sintéticas o líneas *elite* representativas del

germoplasma utilizado en los programas de mejoramiento (Breseghello and Sorrells, 2006). Las colecciones núcleo están conformadas por un número limitado de entradas que representan la variabilidad genética presente en la especie en estudio, resultan interesantes para la asociación de caracteres a través de la estrategia de genes candidatos por la corta extensión del DL, pero es probable que algunas entradas no estén adaptadas a las condiciones de crecimiento en el campo ni a las enfermedades más comunes del cultivo, lo cual repercutirá en una medida fenotípica de baja calidad (Breseghello and Sorrells, 2006). Las poblaciones sintéticas constituirían la PMA ideal debido a la existencia de panmixia, lo cual minimiza los riesgos de encontrar falsos positivos. Sin embargo, si la población se encuentra sometida a selección de forma continua, los análisis de asociación deben contemplarse con la genotipificación de los individuos en sucesivas generaciones para poder reflejar correctamente la constitución genética de la población en estudio, siendo ésta su principal desventaja de utilización. Las líneas *elite* son estables y están adaptadas a las condiciones de crecimiento normales del cultivo (Breseghello and Sorrells, 2006), pero muestran patrones de estructura poblacional, lo cual implica que se debe incorporar la corrección de este efecto.

Identificación y caracterización de genes candidatos

Para desarrollar una estrategia de mapeo por asociación bajo el enfoque de genes candidatos, resulta necesario determinar cuáles son los genes o regiones que por su función podrían estar relacionadas con el carácter de interés y conocer los polimorfismos contenidos en ellas.

En base a estudios de expresión en *Arabidopsis thaliana*, arroz y maíz bajo condiciones de sequía se conocen dos categorías de genes candidatos involucrados en la respuesta, los que contribuyen a la protección de membranas y proteínas (chaperonas, proteínas de choque térmico, proteínas abundantes en la embriogénesis tardía y osmoprotectores), y por otro lado los genes que controlan la expresión y las señales de transducción (Bray 2002; Seki *et al.*, 2003; Wang *et al.*, 2003). Diversos genes candidatos han sido identificados en respuesta a déficit hídrico en girasol cultivado,

entre ellos genes que codifican proteínas quinasas (Roche *et al.*, 2007), canales de agua (*aquaporins*) y fosfatasas involucradas en respuesta al estrés oxidativo (Poormahammad Kiani *et al.*, 2007), factores de transcripción (Ariel *et al.*, 2007; Dezar *et al.*, 2005; Gago *et al.*, 2002; Poormahammad Kiani *et al.*, 2007) y dehidrinas (Cellier *et al.*, 1998; Cellier *et al.*, 2000).

En ausencia de conocimiento previo sobre la base genética del carácter, la identificación de genes candidatos también ha sido abordada desde el marco teórico provisto por la genética de poblaciones. En este sentido, los patrones de polimorfismos de nucleótido simple, son informativos en cuanto a los procesos evolutivos que han sufrido las distintas regiones del genoma en las entradas que constituyen la PMA. Por ende, las regiones genómicas que evidencien patrones de sustitución nucleotídica diferenciales, serán potencialmente interesantes para la búsqueda de resistencia frente a estreses bióticos y abióticos.

Las medidas de variabilidad genética más frecuentemente utilizadas para el estudio de secuencias son la diversidad nucleotídica y el número de sitios segregantes. La diversidad nucleotídica (π) (Tajima, 1983) es medida como el número promedio de diferencias por sitio entre pares de secuencias de una muestra. El número de sitios segregantes (S) o también llamado θ_W (Watterson, 1975) se basa en el modelo de mutación de sitios infinitos, es decir donde cada evento mutacional ocurre en un sitio nuevo introduciendo un alelo alternativo al sitio. Se calcula tomando el número de sitios que varían en un total de secuencias analizadas, por lo tanto cada mutación se considera como un sitio diferente. El número esperado de sitios segregantes es el número esperado de mutaciones en base a la teoría de coalescencia (Felsenstein, 2009).

Existen pruebas estadísticas que permiten comprobar la presencia de selección o la ocurrencia de procesos demográficos como cuellos de botella o expansiones poblacionales, teniendo en cuenta los patrones de polimorfismo observados (Nordborg, 2002). La prueba más difundida es la de Tajima (prueba D) (Tajima, 1989) que compara el número de polimorfismos nucleotídicos con la diferencia media de a pares entre secuencias (π y θ_W). En condiciones de neutralidad selectiva, el valor de D

será nulo puesto que π y θ_W son iguales. El estimador θ_W toma en cuenta únicamente el número de sitios segregantes, independientemente de su frecuencia, mientras que π es influenciado por las frecuencias alélicas de cada uno de los sitios.

Los procesos demográficos tales como cuellos de botella, expansiones poblacionales o apareamiento no aleatorio (por ej.: estructuración poblacional) pueden generar valores significativos del estadístico D, aún cuando no está operando la selección natural sobre el locus considerado. Los valores positivos de D indican desvíos de la neutralidad hacia un exceso de variantes en frecuencias intermedias consecuencia de selección equilibradora, estructuración poblacional o cuello de botella moderado. Mientras que, los valores negativos de D, indican desvíos de la neutralidad hacia un exceso de variantes en frecuencias bajas a causa de barrido selectivo reciente, expansión poblacional o cuello de botella severo. La prueba D de Tajima es simple y robusta, además requiere sólo una población individual (Felsenstein, 2009).

Otras pruebas disponibles son los estadísticos D, D*, F y F* de Fu y Li (1993) o su desarrollo más avanzado, la prueba F_S (Fu, 1997). Las primeras comparan el número de variantes nucleotídicas derivadas observadas sólo en una muestra contra el número de variantes totales (D y D*), o la media de diferencias de a pares entre secuencias (F y F*) (Prada, 2009). Estas pruebas también están basadas en la distribución de frecuencias de los haplotipos según el valor de θ estimado por π y, al igual que la de Tajima, son capaces de reflejar cambios demográficos sufridos por una población. La prueba F_S compara los mismos estadísticos que las anteriores y tiende a ser negativa cuando se observa un exceso de mutaciones recientes que generan alelos raros, por lo tanto un valor alto y significativamente negativo de este estadístico puede tomarse como evidencia contra la neutralidad de las mutaciones. De todos modos, los alelos raros podrían ser también el resultado de una expansión poblacional reciente.

Para todas las pruebas descritas, el desarrollo se basa en la teoría neutralista lo cual implica contar con una población panmíctica, es por ello que la interpretación de resultados en conjuntos de individuos que no

conforman poblaciones naturales (por ej. conjuntos de líneas de un programa de mejoramiento) debe ser realizada con cautela.

Diversos estudios han logrado detectar genes candidatos bajo procesos selectivos en poblaciones naturales a través de desvíos en las distribuciones de alelos con respecto a lo esperado bajo la teoría neutralista (Ford, 2002; Kreitman, 2000; Rosenberg and Nordborg, 2002). Por lo tanto, la mayoría de los trabajos suelen incorporar alguna de las pruebas de neutralidad descritas en los párrafos anteriores, aplicadas sobre genes candidatos o sitios de aminoácidos que se suponen bajo selección. Cuando se comprueba que los desvíos al modelo neutral no son debido a estructuración poblacional o a procesos demográficos, pueden adjudicarse a selección balanceadora y selección purificadora o barrido selectivo causado por selección positiva (González-Martínez *et al.*, 2006).

Otra estrategia para identificar genes importantes en el proceso de domesticación y mejoramiento consiste en comparar la pérdida de diversidad nucleotídica entre las especies silvestres, razas nativas y variedades o líneas cultivadas. Se espera que los genes sujetos a selección, durante o después de la domesticación, muestren una disminución localizada más severa de la diversidad nucleotídica comparados con genes sin presión de selección (Wright and Gaut, 2005).

En resumen, el programa de mejoramiento del BAG-IM cuenta con 30 líneas de girasol evaluadas en ensayos de campo para el carácter de tolerancia a sequía en estadio de planta adulta. El desarrollo de un futuro estudio de mapeo por asociación para este carácter requiere como paso previo contar con una evaluación fenotípica robusta, disponer de un conjunto de genes candidatos y conocer sus patrones de variación nucleotídica. El presente Capítulo tiene como objetivos: 1) evaluar la tolerancia a sequía en plántulas bajo condiciones de déficit hídrico controlado en las 30 líneas anteriormente mencionadas, 2) comparar los resultados obtenidos con las evaluaciones realizadas en ensayos de campo en planta adulta, y 3) estudiar la diversidad nucleotídica en 7 regiones genómicas candidatas que han sido vinculadas con la respuesta frente al déficit hídrico en girasol.

MATERIALES Y MÉTODOS

M1. Caracterización fenotípica para el carácter de tolerancia a sequía

M1.1. Material vegetal

La caracterización fenotípica para el carácter de tolerancia a sequía en ensayo de campo durante la campaña 2003-2004, se basó en la evaluación de 30 líneas endocriadas pertenecientes al BAG-IM (Tabla 1). Para la caracterización fenotípica del carácter mediante un ensayo en condiciones controladas, se sumaron 4 líneas a las inicialmente descritas (Tablas 1 y 2). Las líneas R432, B71, B59 y A59 son parentales de poblaciones recombinantes empleadas en el análisis de QTL para el carácter estudiado.

Tabla 1. Líneas endocriadas del BAG-IM caracterizadas fenotípicamente en ensayo de campo y en ensayo bajo condiciones controladas. S: silvestre. USA: Estados Unidos. Arg: Argentina. Yugosl: Yugoslavia. s/d: sin dato (corresponde a un híbrido provisto por la firma Advanta, único material que no fue evaluado en condiciones controladas ni genotificado).

Línea	Orígen	Línea	Orígen	Línea	Orígen
R426-1	S x USA	R437	USA x Arg	R435	S x Arg
R307	USA x Arg	B99	USA x Arg	R422	S x USA x Arg
R417	S x Francia x USA	R073	USA	RHA274	USA
R419	S x USA	HAR4	Arg	R423	S x USA x Arg
B59	Arg x Rusia	R420	S x USA	R049	USA x Arg
B67	Yugosl x Arg	R415	S x Francia x USA	HA89	USA
R023	USA x Arg	R043	USA	R431	S x USA x Arg
R421	S x USA x Arg	R427	S x USA	R433	S x USA x Arg
B10	Arg x Rusia	B71	USA x USA	R429	S x USA x Arg
R413	S x USA	R432	S x USA x Arg	AGUARA	s/d

El “orígen” representa la constitución global a nivel genómico de cada línea y servirá para interpretar los patrones obtenidos al momento de analizar la estructura poblacional de los materiales (ver sección M.2.2).

Tabla 2. Líneas endocriadas del BAG-IM caracterizadas fenotípicamente en ensayo bajo condiciones controladas. Estas líneas fueron agregadas a las descritas en la Tabla 1.

Línea	Origen
430-C	Arg
962-C	S x USA
964-C	S x USA
1063-C	USA
A59	Arg x Rusia

Como se puede observar en las Tablas 1 y 2, 17 de las líneas evaluadas fueron seleccionadas por el aporte a su genoma de materiales silvestres que podrían conferir tolerancia al carácter en estudio (*Ing. Álvarez, comunicación personal*).

M1.2. Análisis de datos del ensayo de campo (campaña 2003-2004): evaluación de planta adulta

Se realizó el análisis de los datos obtenidos en la caracterización fenotípica efectuada sobre un conjunto de líneas del BAG-IM sometidas a condiciones de déficit hídrico en el campo durante la campaña 2003-2004. Este análisis incluyó 30 líneas puras endocriadas (Tabla 1), las cuales fueron caracterizadas para las variables altura de la planta (ALT), peso de cien semillas (P 100), número de semillas por capítulo (NSC), contenido de materia grasa en el grano (MG) y rendimiento de semillas en Kg. por hectárea a 11% de humedad (RENSH). Los materiales empleados corresponden a líneas A, B y R (ver Introducción del Capítulo 1) y fueron evaluadas en el campo con un diseño de ensayo en bloques con dos tratamientos (controles irrigados y condiciones de déficit hídrico) que incluyó dos repeticiones. El déficit hídrico se impuso mediante el empleo de una cubierta plástica sobre la superficie del suelo (Figura 1).

Los datos del ensayo de campo fueron gentilmente cedidos por el Ing. Álvarez (Grupo de Mejoramiento Vegetal de Girasol, INTA EEA Manfredi) y parcialmente publicados en Alemano *et al.* (2005).



Figura 1. Imágen tomada en el ensayo de campo con líneas de girasol sometidas a déficit hídrico (INTA EEA Manfredi, campaña 2003-2004).

M1.3. Ensayo en condiciones controladas: evaluación de plántula

Para este ensayo se analizaron 34 líneas endocriadas pertenecientes al programa de mejoramiento de girasol del BAG-IM (Tablas 1 y 2). La evaluación fenotípica en condiciones controladas consistió en un ensayo de déficit hídrico en germinación, utilizando riego con una solución de manitol en diferentes concentraciones (200mM y 400mM, estrés moderado y severo, respectivamente) y un control, en base al protocolo descrito por Bigo (2005). El diseño experimental consistió en tres bloques completos en tres fechas de siembra independientes. Se colocó un papel absorbente en el fondo de cada bandeja (16 cm x 11,5 cm x 4 cm) previamente perforada para favorecer el escurrimiento, y se adicionaron 700 g de arena esterilizada a 170°C 5 h como indica el método descrito por la Asociación Internacional de Ensayos de Semillas (ISTA, 1976; 1985; 2003). La arena fue hidratada al 60% de su capacidad de campo (CC) con agua bidestilada o solución de manitol según el tratamiento (protocolo de humectación descrito en el Apéndice). Se sembraron 25 semillas (valor mínimo requerido por las normas ISTA) por línea y por bandeja a una profundidad equivalente a tres veces su diámetro para asegurar buen crecimiento y homogeneidad de

altura en las plántulas. Las semillas fueron previamente desinfectadas en una solución de lavandina al 4% durante 3 min y luego se efectuaron 10 enjuagues con agua bidestilada. Se realizaron riegos por capilaridad (100 % de CC) a los 4, 7 y 10 días posteriores a la siembra, con agua para el caso de las bandejas control y con la respectiva solución de manitol para cada uno de los tratamientos de estrés. La incubación se realizó en una cámara de germinación (acreditación RNCyFS-INASE I/708), a temperatura y fotoperíodo constantes (25°C, 16 h de luz y 8 de h de oscuridad).

Evaluación fenotípica de las plántulas

Al momento de realizar el riego del 4to día, se midió el parámetro Energía Germinativa (EG) que fue expresado como porcentaje, teniendo en cuenta sólo las semillas germinadas con altura de plántula de 0,5 cm o mayores.

La cosecha de las plántulas se realizó al día 11 de iniciado el ensayo y se efectuaron las siguientes mediciones: Peso Fresco (PF), Recuento Final de Germinación (RFG) y Peso Seco (PS). Para la evaluación del RFG, se consideraron solamente las plántulas normales, es decir aquellas que ISTA define como plántulas que presentan capacidad para continuar su desarrollo en una planta normal, descartándose las plántulas que presentaron distintas anomalías dado que no presentan valor para la siembra (ISTA, 1976; 1985; 2003). El PS fue medido después de someter las plántulas a 70°C por 48 h.

Análisis estadístico

El análisis estadístico se realizó empleando un modelo de regresión logística, tomando como variable dependiente el RFG y calculando como frecuencia el número total de semillas sembradas por bandeja (25) y como variables clasificadoras el genotipo, el tratamiento y la repetición. El análisis fue llevado a cabo con el programa InfoStat/Profesional v.2008 (Di Rienzo *et al.*, 2008).

El modelo planteado consideró a la variable RFG como categórica, en donde el evento de éxito implicó la germinación de la semilla originando una

plántula normal capaz de generar una planta adulta. El caso opuesto fue el fracaso, considerado cuando la semilla no germinó o la plántula obtenida resultó inviable o anómala. En base a ello, se ajustó un modelo de regresión logístico según el cual los valores predichos (o predictivos) constituyeron estimas de probabilidad de éxito, es decir de que una semilla germine y alcance el estadio de plántula normal, tanto en condiciones de irrigación como de déficit hídrico.

M2. Caracterización genética de líneas pertenecientes al programa de mejoramiento para tolerancia a sequía

La caracterización genética se llevó a cabo mediante el análisis de loci SSR y de un conjunto de regiones candidatas para tolerancia a déficit hídrico. El primer análisis incluyó las 34 líneas descritas en las Tablas 1 y 2. En el caso de las regiones candidatas se agregaron, además, las líneas RHA266 y PAC2 por ser los genotipos parentales del mapa genético de referencia de girasol cultivado (Poormohammad Kiani *et al.*, 2007).

M2.1. Extracción y cuantificación de ADN

La extracción y cuantificación de ADN para las 36 líneas endocriadas se realizó del modo descrito en el Capítulo 1 (ver sección Materiales y Métodos).

M2.2. Análisis de variabilidad para marcadores microsatélites

Como se mencionó anteriormente, la caracterización con loci SSR incluyó las 34 líneas descritas en las Tablas 1 y 2. Se seleccionaron 7 microsatélites nucleares (HA77, HA293, HA928, HA2063, HA2920, HA3239 y HA4103) independientes para estimar los niveles de diversidad en el conjunto de líneas evaluadas para tolerancia frente al déficit hídrico. Este análisis se llevó a cabo en 3 individuos de cada una de las líneas estudiadas, excepto para R073, R423, R437, R413, R023, R043 y R421 en las que se analizaron 2 individuos. Los individuos fueron tratados como se

describió en el Capítulo 1 para la extracción, cuantificación y amplificación del ADN con los respectivos iniciadores. Para evaluar la homogeneidad dentro de cada línea, se empleó el criterio propuesto por Heckenberger *et al.* (2002) descrito en Materiales y Métodos del Capítulo 1. Asimismo, el análisis de diversidad de las líneas se llevó a cabo de modo similar a lo descrito en Materiales y Métodos del Capítulo 1, empleando el índice de similitud de Jaccard y el agrupamiento UPGMA implementados en el programa Infogen (Balzarini *et al.*, 2005).

El análisis de la estructura poblacional se realizó utilizando el método bayesiano implementado en el programa Structure 2.2 (Pritchard *et al.*, 2000). El número de grupos evaluados osciló entre 1 y 10. El análisis se realizó utilizando 5 réplicas por cada valor K con un largo del período de *burn-in* de 100.000 iteraciones y un largo de corrida de 500.000 iteraciones. No se utilizó ninguna información adicional para definir los grupos, se empleó el modelo de frecuencias correlacionadas (Falush *et al.*, 2003). La corrida que mostró la mayor probabilidad de los datos fue elegida como resultado final del valor de K.

M2.3. Selección de regiones candidatas para tolerancia a déficit hídrico

Se seleccionaron 15 regiones asociadas con la respuesta de girasol al déficit hídrico para la identificación de SNP e InDel y para el análisis de la variabilidad nucleotídica y del DL en las 36 accesiones analizadas. Estas regiones fueron elegidas en base a los antecedentes descritos en estudios moleculares, fisiológicos y genético-poblacionales relacionados con el carácter en estudio (Tabla 3). De las 15 regiones, 7 generaron fragmentos a partir de los oligonucleótidos empleados cuyos electroferogramas resultaron adecuados para decidir su inclusión en los estudios de diversidad desarrollados en esta sección.

Dos de las 7 regiones (Hahb4-p y Hahb4) fueron seleccionadas a partir de los trabajos de Dezar *et al.* (2005) y Gago *et al.* (2002). Estos autores caracterizaron al ADNc completo (674 pb) que codifica una proteína con homeo-dominio de cierre de leucinas en girasol conocida como Hahb4,

que deriva de *Helianthus annuus* homeobox-4. La estructura de cierre de leucinas (HD-Zip: *homeodomain-leucine zipper*) está involucrada en la dimerización, por lo tanto estas proteínas desempeñan un rol importante en la regulación de procesos de desarrollo asociados con respuesta de plantas a condiciones ambientales (Manavella *et al.*, 2008). Hahb4 es un gen de girasol que codifica un factor de transcripción de la subfamilia I de las proteínas HD-Zip y su expresión está regulada por ABA, sequía, estrés salino y etileno (Gago *et al.*, 2002; Manavella *et al.*, 2006); sugiriendo que podría intervenir en la cascada de señalización que controla las respuestas mediadas por ABA en déficit hídrico.

Tres regiones candidatas, *Suntip7* (*Suntip*), Fructosa-1,6-bifosfatasa (FB) y *Leafy-cotyledon-like-1protein* (HaL1L), fueron seleccionadas a partir del trabajo de Poormahammad Kiani *et al.* (2007). La primera de ellas es una proteína que funciona como canal de agua (*aquaporin*), que se expresa en células guarda de girasol durante la fase de cierre estomático y está directamente involucrada en el estatus de agua en la planta (Sarda *et al.*, 1997; Sarda *et al.*, 1999). En estudios posteriores de detección de QTL se propone que esta proteína reduciría la permeabilidad de las membranas y permitiría la conservación de agua celular durante los períodos de déficit hídrico (Poormahammad Kiani *et al.*, 2007). La fosfatasa FB es una enzima clave en el ciclo de reducción del carbono fotosintético y es sensible al peróxido de hidrógeno producido bajo estrés (Charles and Halliwell, 1981), lo cual conduce a la disminución de la actividad fotosintética (Zrenner *et al.*, 1996). HaL1L es un regulador de proteínas abundantes en la embriogénesis tardía (LEA: *late-embryogenesis-abundant proteins*) (Fambrini *et al.*, 2006) involucradas en la tolerancia a estrés salino y a sequía (Moons *et al.*, 1995, Parcy *et al.*, 1997).

En base a los trabajos de Cellier *et al.* (1998, 2000) se seleccionó una región conocida como dehidrina1 (*dehydrin1*, HaDhn1), cuyos patrones de expresión fueron descritos por estos autores. Cellier *et al.* (1998, 2000) detectaron acumulación de transcritos de dehidrinas en hojas, asociándolo con mecanismos de tolerancia como el mantenimiento de la turgencia celular para prevenir la deshidratación.

Por último, se incluyó una proteína caseína quinasa-II (CK) tomada del trabajo de Roche *et al.* (2007). Estos autores detectaron niveles de expresión elevados de este gen en un microarreglo de ADN copia para un genotipo tolerante a sequía evaluado en condiciones de déficit hídrico en el campo, comparado con un genotipo susceptible. La expresión de dichos transcriptos fue cuatro veces mayor para el genotipo tolerante que para el susceptible, asimismo la expresión en hojas fue superior a la detectada en embriones.

Tabla 3. Descripción del total de regiones candidatas ensayadas. a. Número de accesos del GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>) o del DFCI-TC: consenso tentativo del índice de genes del DFCI (<http://compbio.dfci.harvard.edu/cgi-bin/tgi/gimain.pl?gudb=sunflower>). b. Longitud del fragmento a amplificar por PCR en pares de bases (pb). Las regiones y combinaciones de iniciadores destacadas en fondo gris corresponden a las empleadas en el análisis de diversidad de este capítulo.

ID de la región	Región candidata	Nº acceso ^a	ID iniciador	Secuencia del iniciador	Largo ^b (pb)	Referencias bibliográficas
Hahb4-p	<i>Transcriptional factor promoter</i>	AF339749.1	Hahb4-p	F: 5' ACCACCTGGCACATCGTATCTT 3'	712	Dezar <i>et al.</i> (2005);
			Hahb4-p	R: 5' ATGATGGGAGATTAGGCCGA 3'		
Hahb4	<i>Homeo-box transcriptional factor</i>	AF339748.1	Hahb4	F: 5' TTCGCCAATGTCTCTGAACCGAAG 3'	1600	Gago <i>et al.</i> (2002)
			Hahb4	R: 5' TGAGGGTCTGGCAGTTGTTCTTC 3'		
Suntip	<i>Suntip (aquaporin)</i>	X95950	Suntip1	F: 5' CCACCTTGCTCTTCGTCTTC 3'	714	Poormahammad Kiani <i>et al.</i> (2007)
			Suntip2	R: 5' GCCTGGAATCATACAATCCAA 3'		
			Suntip3	F: 5' CCATCAAGGCCTACATTGCT 3'	744	
Aq1	<i>Aquaporins1</i>	TC44044	aq1	F: 5' GGTGAGCTCTCCTCATGGTC 3'	886	
			aq2	R: 5' AATCCACCACCAACTCAAGG 3'		
			aq3	F: 5' AGCCAAAGCCTATCCTCACA 3'	889	
			aq4	R: 5' GAATGGCCCTGATGACAATC 3'		
Aq2	<i>Aquaporins2</i>	TC40501	aq5	F: 5' CTTTTGGTGCCTTTGTTGGT 3'	890	Peluffo L. (2010)
			aq6	R: 5' TGTTCCACCTTTTCACACA 3'		
			aq7	F: 5' TGCACTTCTTTGGCATCAG 3'	899	
			aq8	R: 5' GTCCAGCCCAGTAGATCCAA 3'		
Aq2'	<i>Aquaporins2'</i>	TC42670	aq9	F: 5' GCCGCTACCAAGAAGAGGTA 3'	715	
			aq10	R: 5' AACGACTCATGAGAGCGGTTA 3'		
			aq11	R: 5' GAGCGGTTAATGAACACAAGC 3'	703	
FB	<i>Fructose-1,6- biphosphatase</i>	CD850867	FB1	F: 5' CACCTGCTGTTCCAGGAATTA 3'	449	Poormahammad Kiani <i>et al.</i> (2007)
			FB2	R: 5' GTTCGTTTTTGGGGTTGCTA 3'		
			FB3	F: 5' CAGGAGGACAATCTGAAATGG 3'	419	
HaL1L	<i>Leafy-cotiledon-like 1 protein</i>	TC40698	LCL1	F: 5' GTTCCATGGCTACCACAGG 3'	733	Poormahammad Kiani <i>et al.</i>
			LCL2	R: 5' GCCAAAGTTTGACCCAATGT 3'		
			LCL3	F: 5' TACATCTGGGATGCAACAGC 3'	701	

VCD	VCD2 – <i>LytB like gene</i>	TC56288	LCL4	R: 5' TTGCCAAAGTTTGACCCAAT 3'	561	(2007) Liu and Baird
			VCD1	F: 5' TTCAAAGTTGGTGTGGAACC 3'		(2003)
			VCD2	R: 5' GAGATGCAAGCTGCAACATAA 3'		
Qhb	<i>Squamosa promoter binding protein</i> (QhB33b17)	BQ968177	Qhb1	F: 5' CCGAAAGCCACAAAAAGAAA 3'	459	Kane and Rieseberg
			Qhb2	R: 5' ATTCTTGAATGTGGGGGTCA 3'		(2007)
			Qhb3	R: 5' CTTGAATGTGGGGGTCAACT 3'	456	
Hacot	<i>Phosphoenolpyruvate carboxykinase</i> (HACotR016B05)	TC44945	hacot1	F: 5' CTCAGGTACCGGAAAACAA 3'	890	Roche <i>et al.</i>
			hacot2	R: 5' CCGCAAGAATTTCTCAGTC 3'		(2007)
			hacot3	F: 5' AGGTCTCTTCGGTGTGCATGC 3'	840	
			hacot4	R: 5' ACACCCTCAACCTCAGTTGG 3'		
HaDhn1	<i>Dehydrin1</i>	X92647	dhn1.1	F: 5' TACGTTTTGCAGCGATTGTT 3'	916	Cellier <i>et al.</i>
			dhn1.1	R: 5' CACACGGGTGATACAACAGG 3'		(1998, 2000);
			dhn1.2	F: 5' CGACTACGAGAGCCCAATTC 3'	718	Poormahammad
			dhn1.2	R: 5' ATGTCCCGGAAGCTTTTCTT 3'		Kiani <i>et al.</i>
HaDhn2	<i>Dehydrin2</i>	X92650	dhn2.1	F: 5' GCGGGATAGCATTCCATAAC 3'	723	(2007);
			dhn2.2	R: 5' TTGCCGTATCACACATCACA 3'		Giordani <i>et al.</i>
			dhn2.3	R: 5' TACATCCTGCCTTGCCGTAT 3'	734	(2003).
Atp	<i>H⁺- transporting ATPase protein</i> (DH0AMM12ZA02ZZM1)	TC43917	atp1	F: 5' ATCCGACAGGAATACGAACG 3'	754	Roche <i>et al.</i>
			atp2	R: 5' AACATGCGACACGACGTTTA 3'		(2007)
			atp3	F: 5' TTTCCCAAATTCAGATCCA 3'	724	
			atp4	R: 5' TAAGCGCCTACCTGGCTAAA 3'		
Scoa	<i>Succinyl-CoA ligase mitochondrial</i> (DH0ALL30ZD02ZZM1)	TC40837	SCoa1	F: 5' CCAAGATGTCTCGTCAAGCA 3'	898	Roche <i>et al.</i>
			SCoa2	R: 5' ACTATGGCTCCAGCATGACC 3'		(2007)
			SCoa3	F: 5' ACCTTTACACCCGAACAAGC 3'	892	
			SCoa4	R: 5' GCCTGGGTGGGTAGTGTTTA 3'		
CK	<i>Putative caseinkinase II catalytic b</i> <i>unit</i> (DH0AL3ZB03ZM1)	TC54007	ck1	F: 5' GGAGGCATAGCAGAAAACCA 3'	432	Roche <i>et al.</i>
			ck2	R: 5' GCCACATACATCATGAAGTTTTAATC 3'		(2007)
			ck3	F: 5' GCATAGCAGAAAACCATGGAC 3'	428	

M2.4. Diseño de oligonucleótidos iniciadores y amplificación por PCR de las regiones candidatas

El diseño de los iniciadores se realizó con el programa Primer3 (Rozen and Skaletsky, 2000), utilizando los parámetros por defecto del mismo. La síntesis se realizó mediante un servicio externo (AlphaDNA, Canadá).

La mezcla de reacción para PCR se preparó como se indica en la Tabla 4. Las secuencias de los iniciadores para las 15 regiones ensayadas se detallan en la Tabla 3.

Tabla 4. Mezcla para la reacción de PCR

Reactivo	Volumen
Buffer con $(\text{NH}_4)_2\text{SO}_4$ [10X] (Invitrogen)	5,00 μl
MgCl_2 [50 mM]	2,00 μl
Desoxirribonucleótidos trifosfato (dNTPs) [10 mM]	1,00 μl
Cebadores <i>forward</i> + <i>reverse</i> [5 μM]	2,50 μl
<i>Taq</i> Polimerasa Platinum (Invitrogen) [5 U/ μl]	0,20 μl
H_2O Mili-Q o tridestilada estéril	37,30 μl
ADN [30 ng/ μl]	2,00 μl

Las reacciones de PCR se llevaron adelante usando el método de disminución progresiva de la temperatura de hibridación (*touchdown*). Se emplearon tres programas diferentes dado que cada par de oligonucleótidos iniciadores se probó en varias condiciones hasta alcanzar la rampa óptima de temperaturas de hibridación. Los *touchdown* (td) de cada programa de PCR fueron de 55-60°C, 58-65°C y 52-64°C en base a las temperaturas de hibridación de cada par de iniciadores.

Las amplificaciones de los individuos se realizaron en placa de 96 pocillos empleando un termociclador GeneAmp System 9700 (Applied Biosystems) con el programa de amplificación descrito en la Tabla 5.

Tabla 5. Condiciones de amplificación por PCR

Etapas del programa	Tiempo
Desnaturalización inicial a 94°C	2 min
<i>Touchdown (n ciclos dependiendo del td)</i>	
Desnaturalización a 94°C	45 s
Hibridación a 65°C, 64°C o 60°C (-1°C/ciclo)	45 s
Elongación a 72 °C	2 min
Amplificación (32 ciclos)	
Desnaturalización a 94°C	45 s
Hibridación a 58°C, 55°C o 52°C	45 s
Elongación a 72 °C	2 min
Extensión final a 72 °C	10 min
Mantenimiento a 15°C	

La presencia de producto de amplificación luego de la reacción de PCR fue verificada mediante electroforesis horizontal en geles de agarosa al 2% (p/v) y tinción con bromuro de etidio (0,5 µg/ml). La metodología empleada fue coincidente con la descrita en el Capítulo 1. Como marcador de peso molecular se utilizó un ladder de 100 pb (Invitrogen).

M2.5. Purificación de los productos de amplificación

La purificación de los productos de amplificación de Hahb4-p y Hahb4 se realizó mediante el empleo del equipo (*kit*) de purificación de fragmentos de PCR QIAquick (QIAGEN). Para los restantes genes se dispuso del método de purificación EXOSAP-IT (Exonuclease I & Shrimp Alkaline Phosphatase, USB, EEUU), empleando 4 µl de la enzima para purificar 10 µl de producto de PCR. Los protocolos se adjuntan en el Apéndice.

M2.6. Secuenciación de los productos de PCR

Los datos de secuencia se obtuvieron usando un protocolo que emplea los reactivos de *BigDye® Terminator v1.1/3.1 Sequencing Buffer* (5X) (Applied Biosystems) por el método de Sanger (Sanger and Coulson, 1975) con terminadores fluorescentes (Tabla 6). Se empleó un termociclador GeneAmp® PCR System 9700 (9600 o 2400). Se realizó una

desnaturalización inicial del ADN a 96 °C por 1 min , 25 ciclos de 96 °C por 10 s, 50 °C por 5 s y 60 °C por 4 min (la rampa térmica desciende 1 °C/s). Luego las muestras se llevaron a 4 °C, se centrifugaron los tubos para hacer descender el contenido de los mismos y se realizó la purificación de los productos. Las electroforesis se realizaron en secuenciadores automáticos de capilares *Genetic Analyzer 3130xl* y *3500xl* (Applied Biosystems, EEUU), en el Servicio de Secuenciación del Instituto de Biotecnología (INTA Castelar). Todos los electroferogramas fueron analizados visualmente para determinar la calidad de la corrida. El ensamblaje de las lecturas *forward* y *reverse* para cada genotipo (*contig*) y los alineamientos para la detección de SNP fueron realizados con el programa *BioEdit Sequence Alignment Editor* 1999-2005 (Hall, 1999).

Tabla 6. Reactivos y proporciones de mezclas para secuenciación. a. La cantidad de templado depende del tamaño del producto de PCR. Para 100-200 bp: 1–3 ng, para 200–500 bp: 3–10 ng, para 500–1000 bp: 5–20 ng, para 1000–2000 bp: 10–40 ng, para más de 2000 bp: 20–50 ng de templado.

Reactivo	Concentración	Volumen
<i>Ready Reaction Premix</i>	2.5X	4 µL
<i>BigDye Sequencing Buffer</i>	5X	2 µL
Iniciador	—	3,2 pmol
Templado	—	(a)
Agua bidestilada	—	hasta 20 µL
Volumen final	1X	20 µL

M2.7. Identificación de SNP y análisis de variabilidad en las regiones candidatas

Las secuencias codificantes y no codificantes de cada región candidata se corroboraron por comparación con los EST (*Expressed Sequence Tags*, secuencias expresadas) de partida y con la base de datos de proteínas del GenBank (<http://www.ncbi.nlm.nih.gov/genbank/>). Los niveles de variabilidad genética se estimaron a través de los parámetros de polimorfismo nucleotídico (θ_W ; Watterson, 1975), de diversidad nucleotídica (π ; Tajima, 1983), y de diversidad haplotípica (DH; Nei, 1987). El valor θ_W de Watterson se basa en la detección del número de sitios segregantes en la

región candidata analizada, mientras que el valor π de Tajima se basa en las diferencias entre pares de secuencias de la muestra total. Estos estimadores se calcularon en las regiones codificantes y no codificantes, correspondiendo estas últimas tanto a intrones como a los extremos UTR. En las regiones codificantes se determinó la diversidad nucleotídica de las sustituciones sinónimas (π_{sin}) y no sinónimas ($\pi_{\text{no-sin}}$), mientras que en las regiones no codificantes la diversidad nucleotídica correspondió a sustituciones denominadas silenciosas (π_{sil}). La diversidad haplotípica (DH) es una medida de cuantificación de la variabilidad genética que considera la probabilidad de que dos haplotipos elegidos al azar en una población sean diferentes. Esta probabilidad es equivalente a la heterocigosis en una población diploide panmíctica y se define en función de las frecuencias alélicas de la misma manera que la heterocigosis, pero a diferencia de ésta, puede utilizarse para cualquier organismo (haploide, diploide o poliploide). Cabe destacar que la identificación de polimorfismos SNP e InDel se llevó a cabo en base a las secuencias de un sólo individuo por línea analizada.

Los parámetros de variabilidad genética descritos fueron obtenidos usando el programa DNAsp v.5.10.01 (Librado and Rozas, 2009), con el cual se realizaron también las pruebas de neutralidad de Tajima (1989) y de Fu y Li (1993). La prueba de neutralidad D de Tajima se llevó a cabo para todas las líneas como un conjunto y además, dividiendo a las mismas en dos grupos, de acuerdo a los resultados obtenidos mediante el empleo del método bayesiano.

Finalmente, se emplearon los datos de genotipificación con marcadores SSR para detectar la posible existencia de un cuello de botella poblacional en las regiones estudiadas. Este análisis se realizó con el programa Bottleneck (Cornuet and Luikart, 1996), para el conjunto de las 34 líneas y dividiendo a las mismas en dos grupos en base a la estructura obtenida mediante el método bayesiano. Para el ajuste se empleó el modelo de alelos infinitos (IAM: *Infinite Allele Model*).

M2.8. Análisis del desequilibrio de ligamiento en las regiones candidatas

El DL es una medida de la desviación de las frecuencias alélicas observadas para alelos presentes en dos loci respecto de las frecuencias esperadas si la combinación de alelos se diera al azar. Por ello, cuando existe DL entre dos loci SNP, las variantes alélicas de ambos segregan asociadas y originan un haplotipo definido. El DL se puede medir con tres parámetros diferentes: el coeficiente de disequilibrio (D), el coeficiente estandarizado de disequilibrio (D'; Lewontin, 1964) y la correlación de la frecuencia alélica al cuadrado para pares de loci (r²; Hill and Weir, 1994). Cuando el valor de D es significativamente diferente de 0 se asume que existe DL. La expresión matemática de este parámetro es:

$$D_{AG} = \pi_{AG} - \pi_A \pi_G$$

Siendo π_{AG} la frecuencia haplotípica de la combinación del alelo A en el sitio 1 con el alelo G en el sitio 2 (por ej.), mientras que π_A es la frecuencia alélica de A y π_G la frecuencia alélica de G. El parámetro D' surgió para lograr una normalización de la medida de D puesto que se expresa en valor absoluto, por lo tanto su resultado varía entre 0 y 1. Su expresión matemática se basa en el valor de D y en el máximo disequilibrio teórico que dos loci podrían alcanzar ($D' = D/D_{m\acute{a}x}$). El estadístico D' mide sólo la historia de recombinación y es altamente afectado por muestras pequeñas, lo cual origina errores cuando se comparan loci con frecuencias alélicas bajas (Flint-García *et al.*, 2003). El tercer parámetro descrito es r² que incluye la historia de mutación y recombinación, además es menos sensible al tamaño de muestra que el estadístico D' (Flint-García *et al.*, 2003). La expresión matemática de r² incluye el tamaño efectivo poblacional (N_e) y la tasa de recombinación (c), reflejando dos propiedades importantes del DL, que a mayor recombinación entre dos sitios disminuye el DL y que éste es función del N_e. Por lo tanto, los factores históricos que afecten al N_e también modificarán el DL, entre ellos los eventos de cuello de botella poblacionales y la selección direccional. Si la tasa de recombinación es alta, la duración del disequilibrio originado por selección será de corto plazo (Przeworski, 2002). En base a lo expuesto, se calculó en este análisis el valor de r² como una medida del DL por medio del programa DNAsp v.5.10.01 (Librado and

Rozas, 2009). Los niveles de significación fueron corregidos según el procedimiento de Bonferroni para comparaciones múltiples con un α global del 5%.

Dado que los estimadores del DL son medidas entre pares de sitios polimórficos, la extensión del DL se determinó a través de una curva de regresión no lineal entre la distancia y el r^2 para cada una de las regiones por separado y para su conjunto. Las gráficas de decaimiento de r^2 a lo largo de todos los loci para las regiones estudiadas fueron obtenidas con el programa descrito anteriormente. La gráfica del decaimiento general del DL para el conjunto total de regiones candidatas estudiadas se realizó mediante el empleo del programa Excel (Microsoft).

RESULTADOS

R1. Caracterización fenotípica para el carácter de tolerancia a sequía

R1.1. Análisis de datos de ensayo de campo: evaluación de planta adulta

Los datos registrados en el ensayo de campo (campaña 2003-2004) para la clasificación de las líneas de acuerdo a su comportamiento frente a condiciones de déficit hídrico se adjuntan en la Tabla A.4 del Apéndice.

Los resultados obtenidos, tanto en condiciones de irrigación como de déficit hídrico, para cada uno de los caracteres agronómicos evaluados en el campo (ALT: altura de planta, P100: peso de cien semillas, NSC: número de semillas por capítulo, MG: porcentaje de materia grasa y RENSH: rendimiento), se muestran en las Figuras 2, 3, 4, 5 y 6. Todas las líneas mostraron mayores valores de ALT en la condición de riego comparada con la de sequía (Figura 2), mientras que para P100, NSC y RENSH el comportamiento fue más variable. Las líneas B67, HA89, B59 y R420 mostraron las mayores diferencias entre los valores de peso de cien semillas en la condición de riego con respecto a los bajos valores de peso en sequía, mientras que el caso inverso se observó para las líneas R437, R431, B71 y R433, dado que se comportaron mejor en sequía (Figura 3). Por otro lado, las líneas B59, B99 y Aguará son las que mostraron mayores diferencias entre los valores de NSC bajo riego, comparado con los valores para este carácter en sequía (Figura 4). Del mismo modo, las líneas B59, B99 y Aguará mostraron, también, las mayores diferencias entre los valores de RENSH bajo riego y en sequía (Figura 6), indicando un comportamiento de susceptibilidad marcado en estas tres entradas. En cuanto al carácter MG, 14 líneas (R427, R413, R073, B71, R417, R433, R435, RHA274, R429, Aguará, R420, R421, B99 y R419) arrojaron mayores valores de MG en condición de sequía que bajo riego (Figura 5).

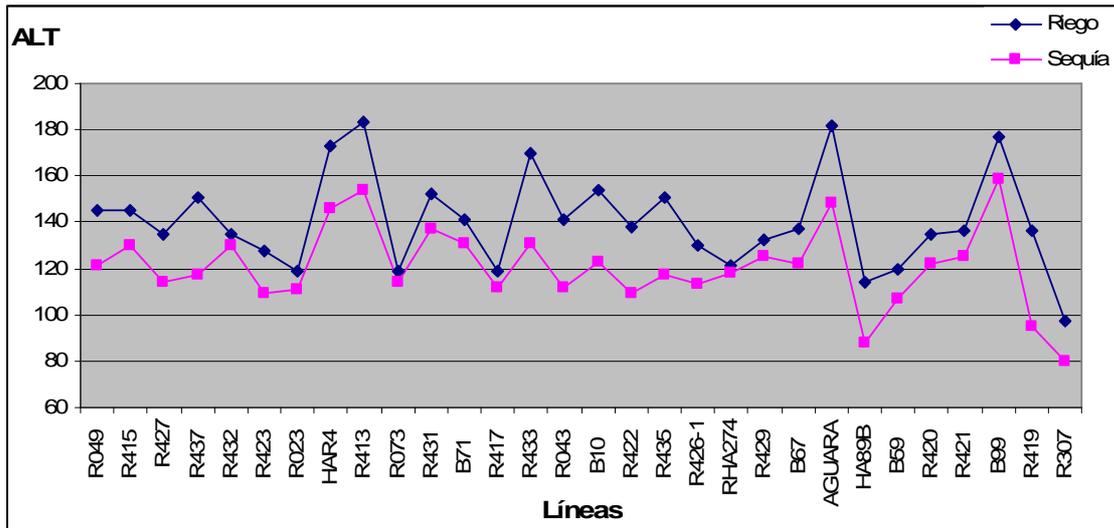


Figura 2. Altura de planta para controles irrigados y plantas en condiciones de déficit hídrico. ALT: altura de planta en cm.

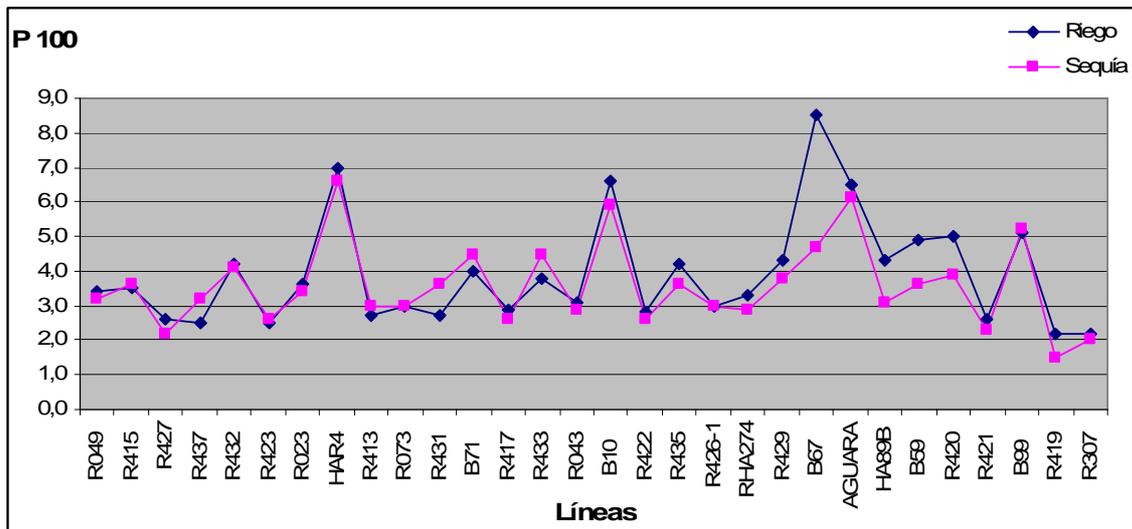


Figura 3. Peso de cien semillas para controles irrigados y plantas en condiciones de déficit hídrico. P 100: peso de cien semillas en gramos.

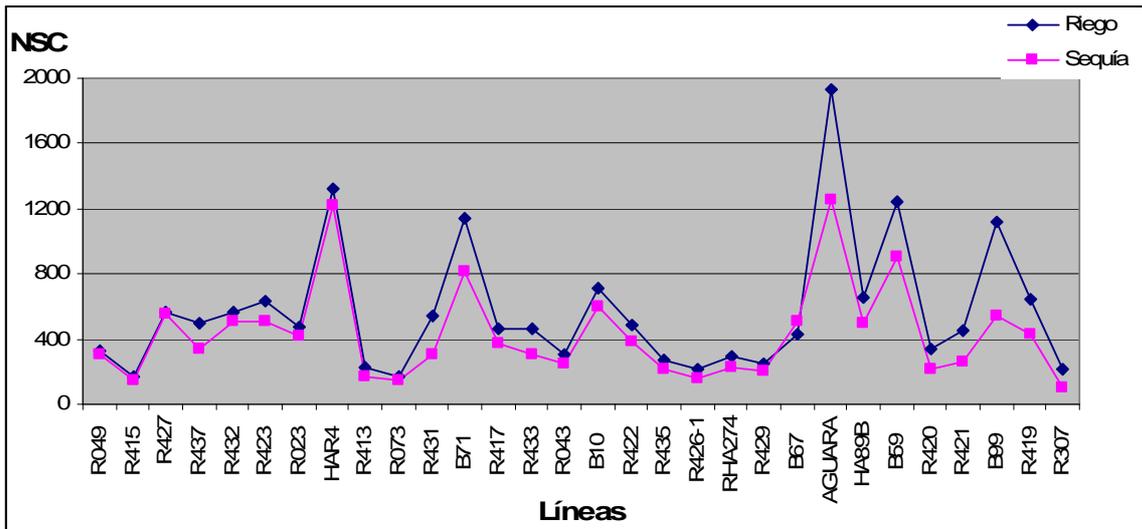


Figura 4. Número de semillas por capítulo para controles irrigados y plantas en condiciones de déficit hídrico. NSC: número de semillas por capítulo.

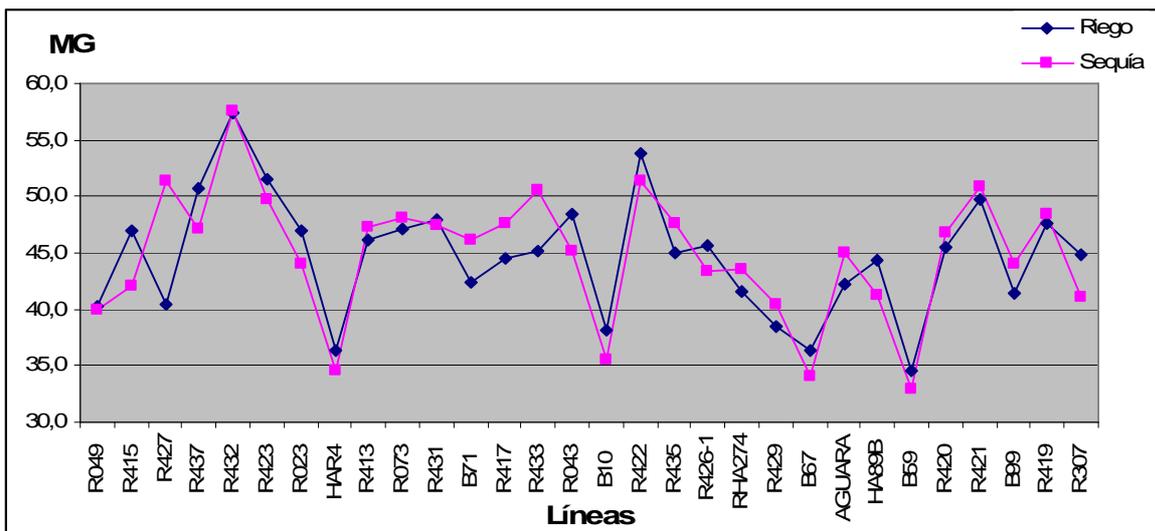


Figura 5. Contenido de materia grasa en grano para controles irrigados y plantas en condiciones de déficit hídrico. MG: materia grasa en porcentaje.

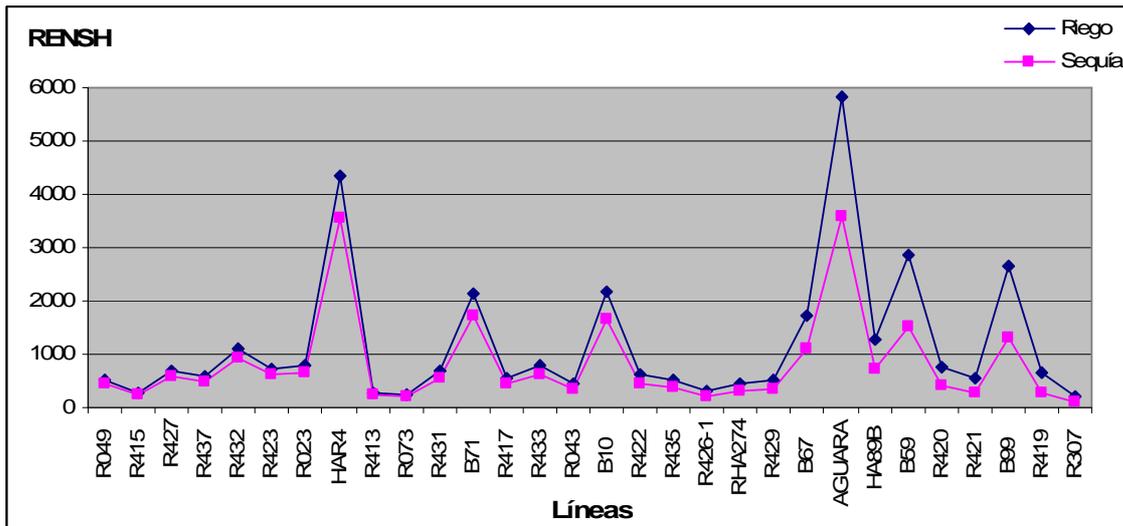


Figura 6. Rendimiento de semillas para controles irrigados y plantas en condiciones de déficit hídrico. RENS: rendimiento de semillas en Kg/ha sobre granos con 11% de humedad.

R1.2. Ensayo en condiciones controladas: evaluación de plántula

Los resultados obtenidos en este ensayo se exponen de forma completa en el Apéndice (Tabla A.5). Las Figuras 7, 8 y 9 muestran fotografías tomadas durante los riegos realizados a los 4, 7 y 10 días posteriores a la siembra, respectivamente. En todos los casos pueden observarse plántulas representativas de las líneas más y menos afectadas en ambas condiciones de déficit hídrico.



Figura 7. Comportamiento de plántulas en ensayo bajo condiciones controladas al 4 día posterior a la siembra. Fotos 1 y 2: líneas que mostraron crecimiento tanto en controles (izq.) como bajo tratamiento con manitol 200mM (der.). En 400mM no se obtuvo crecimiento para ninguna línea al 4 día posterior a la siembra.



Figura 8. Comportamiento de plántulas en ensayo bajo condiciones controladas al 7 día posterior a la siembra. Fotos 1 y 2: líneas menos afectadas a 200mM y 400 mM, respectivamente (izq: plantas tratadas, der: controles). Fotos 3 y 4: líneas más afectadas a 200 mM y 400mM, respectivamente (izq: plantas tratadas, der: controles).



Figura 9. Comportamiento de plántulas en ensayo bajo condiciones controladas al 10 día posterior a la siembra. Fotos 1 y 2: líneas menos afectadas a 200mM y 400 mM, respectivamente (izq: controles, der: plantas tratadas). Fotos 3 y 4: líneas más afectadas a 200 mM y 400mM, respectivamente (izq: controles, der: plantas tratadas).

Las Figuras 10, 11 y 12 muestran imágenes tomadas el día de la cosecha. La primera incluye plántulas normales obtenidas tanto para controles como para ambos tratamientos, mientras que las dos Figuras restantes incluyen fotografías de raíces y plántulas normales en comparación con raíces atrofiadas o necróticas y plántulas anómalas.



Figura 10. Plántulas normales obtenidas en ensayo bajo condiciones controladas al 11° día. Foto 1: plántulas normales crecidas en condición control. Foto 2: plántulas normales crecidas bajo tratamiento moderado (200mM). Foto 3: plántulas normales crecidas bajo tratamiento severo (400mM).



Figura 11. Fenotipos anómalos detectados en raíces de plántulas sometidas a déficit hídrico en ensayo bajo condiciones controladas al 11° día. Foto 1: sistema radicular normal. Fotos 2 y 3: raíces secundarias atrofiadas y necrosadas. Foto 4: raíz primaria atrofiada, necrótica y con ausencia de raíces laterales.

La Figura 12 muestra plántulas con anomalías, motivo de eliminación al momento del recuento final. Los signos de necrosis radicular representaron una parte importante (60%) de los defectos totales hallados en laboratorio, mientras que el porcentaje de necrosis foliar fue del 30%. Otro fenotipo anómalo detectado en este análisis correspondió al estrangulamiento del tallo cercano a los cotiledones (Figura 12: foto 2) principalmente en la línea R307, representando un 10% del total de plántulas con defectos.



Figura 12. Fenotipos anómalos detectados en plántulas sometidas a déficit hídrico en ensayo bajo condiciones controladas al 11° día. Foto 1: plántula normal. Foto 2: plántula anómala por estrangulamiento y escaso desarrollo de raíces laterales. Foto 3: plántula anómala por nulo desarrollo radicular y signos de necrosis foliar.

El análisis de los datos obtenidos en el ensayo bajo condiciones controladas, se realizó en base a un modelo de regresión logística, mediante el cual se logró determinar agrupamientos de las líneas en función de la variable RFG. Con las demás variables (EG, PF y PS) no se logró clasificar a los genotipos en función de su respuesta frente al déficit hídrico. Los resultados obtenidos para cada una de las variables mencionadas se adjuntan en el Apéndice (Tabla A.5). Las Figuras 13 y 14 muestran el comportamiento de las líneas durante la germinación frente a una solución de manitol de 200mM y de 400mM, respectivamente. El comportamiento se expresó como una estima de la probabilidad del éxito de germinación,

constituyendo así valores predictivos o predichos. En el Apéndice se adjuntan los valores predichos para cada línea y tratamiento (Tabla A.6).

En la condición de déficit hídrico moderado (manitol 200mM, Figura 13) las líneas con valores predichos entre 0,09 y 0,18 se clasificaron como tolerantes (R426-1, R435, R429, B71, B99, RHA274, R431, R421, HA89, R432 y B10). Las 11 líneas, cuyos valores predichos fueron de 0,03 o inferiores, se clasificaron como susceptibles, mostrando una menor probabilidad de éxito. Este grupo estuvo conformado por R073, R043, R422, R420, B67, 964-C, 1063-C, R307, R423, B59 y A59.

En la condición de déficit hídrico severo (manitol 400mM, Figura 14) las líneas con valores predichos entre 0,07 y 0,25 (R426-1, R420, R431, R437, R073, R433 y R435) se clasificaron como tolerantes. Mientras que los materiales con valores predichos cercanos a 0 (1063-C, HAR4, 430-C, R417, R422, R049, R043, R421, R413, R307, 962-C, RHA274, HA89, B67, 964-C, B71, R415 y B99) se clasificaron como susceptibles.

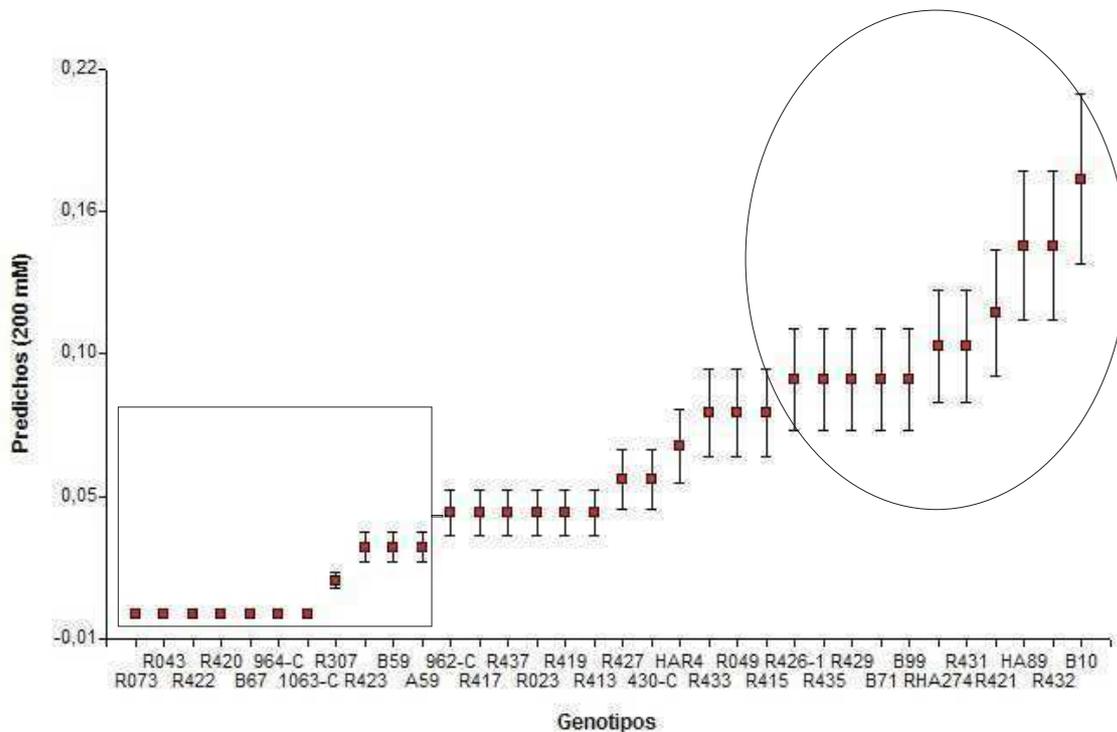


Figura 13. Comportamiento de las líneas frente a déficit hídrico moderado (manitol 200mM). El óvalo incluye el grupo de líneas tolerantes, mientras que el rectángulo el grupo de líneas susceptibles.

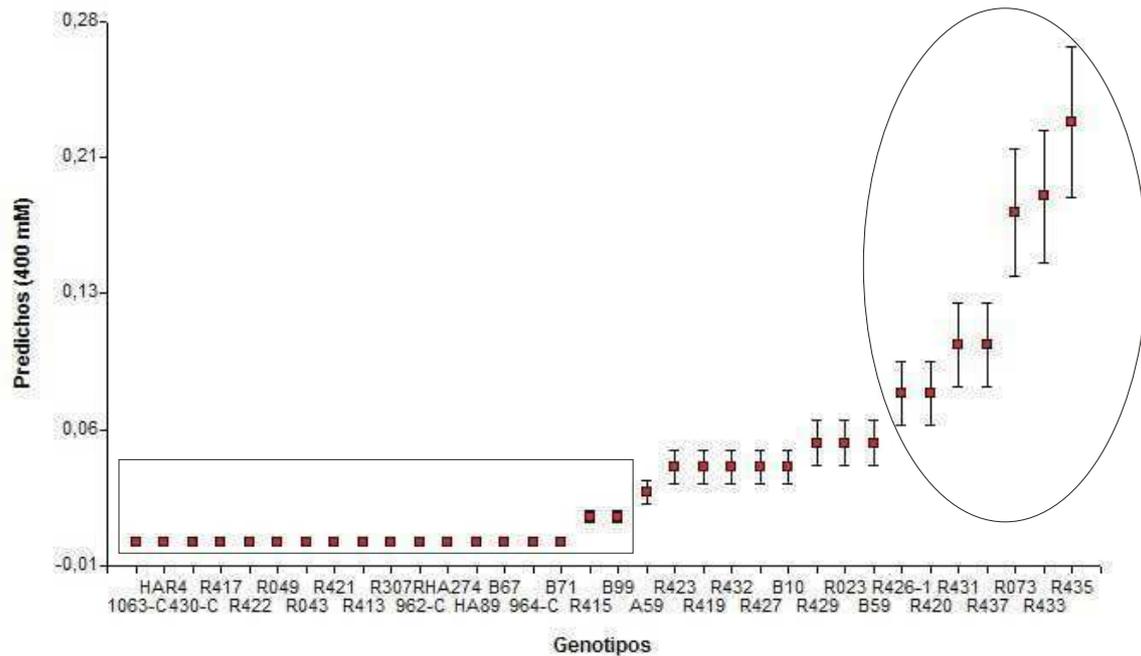


Figura 14. Comportamiento de las líneas frente a déficit hídrico severo (manitol 400mM). El óvalo incluye el grupo de líneas tolerantes, mientras que el rectángulo las líneas susceptibles.

En cuanto a la comparación entre los resultados obtenidos en la condición de déficit hídrico moderado y severo, las líneas que se clasificaron como tolerantes en ambas condiciones fueron R426-1, R435 y R431, mientras que las líneas susceptibles fueron R043, R422, B67, 964-C, 1063-C y R307. La comparación con los resultados del análisis de los datos del ensayo de campo y en condiciones controladas mostró que la línea R431 se clasificó como tolerante en todos los casos, mientras que B67 se caracterizó como susceptible.

R2. Caracterización genética de las líneas para tolerancia a sequía

R.2.1. Análisis de variabilidad microsatélite

Se genotipificaron 2-3 individuos por línea para las 34 entradas en estudio, empleando 7 marcadores SSR. Luego de realizar las amplificaciones por triplicado para los casos de no amplificación, el rango de

fallas osciló entre 0 y 9,47% para HA2063 y HA3239, respectivamente. El promedio de datos faltantes fue de 4,36%.

Los resultados obtenidos mostraron la existencia de diferencias entre individuos dentro de una misma línea, sólo 9 de las 34 fueron completamente homogéneas (R435, R423, 962-C, R413, R426-1, B99, R023, R419 y R043). Las líneas con mayor variación en el conjunto completo de microsatélites estudiados fueron: HA89 con 5 SSR variables entre individuos de la misma línea, y R415, A59 y R429, que mostraron 4 de los 7 SSR con variación (Tabla 7). Los datos completos de genotipificación se adjuntan en la Tabla A.7 del Apéndice.

Con respecto al comportamiento de los marcadores, todos ellos mostraron heterogeneidad, con un rango de líneas heterogéneas que varió entre 4 y 9 (Tabla 7). Para los marcadores HA77, HA928 y HA3239, 4 de las 34 líneas (11,7%) presentaron alelos diferentes en los tres individuos evaluados. Para los marcadores HA2063 y HA2920 se detectaron 5 líneas variables entre los 3 individuos estudiados (14,7%), para HA293 se hallaron 8 líneas con variación entre individuos (23,5%), mientras que para HA4103 se detectaron 9 líneas variables entre individuos (26,5%).

Tabla 7. Esquema de la variación entre individuos de una misma línea para cada marcador SSR. Los cuadros grises representan los casos de heterogeneidad.

Líneas	HA77	HA293	HA928	HA2063	HA2920	HA3239	HA4103
B10							
A59							
R073							
R422							
R435							
B59							
430-C							
R307							
R423							
R437							
B67							
962-C							
R413							
R426-1							
B71							
964-C							
R415							
R427							
B99							
1063-C							
R417							
R429							
RHA274							
R023							
R419							
R431							
HAR4							
R043							
R420							
R432							
HA89							
R049							
R421							
R433							

El análisis de los casos de heterogeneidad entre individuos de la misma línea en base a los criterios propuestos por Heckenberger *et al.* (2002), descritos en el Capítulo 1, se expone en la Tabla 8. Las comparaciones de a pares entre individuos de una misma línea, mostraron una ocurrencia de casos normales entre 42,85 y 100%, con un valor promedio de 80,67. Las 4 líneas más variables descritas anteriormente (HA89, R415, A59 y R429; Tabla 7) son las que presentaron los valores más bajos de ocurrencia para los casos normales (Tabla 8). El caso de heterogeneidad más frecuente en el conjunto de todas las líneas estudiadas correspondió al caso 8 (con una ocurrencia del 7,98), donde un individuo fue homocigota y para el otro individuo no se obtuvo dato de genotipificación. En segundo lugar, de acuerdo con la frecuencia de aparición, se ubicó el caso 10 (con una ocurrencia del 6,86), donde un individuo mostró un genotipo homocigota para un alelo y el otro individuo resultó homocigota para otro alelo diferente al primero.

Tabla 8. Ocurrencia de diferentes casos de heterogeneidad entre los individuos de una misma línea evaluada con 7 SSR. Ncp: número de comparaciones de a pares.

Línea	Ncp	Casos									
		1	2	3	4	5	6	7	8	9	10
B10	3	76,19	0	4,76	0	0	0	4,76	14,28	0	0
A59	3	52,38	0	23,81	0	0	0	4,76	14,28	0	4,76
R073	1	85,71	0	0	0	0	0	0	14,28	0	0
R422	3	71,43	0	9,52	0	0	0	0	9,52	0	9,52
R435	3	100	0	0	0	0	0	0	0	0	0
B59	3	71,43	0	0	0	0	0	0	28,57	0	0
430-C	3	66,66	0	0	0	0	0	0	9,52	0	23,81
R307	3	80,95	0	0	0	0	0	0	0	0	19,05
R423	1	100	0	0	0	0	0	0	0	0	0
R437	1	85,71	0	0	0	0	0	0	0	0	14,28
B67	3	90,48	0	0	0	0	0	0	0	0	9,52
962-C	3	100	0	0	0	0	0	0	0	0	0
R413	1	100	0	0	0	0	0	0	0	0	0
R426-1	3	100	0	0	0	0	0	0	0	0	0
B71	3	76,19	14,28	0	0	0	0	0	9,52	0	0
964-C	3	90,48	0	0	0	0	0	0	9,52	0	0
R415	3	47,62	0	9,52	0	0	0	0	9,52	0	33,33
R427	3	90,48	0	0	0	0	0	0	0	0	9,52
B99	3	100	0	0	0	0	0	0	0	0	0
1063-C	3	85,71	4,76	9,52	0	0	0	0	0	0	0
R417	3	80,95	0	0	0	0	0	0	19,05	0	0
R429	3	42,85	0	0	0	0	0	0	28,57	0	28,57
RHA274	3	71,43	0	0	0	0	0	0	19,05	0	9,52
R023	1	100	0	0	0	0	0	0	0	0	0
R419	3	100	0	0	0	0	0	0	0	0	0
R431	3	66,66	4,76	9,52	0	0	0	0	19,05	0	0
HAR4	3	85,71	0	0	0	0	0	0	9,52	0	4,76
R043	1	100	0	0	0	0	0	0	0	0	0
R420	3	61,9	0	0	0	0	0	0	19,05	0	19,05
R432	3	71,43	0	0	0	0	0	0	9,52	0	19,05
HA89	3	47,62	4,76	19,05	0	0	0	0	0	0	28,57
R049	3	90,48	0	0	0	0	0	0	9,52	0	0
R421	1	85,71	14,28	0	0	0	0	0	0	0	0
R433	3	66,66	4,76	9,52	0	0	0	0	19,05	0	0
Promedio		80,67	1,4	2,81	0	0	0	0,28	7,98	0	6,86

De las 14 líneas homogéneas según el esquema presentado en la Tabla 7, 9 de ellas mostraron un 100% de ocurrencia de casos normales de acuerdo con el criterio propuesto por Heckenberger *et al.* (2002) (Tabla 8), mientras que 5 líneas mostraron la ocurrencia de casos 8 (un individuo homocigota y otro sin dato de genotipificación). Por lo tanto, no sería correcto clasificar estas 5 líneas como homogéneas porque no se tiene conocimiento completo en cuanto a los genotipos de sus individuos.

Siguiendo con lo propuesto por Heckenberger *et al.* (2002), se construyó un árbol de conglomerados en base al índice de similitud de Jaccard entre pares de individuos. Las líneas R417, B99, R307, R431, R049, R423, R426-1, R023, R073, R437, R043, HAR4, 964-C, 962-C, R435 y B67 mostraron cohesión puesto que los 3 individuos de cada una de ellas constituyeron un grupo único, con un rango de distancias entre 0 y 0,6 (Figura 15). Las restantes líneas (18) presentaron un patrón disperso, dado que sus individuos se agruparon con los de otra/s línea/s (Figura 15).

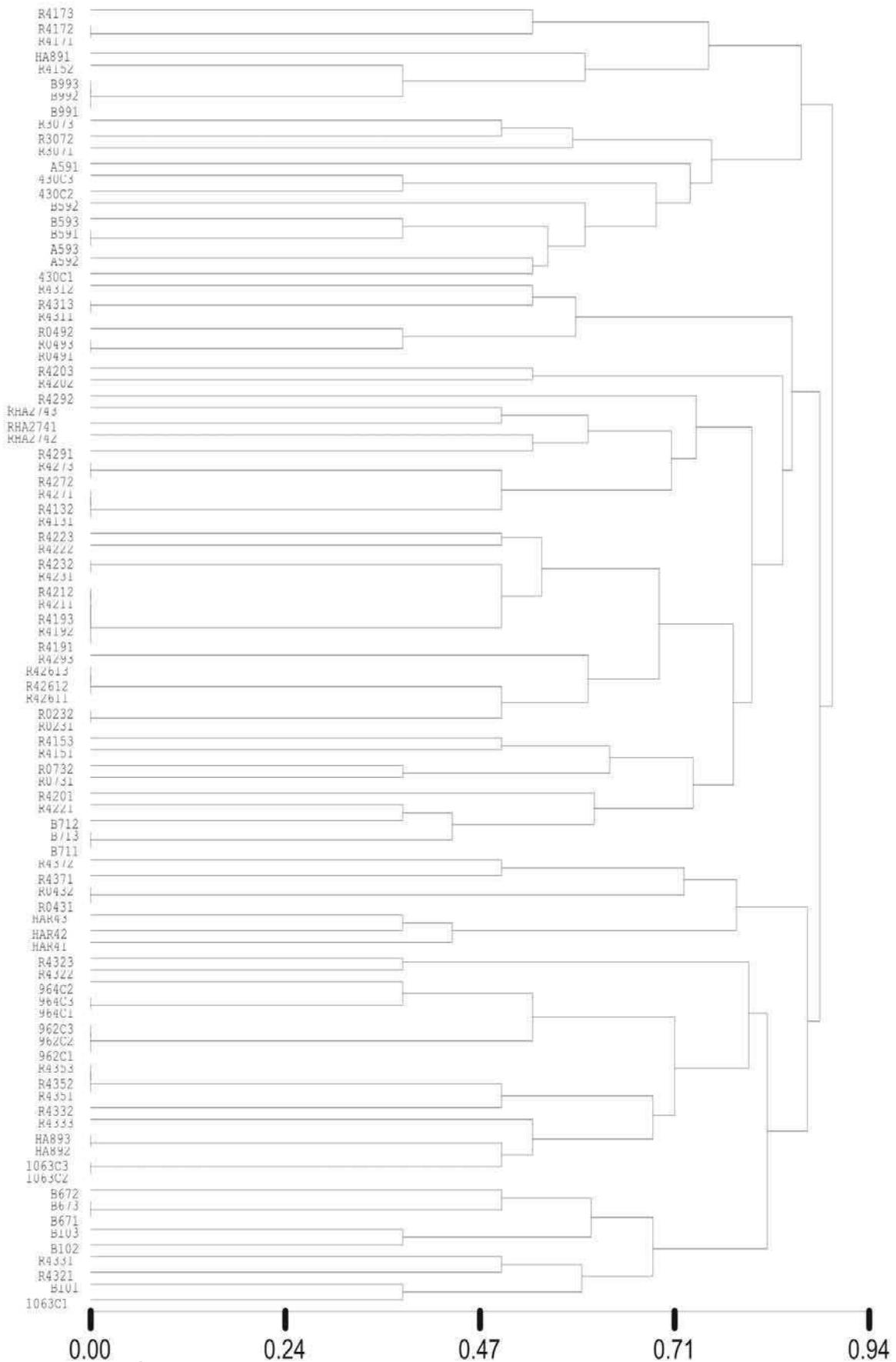


Figura 15. Árbol de conglomerados en base al índice de Jaccard para las líneas analizadas. Se identificó a cada individuo con el nombre de la línea a la que pertenece, seguido del respectivo número de individuo.

Los resultados de las estimas de variabilidad considerando a las 34 líneas como una población global fueron de 0,57 para la heterocigosis promedio esperada, mientras que se obtuvo un valor de 0,031 para la heterocigosis promedio observada. El número promedio de alelos por locus fue de 3,85; con un rango que varió entre 3 y 6.

Se analizó la estructura poblacional mediante un modelo bayesiano (Pritchard *et al.*, 2000). El número de poblaciones ideales fue difícil de determinar dado que se alcanzó una meseta recién a partir de $K=9$, con una leve disminución de la pendiente de la curva $\ln P(D)$ vs. K (Figura 16-A). En base el análisis propuesto por Evanno *et al.* (2005), se detectaron tres picos de ΔK en función de K , en $K=2$, en $K=6$ y en $K=9$ (Figura 16-B). Esto indicaría, coincidentemente lo observado para las poblaciones y compuestos analizados en el Capítulo 1, distintos niveles de sub-estructuración en las entradas estudiadas. La detección de 9 acervos génicos da indicios de un alto nivel de estructuración, superior al esperado de acuerdo con información previa provista por el mejorador.

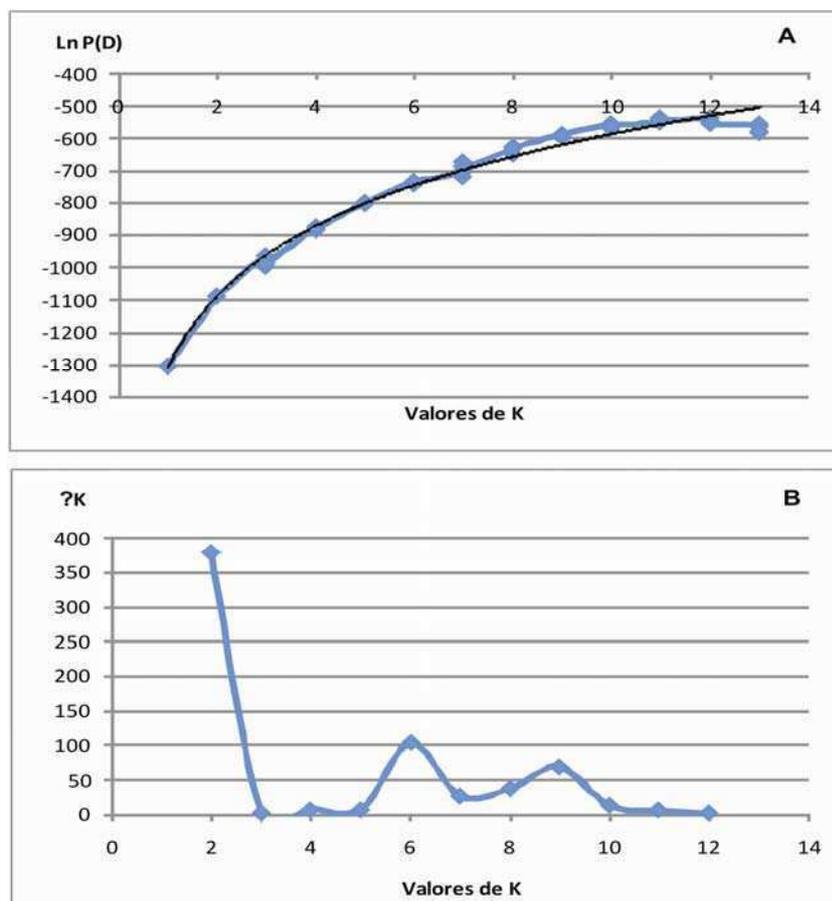


Figura 16. Análisis de estructura poblacional con el método bayesiano de Pritchard et al. (2000) (página anterior). A. Logaritmo de las probabilidades para cada K. B. Gráfico de ΔK vs. K.

Las figuras 17, 18 y 19 muestran la distribución de los individuos en grupos genómicos o acervos génicos para K=2, K=6 y K=9. Las asignaciones de los individuos a cada grupo en las 5 repeticiones fueron estables. El criterio empleado para asignar un individuo a un grupo u otro fue que más del 50% del genoma del individuo estuviera constituido por la contribución de ese grupo. La tendencia general observada correspondió a un patrón de cohesión entre individuos de la misma línea, excepto para HA89, A59, R429 y R415 que, como ya se mencionó, fueron las líneas más variables en la genotipificación con SSR.

Para K=2, las líneas B10, A59, R435, B59, 430-C, R307, B67, 964-C, B99, 1063-C, R417, HAR4, R432, HA89, R433 y un individuo de R415, mostraron la contribución mayoritaria del mismo acervo genético (grupo rojo). Por otro lado, los genotipos R073, R422, R423, R437, R413, R426-1, B71, R427, R429, RHA274, R023, R419, R431, R043, R420, R049 y R421 y dos individuos de R415, exhibieron contribución mayoritaria del acervo verde. Sólo el material 962-C no logró ser adjudicado a ninguno de los dos grupos puesto que sus individuos (N°31, 32 y 33; Figura 17) presentaron contribuciones iguales de ambos. Los orígenes geográficos de las líneas son diversos dentro de cada grupo (Tabla 1, Tabla 2).

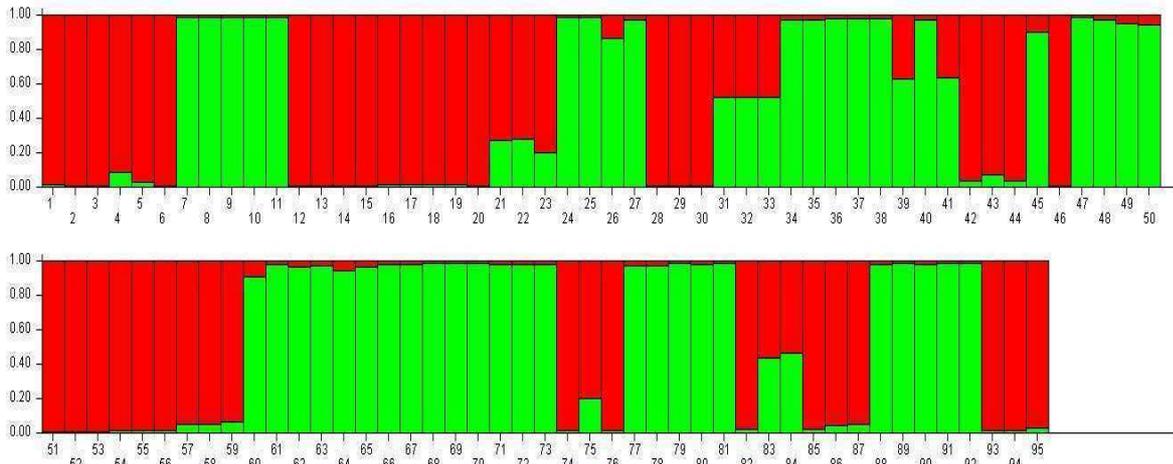


Figura 17. Distribución del total de individuos estudiados de acuerdo con la inferencia de estructura poblacional en $K=2$. B10 (1-2-3), A59 (4-5-6), R073 (7-8), R422 (9-10-11), R435 (12-13-14), B59 (15-16-17), 430-C (18-19-20), R307 (21-22-23), R423 (24-25), R437 (26-27), B67 (28-29-30), 962-C (31-32-33), R413 (34-35), R426-1 (36-37-38), B71 (39-40-41), 964-C (42-43-44), R415 (45-46-47), R427 (48-49-50), B99 (51-52-53), 1063-C (54-55-56), R417 (57-58-59), R429 (60-61-62), RHA274 (63-64-65), R023 (66-67), R419 (68-69-70), R431 (71-72-73), HAR4 (74-75-76), R043 (77-78), R420 (79-80-81), R432 (82-83-84), HA89 (85-86-87), R049 (88-89-90), R421 (91-92), R433 (93-94-95).

El pico de ΔK en $K=6$ mostró una distribución en grupos de acuerdo con la Figura 18. Las líneas B10, B67, 1063-C y R433 mostraron contribución mayoritaria del grupo rojo; mientras que el grupo verde conformó los genotipos A59, B59, 430-C, R307, B99, R417 y un individuo de R415. Por su parte las líneas R073, R423, R426-1, R023, R419, R420, R421, dos individuos de R422 y un individuo de R429, mostraron contribución del conglomerado amarillo; mientras que el grupo azul incluyó los genotipos R437, HAR4, R043 y un individuo de R432. El grupo coloreado de fucsia contribuyó a las entradas R435, 962-C y 964-C; mientras que el celeste comprendió a las líneas R413, B71, RHA274, R431, R049, R427, un individuo de R422, dos individuos de R415 y dos individuos de R429. Al igual que lo descrito para $K=2$, no se detectó un patrón claro de estructuración en base a los orígenes geográficos (Figura 18).

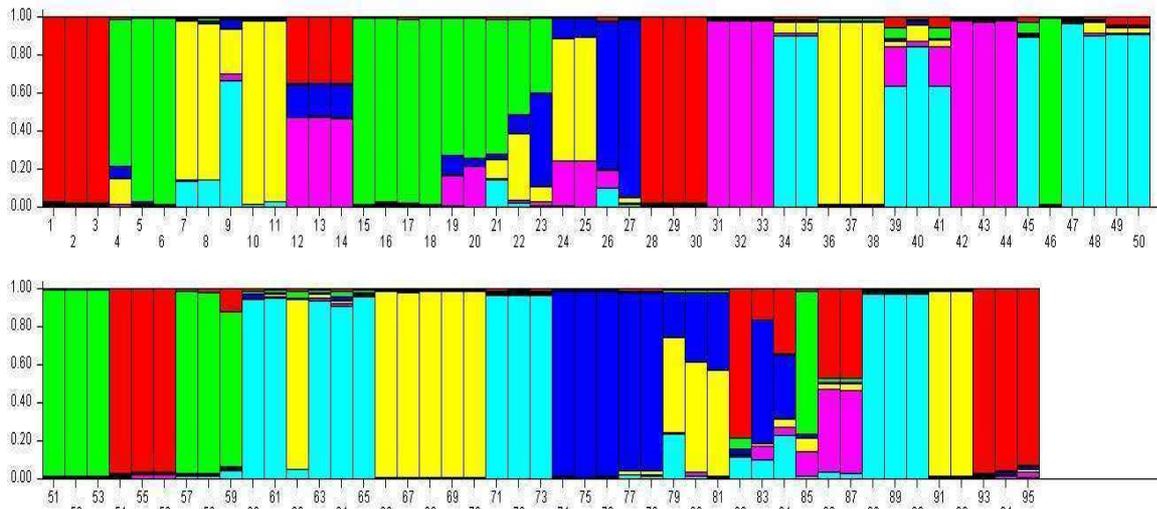


Figura 18. Distribución del total de individuos estudiados de acuerdo con la inferencia de estructura poblacional en K=6. B10 (1-2-3), A59 (4-5-6), R073 (7-8), R422 (9-10-11), R435 (12-13-14), B59 (15-16-17), 430-C (18-19-20), R307 (21-22-23), R423 (24-25), R437 (26-27), B67 (28-29-30), 962-C (31-32-33), R413 (34-35), R426-1 (36-37-38), B71 (39-40-41), 964-C (42-43-44), R415 (45-46-47), R427 (48-49-50), B99 (51-52-53), 1063-C (54-55-56), R417 (57-58-59), R429 (60-61-62), RHA274 (63-64-65), R023 (66-67), R419 (68-69-70), R431 (71-72-73), HAR4 (74-75-76), R043 (77-78), R420 (79-80-81), R432 (82-83-84), HA89 (85-86-87), R049 (88-89-90), R421 (91-92), R433 (93-94-95).

El pico de ΔK en $K=9$ mostró una distribución en grupos de acuerdo con la Figura 19. Las líneas B10, B67 y 1063-C mostraron contribución mayoritaria del grupo amarillo; mientras que el grupo fucsia conformó los genotipos A59, B59, 430-C y R307, coincidente con lo observado en la Figura 15. Por su parte las líneas R073, R420, B71, dos individuos de R422, dos individuos de R415 y dos de RHA274, mostraron contribución del conglomerado marrón claro; mientras que el grupo rojo incluyó los genotipos R435, R432 y R433. El grupo coloreado de verde contribuyó a los materiales R423, R413, R426-1, R023, R419 R421, un individuo de R427, un individuo de R429, uno de RHA274 y uno de R422; mientras que el azul comprendió a las líneas B99, R417, un individuo de HA89 y uno de R415, coincidente con el árbol en base al índice de similitud de Jaccard. Por su parte las líneas 962-C y 964-C mostraron contribución del conglomerado celeste. Dos individuos de HA89 presentaron el aporte del acervo celeste y del amarillo en la misma proporción. El grupo coloreado de marrón oscuro contribuyó a las

líneas HAR4, R043 y R437, coincidente con la Figura 15, mientras que el naranja aportó a R431, R049, dos individuos de R427 y dos de R429. Al igual que lo descrito para K=2 y para K=6, no se detectó un patrón claro de estructuración en base a los orígenes geográficos (Figura 19).

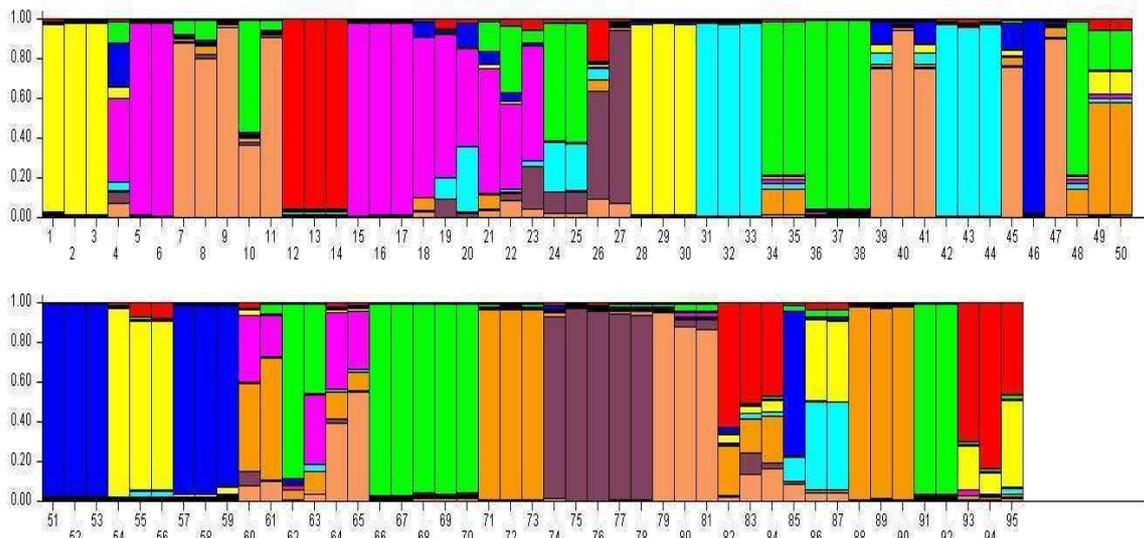


Figura 19. Distribución del total de individuos estudiados de acuerdo con la inferencia de estructura poblacional en K=9. B10 (1-2-3), A59 (4-5-6), R073 (7-8), R422 (9-10-11), R435 (12-13-14), B59 (15-16-17), 430-C (18-19-20), R307 (21-22-23), R423 (24-25), R437 (26-27), B67 (28-29-30), 962-C (31-32-33), R413 (34-35), R426-1 (36-37-38), B71 (39-40-41), 964-C (42-43-44), R415 (45-46-47), R427 (48-49-50), B99 (51-52-53), 1063-C (54-55-56), R417 (57-58-59), R429 (60-61-62), RHA274 (63-64-65), R023 (66-67), R419 (68-69-70), R431 (71-72-73), HAR4 (74-75-76), R043 (77-78), R420 (79-80-81), R432 (82-83-84), HA89 (85-86-87), R049 (88-89-90), R421 (91-92), R433 (93-94-95).

R.2.2. Caracterización de regiones candidatas para el carácter de tolerancia a sequía

Para la identificación de SNP y el análisis de la diversidad nucleotídica se utilizaron como punto de partida las 15 regiones genómicas descritas en la sección M2.3 (Tabla 3). Los iniciadores diseñados amplificaron un único producto de PCR en 8 de las 15 regiones candidatas, en 2 (Qhb y VCD) se obtuvieron productos múltiples y 5 regiones no generaron producto de amplificación. De las 8 regiones que originaron un producto único, una se

descartó por mostrar un alto número de indeterminaciones en la secuencia de ADN (HaCot), las 7 restantes fueron empleadas para el análisis de diversidad nucleotídica (Hahb4-p, Hahb4, *Suntip*, FB, HaL1L, HaDhn1 y CK; Tabla 3) en el conjunto de 36 líneas detalladas en la sección M.2. Para las 7 regiones candidatas seleccionadas se obtuvieron datos de secuenciación de alta calidad, pudiendo determinarse la longitud de las regiones codificantes y no codificantes y los polimorfismos presentes en las mismas (Tabla 9).

R.2.2.1. Frecuencia de SNP y variabilidad nucleotídica

Las siete regiones candidatas estudiadas comprendieron 5646 pb por línea, el largo de las secuencias por región osciló entre 585 y 1186 pb. Se encontró un total de 117 SNP, con un valor promedio por locus de 16,71 y una frecuencia de 1 SNP cada 48,25 pb. En cuanto a los polimorfismos de tipo InDel, se hallaron un total de 17, sumando 291 pb, con un valor promedio por locus de 2,43 y una frecuencia de 1 InDel cada 332,12 pb. Se observó la ocurrencia tanto de SNP como de InDel en 6 de las 7 regiones evaluadas, puesto que en Hahb4 no se detectaron InDel. El rango de SNP varió entre 6 (CK) y 39 (Hahb4-p), mientras que el rango de InDel osciló entre 0 (Hahb4) y 6 (Hahb4-p) (Tabla 9).

El número de SNP resultó también variable al comparar las regiones codificantes y no codificantes, puesto que se encontraron 33 SNP en 2976 pb de zona codificante, en comparación con 82 SNP en 2670 pb de secuencia no codificante. En base a estos resultados, la frecuencia promedio de SNP para las regiones codificantes fue de 1 SNP/90,18 pb, mientras que para las regiones no codificantes fue de 1 SNP/32,56 pb.

Por definición los InDel son pequeñas inserciones, en general todos los polimorfismos de este tipo cumplieron con la mencionada condición excepto uno de los InDel de la región Hahb4-p cuya longitud fue de 196 pb.

Tabla 9. Descripción de los polimorfismos detectados en las regiones candidatas analizadas. (cod/no-cod): polimorfismos encontrados en las regiones codificantes y no codificantes, respectivamente.

Región	Longitud total (cod/no-cod)	SNP totales (cod/no-cod)	Nº InDel totales (cod/no-cod)
Hahb4-p	1186 (0/1186)	39 (0/39)	6 (0/6: 211 pb)
Hahb4	614 (455/159)	16 (9/7)	0 (0/0)
<i>Suntip</i>	788 (604/184)	28 (14/12)	2 (0/2: 43 pb)
FB	1070 (381/689)	7 (0/7)	3 (0/3: 20 pb)
HaL1L	642 (543/99)	7 (5/2)	1 (0/1: 1 pb)
HaDhn1	761 (612/149)	14 (4/10)	4 (1/3: 11 pb)
CK	585 (381/204)	6 (1/5)	1 (0/1: 5 pb)
Total	5646 (2976/2670)	117 (33/82)	17 (1/16: 291 pb)
Promedio/locus		16,71	2,43
Frecuencia		1/48,25	1/332,12

Con respecto a la región promotora de Hahb4 (Hahb4-p) se analizaron las secuencias correspondientes a 29 líneas puesto que, a pesar de reiterados intentos, no pudo obtenerse producto de amplificación para las líneas R426-1, R023, B99, 430-C, R431, PAC2 y RHA266. Se detectaron 39 SNP y 6 InDel ($\pi_{\text{sil}}=0,00276$; $\theta_{\text{sil}}=0,01018$). Uno de los InDel correspondió a 196 pb, el más largo dentro del conjunto total de regiones candidatas estudiadas. Se hallaron 2 haplotipos, uno de ellos se detectó únicamente en la línea B67 (haplotipo minoritario). Para el caso de la región Hahb4 también se analizaron 29 líneas. Se detectaron 7 SNP en la región no codificante ($\pi_{\text{sil}}=0,00350$; $\theta_{\text{sil}}=0,01291$), mientras que en la región codificante se hallaron 9 SNP ($\pi_{\text{sin}}=0,00426$; $\theta_{\text{sin}}=0,01567$; $\pi_{\text{no-sin}}=0,00058$; $\theta_{\text{no-sin}}=0,00215$). Del total de SNP en exones, 6 correspondieron a sustituciones sinónimas y 3 a mutaciones no sinónimas. Los cambios fueron de treonina (ACA) a serina (TCA), de alanina (GCA) a ácido glutámico (GAA) y de éste último (GAA) a glutamina (CAA). En concordancia con lo observado para la región promotora, se detectaron 2 haplotipos y sólo la línea B67 mostró el haplotipo minoritario diferenciándose del resto de las entradas (Tabla 10).

Para la región *Suntip* se analizaron el total de las líneas (36). Se hallaron 12 SNP y 2 InDel en la región no codificante ($\pi_{\text{sil}}=0,01874$; $\theta_{\text{sil}}=0,01777$), mientras que en la región codificante se encontraron 14 SNP y ningún InDel ($\pi_{\text{sin}}=0,01987$; $\theta_{\text{sin}}=0,01786$; $\pi_{\text{no-sin}}=0,00091$; $\theta_{\text{no-sin}}=0,00109$). En el primer exón se observaron 5 SNP, mientras que en el segundo exón se hallaron los 9 restantes, de los cuales 2 fueron mutaciones no sinónimas

de isoleucina (ATT) a valina (GTT) y de valina (GTC) a isoleucina (ATC). Se detectaron 6 haplotipos, el mayoritario (Hap 1) fue encontrado en 26 líneas (R426-1, R419, B67, R023, R421, R413, R437, B99, 962-C, HAR4, R420, R415, 430-C, R427, R422, R423, R049, R431, HA89, 1063-C, R043, B71, R429, R432, RHA266, PAC2), mientras que 4 líneas (R307, R073, RHA274, 964-C) mostraron el haplotipo 2 (Hap 2), 1 línea (R417) el haplotipo 3 (Hap 3), 1 línea (B59) el haplotipo 4 (Hap 4), 3 líneas (R435, R433, B10) presentaron el haplotipo 5 (Hap 5) y 1 línea (A59) el haplotipo 6 (Hap 6) (Tabla 10). En el contexto de las líneas empleadas como parentales de mapeo para tolerancia a sequía, los parentales susceptibles, A59 y B59, presentaron haplotipos distintos.

En cuanto a la región FB, el número total de líneas analizadas fue de 33, se encontraron 7 SNP y 3 InDel en la región no codificante ($\pi_{sil}=0,00056$; $\theta_{sil}=0,00229$). En la región codificante no se halló ninguno de los dos tipos de polimorfismo. Del análisis pudieron determinarse dos haplotipos, la única línea que mostró el haplotipo minoritario fue 430-C (Tabla 10).

Para la región candidata HaL1L se analizaron las secuencias correspondientes a 35 líneas y se detectaron 2 haplotipos. La única línea que mostró un haplotipo diferente al del resto fue R437 (Tabla 10). Se hallaron 2 SNP y 1 InDel en la región no codificante ($\pi_{sil}=0,00155$; $\theta_{sil}=0,00660$), mientras que se detectaron 5 SNP en la región codificante. De estos últimos, cuatro correspondieron a sustituciones sinónimas y ($\pi_{sin}=0,00186$; $\theta_{sin}=0,00791$) y 1 SNP a una mutación no sinónima ($\pi_{no-sin}=0,00014$; $\theta_{no-sin}=0,00058$) que correspondió a un cambio de treonina (ACC) por asparagina (AAC).

Se analizaron el total de las líneas (36) para la región HaDhn1, detectándose 10 SNP y 3 InDel en la región no codificante ($\pi_{sil}=0,0096$; $\theta_{sil}=0,00795$), mientras que se hallaron 4 SNP y 1 InDel en la región codificante ($\pi_{sin}=0,00320$; $\theta_{sin}=0,00152$; $\pi_{no-sin}=0,00295$; $\theta_{no-sin}=0,00163$). De los 4 SNP detectados, 3 fueron mutaciones no sinónimas de treonina (ACT y ACA) a isoleucina (ATT y ATA) y de isoleucina (ATA) a valina (GTA). El InDel en la región codificante comprendió 6 pb e involucró dos aminoácidos iguales a los dos anteriores adyacentes al InDel. Se detectaron 5 haplotipos, el mayoritario (Hap 1) estuvo presente en 17 líneas (R417, B67, R073,

HAR4, 430-C, R427, R435, R422, R049, HA89, R431, R433, R043, B10, B71, R429, R432), el haplotipo 2 (Hap 2) fue detectado en 9 líneas (R307, R419, R023, R421, R413, R415, RHA274, R423, RHA266), el haplotipo 3 (Hap 3) en 7 (R437, 962-C, R420, 1063-C, 964-C, PAC2, R426-1), el haplotipo 4 (Hap 4) fue hallado en 2 líneas (B59 y B99), mientras que el haplotipo 5 (Hap 5) sólo en 1 línea (A59). Cabe destacar que los genotipos de referencia PAC2 y RHA266 presentaron haplotipos distintos (Hap 3 y Hap 2, respectivamente). En el contexto de las líneas empleadas como parentales de mapeo para tolerancia a sequía, los parentales tolerantes B71 y R432 mostraron el haplotipo 1, mientras que el parental susceptible (B59) presentó el haplotipo 4. Asimismo, las líneas B59 y A59 también exhibieron haplotipos distintos (Hap 4 y Hap 5, respectivamente) (Tabla 10).

En cuanto a la región CK se analizaron las secuencias de 35 líneas y se encontraron 5 SNP y 1 InDel en la región no codificante ($\pi_{sil}=0,00235$; $\theta_{sil}=0,00514$), mientras que sólo 1 SNP en la región codificante. Este único polimorfismo en exones correspondió a una mutación sinónima ($\pi_{sin}=0,00134$; $\theta_{sin}=0,00294$). El InDel abarcó una extensión de 5 pb. Se detectaron 2 haplotipos, uno mayoritario y el otro sólo presente en dos líneas (R417 y B99) (Tabla 10).

Tabla 10. Descripción de los haplotipos detectados en cada región analizada. Sitios SNP en rojo: presentaron tres variantes alélicas. I: presencia del InDel (inserción), D: ausencia del InDel (delección). **Hahb4-p**: 199: A, 276-472: 196 pb, 844-845: GATCCG, 875-880: GTCT, 913-914: TCA, 1126: T. **Suntip**: 373-374: 30 pb, 8 pb (AGTGGTGT), 4 pb (AGTG), 397: A. **FB**: 62: T, 381-386: TGAC, 441-442: AACATAAATGAACA. **HaL1L**: 583-584: T. **HaDhn1**: 164-165: ACAGGA, 530: CT, 564: AAA, 565: AA. **CK**: 57-58: ACAAG. Los sitios InDel están tomados en base al haplotipo mayoritario.

Hahb4-p	199	209	230	276-472	491	550	563	588	620	704	716	735	748	758	793	794	815	835	838	840	842	844	844-845	845	871	873	874	875-880
Hap 1	I	G	T	I	G	G	G	G	C	C	G	A	T	A	G	A	C	G	C	A	G	G	D	T	C	T	T	I
Hap 2	D	A	G	D	A	A	C	A	T	T	A	G	C	T	A	G	T	T	T	C	A	A	I	C	T	C	C	D
Hahb4-p (cont.)	881	895	907	911	913-914	915	916	919	923	924	1116	1118	1119	1121	1123	1126	1133											
Hap 1	G	T	C	C	D	T	C	G	C	G	T	A	G	A	C	I	A											
Hap 2	A	C	G	T	I	G	A	C	T	A	A	G	A	C	T	D	G											
Hahb4	26	39	47	82	93	333	366	398	406	437	443	506	508	534	570	612												
Hap 1	C	G	G	A	C	G	G	A	G	T	C	C	G	T	G	G												
Hap 2	T	T	A	T	A	A	A	G	A	A	G	A	C	C	T	A												
Suntip	98	125	137	272	287	315	319	328	331	335	340	373-374	374	375	377	386	396	397	401	465	472	492	570	585	597	666	697	699
Hap 1	C	T	C	C	C	G	C	A	T	G	C	D	A	G	G	A	A	I	T	C	A	T	A	C	C	T	G	T
Hap 2	A	C	C	C	C	A	T	C	C	A	A	130	A	G	G	T	A	D	T	T	A	G	T	T	T	T	T	C
Hap 3	C	T	G	T	C	A	A	A	T	G	A	D	A	G	G	G	A	D	T	T	A	G	T	C	C	G	G	C
Hap 4	A	C	C	C	T	A	T	A	T	G	A	D	A	G	G	G	A	D	T	T	A	G	T	T	C	T	A	C
Hap 5	A	C	C	C	T	A	T	A	T	G	A	18	T	T	T	G	G	D	C	T	G	G	T	C	C	T	A	C
Hap 6	C	T	C	C	C	A	T	A	T	G	A	14	A	G	G	G	A	D	T	T	A	G	T	T	C	T	A	C

FB	12	26	62	174	374	381-386	441-442	442	447	659								
Hap 1	G	T	I	T	A	I	D	T	A	T								
Hap 2	A	C	D	A	G	D	I	C	G	G								
HaL1L	8	97	111	120	156	557	583-584	591										
Hap 1	C	T	G	T	C	T	D	T										
Hap 2	A	C	A	C	T	C	I	C										
HaDhn1	164-165	169	175	219	489	497	501	507	528	530	530	560	564	564	565	573	589	729
Hap 1	D	T	T	G	A	T	T	T	T	C	I	A	I	A	D	T	C	G
Hap 2	I	T	C	G	G	T	T	A	T	C	I	A	D	A	D	T	A	T
Hap 3	I	C	C	A	G	T	C	T	T	C	I	A	I	A	I	C	C	T
Hap 4	I	C	C	A	A	C	T	T	G	T	D	T	I	C	D	T	A	G
Hap 5	I	C	C	G	A	T	T	T	T	C	I	A	I	A	D	T	C	G
CK	55	57-58	78	80	112	184	251											
Hap 1	C	D	T	A	G	C	G											
Hap 2	T	I	C	G	C	T	C											

Como ya se mencionó, en la región *Suntip*, se detectaron sólo dos InDel en la región no codificante. Uno de ellos de longitud variable (número diverso de eventos de inserción/delección) se muestra en la Figura 20-A. La misma incluye 3 líneas: R043 que posee el haplotipo mayoritario, 964-C que presentó el haplotipo 2 y B10 que mostró el haplotipo 5 (Tabla 10).

En la Figura 20-B (región HaDhn1) se observa claramente la distinción entre la secuencia del parental de mapeo susceptible a sequía (A59) (Hap 5) y las líneas parentales tolerantes (R432 y B71) que poseen un haplotipo diferente al primero (Hap 1). Este InDel correspondió al único caso donde se registró un polimorfismo de este tipo en la región codificante, mostrando un evento de inserción en el que se repitieron los 6 nucleótidos anteriores y se codificaron los mismos dos aminoácidos adyacentes (treonina y glicina). Esta inserción fue detectada en 4 de los 5 haplotipos hallados para la región HaDhn1 (Tabla 10).



Figura 20. Estructura de InDel encontrados en una porción de las regiones *Suntip* y *HaDhn1*. A: *Suntip*. B: *HaDhn1*. Los rectángulos indican la posición de los polimorfismos InDel.

En su conjunto, las regiones candidatas analizadas mostraron bajos niveles de variabilidad (Tabla 11). La variación genética a nivel nucleotídico fue estimada como la diversidad nucleotídica en promedio para todas las regiones candidatas evaluadas ($\pi_T=0,00294$) y como el número de sitios segregantes ($\theta_W= 0,00501$). Los valores de diversidad promedio para el total de sitios silenciosos ($\pi_{sil}=0,00558$) y el total de sitios sinónimos

($\pi_{\text{sin}}=0,00508$) fueron superiores a los valores para las mutaciones no sinónimas ($\pi_{\text{no-sin}}=0,00076$). El patrón de diversidad superior para sustituciones sinónimas que para mutaciones no sinónimas se reflejó en el total de las regiones estudiadas. Para FB ambos valores de diversidad (π_{sin} y $\pi_{\text{no-sin}}$) fueron iguales a 0, mientras que para CK sólo el valor de $\pi_{\text{no-sin}}$ correspondió a 0.

La cantidad de haplotipos hallados en el conjunto de regiones estudiadas osciló entre 2 y 6, se pudo observar que la mayoría (71,4%) de las regiones exhibió dos haplotipos. El número de haplotipos promedio fue de 3 (Tabla 11).

La diversidad haplotípica varió entre 0,057 y 0,692, con un promedio de 0,218. En HaL1L, FB y Hahb4 el haplotipo minoritario estuvo presente sólo en una línea. Esta situación lleva a que las frecuencias alélicas de las variantes minoritarias para cada sitio SNP sean muy bajas ($\sim 0,03$). Un patrón similar se observó en la región CK, donde sólo dos líneas del total analizado mostraron el haplotipo minoritario, con lo cual las frecuencias alélicas de los sitios polimórficos para este caso fueron levemente superiores a lo descrito en HaL1L, FB y Hahb4.

Tabla 11. Estimaciones de variabilidad en 7 regiones candidatas para tolerancia a déficit hídrico. a. Sitios informativos usados en las medidas de diversidad nucleotídica. b. N° Hap: Número de haplotipos. c. DH: Diversidad haplotípica. d. El valor de Eta (número total de mutaciones) fue superior (25) al SI, dado que para 2 sitios polimórficos se detectaron dos variantes SNP.

Región	SI ^a	θ_w	π_T	π_{sil}	π_{sin}	$\pi_{\text{no-sin}}$	$\frac{\pi_{\text{no-sin}}}{\pi_{\text{sin}}}$	N° Hap ^b	DH ^c
Hahb4-p	39	0,01018	0,00276	0,00276	-	-	-	2	0,069
Hahb4	16	0,00664	0,00180	0,00350	0,00426	0,00058	0,13615	2	0,069
Suntip	23 ^d	0,00734	0,00827	0,01874	0,01987	0,00091	0,04579	6	0,470
FB	7	0,00164	0,00040	0,00056	0	0	-	2	0,061
HaL1L	7	0,00265	0,00062	0,00155	0,00186	0,00014	0,07526	2	0,057
HaDhn1	13	0,00417	0,00562	0,00960	0,00320	0,00295	0,92187	5	0,692
CK	6	0,00249	0,00114	0,00235	0,00134	0	0	2	0,111
Promedio	15,8	0,00501	0,00294	0,00558	0,00508	0,00076	0,19	3	0,218

R.2.2.2. Pruebas de neutralidad

El estadístico D de Tajima (Tajima, 1989) y los D* y F* de Fu y Li (1993) mostraron resultados consistentes para las 7 regiones estudiadas. La única excepción se observó para el valor de D* en HaDhn1 (1,503) que resultó significativo mientras que los otros dos estimadores no lo fueron (Tabla 12). Tanto el D de Tajima como D* y F* fueron negativos y significativamente diferentes de cero para las regiones Hahb4-p, Hahb4, FB y HaL1L, revelando un exceso de variantes raras o en frecuencias bajas. Las regiones *Suntip* y CK no evidenciaron desvíos a la neutralidad en base a los resultados obtenidos en esta sección. Para el caso de HaDhn1, donde se obtuvo un valor de D* positivo y significativamente diferente de cero, esto indicaría un exceso de variantes en frecuencias intermedias. De todos modos, este resultado no es contundente dado que los otros dos estimadores no fueron significativos.

Tabla 12. Pruebas de neutralidad para las 7 regiones candidatas estudiadas. Los niveles de significación corresponden a: *p<0,05; **p<0,02; ***p<0,01; ****p<0,001.

Región	D de Tajima	D*	F*
Hahb4-p	-2,697****	-4,958**	-4,979**
Hahb4	-2,493****	-4,356**	-4,425**
<i>Suntip</i>	0,126	0,609	0,53
FB	-2,178***	-3,721**	-3,796**
HaL1L	-2,18***	-3,792**	-3,854**
HaDhn1	1,107	1,503*	1,618
CK	-1,494	1,202	0,446

En cuanto al análisis por ventanas de distancias entre sitios polimórficos para detectar desvíos a la neutralidad según el estimador D de Tajima, se hallaron resultados significativos para las regiones Hahb4-p, Hahb4 y HaL1L. Para la primera región se encontró que toda la porción analizada en este trabajo mostró valores de D negativos y significativos (desde D=-2,09; p<0,05 hasta D=-2,62; p<0,001). La región Hahb4 mostró D significativo entre 1-200 pb (D=-2,0068; p<0,05), entre 250-500 pb (D=-2,0984; p<0,05) y entre 550-600 pb (D=-2,17; p<0,05). Las ventanas de análisis para estas dos regiones fueron mayores que para las demás porque se detectaron resultados significativos entre la mayoría de los puntos en estudio. En tanto que para HaL1L se detectó sólo una fracción significativa

para la prueba de neutralidad de Tajima ($D=-1,8863$, $p<0,05$) para 76-175 pb.

Dado que las pruebas de neutralidad han sido desarrolladas en base a la dinámica esperada en poblaciones naturales, el resultado de su aplicación en un conjunto de líneas endocriadas puede llevar a conclusiones erróneas y debe ser interpretado con cautela. Con el objeto de lograr una mejor aproximación a esta condición, se calcularon los D de Tajima para cada una de las regiones considerando los grupos identificados como unidades panmícticas en el análisis de estructura poblacional mediante el método bayesiano (rojo y verde, Figura 17). Los resultados de este análisis se muestran en la Tabla 13. Las regiones Hahb4-p, Hahb4, FB y HaL1L mostraron valores de D negativos y significativos, coincidente con lo reportado en la Tabla 12. Este patrón se mantuvo en los resultados obtenidos considerando los subgrupos de acuerdo al método bayesiano.

Tabla 13. Prueba de neutralidad D de Tajima para las 7 regiones candidatas estudiadas en el conjunto de líneas y en dos grupos. Los niveles de significación corresponden a: * $p<0,05$; ** $p<0,02$; *** $p<0,01$; **** $p<0,001$.

Región	Todas las líneas D Tajima	Grupo verde D Tajima	Grupo rojo D Tajima
Hahb4-p	-2,697****	-2,319***	-
Hahb4	-2,493****	-2,196**	-
<i>Suntip</i>	0,126	1,24	0,9791
FB	-2,178***	-1,94*	-
HaL1L	-2,18***	-	-2,080*
HaDhn1	1,107	0,058	1,69
CK	-1,494	-0,57897	-

Para discriminar entre selección y un cuello de botella severo como posibles causales de los valores de D negativos y significativos para 4 de las 7 regiones candidatas analizadas, se realizó el análisis propuesto por Cornuet and Luikart (1996) para detectar reducciones recientes en el tamaño poblacional, utilizando los datos de la genotipificación de los marcadores SSR, tanto en el conjunto de las 34 líneas, como dividiéndolas en los dos grupos detectados con el método bayesiano (Figura 17). Todas las comparaciones realizadas, con excepción de *Suntip* en el grupo rojo,

mostraron una heterocigosis esperada obtenida en base a las frecuencias alélicas observadas (H_e) mayor a la heterocigosis esperada para una población en equilibrio deriva-mutación dado un determinado número de alelos (H_eq) (Tabla 14). Los valores de significación obtenidos en la prueba de Wilcoxon al considerar la hipótesis alternativa de exceso de heterocigotas fueron de $p=0,00391$ para el conjunto de las líneas, de $p=0,00391$ para el grupo verde y de $p=0,01172$ para el grupo rojo. En base a estos resultados, se podría inferir un cuello de botella moderado, que se manifiesta a través de valores de heterocigosis esperada superiores a los predichos por la teoría neutralista para una población en equilibrio deriva-mutación, es decir, a través del exceso de alelos en frecuencias intermedias. Este resultado permite descartar un cuello de botella severo como responsable de los valores negativos y significativos del D de Tajima.

Tabla 14. Estimaciones de heterocigosis esperada para detectar desvíos de la teoría neutralista en las 7 regiones candidatas estudiadas, para el conjunto de líneas y en dos grupos. NA: número de alelos. Obs: heterocigosis observada. IAM: modelo de alelos infinitos.

Región	Todas las líneas			Grupo verde			Grupo rojo		
	NA	Obs	IAM	NA	Obs	IAM	NA	Obs	IAM
Hahb4-p	3	0,315	0,302	2	0,363	0,187	2	0,21	0,193
Hahb4	4	0,65	0,392	4	0,457	0,439	3	0,63	0,327
<i>Suntip</i>	4	0,602	0,393	4	0,613	0,438	2	0,093	0,195
FB	6	0,771	0,536	3	0,556	0,338	4	0,584	0,426
HaL1L	4	0,537	0,395	4	0,579	0,438	3	0,46	0,318
HaDhn1	3	0,661	0,288	3	0,601	0,343	3	0,586	0,33
CK	3	0,497	0,302	3	0,509	0,33	2	0,483	0,186

R.2.3. Análisis de desequilibrio de ligamiento en las regiones candidatas

El número total de comparaciones de a pares entre sitios polimórficos, el valor promedio de r^2 y el rango de distancias entre cada uno de los sitios pertenecientes a cada región analizada, se muestran en la Tabla 15. Para todos los cálculos de significación, tanto el método de Fisher como el de Chi-cuadrado arrojaron la misma cantidad de comparaciones significativas, excepto para la región HaDhn1 donde el segundo método detectó 13 comparaciones significativas adicionales. Los rangos de distancias entre sitios comparados variaron entre 1 pb y 917 pb.

Tabla 15. Análisis de DL en las 7 regiones candidatas estudiadas. a. Número de comparaciones significativas con la prueba de Fisher. b. Número de comparaciones significativas con la prueba de Chi-cuadrado.

Región	Nº sitios polimorf.	Nº comparac. de a pares	F ^a	X ^b	Distancia	r ² promedio
Hahb4-p	39	741	741	741	1-917	1
Hahb4	16	120	120	120	2-586	1
<i>Suntip</i>	21	210	131	131	2-605	0,349
FB	7	21	21	21	5-647	1
HaL1L	7	21	21	21	9-583	1
HaDhn1	13	78	33	46	4-560	0,279
CK	6	15	15	15	2-196	1

El DL arrojó valores de r^2 promedio de 1 en 4 de las regiones candidatas estudiadas (Tabla 15, Figura A.2), sólo *Suntip* y HaDhn1 mostraron valores menores a las restantes (0,349 y 0,279; respectivamente). Cabe aclarar que, para el cálculo del r^2 en la región Hahb4, se consideraron conjuntamente tanto la porción correspondiente al promotor como la región transcrita. Para *Suntip*, el r^2 promedio decayó hasta el valor de 0,325 a las 207 pb, mientras que para HaDhn1 se observó un decaimiento hasta 0,249 a las 78 pb (Figura 21).

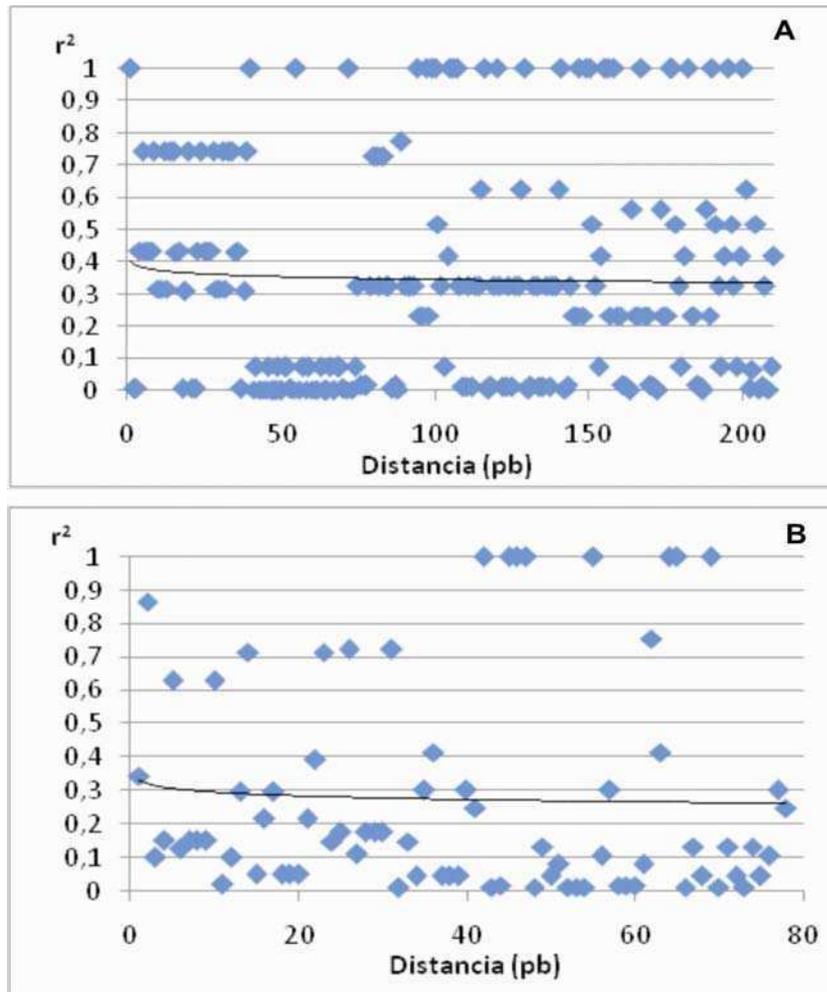


Figura 21. Decaimiento del DL en función de r^2 para dos regiones candidatas. A. Suntip. B. HaDhn1.

La Figura 22 muestra el decaimiento del DL a lo largo de todas las regiones estudiadas en base a los valores promedios de r^2 . Los gráficos de dispersión de los valores de r^2 en función de la distancia física entre todos los pares de SNP para cada una de las regiones, se adjuntan en el Apéndice (Figura A.2). En este trabajo los patrones de DL son expuestos a distancias cortas puesto que las regiones estudiadas no superan los 1500 pb (1,5 Kpb). Para determinar el decaimiento del DL en un amplio rango de distancia, se realizó la extrapolación en base a la ecuación que ajustó la línea de tendencia logarítmica de la Figura 22.

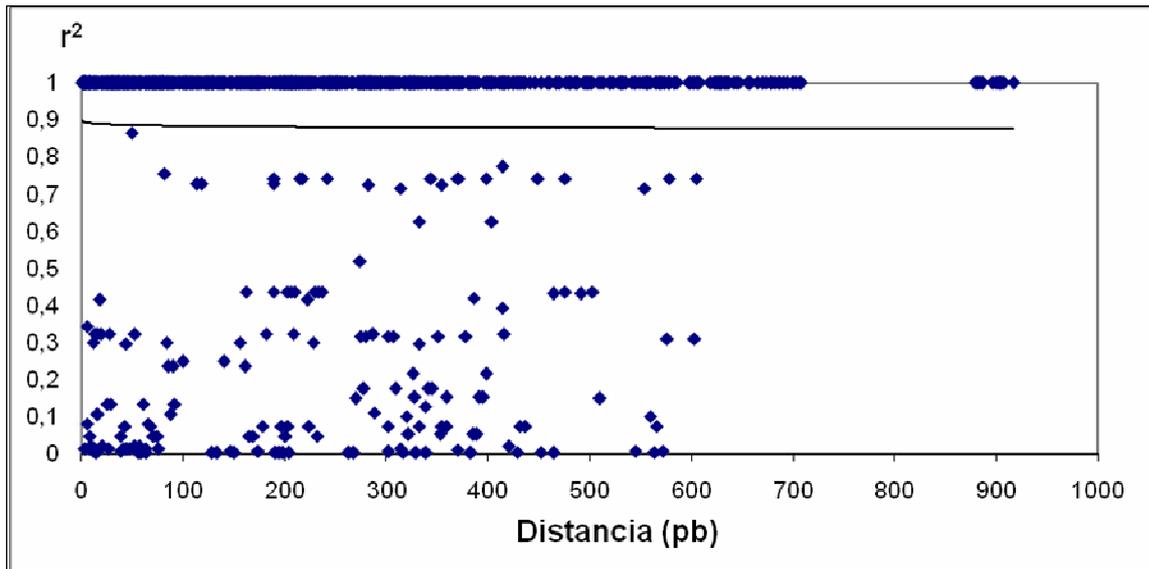


Figura 22. Decaimiento global del DL en función de r^2 . Se representan las 7 regiones candidatas estudiadas.

DISCUSIÓN

D.1. Caracterización fenotípica

La evaluación del carácter de tolerancia al déficit hídrico en ensayos de campo resulta sumamente compleja porque hay diversos factores involucrados tales como las condiciones ambientales no controladas, la homogeneidad insuficiente del suelo, el alto número de materiales vegetales a evaluar, los costos de tiempo y trabajo (Turhan and Baser, 2004) y, principalmente, la escasa definición sobre los caracteres fenotípicos asociados a la tolerancia (Poormohammad Kiani *et al.*, 2007).

Como ya se mencionó, el rendimiento en girasol involucra a tres componentes, el número de capítulos por unidad de superficie, el número de frutos llenos por capítulo y el peso individual de esos frutos. Dado que, los caracteres agronómicos peso de cien semillas (P100) y número de semillas por capítulo (NSC), son componentes del carácter complejo rendimiento (RENSH), éstos mostraron un patrón similar en los diversos genotipos analizados en los ensayos de planta adulta (Figuras 3, 4 y 6). Las líneas que mostraron las mayores diferencias entre los valores obtenidos para los tres caracteres bajo la condición de riego con respecto a valores inferiores en la condición de sequía (B67, HA89, B59, R420, B99 y Aguara), podrían clasificarse como susceptibles. Esto es coincidente con los resultados publicados por Alemano *et al.* (2005), donde HA89, B59, R420 y B99 se reportaron como susceptibles según su rendimiento relativo al control irrigado en ensayo de campo. Mientras que, las líneas R437, R431, B71 y R433 se comportaron mejor en sequía sólo para el carácter P100, por lo tanto podrían clasificarse como tolerantes. Tres de ellas (437, R431 y B71) resultaron clasificadas del mismo modo por Alemano *et al.* (2005). La respuesta diferencial de cada genotipo en sequía o riego para P100, NSC y RENSH puede adjudicarse a la capacidad genética del material para tolerar o no el déficit hídrico.

En concordancia con lo descrito por Roche *et al.* (2004), 14 de las entradas aquí estudiadas, incluyendo líneas tolerantes, intermedias y

susceptibles, mostraron modificaciones en la composición bioquímica de las semillas (mayor porcentaje de materia grasa) en respuesta al déficit hídrico (Figura 5). Esta observación sugiere que este tipo de modificaciones bioquímicas no estarían relacionadas con su tolerancia o susceptibilidad a la sequía.

Dado que el rendimiento es un carácter cuantitativo, son muchos los genes involucrados en la respuesta de la planta, y su evaluación precisa en condiciones de campo resulta dificultosa y lenta. Para acelerar el proceso de mejoramiento es importante utilizar técnicas sencillas, rápidas y económicas que permitan agilizar la obtención de nuevos materiales mejor adaptados a condiciones de sequía. Por ello, se propone el ensayo en condiciones controladas utilizado en esta tesis, como una alternativa para la caracterización rápida de genotipos susceptibles al déficit hídrico simulado (*in vitro*). Para estudiar la tolerancia al déficit hídrico en condiciones controladas se han utilizado diversos agentes osmóticos, cuya función es reducir la disponibilidad de agua en el sustrato donde la semilla germina y crece. La evaluación de genotipos empleando soluciones osmóticas resulta apropiada para diferenciar materiales resistentes que puedan ser utilizados como parentales de poblaciones de mapeo (Blum, 1988; Teruel *et al.*, 2008). Entre los agentes desecantes comúnmente usados se encuentran el polietilenglicol (PEG), el manitol y la sacarosa en distintas concentraciones (Blum and Ebercon, 1981; Farooq and Azam, 2006). Se ha reportado que el empleo de PEG produce una mayor cantidad de semillas anormales que la sacarosa, la cual posee un efecto protector al penetrar al interior de las semillas durante el tratamiento (Hernández, 2005). En este trabajo de tesis se empleó el manitol como agente osmótico dado que, la sacarosa favorece la contaminación con hongos, mientras que el PEG modifica los niveles de oxígeno disueltos en la solución, dificultando su uso en ensayos de más de 72 h. Existen numerosos trabajos realizados en condiciones controladas de laboratorio para estudiar el efecto del déficit hídrico en diversos cultivos, entre ellos arroz (Pérez-Molphe-Balch *et al.*, 1996), trigo (Blum, 1993), maíz (Alemanno *et al.*, 2003; Biasutti and Galiñanes, 2001; Chimenti and Hall, 1993; Teruel *et al.*, 2008; Trapani and Gentinetta, 1984), cebada (González,

2008) y girasol (Alemano *et al.* 2005; Hernández, 2005; Turhan and Baser, 2004).

En el trabajo de Turhan and Baser (2004) se determinó correlación entre la respuesta de la planta a sequía en el campo e *in vitro* para concentraciones bajas de PEG (3, 6 y 9%). A medida que la concentración del agente osmótico aumentó, los genotipos se vieron cada vez más afectados en su supervivencia; mientras que con 12% no sobrevivió ningún explante. Asimismo, estos autores reportaron que el carácter Peso Fresco (PF) fue el más afectado en la evaluación *in vitro*. Por otro lado, el trabajo de Hernández (2005), demostró que a medida que aumentaba el tiempo de imbibición de las semillas y la concentración de soluciones de PEG y sacarosa, se incrementó el porcentaje de plántulas anormales desde 1-2% hasta 15%, a las 25 h posteriores a la imbibición con incubación a 20°C. Asimismo, a medida que el tratamiento se prolongó, principalmente a más de 48 h, la cantidad de semillas con germinación anormal fue mayor.

En los ensayos controlados realizados en esta tesis, la variable Recuento Final de Germinación (RFG), dio lugar a la delimitación de las tres categorías de materiales definidas de acuerdo con la evaluación del rendimiento relativo en el ensayo de campo con plantas adultas. En cambio, las variables Peso Fresco (PF) y Peso Seco (PS) no permitieron discriminar claramente los genotipos en cuanto a su comportamiento frente al déficit hídrico, tanto en condiciones moderadas como severas, coincidentemente con lo reportado por Alemano *et al.* (2005). Por lo tanto, la variable Recuento Final de Germinación (RFG) fue la que más se asemejó al patrón de respuesta obtenido en la evaluación del rendimiento relativo en ensayo de campo. Dado que, las raíces están activamente implicadas en la respuesta al déficit hídrico moderado a través del control de la absorción de agua desde el suelo (Turhan and Baser, 2004), esto podría explicar que la mayoría de los defectos (60%) encontrados en este trabajo se hayan evidenciado en raíces.

El trabajo de Alemano *et al.* (2005) incluyó las 30 líneas de girasol estudiadas en el ensayo de campo detallado en la sección M.1.2, siendo 7 de ellas caracterizadas también en el estadio de plántula bajo condiciones controladas en un ensayo similar al utilizado en este Capítulo. Los resultados

obtenidos en el presente trabajo mostraron sólo algunas coincidencias con lo descrito por Alemano *et al.* (2005) para el ensayo de plántula. Estos autores reportaron que las líneas R432, B71 y HAR4 resultaron tolerantes en ambas condiciones de estrés, mientras que HA89 fue intermedia y B59 se comportó como susceptible. La línea RHA274 se comportó erráticamente, dado que resultó tolerante a 200mM y susceptible con déficit hídrico severo. Los resultados hallados en el presente trabajo de tesis mostraron que R432 y B71 fueron tolerantes a 200mM, pero con déficit hídrico severo R432 fue intermedia y B71 resultó susceptible. Por su parte, HAR4 resultó intermedia a 200mM y susceptible en solución de manitol 400mM, mientras que la línea HA89 se comportó como tolerante con déficit hídrico moderado y susceptible con déficit hídrico severo, al igual que RHA274. Finalmente, la entrada B59 se comportó como susceptible a 200mM e intermedia a 400mM. Estas diferencias en cuanto a la respuesta de los genotipos frente a las condiciones de déficit hídrico simulado en laboratorio, pueden deberse a las condiciones ambientales de crecimiento durante el ensayo. En este trabajo se empleó una cámara de germinación con temperatura y fotoperíodo constantes (25°C, con 16 h de luz y 8 h de oscuridad), mientras que los mencionados autores utilizaron una cámara de crecimiento a 28°C con 16 h de luz y a 18°C con 8 h de oscuridad.

En la condición de déficit hídrico severo se registró una baja probabilidad de éxito de germinación (Figura 14), un alto número de plántulas con defectos (coincidente con Hernández, 2005) y una baja correlación entre la clasificación de los genotipos en ensayo de campo y bajo condiciones controladas con déficit hídrico moderado, hecho que dificulta la interpretación de los resultados y sugiere que se trata de una condición extrema. Por lo tanto se emplearon los resultados obtenidos en la condición de déficit hídrico moderado para hacer las comparaciones con el ensayo a campo de plantas adultas.

La comparación entre las líneas clasificadas como tolerantes en el campo y en el ensayo en condiciones controladas con déficit hídrico moderado arrojó tres coincidencias, es decir que en ambos casos se determinaron como materiales tolerantes a R432, B71 y R431. Esta última línea fue la única que también resultó tolerante en condición de déficit hídrico

severo. En cuanto a la comparación para discriminar las líneas susceptibles al déficit hídrico moderado, se detectaron cuatro coincidencias correspondiendo a los genotipos R420, B59, B67 y R307; de los cuales el último también resultó susceptible en condición de déficit hídrico severo. Es decir que, tanto en la caracterización fenotípica en ensayo de campo (estadío de planta adulta) como en ensayo bajo condiciones controladas con déficit hídrico moderado (estadío de plántula), se determinaron como tolerantes R432, B71 y R431, en concordancia con la elección de dos de ellos como parentales para el desarrollo de poblaciones de mapeo. Del mismo modo, B59 fue elegido como padre susceptible y cruzado con R432 y B71 en el contexto del programa de mejoramiento para tolerancia a sequía del BAG-IM. En el caso de las líneas que no presentaron el mismo comportamiento en el ensayo de campo y bajo condiciones controladas, esto podría atribuirse a que en laboratorio sólo se evalúa el efecto del déficit hídrico en estadío de plántulas (hasta los 11 días posteriores a la siembra), mientras que en campo se considera todo el ciclo de la planta hasta la senescencia. Los efectos de la interacción con el ambiente y el modo de inducción de la sequía son factores que también podrían influir en el comportamiento de las líneas.

De esta etapa del trabajo se concluye que la evaluación de la tolerancia al déficit hídrico moderado en girasol a través del ensayo en plántula bajo condiciones controladas, es una herramienta apropiada para realizar análisis exploratorios que luego deberán corroborarse en ensayos de campo.

D.2. Caracterización genética de las líneas para tolerancia a sequía

D.2.1. Diversidad genética y estructura poblacional en base a marcadores SSR

El análisis de heterogeneidad de las líneas en estudio se llevó a cabo en base a los criterios propuestos por Heckenberger *et al.* (2002). El valor obtenido en este trabajo para la ocurrencia promedio de casos normales

(80,67) fue levemente inferior al reportado por estos autores para líneas endocriadas de maíz (86,6). Sin embargo, el rango de valores entre las líneas analizadas en este trabajo fue mayor (42,85-100) al informado por los mencionados autores (78,7-95), indicando una mayor heterogeneidad global de las entradas aquí estudiadas. Las líneas más variables fueron HA89, R415, A59 y R429 (Tablas 7 y 8). Asimismo, las 7 líneas analizadas en el Capítulo 1 mostraron mayor heterogeneidad que las estudiadas en esta sección, puesto que la ocurrencia promedio de casos normales fue de 67,6. Las posibles causas de heterogeneidad han sido expuestas de manera extensa en el Capítulo 1, por lo tanto sólo se mencionan los motivos genéticos y/o técnicos aplicables en esta sección. Dentro de las causas genéticas se incluyen las mutaciones *de novo* y la presencia de heterocigosis residual (Cooke *et al.*, 2003; Gunjaca *et al.*, 2008), mientras que como causa técnica podría citarse el efecto inesperado de deslizamiento de la ADN polimerasa durante la reacción de PCR. El caso de heterogeneidad con mayor ocurrencia en el conjunto de líneas estudiadas aquí correspondió al caso 8 (7,98; Tabla 8), coincidentemente con lo reportado por Heckenberger *et al.* (2002). Mientras que las 7 líneas del Capítulo 1, mostraron la mayor ocurrencia para el caso 3 (13,8). Estas diferencias podrían corresponder a fallas repetidas en la amplificación por PCR para determinados individuos, con lo cual no es posible genotipificarlos y por ende finalmente se reporta como “dato faltante” (caso 8, ver Materiales y Métodos del Capítulo 1).

El porcentaje de líneas que mostraron cohesión entre sus individuos fue menor (47,06%, Figura 15) en este Capítulo, que para las 7 líneas estudiadas en el Capítulo 1 (85,72%, Figura 18), lo cual puede deberse principalmente, al número de loci SSR analizados para cada caso. Esta interpretación se basa en el hecho que líneas cuyos tres individuos presentaron el mismo genotipo para todos los loci se agruparon en el mismo *cluster* con individuos de otras líneas también homogéneas, lo cual indicaría una deficiente discriminación de los marcadores empleados en este Capítulo.

En cuanto a las estimas de diversidad, la heterocigosis esperada promedio de las 34 líneas (0,57) fue similar al valor hallado para el total de

las entradas analizadas en el Capítulo 1 (0,60). Esto indica una elevada variabilidad para líneas pertenecientes al programa de mejoramiento para tolerancia al déficit hídrico. Esta variabilidad resulta coincidente con la diversidad de orígenes de los materiales incluidos en el programa de mejoramiento. Cabe destacar que las estimas de diversidad deben ser consideradas con cautela, dado el elevado nivel de heterogeneidad detectado en las líneas estudiadas.

El método bayesiano (Pritchard *et al.*, 2000) utilizado para el análisis de estructura poblacional reveló que las líneas endocriadas examinadas en este trabajo muestran sub-estructuración, con una distribución en dos grandes grupos (K=2) (Figura 17) y estructuras más internas o profundas con asignaciones de los materiales a 6 y 9 grupos o acervos genéticos diferentes (K=6, Figura 18 y K=9, Figura 19). Estos resultados podrían ser una de las causas de los desvíos a la neutralidad detectados en el análisis de regiones candidatas. La estructuración hallada en las líneas es esperable, considerando que las mismas han sido seleccionadas como participantes del programa de mejoramiento orientado al carácter de sequía a partir de diversos orígenes. Los resultados obtenidos reafirman la necesidad de disponer de información acerca de la estructura poblacional para futuros estudios de mapeo por asociación que involucren a las líneas empleadas en este trabajo.

D.2.2. Frecuencia de SNP y diversidad nucleotídica

El estudio de diversidad nucleotídica se efectuó en base a un grupo de regiones candidatas sobre las cuales se determinaron los polimorfismos SNP e InDel. Las secuencias estudiadas se eligieron debido a su probable rol en la respuesta al déficit hídrico según diversos reportes previos (Cellier *et al.*, 1998; Cellier *et al.* 2000; Charles and Halliwell, 1981; Dezar *et al.*, 2005; Gago *et al.*, 2002; Moons *et al.*, 1995; Parcy *et al.*, 1997; Poormahammad Kiani *et al.*, 2007; Roche *et al.*, 2007; Sarda *et al.*, 1997; Sarda *et al.*, 1999; Zrenner *et al.*, 1996). De un total de 15 regiones, 7 fueron secuenciadas y caracterizadas en función de los polimorfismos detectados.

El porcentaje de regiones candidatas que lograron amplificar un producto único fue levemente menor (53,30%) a lo reportado por Fusari *et al.* (2008) de 65,5%. Esto puede deberse a diferencias en la calidad de las secuencias utilizadas para el diseño de los oligonucleótidos de partida en cada caso.

La frecuencia de SNP encontrada en este trabajo fue de 1/48,25 pb, cifra superior a lo observado en otros cultivos de interés agronómico como sorgo (*Sorghum bicolor*) con 1 SNP/123 pb (Hamblin *et al.*, 2004), soja (*Glicine max*) con 1 SNP/328 y 1 SNP/536 pb (Van *et al.*, 2005; Zhu *et al.*, 2003) y arroz (*Oryza sativa*) que mostró 1 SNP/113 pb y 1 SNP/100 pb (Garris *et al.*, 2003; Monna *et al.*, 2006). Este hallazgo resulta interesante porque indicaría un mayor nivel de polimorfismo de la especie en estudio. El valor obtenido en este trabajo resultó similar al de Kolkman *et al.* (2007) de 1 SNP/45,7 pb en 10 líneas endocriadas de girasol, mientras que fue superior a la frecuencia obtenida por Fusari *et al.* (2008) de 1/69 pb para 19 líneas de girasol analizadas. Mientras que Liu y Burke (2006) reportaron una frecuencia más elevada (1 SNP/39 pb) para 16 cultivares primitivos y mejorados de girasol, similar a lo informado por Giordani *et al.* (2010) de 1SNP/34 pb en 8 líneas endocriadas. La misma tendencia se observó al comparar la frecuencia de polimorfismos InDel obtenida en este trabajo (1/332,12 pb) con lo reportado por Kolkman *et al.* (2007) (1/277 pb), Fusari *et al.* (2008) (1/377,6 pb) y Liu y Burke (2006) (1/264,7 pb).

Las estimas de diversidad obtenidas en este trabajo, $\theta_W=0,00501$ y $\pi_T=0,00294$, fueron menores a lo reportado por Kolkman *et al.* (2007) ($\theta_W=0,0094$, $\pi_T=0,0107$), por Fusari *et al.* (2008) ($\theta_W=0,0056$, $\pi_T=0,0061$) y por Liu y Burke (2006) para los materiales cultivados ($\theta_W=0,0072$, $\pi_T=0,0056$). Estos resultados indicarían que la muestra analizada exhibió una moderada proporción de sitios segregantes similar a lo reportado por Fusari *et al.* (2008), mientras que las diferencias alélicas entre pares de secuencias fueron bajas por ende la diversidad nucleotídica hallada en este trabajo fue menor que la reportada en la bibliografía. Estas marcadas diferencias en los valores de π pueden explicarse porque en Hahb4-p, Hahb4, FB y HaL1L se encontró sólo una línea que mostró el haplotipo minoritario y en CK sólo dos entradas presentaron la variante menos

frecuente (Tabla 10). Por lo tanto, la diferenciación promedio entre pares de secuencias fue baja.

La frecuencia de SNP para la región codificante total fue de 1 SNP cada 90,18 pb; mientras que para la región no codificante fue de 1 SNP cada 32,56 pb. La frecuencia en la región codificante total resultó menor a lo reportado por Kolkman *et al.* (2007) (1/62,8 pb), mientras que similar para la frecuencia de la región no codificante (1/32,1 pb). La comparación de estos resultados con lo obtenido por Fusari *et al.* (2008) (1/87 pb para la región codificante y 1/40,5 pb para la región no codificante) mostró un patrón similar. El 68,42% de los SNP en las regiones codificantes correspondieron a sustituciones sinónimas, lo cual refleja la tendencia de conservación de sitios.

Para el caso de los polimorfismos InDel, la tendencia fue la misma que lo detectado para SNP. La mayoría de ellos se hallaron en las regiones no codificantes de los genes candidatos analizados, siendo la región HaDhn1 la única que presentó un InDel en la zona codificante.

La diversidad nucleotídica en sitios silenciosos ($\pi_{\text{sil}}=0,00558$) fue similar a la de sitios sinónimos ($\pi_{\text{sin}}=0,00508$) y ambas resultaron muy superiores a la diversidad de sitios no-sinónimos ($\pi_{\text{no-sin}}=0,00076$), de acuerdo a lo esperado puesto que las regiones codificantes suelen ser más conservadas que las no codificantes. Se observó que estos resultados fueron inferiores a los obtenidos por los autores anteriormente mencionados. Kolkman *et al.* (2007) reportaron π_{sil} de 0,0135, π_{sin} de 0,0176, $\pi_{\text{no-sin}}$ de 0,0028. Fusari *et al.* (2008) obtuvieron un valor de $\pi_{\text{sil}}=0,0140$, de $\pi_{\text{sin}}=0,0174$, de $\pi_{\text{no-sin}}=0,0013$, mientras que Liu y Burke (2006) reportaron π_{sil} de 0,0096, π_{sin} de 0,0128, $\pi_{\text{no-sin}}$ de 0,0015. Este patrón indicaría, nuevamente, una menor diversidad nucleotídica de las líneas estudiadas en este trabajo. El valor del cociente entre cambios no-sinónimos y sinónimos ($\pi_{\text{no-sin}}/\pi_{\text{sin}}=0,19$) fue intermedio entre lo que informan Kolkman *et al.* (2007) (0,16) y Liu y Burke (2006) (0,21). 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.

Para la región *Suntip* se detectaron 6 haplotipos. Las líneas parentales susceptibles (B59 y A59) de las poblaciones de mapeo para tolerancia a sequía desarrolladas en el BAG-IM, mostraron haplotipos diferentes (Hap 4 y Hap 6, respectivamente) (Tabla 10). Esta situación podría ser consecuencia de la heterogenidad detectada en las líneas con marcadores neutros, principalmente para A59, dado que ambas comparten el mismo fondo genético (Tablas 1 y 2) y sólo se diferencian en su comportamiento frente a la fertilidad. La línea B59 es mantenedora de la fertilidad, mientras que A59 es andro-estéril (CMS: *Cytoplasmic Male Sterility*), porque su pólen no es funcional. La característica CMS es una disfunción citoplásmica que se relaciona con regiones, llamadas PET1, asociadas a marcos de lectura abiertos quiméricos en el genoma mitocondrial (ORFH522), originando proteínas que interfieren con la funcionalidad de la mitocondria y el desarrollo de pólen (Schnable and Wise, 1998). El ORFH522 presente actualmente en girasol cultivado, proviene del genoma mitocondrial de *H. petiolaris* que fue cruzado con *H. annuus*. Dado que, las regiones candidatas analizadas son, probablemente, de localización nuclear, la característica CMS diferencial para las líneas B59 y A59 no explicaría los haplotipos diversos entre ellas.

Existen estudios previos acerca del rol de *Suntip* en la respuesta de las plantas de girasol al déficit hídrico en cuanto a análisis de transcritos y funcionales (Poormohammad Kiani *et al.*, 2007; Poormohammad Kiani *et al.*, 2009; Sarda *et al.*, 1997; Sarda *et al.*, 1999), pero no se han publicado trabajos en cuanto a su variabilidad nucleotídica. En este aspecto, se dispone de información para proteínas canales de agua (*aquaporins*) de la familia PIP (*Plasma Membrane Intrinsic Protein*) que está directamente relacionada con la familia TIP (*Tonoplast Intrinsic Protein*), a la que pertenece *Suntip*. En este trabajo se detectaron 6 haplotipos, levemente inferior al número promedio (9) reportado para 11 regiones correspondientes a canales de agua de la familia PIP en especies de árboles neotropicales (Audigeos *et al.*, 2010). El mismo patrón se observó al comparar el resultado de diversidad haplotípica hallado en este trabajo (0,47) con respecto al

reportado por los mencionados autores (0,63). Se obtuvieron valores superiores de diversidad para los estimadores π_T (0,00827) y θ_W (0,00734), comparado con lo informado por Audigeos *et al.* (2010) de 0,00273 para π_T y de 0,00235 para θ_W . Esto indicaría mayor variabilidad nucleotídica dentro de las secuencias de la familia TIP analizadas en el presente trabajo, lo cual puede relacionarse con la especie en estudio, con las entradas analizadas y con propiedades intrínsecas de las regiones estudiadas. De acuerdo con estos resultados, la región *Suntip* mostró un alto nivel de diversidad, hecho que la convierte en una candidata promisoría para futuros estudios de mapeo por asociación para el carácter de tolerancia al déficit hídrico.

En la región HaDhn1 se detectaron menores valores de π_T (0,0056), π_{sin} (0,0032) y π_{no-sin} (0,00295) que los reportados por Natali *et al.* (2003) para cultivares de girasol ($\pi_T=0,0084$, $\pi_{sin}=0,00514$, $\pi_{no-sin}=0,00697$). El mismo patrón se determinó al comparar los resultados de esta tesis frente a los reportados en el trabajo de Giordani *et al.* (2010), en base a 8 líneas endocriadas ($\pi_T=0,01498$, $\pi_{sin}=0,02615$, $\pi_{no-sin}=0,00634$). Los polimorfismos detectados en esta tesis para la región HaDhn1 se concentraron en dos zonas de la secuencia, una correspondiente a un dominio rico en glicina (región G1), coincidentemente con lo reportado por los mencionados autores, que definen a este dominio como el más variable en genotipos cultivados (excluyendo la región del intrón). La región G1 estaría involucrada con la capacidad hidrofílica de la proteína (Close *et al.*, 1989), demostrando la importancia que cumple en la función biológica de la misma. La segunda zona que presentó polimorfismos correspondió al intrón, mientras que en los sitios adyacentes al mismo (residuos de serina), no se detectó variabilidad, lo cual puede deberse a que es un sitio fosforilable posiblemente relacionado con la traslocación de estas proteínas desde el citoplasma al núcleo (Giordani *et al.*, 2010; Godoy *et al.*, 1994; Natali *et al.*, 2003; Vilardell *et al.*, 1990).

El InDel detectado en la región codificante de HaDhn1 correspondió a la adición de dos aminoácidos (6 nucleótidos) en la región G1 de la secuencia, coincidentemente con lo reportado por Natali *et al.* (2003), quienes detectaron estos dos aminoácidos extra para entradas de girasol silvestre únicamente. Esta coincidencia puede deberse a que 17 de las

líneas empleadas en esta tesis poseen aporte de materiales silvestres. Los 6 nucleótidos involucrados en este InDel se encuentran presentes en los haplotipos 2, 3, 4 y 5 (Tabla 10), los cuales fueron encontrados en líneas clasificadas como tolerantes, intermedias y susceptibles. El hecho de que las estimas de diversidad para HaDhn1 sean inferiores a lo reportado en la bibliografía, podría deberse a que muchas de las variantes alélicas se hayan perdido en el proceso de mejoramiento. De acuerdo con estos resultados, aunque el nivel de diversidad de la región HaDhn1 en las líneas aquí estudiadas sea menor que en los trabajos citados, la multiplicidad y frecuencia de los alelos detectados para esta región, hacen de ella una candidata interesante para futuros estudios de mapeo por asociación para el carácter de tolerancia al déficit hídrico.

Las regiones Hahb4-p, Hahb4, FB, HaL1L y CK mostraron solamente dos haplotipos. Para las dos primeras regiones la única línea que mostró el haplotipo menos frecuente fue B67, para la tercera fue 430-C, para la cuarta fue R437 y para la última fueron dos genotipos, R417 y B99. Por su parte, la línea B67 se comportó como susceptible en ensayo de campo en base al carácter de peso de cien semillas y en ensayo bajo condiciones controladas con déficit hídrico moderado, y es la única línea que proviene del cruzamiento entre materiales argentinos y yugoslavos, lo cual podría ser la causa de que esta línea posea un haplotipo distinto a las demás en la región Hahb4. Dado que los genes que codifican proteínas regulatorias muestran una menor diversidad que los genes que codifican enzimas o proteínas de defensa (Giordani *et al.*, 2010), esto podría explicar la escasa variabilidad detectada en esta región. La línea 430-C es de origen argentino y se comportó como intermedia en el ensayo de fenotipificación en condiciones controladas con déficit hídrico moderado, mientras que la entrada R437 fue tolerante en ensayo de campo en base al carácter de rendimiento relativo (Alemano *et al.*, 2005) y es proveniente de materiales argentinos y estadounidenses. Una de las funciones usualmente atribuidas a la caseína quinasa (CK) es como regulador de la transcripción génica a través de factores de transcripción de la fosforilación (Lee *et al.*, 1999), por ello se la relaciona con genes involucrados en la respuesta a sequía. Por otro lado, las proteínas quinasas se han propuesto como reguladores de la división celular

durante el estrés impuesto por el ambiente. Éstas son sub-reguladas en genotipos tolerantes al estrés sugiriendo una estrategia o vía de escape al daño producido por el déficit hídrico (Roche *et al.*, 2007). Las dos líneas diferentes al resto para el caso de la región CK, una de ellas (B99) se comportó como susceptible en ensayo de campo y como tolerante en condición moderada de déficit hídrico y su genoma incluye acervos argentino y estadounidense, mientras que la otra línea (R417) se comportó como intermedia en ensayo de campo y en condición moderada de déficit hídrico y posee material silvestre, francés y estadounidense. Esto sugiere que las mezclas genómicas (*admixture*) (Figuras 17, 18 y 19) que conforman las líneas de acuerdo con la información de su origen, no parecen estar asociadas con su comportamiento en relación al déficit hídrico. Teniendo en cuenta sus bajos niveles de variabilidad, estas regiones no serían las más adecuadas para un estudio de mapeo por asociación orientado al carácter de tolerancia al déficit hídrico. Esto se debe a que las variantes alélicas en frecuencias bajas tienen menor probabilidad de ser detectadas en las asociaciones. Además, el poder o potencia del mapeo por asociación no sólo se ve influenciado por las variantes alélicas en frecuencias bajas, sino también por el efecto de las mismas sobre el carácter agronómico en estudio (Hall *et al.*, 2010).

En resumen, a pesar de los diferentes orígenes geográficos de las líneas empleadas en este trabajo, la diversidad nucleotídica fue baja, al igual que el número de haplotipos y las frecuencias de los mismos. Este hecho se contrapone con los altos niveles de variación detectados por los marcadores microsatélites para el mismo conjunto de líneas.

D.2.3. Pruebas de neutralidad

Las pruebas de neutralidad se han convertido en una poderosa herramienta para identificar regiones interesantes desde el punto de vista agronómico (por ej.: González-Martínez *et al.*, 2006; Schmid *et al.*, 2005; Wright and Gaut, 2005). Sin embargo, como se mencionara previamente, los resultados de su aplicación sobre conjuntos de individuos que no constituyen

una población natural deben ser interpretados con cautela, sin olvidar sus limitaciones.

Los resultados obtenidos en las pruebas de neutralidad para las regiones Hahb4, Hahb4-p, FB y HaL1L, revelaron un exceso de variantes en bajas frecuencias, tanto al considerar el conjunto total de líneas, como los dos grupos correspondientes al nivel de estructuración inferido en base al método bayesiano (Tabla 14). En estas regiones candidatas, se detectó un valor de θ_w superior al de π_T , originando un resultado de D negativo y significativo (Tablas 12 y 13), coincidente con lo reportado por Giordani *et al.* (2010), en 4 de los 8 genes relacionados con la tolerancia a sequía que los autores estudiaron. Es decir que, los valores negativos y significativos de D son consistentes con la existencia de un patrón de exceso de variantes alélicas raras con respecto a las predicciones de la teoría neutralista (Braverman *et al.*, 1995). Este patrón de polimorfismo está comúnmente asociado con un aumento reciente en el tamaño poblacional, con un cuello de botella severo, o con un barrido selectivo que reduce la variación nucleotídica. (González-Martínez *et al.*, 2006). Las últimas dos alternativas son las más viables de acuerdo con el modo de manejo de las líneas endocriadas pertenecientes al programa de mejoramiento.

Los efectos demográficos, tales como un cuello de botella poblacional, influyen todas las porciones del genoma nuclear, tanto neutras como funcionales (Audigeos *et al.*, 2010). Las poblaciones que han experimentado una reducción severa de su tamaño poblacional efectivo, exhiben una disminución en el número de alelos y en la diversidad genética. Los resultados provistos por el análisis de loci SSR en las líneas aquí analizadas mostraron una tendencia contraria a estas predicciones. La comparación de los valores de H_e obtenidos con los esperados en el equilibrio deriva-mutación mostró un exceso de alelos en frecuencias intermedias (Tabla 14), lo cual sería evidencia de un cuello de botella moderado. Dado que los resultados fueron los mismos, tanto al analizar el conjunto de las 34 líneas, como al considerar separadamente los dos grupos principales detectados por el método bayesiano, el exceso de variantes en frecuencias intermedias no sería consecuencia de la subestructuración. Por lo tanto, los valores de D negativos y significativos obtenidos para las regiones Hahb4, Hahb4-p, FB y

HaL1L no pueden ser atribuidos a fenómenos demográficos, sino que responderían a causas selectivas. De todos modos, no es posible determinar si la selección está relacionada con el hecho de que estas líneas pertenecen al programa de mejoramiento, o si los efectos de la misma se reflejan de manera general en el cultivo como consecuencia de la domesticación y subsiguientes ciclos de mejoramiento.

La única región que mostró valores positivos y significativos para alguna de las pruebas de neutralidad fue HaDhn1 (D^* , Tabla 12). Si bien, esto es concordante con el cuello de botella moderado inferido en base a los marcadores microsatélites, los restantes estimadores no parecen haber sido capaces de detectar desvíos respecto de lo esperado por la teoría neutralista.

En resumen, se propone que de acuerdo con los valores de D negativos y significativos, las regiones candidatas Hahb4, Hahb4-p, FB y HaL1L estarían sujetas a selección, mientras que las regiones *Suntip*, CK y HaDhn1 no evidenciaron este patrón. El valor positivo y significativo del estimador D^* para la última región mencionada podría estar generado por el cuello de botella moderado inferido en base a los marcadores SSR.

D.2.4. Determinación del desequilibrio de ligamiento

Para el estudio del DL se debe tener en cuenta que existen factores asociados como la subdivisión y la estructuración de la población que pueden incrementar los valores de desequilibrio (Gaut and Long, 2003), sumado a la compleja historia de obtención de líneas endocriadas y a la falta de conocimiento acabado acerca de la estructura del DL en la mayoría de las especies vegetales (Flint-Garcia *et al.*, 2005).

Se ha determinado que el DL en girasol tendría una extensión aproximada de 100 Kpb (Fusari *et al.*, 2010). Existen antecedentes en especies de autofecundación que muestran patrones comparables a lo que sucede con girasol, como por ej. en soja con DL a más de 50 Kpb (Zhu *et al.*, 2003), en arroz con DL hasta 100 Kpb (Garris *et al.*, 2003) y en sorgo con DL hasta 10 Kpb (Hamblin *et al.*, 2004).

Los resultados obtenidos en esta tesis mostraron que sólo 2 regiones (*Suntip* y HaDhn1) de las 7 analizadas evidenciaron decaimiento del DL a lo largo de la secuencia, mientras que para las restantes el valor promedio de r^2 se mantuvo constante (Figura A.2). La región *Suntip* mostró un valor de r^2 promedio de 0,349, con un decaimiento hasta 0,325 a las 207 pb (Figura 21-A); similar a lo reportado por otros autores para materiales cultivados (Fusari *et al.*, 2008; Kolkman *et al.*, 2007; Liu and Burke, 2006). El valor de r^2 promedio para la región HaDhn1 fue de 0,279 (con decaimiento hasta 0,249 a las 78 pb) (Figura 21-B), levemente superior a lo reportado por Giordani *et al.* (2010) de 0,204, con un decaimiento completo a las 1911 pb.

La línea de tendencia global del DL (para el total de las regiones) evidenció un decaimiento escaso, ya que se alcanzó un valor de r^2 promedio de 0,872 a las 5500 pb. Este resultado fue superior a lo reportado en la bibliografía, con valores que alcanzan 0,32 a las 5500 pb (Fusari *et al.*, 2008; Kolkman *et al.*, 2007), o un decaimiento a las 1000 bp para un grupo de genes involucrados en la respuesta a sequía (Giordani *et al.*, 2010).

Los altos niveles de DL detectados para 5 de las 7 regiones analizadas (Tabla 15), coinciden con lo esperado para regiones sujetas a selección. El alcance del DL en especies de polinización cruzada es menor que en especies autógamas (Oraguzie *et al.*, 2007), pero en este estudio también debe considerarse la posible influencia de la selección sobre los patrones de DL de las regiones Hahb4, Hahb4-p, FB y HaL1L.

En el contexto de un estudio de mapeo por asociación, los elevados niveles de DL detectados constituyen una ventaja, ya que permiten reducir el número de polimorfismos a genotipificar a través del uso de Tag SNP (Zhang *et al.*, 2004).

En base a los resultados obtenidos en este Capítulo, las regiones *Suntip* y HaDhn1 serían las mejores candidatas para una futura estrategia de mapeo por asociación para tolerancia al déficit hídrico en girasol por presentar múltiples alelos en frecuencias superiores al 5%. Por su parte, las regiones Hahb4, Hahb4-p, FB y HaL1L resultan interesantes debido a la posible incidencia de fuerzas selectivas en los patrones de sustitución nucleotídica hallados. Si bien, su baja diversidad haplotípica puede resultar limitante para la detección de asociaciones fenotipo-genotipo

estadísticamente significativas, particularmente considerando la compleja naturaleza del carácter tolerancia al déficit hídrico, la frecuencia del alelo minoritario para dichas regiones estuvo en torno al 3% (2,86-3,45%), un valor ligeramente inferior al 5% que suele considerarse apropiado para los estudio de mapeo por asociación (Thornsberry *et al.*, 2001).

DISCUSIÓN GENERAL

Este trabajo de tesis constituye el primer aporte al análisis genético de entradas del BAG-IM en base a marcadores moleculares. El objetivo general del mismo fue sentar las bases metodológicas y conceptuales para generar conocimientos esenciales sobre el nivel de diversidad genética en las entradas del banco, con el objeto de orientar su caracterización exhaustiva, asociarla a caracteres agronómicos complejos y asistir a los programas de mejoramiento de girasol de INTA.

Los niveles de diversidad genética en base a los marcadores microsatélites descritos en el Capítulo 1 resultaron consistentes con el espectro de diversidad morfológica y agronómica de los materiales analizados. Este análisis es el primer paso hacia la caracterización exhaustiva del banco, aportando conocimientos básicos para delinear futuras estrategias de muestreo. Los resultados obtenidos en el análisis de muestra mínima para diversidad, de constitución de grupos núcleo y de diferenciación genética sugieren que para optimizar la relación entre costo y número de alelos detectados, sería más conveniente analizar un mayor número de entradas, a expensas de reducir el número de individuos dentro de cada una de ellas, y no a la inversa.

La heterogeneidad detectada en este trabajo entre individuos de una misma línea destaca la necesidad de realizar un seguimiento de las mismas en generaciones sucesivas de autofecundación mediante marcadores moleculares. Esta falta de homogeneidad trae aparejadas consecuencias negativas tanto para los aspectos legales, como para el avance de los programas de mejoramiento en los cuales participan. Por ejemplo, la selección de los marcadores moleculares a incorporar en la prueba de Distinguibilidad, Homogeneidad y Estabilidad (DHE) requiere el empleo de líneas endocriadas homogéneas y estables. Con respecto a los programas de mejoramiento, la utilización de líneas heterogéneas puede generar severas distorsiones en estudios de mapeo en poblaciones biparentales, y especialmente, en el mapeo por asociación.

El carácter de tolerancia a déficit hídrico es sumamente complejo y su evaluación se ve marcadamente afectada por las condiciones ambientales. Los resultados presentados en el Capítulo 2 sugieren que el ensayo de evaluación de respuesta al déficit hídrico moderado en condiciones controladas provee una herramienta confiable para la fenotipificación rápida de líneas a partir de las cuales se seleccionarán aquellas más promisorias para ser caracterizadas en el ensayo de campo. Esto se debe a que el estadio de implantación no necesariamente refleja lo que sucederá en el campo, en donde la floración es el período más susceptible al déficit hídrico.

El análisis de los patrones de variación en regiones candidatas asociadas a la respuesta al déficit hídrico representa un primer aporte para el abordaje de futuros estudios de mapeo por asociación para este carácter. Las regiones *Suntip* y HaDhn1 serían las más apropiadas para seleccionarlas como candidatas para este tipo de estudios, mientras que las regiones Hahb4, Hahb4-p, FB y HaL1L resultan interesantes debido a la posible incidencia de procesos selectivos y podrían emplearse en un futuro, complementando el análisis con un mayor número de entradas en la población de mapeo.

En base al grado de DL detectado, se estima que posiblemente con la genotipificación de sólo un SNP por región, se podría identificar el haplotipo de la línea analizada. De este modo, la genotipificación sería más económica y rápida a los fines de la identificación de haplotipos asociados a características fenotípicas relacionadas con la tolerancia al déficit hídrico.

CONCLUSIONES

Los resultados de este trabajo de tesis ponen de manifiesto la importancia del análisis molecular de los recursos genéticos de girasol almacenados en el BAG-IM y la necesidad del conocimiento de los mismos para el programa de mejoramiento del cultivo. Las conclusiones más importantes que derivan de este trabajo son:

- I. Las entradas del BAG-IM analizadas en este trabajo mostraron valores moderados de diversidad genética y un alto grado de diferenciación.
- II. La diversidad genética detectada en base a caracteres morfológicos y agronómicos fue consistente con la diversidad determinada a nivel de ADN mediante el empleo de marcadores microsatélites.
- III. La heterogeneidad detectada en las líneas del BAG-IM determina la necesidad de realizar un seguimiento de las mismas en sucesivas generaciones de autofecundación, mediante marcadores moleculares.
- IV. Los tamaños muestrales de las entradas correspondientes a las categorías de poblaciones y compuestos resultaron adecuados para el estudio de su diversidad genética. De acuerdo a ello, futuros muestreos del BAG-IM deberán tener en cuenta el incremento del número de entradas y la reducción del número de individuos dentro de cada una de ellas.
- V. El ensayo bajo condiciones controladas en déficit hídrico moderado resultó eficiente para la fenotipificación rápida de líneas pertenecientes al programa de mejoramiento para tolerancia a sequía.
- VI. Las regiones candidatas *Suntip* y *HaDhn1* mostraron mayor diversidad nucleotídica y número de haplotipos en frecuencias similares, por ello son las más adecuadas para la implementación de una estrategia de mapeo por asociación para el carácter de tolerancia al déficit hídrico.

- VII. Las regiones candidatas Hahb4, Hahb4-p, FB y HaL1L mostraron indicios de estar sujetas a selección, por lo tanto resultan interesantes para futuros estudios de mapeo por asociación.

Se considera que los resultados aquí presentados serán útiles para diseñar estrategias con el objeto de efficientizar el empleo de los recursos genéticos de girasol y para una futura caracterización molecular del banco. A partir de este trabajo se inicia el estudio y la caracterización basada en microsatélites de los recursos genéticos del BAG-IM, para su utilización en los programas y el manejo en base, no sólo a la caracterización fenotípica (morfológica, fenológica, fisiológica y agronómica), sino también a la caracterización genética de los mismos. Sumado a ello, como otro aporte de este trabajo de tesis se incorpora conocimiento sumamente importante para el diseño y la implementación de estrategias de mapeo por asociación para el carácter de tolerancia al déficit hídrico.

Doctorando
M. Valeria Moreno

Director
Dra. Verónica V. Lia

Director Asistente
Dra. Norma B. Paniego

BIBLIOGRAFÍA

1. Aguirrezábal, L.A.N.; Orioli, G.A.; Hernández, L.F.; Pereyra, V.R.; Mirave, J.P. 2001. **Girasol. Aspectos fisiológicos que determinan el rendimiento**. Balcarce (AR). ISBN: 950-9853 71-2. p 111.
2. Akagi, H.; Yokozeki, Y.; Inagaki, A.; Fujimura, T. 1996. **Microsatellite DNA markers for rice chromosomes**. Theoretical Applied Genetics 94: 61–67.
3. Al-Chaarani, G.R.; Gentzbittel, L.; Huang, X.; Sarrafi, A. 2004. **Genotypic variation and identification of QTLs for agronomic traits using AFLP and SSR in recombinant inbred lines of sunflower (*Helianthus annuus* L.)**. Theoretical Applied Genetics 109: 1353-1360.
4. Al-Chaarani, G.R.; Gentzbittel, L.; Wedzony, M.; Sarrafi, A. 2005. **Identification of QTLs for germination and seedling development in sunflower (*Helianthus annuus* L.)**. Plant Science 169: 221-227.
5. Alemano, S.G.; Álvarez, D.; Bigo, V.; Ferragutti, S.; Vigliocco, A.E.; Abdala, G. 2005. **Caracterización de Germoplasma de girasol en condiciones de estrés hídrico**. Buenos Aires (AR): ASAGIR. Disponible en: <http://www.asagir.org.ar/3erCongreso/Word/Resumen%20ALEMANO%20ASAGIR%202005%20Final.doc>. Consultado: Diciembre 2009.
6. Alemanno, G.; Lerda, S.; Oviden, E.; Vagliendo, C.; Valiente, P.; Biasutti, C.A. 2003. **Heterosis en ensayos de calidad de semilla en híbridos experimentales de maíz**. Basic and Applied Genetics 15(2) p 124.
7. Allinne, C.; Maury, P.; Sarrafi, A.; Grieu, P. 2009. **Genetic control of physiological traits associated to low temperature growth in sunflower under early showing conditions**. Plant Science 177: 349-359.
8. Anderson, J.A.; Churchill, G.A.; Autrique, J.E.; Tanksley, S.D.; Sorrells, M.E. 1993. **Optimizing parental selection for genetic linkage maps**. Genome. 36: 181-186.
9. Anderson, P.K.; Cunningham, A.A.; Patel, N.G.; Morales, F.J.; Epstein, P.R.; Daszak, P. 2004. **Emerging infectious diseases of plants: pathogen pollution, climate change and agrotechnology drivers**. Trends in Ecology and Evolution 19 (10): 535-544.
10. Ariel, F.D.; Manavella, P.A.; Dezar, C.A.; Chan, R.L. 2007. **The true story of the HD-Zip family**. Trends in Plant Science 12: 419-426.

11. Arumuganathan, K.; Earle, E.D. 1991. **Nuclear DNA content of some important plant species**. *Plant Molecular Biology*. Rep. 9: 208-219.
12. Audigeos, D.; Buonamici, A.; Belkadi, L.; Rymer, P.; Boshier, D.; Scotti-Saintagne, C.; Vendramin, G.G.; Scotti, I. 2010. **Aquaporins in the wild: natural genetic diversity and selective pressure in the PIP gene family in five Neotropical tree species**. *BMC Evolutionary Biology* 10: 202. Disponible en: <http://www.biomedcentral.com/1471-2148/10/202>. Consultado: Enero 2011.
13. Balfourier, F.; Roussel, V.; Strelchenko, P.; Exbrayat-Vinson, F.; Sourdille, P.; Boutet, G.; Koenig, J.; Ravel, C.; Mitrofanova, O.; Beckert, M.; Charmet, G. 2007. **A worldwide bread wheat core collection arrayed in a 384-well plate**. *Theoretical Applied Genetics* 114(7): 1265-1275.
14. Balzarini, M.; Bruno, C.; Arroyo, A. 2005. **Análisis de ensayos agrícolas multiambientales. Ejemplos en Info-Gen**. 1era edición. Córdoba (AR). Universidad Nacional de Córdoba. Facultad de Ciencias Agropecuarias. Cátedra de Estadística y Biometría. 105 p. ISBN: 987-05-0349-7.
15. Baxter, C.J.; Sabar, M.; Quick, W.P.; Sweetlove, L.J. 2005. **Comparison of changes in fruit gene expression in tomato introgression lines provides evidence of genome-wide transcriptional changes and reveals links to mapped QTLs and described traits**. *Journal of Experimental Botany* 56: 1591-1604.
16. Bert, P.; Dechamp-Guillaume, G.; Seere, F.; Jouan, I.; Tourvieille De Labrouhe, D.; Nicolas, P.; Vear, F. 2004. **Comparative genetic analysis of quantitative traits in sunflower (*Helianthus annuus* L.) 3. Characterization of QTL involved in resistance to *Sclerotinia sclerotiorum* and *Phoma macdonaldii***. *Theoretical and Applied Genetics* 109: 865-874.
17. Bertero de Romano, A.; Vázquez, A.N. 2003 **Origin of the Argentine sunflower varieties**. *Helia* 26 (38): 127-136. Disponible en: <http://www.doiserbia.nb.rs/Article.aspx?ID=1018-18060338127d>. Consultado: Noviembre 2008.
18. Biasutti, C.A.; Galiñanes, V. 2001. **Influencia del ambiente de selección sobre la germinación de semillas de maíz (*Zea mays* L.) bajo estrés hídrico. Relaciones entre caracteres de plántula con el rendimiento a campo**. *Agriscientia* 18: 37-44.
19. Bigo, V. 2005. **Caracterización de germoplasma de girasol (*Helianthus annuus*), en condiciones de estrés hídrico**. Tesis de grado. Río Cuarto, Córdoba (AR). Universidad Nacional de Río Cuarto. Facultad de Ciencias Exactas, Físico-Químicas y Naturales. Laboratorio de Fisiología Vegetal. 19 p.

20. Blum, A. 1988. **Plant breeding for stress environments**. Boca Raton, Florida (USA) CRC Press Inc p. 1-232.
21. Blum, A. 1993. **Selection for sustained production in water deficit environments**. *In: International Crop Science I*: 343-347.
22. Blum, A.; Ebercon, A. 1981. **Cell membrane stability as a measure of drought and heat tolerance in wheat**. *Crop Science* 21(1): 43-47.
23. Bracco, M.; Lia, V.V.; Gottlieb, A.M.; Cámara Hernández, J.; Poggio, L. 2009. **Genetic diversity in maize landraces from indigenous settlements of Northeastern Argentina**. *Genetica* 135 (1): 39-49.
24. Braverman, J.M.; Hudson, R.R.; Kaplan, N.L.; Langley, C.H.; Stephan, W. 1995. **The hitchhiking effect on the site frequency spectrum of DNA polymorphisms**. *Genetics* 140: 783-796.
25. Bray, E.A. 2002. **Classification of genes differentially expressed during water-deficit stress in *Arabidopsis thaliana*: an analysis using microarray and differential expression data**. *Annals of Botany (London)* 89: 803-811.
26. Breseghello, F.; Sorrells, M.E. 2006. **Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars**. *Genetics* 172: 1165-1177.
27. Brickell, C.D.; Baum, B.R.; Hetterscheid, W.L.A.; Leslie, A.C.; McNeill, J.; Trehane, P.; Vrugtman, F.; Wiersema, J.H. 2004. **International Code of Nomenclature of Cultivated Plants**. *Acta Horticulturae* 647: 1-123.
28. Brookes, A.J. 1999. **The essence of SNPs**. *Gene* 234: 177-186.
29. Brown, A.H.D. 1995. **The core collection at the crossroads**. *In: Hodgkin, T.; Brown, A.H.D.; van Hintum, Th.J.L.; Morales, E.A.V. (Ed) Core collections of plant genetic resources*. New York (USA): John Wiley. p. 3-20.
30. Brownstein, M.J.; Carpten, J.D.; Smith, J.R. 1996. **Modulation of non-templated nucleotide addition by *Taq* DNA polymerase: primer modifications that facilitate genotyping**. *Biotechniques* 20: 1004-1010.
31. Brunel, D. 1994. **A microsatellite marker in *Helianthus annuus* L.** *Plant Molecular Biology* 24: 397-400.
32. Bruniard, J.M. 2002. **Presente y futuro de la investigación en girasol en Argentina**. *IDIA XXI (AR)* 3: 116-118. Disponible en:

- <http://www.inta.gov.ar/ediciones/idia/oleaginosa/girasol.htm>. Consultado: Agosto 2010.
33. Buckler, E.S.; Thornsberry, J.M. 2002. **Plant molecular diversity and applications to genomics**. *Current Opinion in Plant Biology* 5: 107-111.
34. Burke, J.; Tang, S.; Knapp, S.J.; Rieseberg, L.H. 2002. **Genetic analysis of sunflower domestication**. *Genetics* 161: 1257-1267.
35. Cai, Q.; Aitken, K.; Deng, H.H.; Chen, X.W.; Fu, C.; Jackson, P.A.; McIntyre, C.L. 2005. **Verification of the introgression of *Erianthus arundinaceus* germplasm into sugarcane using molecular markers**. *Plant Breeding* 124 (4): 322-328.
36. Cantamutto, M.; Poverene, M. 2007. **Genetically modified sunflower release: Opportunities and risks**. *Field Crop Research* 101(2):133-144.
37. Casini, C. 2010. **Argentina hacia la industrialización del campo**. *En: Agroindustria en origen. PRECOP, INTA (AR). Actualización técnica (52) Junio 2010. 26 p.*
38. Causse, M.; Duffe, P.; Gómez, M.C.; Buret, M.; Damidaux, R.; Zamir, D.; Chevalier, C.; Lemaire-Chamley, M.; Rothan, C. 2004. **A genetic map of candidate genes and QTLs involved in tomato fruit size and composition**. *Journal of Experimental Botany* 55: 1671-1685.
39. Cellier, F.; Conejero, G.; Breitler, J.C.; Casse, F. 1998. **Molecular and physiological responses to water deficit in drought-tolerant and drought-sensitive lines of sunflower. Accumulation of dehydrin transcripts correlates with tolerance**. *Plant Physiology* 116: 319-328.
40. Cellier, F.; Conejero, G.; Casse, F. 2000. **Dehydrin transcript fluctuations during a day/night cycle in drought stressed sunflower**. *Journal of Experimental Botany* 51: 299-304.
41. Chao, S.; Dubcovsky, J.; Dvorak, J.; Luo, MC.; Baenziger, S.; Matnyazov, R.; Clark, D.; Talbert, L.; Anderson, J.; Dreisigacker, S.; Glover, K.; Chen, J.; Campbell, K.; Bruckner, P.; Rudd, J.; Haley, S.; Carver, B.; Perry, S.; Sorrells, M.; Akhunov, E. 2010. **Population and genome-specific patterns of linkage disequilibrium and SNP variation in spring and winter wheat (*Triticum aestivum* L.)**. *BMC Genomics* 11: 727.
42. Charles, S.A.; Halliwell, B. 1981. **Light activation of fructose biphosphatase in isolated spinach chloroplasts and deactivation by hydrogen peroxide**. *Planta* 151: 242-246.

43. Chimenti, C.A.; Hall, A.J. 1993. **Genetic variation and changes with ontogeny of osmotic adjustment in sunflower (*Helianthus annuus* L.)**. *Euphytica* 71(3): 201-210.
44. Close, T.J.; Kortt, A.A.; Chandler, P.M. 1989. **A cDNA-based comparison of dehydration-induced proteins (dehydrins) in barley and corn**. *Plant Molecular Biology* 13: 95-108.
45. Connor, D.J.; Hall, A.J. 1997. **Sunflower physiology**. In Schneiter, A.A. (Ed.) *Sunflower Technology and Production*. Madison (USA) American Society of Agronomy 35: 113-182.
46. Conrero, J.M. 2007. **Métodos de mejoramiento en plantas autógamas**. Córdoba (AR): Universidad Nacional de Córdoba. Facultad de Ciencias Agropecuarias. Cátedra de Mejoramiento Genético Vegetal, Departamento de Producción Vegetal. 100 p.
47. Cooke, R.J.; Bredemeijer, G.M.M.; Ganai, M.W.; Peeters, R.; Isaac, P.; Rendell, S.; Jackson, J.; Roder, M.S.; Korzun, V.; Wendehake, K.; Areshchenkova, T.; Dijcks, M.; Laborie, D.; Bertrand, L.; Vosman, B. 2003. **Assessment of the uniformity of wheat and tomato varieties at DNA microsatellite loci**. *Euphytica* 132: 331-341.
48. Cornuet, J.M.; Luikart, G. 1996. **Description and power analysis of two tests for detecting recent population bottlenecks from allele frequency data**. *Genetics* 144: 2001-2014.
49. Cox, T.S.; Kiang, Y.T.; Gorman, M.B.; Rodgers, D.M. 1985. **Relationships between coefficient of parentage and genetic similarity indices in soybean**. *Crop Science* 25: 529-532.
50. Crisci, J.V. 1983. **Introducción a la teoría y práctica de la taxonomía numérica**. Washington (USA). OEA. Secretaría General. Monografías Científicas. Serie Biología. Nº 26. 132 p.
51. Cubero, J. 1999. **Introducción al mejoramiento genético vegetal**. Madrid (ES). Mundiprensa. 365 p.
52. Daszak, P.; Cunningham, A.A.; Hyatt, A.D. 2000. **Emerging infectious diseases of wildlife- threats to biodiversity and human health**. *Science* 287: 443-449.
53. Dean, R.E.; Dahlberg, J.A.; Hopkins, M.S.; Mitchell, S.E.; Kresovich S. 1999. **Genetic redundancy and diversity among "orange" accessions in the US National Sorghum Collection as assessed with simple sequence repeat (SSR) markers**. *Crop Science* 39: 1215-1221.

54. Dehmer, K.J.; Friedt, W. 1998. **Evaluation of different microsatellite motifs for analysing genetic relationships in cultivated sunflower (*Helianthus annuus* L.)** Plant Breeding 117: 45-48.
55. Dellaporta, S.; Woods, J.; Hicks, J. 1983. **A plant DNA miniprep, version II.** Plant Molecular Biology Reporter 1: 19-21.
56. Dezar, C.A.; Gago, G.M.; González, D.H.; Chan, R.L. 2005. **Hahb-4, a sunflower homeobox-leucine zipper gene, is a developmental regulator and confers drought tolerance to *Arabidopsis thaliana* plants.** Transgenic Research 14: 429-440.
57. Di Rienzo, J.A.; Casanoves, F.; Balzarini, M.G.; González, L.; Tablada, M.; Robledo, C.W. 2008. **InfoStat, Versión 2008.** Córdoba (AR). Universidad Nacional de Córdoba. Facultad de Ciencias Agropecuarias. Grupo InfoStat.
58. Doebley, J.F.; Gaut, B.S.; Smith, B.D. 2006. **The molecular genetics of crop domestication.** Cell 127: 1309-1321.
59. Ebrahimi, A.; Maury, P.; Berger, M.; Poormohammad Kiani, S.; Nabipour, A.; Shariati, F.; Grieu, P.; Sarrafi, A. 2008. **QTL mapping of seed-quality traits in sunflower recombinant inbred lines under different water regimes.** Genome 51: 599-615.
60. Ehrenreich, I.M.; Hanzawa, Y.; Chou, L.; Roe, J.L.; Kover, P.X.; Purugganan, M.D. 2009. **Candidate gene association mapping of *Arabidopsis* flowering time.** Genetics 183: 325-335.
61. El Mousadik, A.; Petit, R. 1996. **High level of genetic differentiation for allelic richness among populations of the argan tree (*Argania spinosa* (L.) Skeels) endemic to Morocco.** Theoretical and Applied Genetics 92: 832-839.
62. Ellis, J.R.; Pashley, C.H.; Burke, J.M.; McCauley, D.E. 2006. **High genetic diversity in a rare and endangered sunflower as compared to a common congener.** Molecular Ecology 15: 2345-2355.
63. Ersoz, E.; Yu, J.; Buckler, E.S. 2007. **Applications of linkage disequilibrium and association mapping in crop plants.** In: Genomics-assisted crop improvement. Varshney, Tuberosa (Ed). Springer Netherlands (AL). Vol 1: Genomics approaches and platforms: 97-119 p.
64. Escande, A.R.; Laich, F.S.; Pedraza, M.V. 2002. **Field testing of honeybee dispersed *Trichoderma* spp. to manage sunflower head rot (*Sclerotinia sclerotiorum*).** Plant Pathology 51: 346-351.

65. Escande, A.; Pereyra, V.; Pedraza, M.V.; Troglia, C.; Quiroz, F. 2006. **Sclerotinia en girasol**. Balcarce (AR). IDIA 21: 140-143.
66. 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.
67. Falush, D.; Stephens, M.; Pritchard, J.K. 2003. **Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies**. Genetics 164: 1567-1587.
68. Fambrini, M.; Durante, C.; Cionini, G.; Geri, C.; Giorgetti, L.; Michelotti, V.; Salvini, M.; Pugliesi, C. 2006. **Characterization of LEAFY COTYLEDON1-LIKE gene in *Helianthus annuus* and its relationship with zygotic and somatic embryogenesis**. Development Genes and Evolution 216: 253-264.
69. Farooq, S.; Azam, F. 2006. **The use of cell membrane stability (CMS) technique to screen for salt tolerant wheat varieties**. Journal of Plant Physiology 163: 629-637.
70. Felsenstein, J. 2009. **Molecular population genetics** In: Theoretical Evolutionary Genetics. Washington (USA). 359-383 p.
71. Felsenstein, J. 1991-2004. **Phylip (Phylogeny Inference Package)**. University of Washington, Washington.
72. Feltus, F.A.; Singh, H.P.; Lohithaswa, H.C.; Schulze, S.R.; Silva, T.D.; Paterson, A.H. 2006. **A comparative genomics strategy for targeted discovery of single-nucleotide polymorphisms and conserved non-coding sequences in orphan crops**. Plant Physiology 140: 1183-1191.
73. Fernandez, P.; Paniago, N.; Lew, S.; Hopp, H.E.; Heinz, R.A. 2003. **Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project**. BMC Genomics 4: 40.
74. Fernie, A.R.; Tadmor, Y.; Zamir, D. 2006. **Natural genetic variation for improving crop quality**. Current Opinion in Plant Biology 9(2): 196-202. Disponible en: <http://www.sciencedirect.com/science/journal/13695266>. Consultado: Mayo 2008.
75. Flint-Garcia, S.A.; Thornsberry, J.M.; Buckler, E.S. 2003. **Structure of linkage disequilibrium in plants**. Annual Review of Plant Biology 54: 357-374.
76. Flint-Garcia, S.A.; Thuillet, A.C.; Yu, J.; Pressoir, G.; Romero, S.M.; Mitchell, S.E.; Doebley, J.; Kresovich, S.; Goodman, M.M.; Buckler,

- E.S. 2005. **Maize association population: a high-resolution platform for quantitative trait locus dissection.** *Plant Journal* 44: 1054-1064.
77. Ford, M.J. 2002. **Applications of selective neutrality tests to molecular ecology.** *Molecular Ecology* 11: 1245-1262.
78. Frankel, O.H. 1984. **Genetic perspectives of germplasm conservation.** In: Arber, W.; Llimensee, K.; Peacock, W.J.; Starlinger, P. Eds. *Genetic manipulation: impact on man and society.* Cambridge (UK) University Press p. 161–170.
79. Fu, Y.X. 1997. **Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection.** *Genetics* 147: 915-925.
80. Fu, Y.X.; Li, W.H. 1993. **Statistical tests of neutrality of mutations.** *Genetics* 133: 693-709.
81. Furini, A.; Wunder, J. 2004. **Analysis of eggplant (*Solanum melongena*)-related germplasm: morphological and AFLP data contribute to phylogenetic interpretations and germplasm utilization.** *Theoretical and Applied Genetics* 108: 197-208.
82. Fusari, C.M.; Lia, V.V.; Hopp, H.E.; Heinz, R.A.; Paniego, N.B. 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.
83. Fusari, C.M.; Lia, V.V.; Nishinakamasu, V.; Zubrzycki, J.E.; Puebla, A.F.; Maligne, Hopp A.E.; Heinz, R.A.; Paniego, N.B. 2010. **Single nucleotide polymorphism genotyping by heteroduplex analysis in sunflower (*Helianthus annuus* L.).** *Molecular Breeding* 1-17.
84. Fusari, C.M. 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. Buenos Aires (AR). Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. 251 p.
85. Gago, G.M.; Almoguera, C.; Jordano, J.; González, D.H.; Chan, R.L. 2002. **Hahb4, a homeobox-leucine zipper gene potentially involved in abscisic acid-dependent responses to water stress in sunflower.** *Plant Cell and Environment* 25: 633-640.
86. Garris, A.J.; McCouch, S.R.; 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.

87. Gaut, B.S.; Long, A.D. 2003. **The lowdown on linkage disequilibrium**. *Plant Cell* 15: 1502-1506.
88. Ghislain, M.; Spooner, D.M.; Rodríguez, F.; Villamon, F.; Núñez, C.; Vásquez, C.; Bonierbale, M. 2004. **Selection of highly informative and user-friendly microsatellites (SSR) for genotyping of cultivated potato**. *Theoretical and Applied Genetics* 108: 881-890.
89. Giancola, S.; McKhann, H.I.; Berard, A.; Camilleri, C.; Durand, S.; Libeau, P.; Roux, F.; Reboud, X.; Gut, I.G.; Brunel, D. 2006. **Utilization of the three high-throughput SNP genotyping methods, the GOOD assay, Amplifluor and TaqMan, in diploid and polyploid plants**. *Theoretical and Applied Genetics* 112: 1115-1124.
90. Giordani, T.; Buti, M.; Natali, L.; Pugliesi, C.; Cattonaro, F.; Morgante, M.; Cavallini, A. 2010. **An analysis of sequence variability in eight genes putatively involved in drought response in sunflower (*Helianthus annuus* L)**. *Theoretical and Applied Genetics*. DOI 10.1007/s00122-010-1509-0. Consultado: Enero 2011.
91. Giordani, T.; Natali, L.; Cavallini, A. 2003. **Analysis of a dehydrin encoding gene and its phylogenetic utility in *Helianthus***. *Theoretical and Applied Genetics* 107: 316-325.
92. Godoy, J.A.; Lunar, R.; Torres-Schumann, S.; Moreno, J.; Rodrigo, R.M.; Pintor-Toro, J.A. 1994. **Expression, tissue distribution and subcellular localization of dehydrin TAS14 in salt-stressed tomato plants**. *Plant Molecular Biology* 26: 1921-1934.
93. González, A. 2008. **Técnicas de interés en la mejora de cebadas tolerantes a la sequía y temperaturas elevadas. I. Estabilidad de las membranas celulares**. *Agroecología (ES)* 3: 55-60.
94. González-Martínez, S.C.; Ersoz, E.; Brown, G.R.; Wheeler, N.C.; Neale, D.B. 2006. **DNA Sequence variation and selection of Tag Single-Nucleotide polymorphisms at candidate genes for drought stress response in *Pinus taeda* L**. *Genetics*. 172: 1915-1926.
95. Goudet, J. 2001. **FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3)**. Disponible en: <http://www.unil.ch/izea/software/fstat.html>. Updated from Goudet (1995) FSTAT version 1.2: a computer program to calculate F-statistics. *Journal of Heredity* 86: 485-486.
96. Gross, B.L.; Schwarzbach, A.E.; Riesberg, L.H. 2003. **Origin(s) of the diploid hybrid species *Helianthus deserticola* (Asteraceae)**. *American Journal of Botany* 90: 1708-1719.

97. Gunjaca, J.; Buhinicek, I.; Jukic, M.; Sarcevic, H.; Vragolovic, A.; Kozic, Z.; Jambrovic, A.; Pejic, I. 2008. **Discriminating maize inbred lines using molecular and DUS data**. *Euphytica* 161: 165-172.
98. Haddadi, P.; Yazdi-samadi, B.; Langlade, N.B.; Naghavi, M.R.; Berger, M.; Kalantari, A.; Calmon, A.; Maury, P.; Vincourt, P.; Sarrafi, A. 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(40): 6768-6782.
99. Hall, T.A. 1999. **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT**. *Nucleic Acids Symposium Series* 41: 95-98.
100. Hall, A.J.; Chimenti, C.A.; Vilella, F.; Freier, G. 1985. **Timing of water stress effects on yield components in sunflower**. *Proceedings XI International Sunflower Conference, Mar del Plata (AR)* p. 131-6.
101. Hall, D.; Tegstrom, C.; Ingvarsson, P.K. 2010. **Using association mapping to dissect the genetic basis of complex traits in plants**. *Brifings in functional genomics* 9 (2): 157-165.
102. Hall, A.J.; Whitfield, D.M.; Connor, D.J. 1990. **Contribution of preanthesis assimilates to grain-filling in irrigated and water-stressed sunflower crops. II. Estimates from a carbon budget**. *Field Crop Research* 24:273–294.
103. Hamblin, M.T.; Mitchell, S.E.; White, G.M.; Gallego, J.; Kukatla, R.; Wing, R.A.; Paterson, A.H.; Kresovich, S. 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.
104. Hardy, O.J.; Charbonnel, N.; Freville, H.; Heuertz, M. 2003. **Microsatellite allele sizes: A simple test to assess their significance on genetic differentiation**. *Genetics* 163: 1467-1482.
105. Harlan, J.R. 1995. **The living fields**. Cambridge (UK): University Press. 271 p.
106. Heckenberger, M.; Bohn, M.; Melchinger, A.E. 2005. **Identification of essentially derived varieties obtained from biparental crosses of homozygous lines: I. Simple sequence repeat data from maize inbreds**. *Crop Science* 45: 1120-1131.
107. Heckenberger, M.; Bohn, M.; Ziegler, J.S.; Joe, L.K.; Hauser, J.D.; Hutton, M.; Melchinger, A.E. 2002. **Variation of DNA fingerprints among accessions within maize inbred lines and**

- implications for identification of essentially derived varieties. I. Genetic and technical sources of variation in SSR data.** *Molecular Breeding* 10: 181-191.
108. Heckenberger, M.; Van der Voort, J.R.; Peleman, J.; Bohn, M. 2003. **Variation of DNA fingerprints among accessions within maize inbred lines and implications for identification of essentially derived varieties: II. Genetic and technical sources of variation in AFLP data and comparison with SSR data.** *Molecular Breeding* 12: 97-106.
109. Hernández, C.R. 2005. **Acondicionamiento osmótico de simientes de girasol (*Helianthus annuus* L.) para el avance de la germinación en siembras precoces para zonas áridas.** Tesis Doctoral. Madrid (ES). Universidad Politécnica de Madrid. 210 p.
110. Hervé, D.; Fabre, F.; Berrios, E.F.; Leroux, N.; Al-Chaarani, G.R.; Planchon, C.; Sarrafi, A.; Gentzbittel, L. 2001. **QTL analysis of photosynthesis and water status traits in sunflower (*Helianthus annuus* L.) under green house condition.** *Journal of Experimental Botany* 362: 1857-1864.
111. Hill, W.G.; Weir, B.S. 1994. **Maximum-likelihood estimation of gene location by linkage disequilibrium.** *American Journal of Human Genetics* 54: 705-714.
112. Hirschhorn, J.N.; Daly, M.J. 2005. **Genome-wide association studies for common diseases and complex traits.** *Nature Reviews: Genetics* 6: 95-108.
113. Hokanson, S.C.; Szewc-McFadden, A.K.; Lamboy, W.F.; McFerson, J.R. 1998. **Microsatellite (SSR) markers reveal genetic identities, genetic diversity and relationships in a *Malus x domestica* Borkh. core subset collection.** *Theoretical and Applied Genetics* 97:671-683.
114. Hvarleva, T.; Bakalova, A.; Chepinski, I.; Hristova-Cherbadji, M.; Hristov, M.; Atanasov, A. 2007. **Characterization of bulgarian sunflower cultivars and inbred lines with microsatellite markers.** *Biotechnology and Biotechnology* 21: 408-412.
115. Hyten, D.L.; Pantalone, V.R.; Sams, C.E.; Saxton, A.M.; Landau-Ellis, D.; Stefaniak, T.R.; Schmidt, M.E. 2004. **Seed quality QTL in a prominent soybean population.** *Theoretical and Applied Genetics* 109: 552-561.
116. Iriondo Alegría, J.M. 2001. **Conservación de germoplasma de especies raras y amenazadas (Revisión).** *Invest. Agr.: Prod. Prot. Veg (ES)* 16 (1): 5-24.

117. Internacional Seed Testing Association. 1976. **International Rules for Seed Testing**. Seed Science and Technology 4 (1): 1-180.
118. Internacional Seed Testing Association. 1985. **International Rules for Seed Testing**. Seed Science and Technology 13 (2): 1-520.
119. International Seed Testing Association. 2003. **International Rules for Seed Testing**. Edition 2004. Zurich. ISBN: 3-906549-38-0.
120. Jansson, S.; Buckler, E.S. 2007. **Genome studies and molecular genetics. Genomics-deeper and wider in order to understanding plant diversity**. Plant Biology 10: 107-108.
121. Kane, N.C.; Rieseberg, L.H. 2007. **Selective sweeps reveal candidate genes for adaptation to drought and salt tolerance in common sunflower, *Helianthus annuus***. Genetics 175: 1823–1834.
122. Kojima, J.; Eban, K.; Fukuoka, S.; Nagamine, T.; Kawase, M. 2005. **Development of RFLP-based rice diversity research set of germplasm**. Breeding Science 55: 431-440.
123. Kolkman, J.M.; Berry, S.T.; Leon, A.; Slabaugh, M.B.; Tang, S.; Gao, W.; Shintani, D.K.; Burke, J.M.; Knapp, S.J. 2007. **Single nucleotide polymorphisms and linkage disequilibrium in sunflower**. Genetics 177 (1): 457-468.
124. Koornneef, M.; Stam, P. 2001. **Changing paradigms in plant breeding**. Plant Physiology 125: 156-159. Disponible en: www.plantphysiol.org. Consultado: Marzo 2007.
125. Kreitman, M. 2000. **Methods to detect selection in populations with applications to the human**. Annual Review of Genomics and Human Genetics 1: 539-559.
126. Kresovich, S.; Lamboy, W.; Mc. Ferson, J. and Forsline, P. 1995. **Integrating different types of information to develop core collections, with particular reference to *Brassica oleracea* and *Malus X domestica***. In: Hodgkin, T.; Brown, A.; Van Hintum, Th.; Morales, E. (Ed.) Core Collections of Plant Genetic Resources. CAB International, Wallingford, Oxon. P. 147-168.
127. Kumar, R.S.; Parthiban, K.T.; Govinda Rao, M. 2009. **Molecular characterization of *Jatropha* genetic resources through inter-simple sequence repeats (ISSR) markers**. Molecular Biology Reports 36: 1951-1956.
128. Lai, Z.; Livingstone, K.; Zou, Y.; Church, S.A.; Knapp, S.J.; Andrews, J.; Rieseberg L.H. 2005. **Identification and mapping of SNP from ESTs in sunflower**. Theoretical and Applied Genetics 111 (8): 1532-1544.

129. Laurentin, H. 2009. **Data analysis for molecular characterization of plant genetic resources**. Genetic Resources Crop Evolution. DOI: 10.1007/s10722-008-9397-8. Consultado: Noviembre 2010.
130. Lee, Y.; Lloyd, A.M.; Roux, S.J. 1999. **Antisense expression of the CK2 α -subunit gene in *Arabidopsis*: effects on light-regulated gene expression and plant growth**. Plant Physiology 119: 989–1000.
131. Lewontin, R.C. 1964. **The interaction of selection and linkage. I. General considerations: heterotic models**. Genetics 49: 49-67.
132. Li, S.B.; Zhang, Z.H.; Hu, Y.; Li, C.Y.; Jiang, X.; Mao, T.; Li, Y.S.; Zhu, Y.G. 2006. **Genetic dissection of development behavior of crop growth rate and its relationships with yield yield related traits in rice**. Plant Science 170: 911-917.
133. Lia, V. 2004. **Diversidad genética y estructura poblacional en razas nativas de maíz (*Zea mays* spp. *mays*) del Noroeste Argentino: presente y pasado del germoplasma autóctono**. Tesis Doctoral. Buenos Aires (AR): Universidad de Buenos Aires. Facultad de Ciencias Exactas y Naturales. Departamento de Ecología, Genética y Evolución. 186 p.
134. Lia, V.V.; Bracco, M.; Gottlieb, A.M.; Poggio, L.; Confalonieri, V.A. 2007. **Complex mutational patterns and size homoplasy at maize microsatellite loci**. Theoretical and Applied Genetics 115: 981-991.
135. Librado, P.; Rozas, J. 2009. **DNAsp v5: A software for comprehensive analysis of DNA polymorphism data**. Bioinformatics 25: 1451-1452.
136. Liu, X.; Baird, W.V. 2003. **Differential expression of genes regulated in response to drought or salinity stress in sunflower**. Crop Science 43: 678-687.
137. Liu, A.; Burke, J.M. 2006. **Patterns of nucleotide diversity in wild and cultivated sunflower**. Genetics 173: 321-330.
138. Liu, K.; Goodman, M.; Muse, S.; Smith, J.S.; Buckler, E.; Doebley, J. 2003. **Genetic structure and diversity among maize inbred lines as inferred from DNA microsatellites**. Genetics 165: 2117-2128.

139. Liu, K.; Muse, S. 2005. **PowerMarker: an integrated analysis environment for genetic marker analysis**. *Bioinformatics* 21: 2128-2129.
140. Lobell, D.B.; Burke, M.; Tebaldi, C.; Mastrandrea, M.C.; Falcon, W.P.; Naylor, R.S. 2008. **Prioritizing climate change adaptation needs for food security in 2030**. *Science* 319: 607-610.
141. Long, A.D.; Langley, C.H. 1999. **The power of association studies to detect the contribution of candidate genetic loci to variation in complex traits**. *Genome Research* 9: 720-731.
142. López Pereira, M.; Berney, A.; Hall, A.J.; Trápani, N. 2008. **Contribution of pre-anthesis photoassimilates to grain yield: Its relationship with yield in Argentine sunflower cultivars released between 1930-1955**. *Field Crops Research* 105: 88-96.
143. Mago, R.; Bariana, H.S.; Dundas, I.S.; Spielmeyer, W.; Lawrence, G.J.; Pryor, A.J.; Ellis, J.G. 2005. **Development of PCR markers for the selection of wheat stem rust resistance genes *Sr24* and *Sr26* in diverse wheat germplasm**. *Theoretical and Applied Genetics* 111: 496-504.
144. Malosetti, M.; Van der Linden, C.G.; Vosean, B.; Van Eeuwijk, F.A. 2007. **A mixel-model approach to association mapping using pedigree information with an illustration of resistance to *Phytophthora infestans* in potato**. *Genetics* 175: 879-889.
145. Manavella, P.A.; Arce, A.L.; Dezar, C.A.; Bitton, F.; Renou, J.P.; Crespi, M.; Chan, R.L. 2006. **Cross-talk between ethylene and drought signaling path ways is mediated by the sunflower *Hahb4* transcription factor**. *The Plant Journal* 48: 125-137.
146. Manavella, P.A.; Dezar, C.A.; Ariel, F.D.; Chan, R.L. 2008. **Two ABREs, two redundant root-specific and one w-box cis-acting elements are functional in the sunflower *Hahb4* promotor**. *Plant Physiology and Biochemistry* 46: 860-867.
147. Maringolo, C.A. 2007. **Regiones cromosómicas asociadas a resistencia a Podredumbre Húmeda del Capítulo de girasol (*Sclerotinia sclerotiorum* (Lib.) de Bary)**. Tesis de Magister Scientiae. Balcarce, Buenos Aires (AR). Universidad Nacional de Mar del Plata. Facultad de Ciencias Agrarias. 133 p.
148. Matsuoka, Y.; Mitchell, S.E.; Kresovich, S.; Goodman, M.M.; Doebley, J.F. 2002. **Microsatellites in *Zea* – variability, patterns of mutation, and use for evolutionary studies**. *Theoretical and Applied Genetics* 104: 436-450.

149. Meksem, K.; Kahl, G. 2005. **The handbook of plant genome mapping. Genetic and physical mapping.** Alemania: WILEY-VCH Verlag, Weinheim 406 p.
150. Micic, Z.; Hahn, V.; Bauer, E.; Melchinger, A.E.; Knapp, S.J.; Tang, G.S.; Schön, C.C. 2005. **Identification and validation of QTL for *Sclerotinia* midstalk rot resistance in sunflower by selective genotyping.** Theoretical and Applied Genetics 111: 233–242.
151. Micic, Z.; Hahn, V.; Bauer, E.; Schön, C.C.; Knapp, S.J.; Tang, S.; Melchinger, A.E. 2004. **QTL mapping of *Sclerotinia* midstalk rot resistance in sunflower.** Theoretical and Applied Genetics 109: 1474–1484.
152. Micic, Z.; Hahn, V.; Bauer, E.; Schön, C.C.; Melchinger, A.E. 2005. **QTL mapping of resistance to *Sclerotinia* midstalk rot in RIL of sunflower population NDBLOSSel CM625.** Theoretical and Applied Genetics 110: 1490–1498.
153. Minch, E.; Ruiz Linares, A.; Goldstein, D.; Feldman, M.; Cavalli-Sforza, L.L. 1995-1996. **Microsat (version 1.5b): a computer program for calculating various statistics on microsatellite allele data.** Disponible en: <http://www.btkastanfordedu/microsat.html>. Consultado: Mayo 2008.
154. Mogg, R.; Batlely, J.; Hanley, S.; Edwards, D.; O Sullivan, H.; Edwards, K.J. 2002. **Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms.** Theoretical and Applied Genetics 105: 532-543.
155. 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 13: 43-51.
156. Moons, A.; Bauw, G.; Dekeyser, R.; Von Montagu, M.; Van Der Straeten, D. 1995. **Novel ABA responsive proteins in vegetative rice tissue.** Current Topics in Plant Physiology 10: 288-289.
157. Monte, L.M. 1986. **Breeding plants for drought resistance-the problem and its relevance. Drought resistance in plants-physiological and genetic aspects.** Congress of Breeding Plant, Amalfi (IT) p 1-11.
158. Morgante, M.; Hanafey, H.; Powell, W. 2002. **Microsatellites are preferentially associated with nonrepetitive DNA in plant genome.** Nature Genetics 30: 194-200.

159. Natali, L.; Giordani, L.; Cavallini, A. 2003. **Sequence variability of a dehydrin gene within *Helianthus annuus***. Theoretical and Applied Genetics 106: 811-818.
160. Nei, M. 1972. **Genetic distance between populations**. American Naturalist 106: 283-292.
161. Nei, M. 1977. **F statistics and analysis of gene diversity in subdivided populations**. Annals of Human Genetics 41: 225-233.
162. Nei, M. 1987. **Molecular evolutionary genetics**. New York (USA). Columbia University Press.
163. Nordborg, M.; Borevitz, J.O.; Bergelson, J.; Berry, C.C.; Chory, J.; Hagenblad, J.; Kreitman, M.; Maloof, J.N.; Noyes, T.; Oefner, P.J.; Stahl, E.A.; Weigel, D. 2002. **The extent of linkage disequilibrium in *Arabidopsis thaliana***. Nature Genetics 30: 190-193.
164. Oraguzie, N.C.; Rikkerink, E.H.A.; Gardiner, S.E.; De Silva, H.N. 2007. **Association mapping in plants**. Oraguzie, N.C.; Rikkerink, E.H.A.; Gardiner, S.E.; De Silva, H.N. New York (USA). Springer Science+Business Media.
165. Paniego, N.; Echaide, M.; Muñoz, M.; Fernandez, L.; Torales, S.; Faccio, P.; Fuxan, I.; Carrera, M.; Zandomeni, R.; Syarez, E.Y.; Hopp, E.H. 2002. **Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.)**. Genome 45: 34-43.
166. Parcy, F.; Valon, C.; Kohara, A.; Misera, S.; Giraudat, J. 1997. **The ABSCISIC ACIDINSENSITIVE3, FUSCA3 and LEAFY COTYLEDON1 genes act in concert to control multiple aspects of *Arabidopsis* seed development**. Plant Cell 9: 1265-1277.
167. Pedraza, M.V.; Pereyra, V.R.; Aguirrezábal, L.A.N.; Laurlund, A. 2000. **Manual de estimación de pérdidas de rendimiento en girasol: efectos de reducciones en el área foliar, en la densidad de plantas y en el número de capítulos**. Balcarce (AR). 150 p.
168. Peluffo, L. 2010. **Caracterización de los mecanismos de defensa a *Sclerotinia sclerotiurom*, agente causal de la podredumbre húmeda de capítulo de girasol a través del estudio de perfiles metabólicos y transcripcionales**. Tesis Doctoral. Buenos Aires (AR): Universidad de Buenos Aires. Facultad de Ciencias Exactas. 122 p.
169. Pereyra, V.; Escande, A. 1994. **Enfermedades del girasol en la Argentina**. Manual de reconocimiento. [En línea]. Balcarce, Buenos Aires (AR): Unidad Integrada INTA-Universidad de Mar del Plata. 113 p.
Disponibile en:

- <http://www.inta.gov.ar/balcarce/actividad/divulga/sincargo/enfgira.htm>. Consultado: Abril 2007.
170. Pereyra-Irujo, G.A.; Velásquez, L.; Lechner, L.; Aguirrezábal, L.A.N. 2008. **Genetic variability for leaf growth rate and duration under water deficit in sunflower: analysis of responses at cell, organ and plant level**. *Journal of Experimental Botany* 59(8): 2221-2232.
171. Pérez-Molphe-Balch, E.; Gidekel, M.; Segura-Nieto, M.; Herrera-Estrella, L.; Ochoa Alejo, N. 1996. **Effects of water stress on plant growth and root proteins three cultivars of rice (*Oryza sativa* L.) with different levels of drought tolerance**. *Physiology Plant* 96: 284-290.
172. Petit, R.J.; Duminil, J.; Fineschi, S.; Hampe, A.; Salvini, D.; Vendramin, G.G. 2005. **Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations**. *Molecular Ecology* 14: 689-701.
173. Plunknett, D.L.; Smith, N.J.H.; Williams, J.T.; Anishetty, N.M. 1987. **A case study in rice germplasm**. IR 36 p. 171-185. *In: Gene banks and the world's food*. (Ed). Plunknett, D.L.; Smith, N.J.H.; Williams, J.T.; Anishetty, N.M. New Jersey (USA) Princeton University Press.
174. Poormohammad Kiani, S.; Grieu, P Maury, P.; Hewezi, T.; Gentzbittel, L.; Sarrafi, A. 2007. **Genetic variability for physiological traits under drought conditions and differential expression of water stress-associated genes in sunflower (*Helianthus annuus* L.)**. *Theoretical and Applied Genetics* 114: 193-207.
175. Poormohammad Kiani, S.; Maury, P.; Nouri, L.; Ykhlef, N.; Grieu, P.; Sarrafi, A. 2009. **QTL analysis of yield-related traits in sunflower under different water treatments**. *Plant Breeding* 128(4): 363-373.
176. Poormohammad Kiani, S.; Talia, P.; Maury, P.; Grieu, P.; Heinz, R.; Perrault, A.; Nishinakamasu, V.; Hopp, E.; Gentzbittel, L.; Paniego, N.; Sarrafi, A. 2007. **Genetic analysis of plant water status and osmotic adjustment in recombinant inbred lines of sunflower under two water treatments**. *Plant Science* 172: 773-787. Disponible en: <http://www.sciencedirect.com/science/journal/01689452>. Consultado: Septiembre 2008.
177. Prada, D. 2009. **Molecular population genetics and agronomic alleles in seed banks: searching for a needle in a haystack?**. *Journal of Experimental Botany* 60(9) p 2541–2552.

178. Priano, J.; Pilatti, M. 1989. **Tolerancia a la salinidad de forrajeras cultivadas**. *Ciencia del suelo* 7: 1-2.
179. Pritchard, J.K.; Stephens, M.; Donnelly, P. 2000. **Inference of population structure using multilocus genotype data**. *Genetics* 155: 945-959.
180. Pritchard, J.K.; Wen, W. 2003. **Documentation for STRUCTURE software: version 2**. Disponible en: <http://pritch.bsd.uchicago.edu>.
181. Przeworski, M. 2002. **The signature of positive selection at randomly chosen loci**. *Genetics* 160: 1179-1189.
182. Radonic, L.M.; Zimmermann, J.M.; Zavallo, D.; López, N.; López Bilbao, M. 2008. **Introduction of antifungal genes in sunflower via *Agrobacterium***. *Electronic Journal of Biotechnology* 11(5): 1-5.
183. Raymond, M.; Rousset, F. 1995. **Genepop (version 1.2): Population genetics software for exact tests and ecumenicism**. *Journal of Heredity* 86: 248-249.
184. Reid, W.V.; Miller, K.R. 1989. **Keeping options alive. The scientific basis for conserving biodiversity**. Washington (USA): World Resources Institute. 128 p.
185. Rice, W. 1989. **Analysing tables of statistical tests**. *Evolution*. 43: 223-225.
186. Risch, N.J. 2000. **Searching for genetic determinants in the new millennium**. *Nature* 405: 847-856.
187. Robel In, M. 1967. **Periodes critiques du tournesol Vis-a-Vis de la secheresse**. Francia: CETIOM. *Informations Techniques* N° 12, p. 1-9.
188. Roche, J.; Essahat, A.; Bouniols, A.; El Asri, M.; Mouloungui, Z.; Mondières, M.; Alghoum, M. 2004. **Diversified composition of sunflower (*Helianthus annuus* L.) seeds within cultural practices and genotypes (hybrids and populations)**. *Helia* 27:73-98.
189. Roche, J.; Hewezi, T.; Bouniols, A.; Gentzbittel, L. 2007. **Transcriptional profiles of primary metabolism and signal transduction-related genes in response to water stress in field-grown sunflower genotypes using a thematic cDNA microarray**. *Planta* 226: 601-617.
190. Rohlf, F.J.; Sokal, R.R. 1981. **Comparing numerical taxonomic studies**. *Systematic Zoology* 30:459-490.

191. Rongwen, J.; Akkaya, M.S.; Bhagwat, A.A.; Lavi, U.; Cregan, P.B. 1995. **The use of microsatellite DNA markers for soybean genotype identification**. Theoretical and Applied Genetics 90:43–48.
192. Rönicke, S.; Hahn, V.; Vogler, A.; Freidt, W. 2005. **Quantitative trait loci analysis of resistance to *Sclerotinia sclerotiorum* in sunflower**. Phytopathology 95: 834-839.
193. Rosenberg, N.A.; Nordborg, M. 2002. **Genealogical trees, coalescent theory and the analysis of genetic polymorphisms**. Nature Reviews: Genetic 3: 380-390.
194. Rosenberg, N.A.; Pritchard, J.K.; Weber, J.L.; Cann, H.M.; Kidd, K.K.; Zhivotovskiy, L.A. 2002. **Genetic structure of human populations**. Science 298 (5602): 2381-2385.
195. Rostoks, N.; Mudie, S.; Cardle, L.; Russell, J.; Ramsay, L.; Booth, A.; Svensson, J.T.; Wanamaker, S.I.; Walia, H.; Rodriguez, E.M.; Hedley, P.E.; Liu, H.; Morris, J.; Close, T.J.; Marshall, D.F.; Waugh, R. 2005. **Genome-wide SNP discovery and linkage analysis in barley based on genes responsive to abiotic stress**. Molecular and General Genetics 274: 515-527.
196. Rousset, F.; Raymond, M. 1995. **Testing heterozygote excess and deficiency**. Genetics 140: 1413-1419.
197. Rozen, S.; Skaletsky, H. 2000. **Primer3 on the WWW for general users and for biologist programmers** [<http://frodo.wi.mit.edu/>] In: Bioinformatics methods and protocols: methods in molecular biology. Krawetz, Misener (Ed). New York (USA). Humana Press. 365-386 p.
198. Saghai-Marooif, M.A.; Soliman, K.M.; Jorgensen, R.A.; Allard, R.W. 1984. **Ribosomal DNA spacer-length polymorphisms in barley: Mendelian inheritance, chromosomal location, and population dynamics**. Proceedings of the National Academy Science 81: 8014-8018.
199. SAGPYA, 2009. http://www.minagri.gob.ar/SAGPyA/agricultura/cultivos_en_la_argentina/01mapa_principales_cultivos/index.php?PHPSESSID=8e7cb4d78f31c04a9dd9a800f4ba3e14. Consultado: Diciembre 2010.
200. Saitou, N.; Nei, M. 1987. **The neighbor-joining method: a new method for reconstructing phylogenetic trees**. Molecular Biology and Evolution 4: 406-425.
201. Sala, C.A.; Bulos, M.; Echarte, M.; Whitt, S.R., Ascenzi, R. 2008. **Molecular and biochemical characterization of an induced**

- mutation conferring imidazolinone resistance in sunflower.** Theoretical and Applied Genetics 118(1): 105-112.
202. Sanger, F.; Coulson, A.R. 1975. **A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase.** Journal of Molecular Biology. 94 (3): 441-446.
203. Sarda, X.; Tousch, D.; Ferrare, K.; Cellier, F.; Alcon, C.; Dupuis, J.M.; Casse, F., Lamaze, T. 1999. **Characterization of closely related δ -TIP gene encoding aquaporins which are differentially expressed in sunflower roots upon water deprivation through exposure to air.** Plant Molecular Biology 40: 179-191.
204. Sarda, X.; Tousch, D.; Ferrare, K.; Legrand, E.; Dupuis, J.M.; Casse, F.; Lamaze, T. 1997. **Two TIP-like genes encoding aquaporins are expressed in sunflower guard cells.** Plant of Journal 12: 1103-1111.
205. Schmid, K.J.; Ramos-Onsins, S.; Ringys-Beckstein, H.; Weisshaar, B.; Mitchell-Olds, T. 2005. **A multilocus sequence survey in *Arabidopsis thaliana* reveals a genome-wide departure from a neutral model of DNA sequence polymorphism.** Genetics 169: 1601-1615.
206. Schmid, K.J.; Sorensen, T.R.; Stracke, R.; Torjek, O.; Altmann, T.; Mitchell-Olds, T.; Weisshaar, B. 2003. **Large-scale identification and analysis of genome-wide single nucleotide polymorphisms for mapping in *Arabidopsis thaliana*.** Genome Research 13(6A): 1250-1257.
207. Schnable, P.S.; Wise, R.P. 1998. **The molecular basis of cytoplasmic male sterility and fertility restoration.** Elsevier Science 3(5): 175-180.
208. Seiler, J.G.; Rieseberg, L.H. 1997. **Systematics, origin, resources of the wild and domesticated sunflower.** In: Sunflower technology and production. Madison. Wisconsin (USA). (Ed) Agronomy series Agronomy. 21-65 p.
209. Seki, M.; Kameiy, A.; Yamaguchi-Shinozakiz, K.; Shinozaki, K. 2003. **Molecular responses to drought, salinity and frost: common and different paths for plant protection.** Current Opinion Biotechnology 14: 194-199.
210. Sharbel, T.F.; Haubold, B.; Mitchell-Olds, T. 2000. **Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe.** Molecular Ecology 9: 2109-2118.

211. Singh, B.D. 2000. **Plant Breeding-Principles and Methods**. Ludhiana. New Delhi (IN): Kalyani Publishers 896 p.
212. Sitalov, F.V.; Ikonikov, P.A. 1969. **Fiziologija rastenij**. Moskva. (In Russian). Tom. 16. Vip. 6.
213. Skoric, D. 2009. **Sunflower breeding for resistance to abiotic stresses**. *Helia* 32 (50): 1-16. Disponible en: <http://www.doiserbia.nb.rs/issue.aspx?issueid=859>. Consultado: Julio 2010.
214. Smith, K.; Petley, D.N. 2009. **Hydrological hazards: drought**. In: Environmental hazards. Assessing risk and reducing disaster. New York (USA) p 262-285.
215. Sneath, P.H.A.; Sokal, R.R. 1973. **Numerical taxonomy**. San Francisco (USA): Freeman. 573 p.
216. Sneller, C.H. 1994. **Pedigree analysis of elite soybean lines**. *Crop Science* 34: 1515-1522.
217. Sokal, R.R.; Sneath, P.H.A. 1963. **Numerical taxonomy. The principles and practice of numerical classification**. San Francisco (USA): Freeman p 121-157.
218. Solodenko, A.E.; Sanalattii, A.B.; Sivolap, Y.M. 2004. **Identification of sunflower genotypes using sunflower markers**. *Citologia i genetika* 38(2): 15-19.
219. Solodenko, A.; Sivolap, Y. 2005. **Genotyping of *Helianthus* based on microsatellite sequences**. *HELIA* 28(42): 19-26.
220. Spooner, D.; Van Treuren, R.; De Vicente, M.C. 2005. **Molecular markers for genebank management**. Rome (IT): IPGR. IPGRI Technical Bulletin N° 10. 127 p. ISBN-13: 987-92-9043-684-3. ISBN-10: 92-9043-684-0.
221. Slatkin, M.; Voelm, L. 1991. **F_{ST} in hierarchical island model**. *Genetics* 127: 627-629.
222. Stocking, M.A. 2003. **Tropical soils and food security: the next 50 years**. *Science* 302: 1356-1359.
223. Syvanen, A.C. 2001. **Assessing genetic variation: genotyping single nucleotide polymorphisms**. *Nature Reviews. Genetics* 2: 930-942.
224. Tajima, F. 1983. **Evolutionary relationship of DNA sequences in finite populations**. *Genetics* 105(2): 437-460.

225. Tajima, F. 1989. **Statistical method for testing the neutral mutation hypothesis by DNA polymorphism**. *Genetics* 123: 585-595.
226. Talia, P.; Nishinakamasu, V.; Hopp, H.E.; Heinz, R.A.; Paniego, N.B. 2010. **Genetic mapping of EST-SSR, SSR and InDel to improve saturation of genomic regions in a previously developed sunflower map**. *Electronic Journal of Biotechnology* 13(6).
227. Tang, S.; Kishore, V.K.; Knapp, S.J. 2003. **PCR-multiplexes for a genome-wide framework of simple sequence repeat marker loci in cultivated sunflower**. *Theoretical and Applied Genetics* 107: 6-19.
228. Tang, S.; Knapp, S.J. 2003. **Microsatellites uncover extraordinary diversity in native American land races and wild populations of cultivated sunflower**. *Theoretical and Applied Genetics* 106: 990-1003.
229. Tang, S.; Yu, J.K.; Slabaugh, M.B.; Shintani, D.K.; Knapp, S.J. 2002. **Simple sequence repeat map of the sunflower genome**. *Theoretical and Applied Genetics* 105: 1124-1136.
230. Tanksley, S.D.; McCouch, S.R. 1997. **Seed banks and molecular maps: unlocking genetic potential from the wild**. *Science* 277: 1063-1066. Disponible en: www.sciencemag.org.
231. Takeda, S.; Matsuoka, M. 2008. **Genetic approaches to crop improvement: responding to environmental and population changes**. *Nature* 9: 444-457. Disponible en: <http://www.nature.com/nrg/journal/v9/n6/full/nrg2342.html>. Consultado: Diciembre 2007.
232. Teruel, M.E.; Biasutti, C.A.; Nazar, M.C.; Peiretti, D.A. 2008. **Efectos de aptitud combinatoria para vigor híbrido de plántula bajo estrés hídrico en maíz**. *Agriscientia* 25(1): 27-34.
233. Thomas, M.R.; Cain, P.; Scott, N.S. 1994. **DNA typing of grapevine: a universal methodology and database for describing cultivars and evaluating genetic relatedness**. *Plant Molecular Biology* 25:939-949.
234. Thornsberry, J.M.; Goodman, M.M.; Doebley, J.; Kresovich, S.; Nielsen, D.; Buckler, E.S. 2001. **Dwarf8 polymorphisms associate with variation in flowering time**. *Nature Genetics* 28: 286-289. Disponible en: <http://genetics.nature.com>. Consultado: Noviembre 2010.
235. Trapani, N.; Gentinetta, E. 1984. **Screening of maize genotypes using drought tolerance tests**. *Maydica* 29: 89-100.

236. Turhan, H.; Baser, I. 2004. ***In vitro* and *in vivo* water stress in sunflower (*Helianthus annuus* L.)**. HELIA 27 (40): 227-236.
237. Van, K.; Hwang, E.Y.; Kim, M.Y.; Park, H.J.; Lee, S.H.; Cregan, P.B. 2005. **Discovery of SNPs in soybean genotypes frequently used as the parents of mapping populations in the United States and Korea**. Journal of Heredity 96: 529-535.
238. Van de Wouw, M.; Van Hintum, T.; Kik, C.; Van Treuren, R.; Visser, B. 2010. **Genetic diversity trends in twentieth century crop cultivars: a meta analysis**. Theoretical and Applied Genetics 120: 1241-1252.
239. Van de Wouw, M.; Van Hintum, T.; Kik, C.; Van Treuren, R.; Visser, B. 2009. **Genetic erosion in crops: concept, research results and challenges**. Plant Genetics Resources 8: 1-15. Disponible en: <http://journals.cambridge.org/action/displayAbstract?fromPage=online&aid=7322332>. Consultado: Junio 2010.
240. Van Hintum, T.; Van de Wiel, C.; Visser, D.L.; Van Treuren, R.; Vossman, B. 2007. **The distribution of genetic diversity in a *Brassica oleracea* gene bank collection related to the effects on diversity of regeneration, as measured with AFLPs**. Theoretical and Applied Genetics 114: 777-786.
241. Vega, P. 2008. **Agrarias avanza en mejoramiento de girasol**. Rosario (AR). Universidad Nacional de Rosario. Disponible en: <http://www.unr.edu.ar/noticia/530/agrarias-avanza-en-mejoramiento-en-girasol>. Consultado: Noviembre 2010.
242. Vigliocco, A.; Alemano, S.; Miersch, O.; Alvarez, D.; Abdala, G. 2007. **Endogenous jasmonates during sunflower germination in seeds from plants grown under different soil moisture content**. Seed Science Research 17: 91-98.
243. Vigouroux, Y.; Glaubitz, J.C.; Matsuoka, Y.; Goodman, M.M.; Sánchez G.J.; Doebley, J. 2008. **Population structure and genetic diversity of new world maize races assessed by DNA microsatellites**. American Journal of Botany 95(10): 1-15.
244. Vilardell, J.; Godoy, A.; Freire, M.A.; Torrent, M.; Martínez, M.C.; Torne, J.M.; Pagès, M. 1990. **Gene sequence, developmental expression and protein phosphorylation of RAB-17 in maize**. Plant Molecular Biology 14: 423-432.
245. Vranceanu, A.V. 2000. **Floarea-soarelui hibrida**. Bucuresti (RO). Editura Ceres. 1147 p.

246. Wang, W.; Vinocur, B.; Altman, A. 2003. **Plant responses to drought, salinity and extreme temperatures: towards genetic engineering for stress tolerance**. *Planta* 218: 1–14.
247. Watterson, G.A. 1975. **On the number of segregating sites in genetical models without recombination**. *Theoretical Population Biology* 7: 256-276.
248. Weir, B.; Cockerham, C.C. 1984. **Estimating F-statistics for the analysis of population structure**. *Evolution*. 38: 1358-1370.
249. Wright, S. 1965. **The interpretation of population structure by F-statistics with special regard to systems of mating**. *Evolution*. 19: 395-420.
250. Wright, S. 1978. **Variability within and among natural populations in evolution and the genetics of populations**. Chicago (USA): University of Chicago. 580 p.
251. Wright, S.I.; Gaut, B.S. 2005. **Molecular population genetics and the search for adaptive evolution in plants**. *Molecular Biology Evolution* 22: 506-519
252. Yan, J.; Shah, T.; Warburton, M.L.; Buckler, E.S.; McMullen, M.D.; Crouch, J. 2009. **Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers**. *Plos One* 4(12): e8451.
253. Yang, G.P.; Maroof, M.A.S.; Xu, C.G.; Zhang, G.; Biyashev, R.M. 1994. **Comparative analysis of microsatellite DNA polymorphism in landraces and cultivars of rice**. *Molecular and General Genetics* 245:187–194.
254. Yu, J.; Buckler, E.S. 2006. **Genetic association mapping and genome organization of maize**. *Biotechnology* 17: 155–160.
255. Yu, J.K.; Mangor, J.; Thompson, L.; Edwards, K.J.; Slabaugh, M.B.; Knapp, S.J. 2002. **Allelic diversity of simple sequence repeat markers among elite inbred lines in cultivated sunflower**. *Genome* 45: 652-660.
256. Yu, J.K.; Tang, S.; Slabaugh, M.B.; Heesacker, A.; Cole, G.; Herring, M.; Soper, J.; Han, F.; Chu, W-C.; Webb, D.M.; Thompson, L.; Edwards, K.J.; Berry, S.; Leon, A.J.; Grondona, M.; Olungu, C.; Maes, N.; Knapp, S.J. 2003. **Towards a saturated molecular genetic linkage map for cultivated sunflower**. *Crop Science* 43: 367–387.
257. Yu, Y.; Yan, J.; Guimaraes, C.T.; Taba, S.; Hao, Z.; Gao, S.; Chen, S.; Li, J.; Zhang, S.; Vivek, B.S.; Magorokesho, C.; Mugo, S.;

- Makumbi, D.; Parentoni, S.N.; Shah, T.; Rong, T.; Crouch, J.H.; Xu, Y. 2009. **Molecular characterization of global maize breeding germplasm based on genome-wide single nucleotide polymorphisms**. *Theoretical Applied Genetics* 120(1): 93-115.
258. Zavallo, D.; López Bilbao, M.; Hopp, H.E.; Heinz, R. 2010. **Isolation and functional characterization of two novel seed-specific promoters from sunflower (*Helianthus annuus* L.)**. *Plant Cell Reports* 29(3): 239-248.
259. Zhang, K.; Qin, Z.S.; Liu, J.S.; Chen, T.; Waterman, M.S.; Sun, F. 2004. **Haplotype block partitioning and Tag SNP selection using genotype data and their applications to association studies**. *Genome Research* 14: 908-916. Disponible en: www.genome.org. Consultado: Diciembre 2010.
260. Zhang, L.S.; Le Clerc, V.; Li, S.; Zhang, D. 2005. **Establishment of an effective set of simple sequence repeat markers for sunflower variety identification and diversity assessment**. *Canadian Journal of Botany* 83: 66-72.
261. Zhu, C.; Gore, M.; Buckler, E.S.; Yu, J. 2008. **Status and prospects of association mapping in plants**. *The Plant Genome* 1: 5–20.
262. Zhu, Y.L.; Song, Q.J.; Hyten, D.L.; Van Tassell, C.P.; Matukumalli, L.K.; Grimm, D.R.; Hyatt, S.M.; Fickus, E.W.; Young, N.D.; Cregan, P.B. 2003. **Single-nucleotide polymorphisms in soybean**. *Genetics* 163: 1123-1134.
263. Zrenner, R.; Krause, K.P.; Apel, P.; Sonnewald, U. 1996. **Reduction of the cytosolic fructose-1,6-bisphosphatase in transgenic potato plants limits photosynthesis sucrose biosynthesis with no impact on plant growth and tuber yield**. *Plant Journal* 9: 671–681.

APÉNDICE

CRITERIOS

Para la construcción de la matriz de datos en base a los caracteres morfológicos y agronómicos provistos por el mejorador se establecieron los siguientes criterios para cada carácter:

Estrías en el grano: es un carácter morfológico de doble estado es decir que se representa como “presencia/ausencia”. La planilla descriptiva empleada en evaluación de campo codifica con 1 para ausencia y con 9 para presencia. En este trabajo se codificó con 0 para ausencia y 1 para presencia.

Superficie de hoja (rugosidad): es un carácter morfológico cualitativo con secuencia lógica. La planilla descriptiva empleada en evaluación de campo codifica con 1 para rugosa, 5 para medianamente rugosa y 9 para lisa. En este trabajo se codificó con 1 para hojas lisas, con 2 para medianamente rugosas y con 3 para rugosas. Esto evita la descomposición del carácter original en tres caracteres de presencia/ausencia, lo cual aumentaría el peso del carácter inicial cuando se calculan los coeficientes de asociación.

Forma de hoja: es un carácter morfológico cualitativo sin secuencia lógica. Según la planilla de evaluación, cada forma es codificada del siguiente modo: 1 para oblonga, 3 para lanceolada, 5 para triangular, 7 para cordada, 9 para redondeada. En este trabajo se codificaron las distintas formas posibles con: 1 para oblonga, 2 para lanceolada, 3 para triangular, 4 para cordada, 5 para redondeada.

Margen de la hoja: es un carácter morfológico cualitativo con secuencia lógica. La planilla de evaluación propone 2 para borde liso, 4 para medianamente aserrado, 6 para fuertemente aserrado y 8 para muy fuertemente aserrado. Dado que los materiales evaluados y caracterizados para esta tesis sólo presentan borde liso o borde medianamente aserrado, se decidió codificar el carácter como de doble estado. Por lo tanto las plantas de bordes liso

corresponden a ausencia de aserrado y toman el valor de 0, mientras que las plantas que presentan aserrado llevan el número 1.

Inclinación de capítulo: es un carácter morfológico cualitativo con secuencia lógica. El modo original de clasificación se basó en diversos grados de acuerdo al ángulo de inclinación del capítulo, 1 para 0°, 3 para 45°, 5 para 90°, 7 para 135° y 9 para 180°. Dado que se evaluaron 80 plantas por cada material (20 plantas por surco, 4 surcos en total), se observó presencia de varios ángulos de inclinación para la misma entrada, por ello se codificó cada ángulo de inclinación como un carácter independiente. Así la planilla final contiene los caracteres IC1, IC2, IC3, IC4, IC5 para 0°, 45°, 90°, 135° y 180°, asignándole el valor de 1 a la presencia de los diversos grados y 0 a la ausencia de los mismos. Es decir que, en los casos en que el material presentó más de un ángulo de inclinación, se codificó como 1 para todos los presentes en esas 80 plantas, siendo 0 la ausencia de los otros IC.

Antocianinas en pálea: es un carácter morfológico cualitativo con secuencia lógica. De acuerdo a la planilla de evaluación de campo sería 1 para ausencia de antocianinas, 5 cuando hay segregación del carácter en algunas plantas y 9 para presencia total del carácter. Para conservar la variabilidad hallada dentro de las plantas del mismo material, principalmente en poblaciones, se decidió codificar con 1 la ausencia de antocianinas, 2 la segregación del carácter y 3 la presencia completa del mismo.

Color del grano: es un carácter morfológico cualitativo sin secuencia lógica. La planilla de evaluación propone los valores de 1 para color negro, 3 para castaño, 5 para blanco, 7 para los granos grises y 9 para otros colores, especificando de qué color se trata. Se codificó con 1, 2, 3, 4, y 5 a los colores negro, castaño, blanco, gris y otro, respectivamente.

Ramificación del tallo: es un carácter morfológico cualitativo sin secuencia lógica. Los posibles tipos de ramificaciones son basal, apical o total con capítulo central. También existe el caso de plantas que no poseen ramificaciones. De acuerdo a ello se codificó este carácter como varios

independientes, transformándolos a cada uno en carácter doble estado. Como algunos materiales mostraron plantas que tenían varios tipos de ramificación, esta codificación permitió conservar la variabilidad morfológica detectada en el campo. Así en la planilla final se codificó como carácter “sin ramificación”, carácter “ramificación basal”, carácter “ramificación apical” y carácter “total con capítulo central”.

PROTOCOLOS

Protocolo de extracción de ADN: *Kit de extracción NucleoSpin Plant II (Macherey-Nagel)*

- a) Pesar entre 20-30 mg de tejido vegetal liofilizado y macerado adecuadamente.
- b) Agregar 400 µl de *buffer* de lisis (PL1 a base de CTAB), agitar intensamente y adicionar 10 µl de *ARNasa A* e incubar a 65°C por 30 min. Agitar periódicamente.
- c) Transferir a una columna *NucleoSpin Filter* con tubo colector y centrifugar por 2 min a 11.000 x g. Si el líquido no ha logrado atravesar completamente el filtro, centrifugar nuevamente.
- d) Adicionar 450 µl de *buffer de binding* (PC), agitar varias veces.
- e) Cargar la muestra en una columna *NucleoSpin Plant II* con tubo colector y centrifugar por 1 min a 11.000 x g.
- f) Adicionar 400 µl de *buffer* de lavado 1 (PW1) en la columna y centrifugar por 1 min a 11.000 x g. Descartar el líquido del tubo colector.
- g) Adicionar 700 µl de *buffer* de lavado 2 (PW2) en la columna y centrifugar por 1 min a 11.000 x g. Descartar el líquido del tubo colector.
- h) Agregar nuevamente 200 µl de *buffer* de lavado 2 (PW2) en la columna y centrifugar por 2 min a 11.000 x g. Descartar el líquido del tubo colector y dejar la columna en reposo para lograr el secado completo de la membrana de sílica.
- i) Remover el tubo colector y reemplazarlo por un tubo *ependorf* estéril de 1,5 ml. Adicionar sobre la membrana 50 µl de *buffer* de elusión (PE)

calentada a 70°C. Incubar a 70°C en baño de agua por 5 min. Centrifugar por 1 min a 11.000 x g.

- j) Repetir el paso i (volumen final de elusión: 100 µl).

Protocolo de Cuantificación por Fluorimetría

- a) Encender el aparato entre 20 min a 1 h antes de la medición.
- b) Armar la curva de calibración descrita en el *ítem i* empleando el estándar de ADN de timo de ternero.
- c) Enjuagar y cargar la cubeta con 2 ml de solución de cuantificación, secarla con papel absorbente. Ajustar en cero y esperar hasta que el aparato lea 0 +/- 5 (tarda 10 s en estabilizarse).
- d) Retirar la cubeta y colocar 2 µl del control, mezclar por inversión varias veces y colocar en el aparato. Ajustar el botón de escala hasta que la lectura corresponda a la concentración correcta.
- e) Lavar la cubeta con H₂O bidestilada y eliminar el exceso de agua varias veces tapando la cuba con un papel absorbente, hasta que no queden rastros de agua.
- f) Colocar 2 ml de solución de cuantificación y 2 µl de muestra en la cubeta.
- g) Colocar en el fluorómetro y leer el valor obtenido. Realizar la mezcla de reacción y la lectura de cada muestra por triplicado.
- h) Repetir el proceso hasta leer todas las muestras. Volver a cero entre cada lectura.
- i) Curva de calibración para medir muestras que tengan concentraciones de ADN entre 20 y 1000 ng:

Cubeta N°:	ADN total (ng):	ADN <i>stock</i> de:	Tomo:	Solución:
1	1000	100 µg/ml (0,1 µg/µl)	10 µl	2 ml
2	500	100 µg/ml (0,1 µg/µl)	5 µl	2 ml
3	200	100 µg/ml (0,1 µg/µl)	2 µl	2 ml
4	100	10 µg/ml (0,01 µg/µl)	10 µl	2 ml
5	50	10 µg/ml (0,01 µg/µl)	5 µl	2 ml
6	20	10 µg/ml (0,01 µg/µl)	2 µl	2 ml
7	0	0	0	0

(Ajustar las concentraciones de los estándares a valores intermedios de la curva en caso de que se observe deficiencia en la cantidad de puntos de la curva).

Con los valores medidos de RFU (unidades de fluorescencia relativa) para cada concentración de ADN de calibración se realiza una regresión lineal (valores de RFU en el eje y; [ADN] en ng/ μ l en el eje x). Se obtiene la ecuación de la recta: $y = m x$; entonces: $y/m = x$; es decir: **valor de RFU/m = [ADN]**

Protocolo de medición de la capacidad de campo (CC) para ensayo de respuesta al déficit hídrico en condiciones controladas

- a) Esterilizar arena fina en estufa a 170°C durante 5 h aproximadamente. (Importante: no reutilizar la arena empleada en ensayos anteriores dado que puede contener trazas de manitol y/o restos vegetales).
- b) Cubrir un embudo con papel de filtro y agregar 100 g de arena estéril, colocar un erlenmeyer abajo.
- c) Adicionar 100 ml de agua bidestilada y dejar percolar el agua una noche o desde el día anterior a la mañana siguiente.
- d) Realizar el cálculo de la CC como se expone a continuación:

(100 ml de agua – volumen de agua percolada) = ml de agua retenida por arena¹ (CC).

100 g de arena _____ volumen de agua retenida¹
 2102 g (por bandeja) _____ x = volumen de agua² al 100 % de CC

100 % de CC _____ volumen de agua²
 60 % de CC _____ x = ml de agua bidestilada por bandeja³

Nota: las bandejas empleadas fueron plásticas, con tapa y de 11,5 cm de ancho, 16 cm de largo y 4 cm de alto.

Protocolo de purificación de productos de PCR: Kit de purificación QIAquick (QIAGEN)

- a) Agregar 5 volúmenes de *buffer* PB a 1 volumen de reacción de PCR y mezclar. No es necesario remover el aceite mineral.

- b) Colocar una columna de centrifugación *QIAquick* en un tubo colector de 2 ml.
- c) Verter la mezcla en la columna y centrifugar a 11.000 rpm por 60 s para adherir el ADN a la columna.
- d) Descartar el líquido del tubo colector y ubicar nuevamente la columna de centrifugación sobre el mismo.
- e) Lavar la columna con 0,75 ml de *buffer* PE y centrifugar por 60 s.
- f) Descartar el líquido del tubo colector y ubicar nuevamente la columna en el mismo. Centrifugar 2 min más con el objeto de remover el etanol residual de la *buffer* PE.
- g) Colocar la columna en un tubo *ependorf* limpio y estéril de 1,5 ml.
- h) Eluir el ADN de la columna agregando 50 µl de agua de alta pureza (Chromasol o calidad HPLC) vertiendo el líquido en el centro de la columna. Para obtener un producto más concentrado adicionar 30 µl.
- i) Dejar la columna abierta por 5 min para permitir que se evaporen las trazas de etanol de la columna.
- j) Centrifugar por 1-2 min a 11.000 rpm.

Protocolo de purificación de productos de PCR: EXOSAP-IT PCR Clean-up (USB)

- a) Agregar 2 µl de *exosap* por cada 5 µl de producto de PCR.
- b) Mezclar e incubar a 37°C durante 15 min.
- c) Inactivar la enzima por calentamiento a 80°C 15 min.
- d) Almacenar a -20°C o en heladera según el uso inmediato posterior del producto de PCR.

SOLUCIONES Y BUFFERS

Método de cuantificación del ADN

Solución de cuantificación

(para 20 muestras)

H ₂ O bidestilada	36 ml
TNE 10 X	4 ml
<i>Hoechst</i> 33258 (1mg/ml)	4 µl

La solución de *Hoescht* (bisbenzimidida) debe guardarse en oscuridad en heladera a 4°C como máximo 6 meses.

Solución TNE 10 X

2 M NaCl	117 gr
100 mM Tris base	12,11 gr
10 mM EDTA disódico dihidratado	3,72 gr
pH 7,4 con HCl	
Volumen final	1000 ml
Almacenar a 4°C máximo 3 meses.	

Buffer TAE 50X

Tris base	242 gr
Acido acético glacial	57,1 ml
EDTA 0.5 M	100 ml
H ₂ O bidestilada	842,9 ml
Volumen final	1000 ml

Buffer TAE 1X:

TAE 50X	20 ml
H ₂ O bidestilada	980 ml
Volumen final	1000 ml

EDTA 0.5M (pH 8)

EDTA	186,1g
H ₂ O bidestilada	1000ml
Ajustar el pH con NaOH.	
Esterilizar y guardar en heladera.	

NaCl 2 M

NaCl	117 g
H ₂ O bidestilada	1000 ml

COLORANTES

Bromo Fenol Blue (BPB)

Hoechst 33258 (1 mg/ml)

Colorante	1 mg
H ₂ O bidestilada	1 ml

Solución de Bromuro de Etidio (0,5 ug/ml)

Bromuro de Etidio	5 ug
H ₂ O bidestilada	10 ml

FLUOROFOROS

6-FAM: 6-Carboxil-fluoresceína

Fórmula molecular: C₂₁H₁₂O₇

Peso molecular: 376.3 g

Pureza (HPLC): ≥ 98.00%

TABLAS

Tabla A.1. Diversidad genética en materiales de girasol cultivado del BAG-IM. Rango alélico: tamaños extremos de los alelos encontrados en pares de bases. A: número de alelos. Entre paréntesis se indica el número de alelos exclusivos para cada material. Rs: riqueza alélica poblacional. He: heterocigosis esperada.

Material	Locus																Promedio
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
KLM																	
Rango	247	151	119	217	137	153	251	170	113	167	137	102	144	119	150	106	-
alélico	259	159	121	227	146	167	257	176	139	173	139	104		122		112	
A (0)	2	3	2	2	3	2	2	2	3	3	2	2	1	2	1	3	2.1875
R _s	1.366	1.904	1.161	1.161	1.996	1.161	1.168	1.840	1.167	1.980	1.302	1.868	1.000	1.889	1.000	1.640	1.475
H _e	0.194	0.499	0.083	0.083	0.534	0.083	0.087	0.484	0.083	0.527	0.159	0.511	0.000	0.518	0.000	0.328	0.26
IAC																	
Rango	247	140	113	217	137	153	233	168	113	170	137	102	140	122	141	109	-
alélico	259	159	121	227	146	167	263	180	143	175	144	104	144	134	150	112	
A (1)	2	5	3	3	2	2	9	5	7	2	2	2	2	3	2	2	3.3125
R _s	1.657	2.467	2.422	1.660	1.659	1.858	3.152	2.515	2.008	1.161	1.573	1.677	1.161	1.449	1.869	1.161	1.841
H _e	0.368	0.659	0.675	0.351	0.361	0.496	0.838	0.684	0.459	0.082	0.313	0.374	0.082	0.227	0.496	0.082	0.409
HAR-4																	
Rango	247	151	113	217	137	153	233	166	107	170	137	102	144	122	141	106	-
alélico	259	153	121	227	146	167	269	170	139	175	139	104		131	150	112	
A (1)	2	2	3	2	3	2	3	2	8	3	2	2	1	2	3	3	2,6875
R _s	1.351	1.422	1.782	1.837	1.932	1.812	1.868	1.606	2.995	1.557	1.122	1.860	1.000	1.063	1.295	1.919	1.651
H _e	0.186	0.224	0.403	0.482	0.524	0.460	0.432	0.332	0.805	0.281	0.063	0.494	0.000	0.031	0.150	0.514	0.336

Tabla A.1. Diversidad genética en materiales de girasol cultivado del BAG-IM (cont.). Rango alélico: tamaños extremos de los alelos encontrados en pares de bases. A: número de alelos. Entre paréntesis se indica el número de alelos exclusivos para cada material. Rs: riqueza alélica poblacional. He: heterocigosis esperada.

Material	Locus																Promedio
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
L-156																	
Rango alélico	259	153	113	225	137	153	260	176	117	170	134	102	144	122	141	112	-
		159	119			167		178		175	144	104			150		
A (0)	1	2	2	1	1	2	1	2	1	2	2	2	1	1	2	1	1.5
R _s	1.000	1.786	1.786	1.000	1.000	1.500	1.000	1.786	1.000	1.786	2.000	1.971	1.000	1.000	1.786	1.000	1.400
H _e	0.000	0.500	0.500	0.000	0.000	0.250	0.000	0.500	0.000	0.500	1.000	0.500	0.000	0.000	0.500	0.000	0.265
P-94																	
Rango alélico	259	140	121	221	137	153	269	176	127	175	144	102	144	119	141	109	-
												104			152		
A (2)	1	1	1	1	1	1	1	1	1	1	1	2	1	1	3	1	1.1875
R _s	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	2.000	1.000	1.000	3.000	1.000	1.188
H _e	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.500	0.000	0.000	1.000	0.000	0.093
F-164																	
Rango alélico	247	153	121	217	146	144	248	170	121	170	144	104	144	122	150	112	-
						153			125								
A (2)	1	1	1	1	1	2	1	1	3	1	1	1	1	1	1	1	1.1875
R _s	1.000	1.000	1.000	1.000	1.000	1.500	1.000	1.000	2.286	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.112
H _e	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.625	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.054

Tabla A.1. Diversidad genética en materiales de girasol cultivado del BAG-IM (cont.). Rango alélico: tamaños extremos de los alelos encontrados en pares de bases. A: número de alelos. Entre paréntesis se indica el número de alelos exclusivos para cada material. Rs: riqueza alélica poblacional. He: heterocigosis esperada.

Material	Locus																Promedio
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
CM-307																	
Rango	247	153	113	205	137	144	248	170	113	170	137	102	144	122	150	112	-
alélico			121	217	146	153	263	178	117	173	139	104					
A (1)	1	1	2	2	2	2	2	2	3	2	2	2	1	1	1	1	1.6875
R _s	1.000	1.000	1.500	1.500	1.500	1.500	1.929	1.786	2.414	1.500	1.786	1.500	1.000	1.000	1.000	1.000	1.432
H _e	0.000	0.000	0.250	0.250	0.250	0.250	0.583	0.500	0.750	0.250	0.500	0.250	0.000	0.000	0.000	0.000	0.239
Colliguay																	
Rango	247	140	113	217	137	153	242	170	107	167	137	102	128	122	141	112	-
alélico	259	159	121	227	146	167	260	178	117	173	144	104	147	131	150		
A (0)	2	5	3	2	2	2	7	3	6	3	2	2	4	2	3	1	3.0625
R _s	1.481	2.911	2.223	1.168	1.305	1.886	3.186	1.966	2.329	1.786	1.891	1.763	2.111	1.886	1.255	1.000	1.884
H _e	0.264	0.794	0.606	0.087	0.160	0.512	0.851	0.531	0.622	0.420	0.515	0.426	0.530	0.510	0.127	0.000	0.433
HAR-1																	
Rango	247	153	119	227	137	153	254	170	135	170	137	102	140	122	141	109	-
alélico	259				146	167	260		139	173		104	144		150		
A (0)	2	1	1	1	3	2	3	1	3	2	1	2	2	1	2	1	1.75
R _s	1.858	1.000	1.000	1.000	1.868	1.757	1.167	1.000	2.152	1.657	1.000	1.791	1.087	1.000	1.826	1.000	1.385
H _e	0.498	0.000	0.000	0.000	0.475	0.424	0.083	0.000	0.575	0.364	0.000	0.439	0.043	0.000	0.467	0.000	0.21

Tabla A.1. Diversidad genética en materiales de girasol cultivado del BAG-IM (cont.). Rango alélico: tamaños extremos de los alelos encontrados en pares de bases. A: número de alelos. Entre paréntesis se indica el número de alelos exclusivos para cada material. Rs: riqueza alélica poblacional. He: heterocigosis esperada.

Material	Locus																Promedio
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
HAR-3																	
Rango	247	140	113	217	137	153	242	170	107	173	134	102	144	122	141	112	-
alélico	259	159	119		146	167	260	171	115	175	144	104					
A (0)	2	2	2	1	2	3	3	2	4	2	2	2	1	1	1	1	1.9375
R _s	1.840	1.711	1.887	1.000	1.312	2.442	2.134	1.887	2.327	1.868	1.791	1.887	1.000	1.000	1.000	1.000	1.630
H _e	0.495	0.386	0.506	0.000	0.167	0.695	0.577	0.527	0.635	0.495	0.449	0.495	0.000	0.000	0.000	0.000	0.339
PAC-1																	
Rango	247	159	121	217	146	153	233	171	117	173	137	102	144	122	141	112	-
alélico	259		123	221				178	119	175		104			150		
A (0)	2	1	2	2	1	1	1	2	2	2	1	2	1	1	2	1	1.5
R _s	1.500	1.000	1.667	1.500	1.000	1.000	1.000	1.500	1.786	1.929	1.000	1.929	1.000	1.000	1.500	1.000	1.332
H _e	0.250	0.000	0.333	0.250	0.000	0.000	0.000	0.250	0.417	0.583	0.000	0.500	0.000	0.000	0.250	0.000	0.177
MP-557																	
Rango	259	153	113	217	146	153	260	178	113	173	137	104	144	122	141	109	-
alélico			119	227		167	263		115	175	144					112	
A (0)	1	1	2	2	1	2	2	1	2	2	3	1	1	1	1	2	1.5625
R _s	1.000	1.000	1.786	1.500	1.000	1.500	1.500	1.000	1.500	1.933	2.286	1.000	1.000	1.000	1.000	1.500	1.344
H _e	0.000	0.000	0.417	0.250	0.000	0.250	0.250	0.000	0.250	0.667	0.667	0.000	0.000	0.000	0.000	0.250	0.187

Tabla A.1. Diversidad genética en materiales de girasol cultivado del BAG-IM (cont.). Rango alélico: tamaños extremos de los alelos encontrados en pares de bases. A: número de alelos. Entre paréntesis se indica el número de alelos exclusivos para cada material. R_s: riqueza alélica poblacional. H_e: heterocigosis esperada.

Material	Locus																Promedio
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
PxS																	
Rango	247	140	113	217	144	153	248	170	111	170	134	102	125	122	141	109	-
alélico	259	155	123	225	146	167	260	180	145	175	144	104	147		150	112	
A (1)	3	3	4	2	2	2	3	3	5	3	4	2	4	1	2	2	2.8125
R _s	1.455	1.307	2.350	1.876	1.791	1.677	1.862	2.058	2.381	2.416	1.713	1.859	1.300	1.000	1.492	1.606	1.759
H _e	0.230	0.155	0.637	0.515	0.450	0.374	0.434	0.559	0.645	0.679	0.350	0.480	0.146	0.000	0.266	0.332	0.39
PraoCo																	
Rango	247	140	113	217	137	153	242	170	109	170	134	102	140	122	141	109	-
alélico	259	159	121	227	146	167	260	180	115	173	144	106	147	131	150	112	
A (1)	3	3	3	3	3	3	5	5	4	2	3	3	3	2	2	2	3.0625
R _s	1.613	1.835	1.989	2.422	1.952	1.914	2.917	2.352	2.117	1.751	2.184	1.973	1.502	1.083	1.806	1.727	1.946
H _e	0.322	0.449	0.487	0.684	0.525	0.484	0.790	0.648	0.537	0.419	0.584	0.532	0.252	0.042	0.457	0.404	0.476
VNIIMK																	
Rango	247	149	113	217	137	153	242	170	113	161	139	102	144	122	141	112	-
alélico	253	153	121		146	167	260	176	115	173	144	104					
A (2)	2	3	2	1	2	4	3	2	2	4	2	2	1	1	1	1	2.0625
R _s	1.886	1.852	1.677	1.000	1.757	2.287	2.009	1.496	1.425	1.613	1.168	1.530	1.000	1.000	1.000	1.000	1.481
H _e	0.518	0.463	0.379	0.000	0.426	0.609	0.540	0.269	0.225	0.309	0.087	0.283	0.000	0.000	0.000	0.000	0.256

Tabla A.1. Diversidad genética en materiales de girasol cultivado del BAG-IM (cont.). Rango alélico: tamaños extremos de los alelos encontrados en pares de bases. A: número de alelos. Entre paréntesis se indica el número de alelos exclusivos para cada material. R_s: riqueza alélica poblacional. H_e: heterocigosis esperada.

Material	Locus																Promedio
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
Júpiter																	
Rango	247	140	113	217	137	142	239	167	109	167	134	102	140	122	141	112	-
alélico	259	157	123	227	146	161	269	178	117	175	144	104	144	131	150		
A (3)	3	6	3	3	3	3	7	3	5	4	3	2	2	2	2	1	3.25
R _s	2.048	2.533	1.658	1.167	1.923	1.702	2.768	2.202	2.371	2.375	2.421	1.783	1.425	1.879	1.513	1.000	1.923
H _e	0.540	0.695	0.328	0.083	0.505	0.375	0.739	0.601	0.645	0.657	0.677	0.442	0.225	0.507	0.279	0.000	0.456
Sunrise																	
Rango	247	140	113	217	137	142	245	170	109	170	134	102	144	122	141	112	-
alélico	262	153	121		146	167	257	180	123	175	144	104		134	150		
A (1)	4	2	2	1	2	4	3	2	3	3	3	2	1	2	2	1	2.3125
R _s	2.131	1.238	1.720	1.000	1.202	2.017	1.987	1.875	1.823	1.883	1.307	1.758	1.000	1.202	1.791	1.000	1.558
H _e	0.579	0.125	0.401	0.000	0.105	0.490	0.547	0.507	0.442	0.450	0.155	0.418	0.000	0.105	0.447	0.000	0.298
Comangir																	
Rango	247	140	113	217	137	153	245	170	111	161	134	102	125	122	141	109	-
alélico	259	159	121		146	167	263	174	143	175	144	104	147	131	150	112	
A (1)	2	5	4	1	2	2	4	2	4	4	3	2	3	2	2	2	2.75
R _s	1.366	2.100	1.694	1.000	1.566	1.757	2.314	1.327	2.511	1.832	2.328	1.496	1.954	1.456	1.659	1.806	1.760
H _e	0.192	0.516	0.340	0.000	0.303	0.424	0.614	0.171	0.681	0.405	0.649	0.265	0.469	0.242	0.362	0.455	0.38

Tabla A.1. Diversidad genética en materiales de girasol cultivado del BAG-IM (cont.). Rango alélico: tamaños extremos de los alelos encontrados en pares de bases. A: número de alelos. Entre paréntesis se indica el número de alelos exclusivos para cada material. R_s : riqueza alélica poblacional. H_e : heterocigosis esperada.

Material	Locus																Promedio
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
HAR-2																	
Rango	247	140	113	217	137	153	257	161	107	170	137	102	140	122	141	109	-
alélico	259	153	121	227	146	167	272	170	115	175		104	144	143	150	112	
A (3)	2	3	3	2	3	2	4	2	5	3	1	2	2	3	2	2	2.5625
R_s	1.552	1.974	1.887	1.842	1.852	1.811	2.080	1.213	2.334	1.916	1.000	1.758	1.143	1.286	1.100	1.825	1.661
H_e	0.303	0.524	0.458	0.491	0.464	0.464	0.551	0.108	0.637	0.498	0.000	0.430	0.071	0.140	0.050	0.466	0.353
Fcambá																	
Rango	244	140	113	203	146	153	230	170	111	173	137	102	140	122	141	109	-
alélico	247	153	121	217		161	242	171	115		144			131	150		
A (3)	2	2	2	2	1	2	2	2	3	1	2	1	1	2	2	1	1.75
R_s	1.253	1.885	1.870	1.143	1.000	1.481	1.676	1.886	2.096	1.000	1.816	1.000	1.000	1.869	1.887	1.000	1.491
H_e	0.133	0.518	0.510	0.071	0.000	0.253	0.376	0.530	0.566	0.000	0.462	0.000	0.000	0.500	0.519	0.000	0.277
HA336																	
Rango	247	153	115	217	146	153	257	171	113	173	134	102	137	122	141	106	-
alélico			121		151	167	260		119	175	137	104	144			112	
A(2)	1	1	3	1	2	3	2	1	3	2	2	2	2	1	1	2	1.8125
R_s	1.000	1.000	2.000	1.000	1.971	2.333	1.500	1.000	2.800	1.786	1.971	1.929	1.500	1.000	1.000	1.500	1.581
H_e	0.000	0.000	0.458	0.000	0.583	0.667	0.250	0.000	0.750	0.500	0.667	0.500	0.250	0.000	0.000	0.250	0.304

Tabla A.2. Ajuste a las proporciones de Hardy-Weinberg, F_{IS} por locus y material. m: locus monomórfico. * F_{IS} : significativo bajo la hipótesis alternativa de defecto de heterocigotas para ese locus en el material evaluado. * F_{IS} con fondo gris: significativo bajo la hipótesis alternativa de exceso de heterocigotas.

Material	Locus																F_{IS} global
	77	95	293	928	991	1402	1848	2063	2077	2920	3239	3555	3581	3582	4103	4239	
KLM	0.785	0.651*	1	1	0.532*	1	1	0.731*	0.5	0.763*	1*	0.86*	m	0.759*	m	0.619*	0.740*
IAC	1*	0.142	0.162	0.881*	-0.04	0.748*	-0.09	0.025	-0.09	-0.02	0.49	0.185	-0.02	-0.09	0.036	-0.02	0.206*
HAR-4	1*	0.44	0.379	0.933*	0.821*	0.184	0.421*	0.529	0.107	0.555	1	0.494	m	m	0.792*	0.635*	0.512*
Colliguay	1	0.204	0.057	1	0.653	0.405	0.029	0.193	0.23	0.242	0.382	0.041	-0.07	0.233	0.318	M	0.226*
HAR-1	0.916*	M	M	m	0.789*	0.214	0.5	m	-0.739*	0.5	m	-0.45	m	m	0.019	M	0.158
HAR-3	1*	-0.29	-0.06	m	1	0.589*	-0.067	1*	0.099	-0.011	0.314	-0.73	m	m	m	M	0.230*
PxS	0.131	0.66	0.738*	1*	0.444	0.297	0.515	0.403	0.457	0.448	0.523	-0.62	-0.03	m	0.624	0.397	0.425*
PraoCo	0.612	0.331	0.218	0.746*	0.307	0.483*	-0.101	0.559	0.224	0.091	0.144	-0.07	0.459	m	0.269	0.072	0.277*
VNIIMK	1*	0.509	0.748*	m	0.413	0.254	0.305	0.838*	0.258	0.437	1	-0.18	m	m	m	m	0.484*
Júpiter	1*	0.542*	0.11	0.5	0.009	0.667*	0.356*	0.667*	0.418*	0.446*	0.358	0.34	0.258	0.425	0.837*	m	0.463*
Sunrise	0.545*	1	0.212	m	1	0.207	0.519	0.452	0.404	0.306	0.66	-0.38	m	1	0.294	m	0.364*
Comangir	0.349	0.031	-0.15	m	-0.2	0.214	0.149	0.468	0.144	0.176	0.4	0.179	0.378	0.25	0.195	0.084	0.174*
HAR-2	0.835*	0.109	0.126	1*	0.88*	0.521	0.456*	-0.03	0.389	0.528	m	0.633	m	-0.02	m	-0.07	0.443*
F. camba	1	0.474	0.738	m	m	-0.13	0.468	1*	0.243	m	0.134	m	m	0.143	0.615	m	0.455*

Tabla A.3. Estimación del número de poblaciones mediante el método bayesiano para los materiales analizados. Ln P(D)=logaritmo natural de la probabilidad de datos dado K.

K	Ln P(D)	K	Ln P(D)	K	Ln P(D)	K	Ln P(D)
1	-10406,5	6	-7462,8	11	-6368,1	16	-5800
1	-10406,1	6	-7640,6	11	-6655,2	16	-5794,9
1	-10406,1	6	-7484,9	11	-6393,5	16	-5950,3
1	-10405,8	6	-7801,3	11	-6198,4	16	-5999,8
1	-10406	6	-7643,2	11	-6237,3	16	-6005,3
1	-10410	6	-7679,6	11	-6366,1	16	-5760
1	-10406,1	6	-7587,8	11	-6370,2	16	-5880,2
1	-10406,2	6	-7708,4	11	-6347	16	-5941
1	-10406,4	6	-7594,9	11	-6399,6	16	-6021
1	-10406	6	-7718,6	11	-6366	16	-5743,6
2	-9160,1	7	-7287,4	12	-6543,1	17	-5795,6
2	-9160,3	7	-7286	12	-6546,2	17	-5562,2
2	-9159,7	7	-7276,9	12	-6022,7	17	-5958,2
2	-9160,3	7	-7250	12	-6021,9	17	-6032,4
2	-9109	7	-7104,9	12	-6024,3	17	-6045,9
2	-9109	7	-7249,1	12	-6201,9	17	-5988
2	-9108,6	7	-7359,5	12	-6231,3	17	-5984,5
2	-9159,6	7	-7223	12	-6391,1	17	-5877
2	-9160	7	-7522	12	-6307	17	-5936
2	-9159,6	7	-7341,3	12	-6261,5	17	-6024,6
3	-8720	8	-7003,7	13	-5969	18	-5566,1
3	-8717,8	8	-7198,5	13	-5908	18	-5742,7
3	-8587,1	8	-6843,2	13	-6146,7	18	-6025,3
3	-8587,5	8	-6798,6	13	-5910,4	18	-6045,1
3	-8660,8	8	-6824,1	13	-6308,8	18	-5988
3	-8703,8	8	-7073,6	13	-6078,1	18	-5763,2
3	-8706,4	8	-6951,6	13	-6458,8	18	-5841,2
3	-8681,4	8	-7034,1	13	-6113,4	18	-5966
3	-8679,9	8	-6937,6	13	-6106,3	18	-6147
3	-8682,2	8	-7176	13	-6107,3	18	-5963,7
4	-8347,8	9	-6840,3	14	-5850,7	19	-5557,4
4	-8250,5	9	-6717,5	14	-6028,2	19	-5776,6
4	-8123,4	9	-6654,9	14	-5924,2	19	-5889,3
4	-8224,9	9	-6700,5	14	-5910,1	19	-5798,1
4	-8296,6	9	-6877,8	14	-5837,5	19	-5911
4	-8227	9	-6768,4	14	-6060,8	19	-6004,2
4	-8203,3	9	-6721,4	14	-5889,4	19	-5587
4	-8229	9	-6989,8	14	-5994,3	19	-5647
4	-8229,9	9	-6820,4	14	-5878,5	19	-5879,6
4	-8394,5	9	-6885,9	14	-6026,6	19	-6007
5	-7966,9	10	-6575	15	-5948,2	20	-5560,8
5	-8007,9	10	-6384,6	15	-5973,2	20	-5698,2
5	-7987	10	-6392	15	-5790,7	20	-5977
5	-7765	10	-6410,6	15	-5790	20	-5697,2
5	-7859,3	10	-6683,7	15	-5790,1	20	-6004,8
5	-7801,7	10	-6557,5	15	-5944,4	20	-5997
5	-7911	10	-6549,8	15	-5944,3	20	-5741,2
5	-7894,2	10	-6573,2	15	-5818,7	20	-5874
5	-8102,9	10	-6482,9	15	-5945,5	20	-6025,7
5	-7906,9	10	-6600,5	15	-6003,5	20	-6055,8

Tabla A.4. Caracteres agronómicos evaluados en ensayo de campo para tolerancia a sequía (campaña 2003-2004). ID: identificación de la línea. ALT: altura de planta (en cm). P 100: peso de 100 granos (en gramos). NSC: número de semillas por capítulo. MG: contenido de materia grasa en el grano sobre materia seca. RENSH: rendimiento de semillas (en Kg/ha a 11% de humedad). RENSHR: rendimiento de semillas relativo al control irrigado (S/R)*100.

ID	ALT		P 100		NSC		MG		RENSH		RENSHR
	RIEGO	SEQUIA									
R049	145	121	3,4	3,2	326	304	40,2	40,0	523	460	87,95
R415	145	130	3,5	3,6	169	146	47,0	42,1	283	247	87,28
R427	135	114	2,6	2,2	568	551	40,5	51,3	682	583	85,48
R437	151	117	2,5	3,2	499	334	50,7	47,1	576	492	85,42
R432	135	130	4,2	4,1	562	507	57,4	57,5	1116	940	84,23
R423	128	109	2,5	2,6	636	508	51,6	49,8	727	612	84,18
R023	119	111	3,6	3,4	478	419	47,0	44,0	810	670	82,72
HAR4	173	146	7,0	6,6	1326	1224	36,3	34,6	4338	3565	82,18
R413	183	154	2,7	3,0	229	169	46,2	47,3	286	235	82,17
R073	119	114	3,0	3,0	174	146	47,1	48,1	248	203	81,85
R431	152	137	2,7	3,6	546	307	48,0	47,4	680	553	81,32
B71	141	131	4,0	4,5	1137	809	42,4	46,2	2132	1721	80,72
R417	119	112	2,9	2,6	462	371	44,5	47,6	554	444	80,14
R433	170	131	3,8	4,5	466	306	45,1	50,6	796	633	79,52
R043	141	112	3,1	2,9	307	248	48,4	45,1	446	345	77,35
B10	154	123	6,6	5,9	712	596	38,1	35,6	2173	1660	76,39
R422	138	109	2,8	2,6	481	384	53,8	51,4	624	464	74,36
R435	151	117	4,2	3,6	269	216	45,0	47,6	533	379	71,11
R426-1	130	113	3,0	3,0	215	157	45,6	43,4	321	218	67,91
RHA274	121	118	3,3	2,9	290	224	41,6	43,6	451	306	67,85
R429	132	125	4,3	3,8	247	198	38,5	40,5	510	339	66,47
B67	137	122	8,5	4,7	433	513	36,3	34,0	1729	1120	64,78
AGUARA	182	148	6,5	6,1	1936	1250	42,3	45,0	5821	3601	61,86
HA89	114	88	4,3	3,1	652	500	44,3	41,2	1270	739	58,19
B59	120	107	4,9	3,6	1241	900	34,5	32,9	2846	1508	52,99
R420	135	122	5,0	3,9	334	213	45,5	46,8	770	404	52,47
R421	136	125	2,6	2,3	453	261	49,7	50,9	562	284	50,53
B99	177	159	5,1	5,2	1119	544	41,4	44,1	2657	1326	49,91
R419	136	95	2,2	1,5	644	431	47,6	48,5	651	289	44,39
R307	97	80	2,2	2,0	210	98	44,8	41,1	219	87	39,73

Tabla A.5. Resultados del ensayo en condiciones controladas para caracterización de líneas frente a déficit hídrico generado con soluciones de manitol. EG: energía germinativa, expresada en número de plántulas. PF: peso fresco en gramos. PS: peso seco en gramos. RFG: recuento final de germinación. RFGR (%): recuento final de germinación relativo en porcentaje.

Genotipo	Tratamiento	Repetición	EG	PF	PS	RFG	RFGR (%)
R426-1	Control	1	13	10,04	0,396	20	0
R426-1	200	1	4	1,53	0,085	3	15
R426-1	400	1	1	0	0	0	0
R426-1	Control	2	22	8,4	0,3338	20	0
R426-1	200	2	4	1,3	0,0879	4	20
R426-1	400	2	4	0,7	0,1033	6	30
R426-1	Control	3	13	3,9	0,2109	13	0
R426-1	200	3	9	0	0	0	0
R426-1	400	3	6	0	0	0	0
R307	Control	1	7	11,64	0,7897	17	0
R307	200	1	7	0	0	0	0
R307	400	1	1	0	0	0	0
R307	Control	2	15	16,4	1,0539	22	0
R307	200	2	12	0,1	0,0447	1	4,545
R307	400	2	0	0	0	0	0
R307	Control	3	11	12,4	0,937	24	0
R307	200	3	2	0	0	0	0
R307	400	3	0	0	0	0	0
R417	Control	1	4	10,71	0,4373	16	0
R417	200	1	1	1,79	0,1225	3	18,75
R417	400	1	0	0	0	0	0
R417	Control	2	5	10,3	0,5482	16	0
R417	200	2	0	0	0	0	0
R417	400	2	2	0	0	0	0
R417	Control	3	3	8,6	0,4638	15	0
R417	200	3	2	0	0	0	0
R417	400	3	0	0	0	0	0
R419	Control	1	5	6,87	0,2755	21	0
R419	200	1	1	0	0	0	0
R419	400	1	0	0	0	0	0
R419	Control	2	7	5,5	0,2305	22	0
R419	200	2	3	0,3	0,0235	3	13,636
R419	400	2	0	0,2	0,0239	3	13,636
R419	Control	3	11	5,3	0,5185	25	0
R419	200	3	2	0	0	0	0
R419	400	3	1	0	0	0	0
B59	Control	1	3	19,23	0,7757	19	0
B59	200	1	0	1,6	0,0907	2	10,526
B59	400	1	0	1,69	0,1836	4	21,053
B59	Control	2	0	24	0,9982	25	0
B59	200	2	0	0	0	0	0
B59	400	2	0	0	0	0	0
B59	Control	3	6	24,1	1,1865	24	0

Apéndice

B59	200	3	0	0	0	0	0
B59	400	3	0	0	0	0	0
B67	Control	1	7	17,41	0,7689	14	0
B67	200	1	0	0	0	0	0
B67	400	1	0	0	0	0	0
B67	Control	2	11	16,3	0,8831	14	0
B67	200	2	0	0	0	0	0
B67	400	2	0	0	0	0	0
B67	Control	3	13	18,7	1,0658	17	0
B67	200	3	3	0	0	0	0
B67	400	3	0	0	0	0	0
R023	Control	1	7	10,76	0,3937	16	0
R023	200	1	0	1,42	0,1018	2	12,5
R023	400	1	0	0	0	0	0
R023	Control	2	20	10,3	0,5175	21	0
R023	200	2	5	0,3	0,0702	1	4,7619
R023	400	2	0	0	0	0	0
R023	Control	3	11	10	0,3761	22	0
R023	200	3	7	0	0	0	0
R023	400	3	0	0,4	0,0999	4	18,1818182
R421	Control	1	16	11,45	0,4853	20	0
R421	200	1	0	1,5	0,0725	3	15
R421	400	1	0	0	0	0	0
R421	Control	2	20	10,3	0,4638	24	0
R421	200	2	5	1	0,1186	5	20,833
R421	400	2	0	0	0	0	0
R421	Control	3	12	9,9	0,3658	18	0
R421	200	3	2	0,2	0,0258	1	5,5555556
R421	400	3	0	0	0	0	0
R413	Control	1	9	16,33	0,6984	24	0
R413	200	1	1	0	0	0	0
R413	400	1	0	0	0	0	0
R413	Control	2	18	14,9	0,6397	24	0
R413	200	2	4	0,7	0,0675	3	12,5
R413	400	2	0	0	0	0	0
R413	Control	3	13	15,1	0,5756	25	0
R413	200	3	2	0	0	0	0
R413	400	3	0	0	0	0	0
R437	Control	1	17	13,71	0,7262	19	0
R437	200	1	3	0	0	0	0
R437	400	1	0	2,26	0,2908	8	42,105
R437	Control	2	20	12,2	0,7181	23	0
R437	200	2	4	0,7	0,1033	3	13,0435
R437	400	2	0	0	0	0	0
R437	Control	3	17	12,7	0,518	22	0
R437	200	3	10	0	0	0	0
R437	400	3	5	0	0	0	0
B99	Control	1	15	25,44	1,2713	24	0
B99	200	1	2	0	0	0	0
B99	400	1	0	0	0	0	0
B99	Control	2	17	26,1	1,358	24	0

Apêndice

B99	200	2	0	1	0,1638	3	12,5
B99	400	2	0	0,3	0,0595	1	4,167
B99	Control	3	24	25,4	1,6658	24	0
B99	200	3	13	1,9	0,2252	4	16,6666667
B99	400	3	0	0	0	0	0
962-C	Control	1	16	24,98	1,6579	22	0
962-C	200	1	0	0	0	0	0
962-C	400	1	0	0	0	0	0
962-C	Control	2	21	24,2	1,492	24	0
962-C	200	2	1	1,4	0,1697	3	12,5
962-C	400	2	0	0	0	0	0
962-C	Control	3	10	27	1,4729	25	0
962-C	200	3	1	0	0	0	0
962-C	400	3	0	0	0	0	0
R073	Control	1	23	16,3	0,6838	23	0
R073	200	1	3	0	0	0	0
R073	400	1	2	1,65	0,1919	5	21,739
R073	Control	2	19	13,2	0,6721	23	0
R073	200	2	14	0	0	0	0
R073	400	2	4	0,2	0,0751	3	13,0435
R073	Control	3	10	9,6	0,4828	17	0
R073	200	3	5	0	0	0	0
R073	400	3	0	0,5	0,1372	5	29,4117647
HAR4	Control	1	5	2,75	0,102	3	0
HAR4	200	1	0	0	0	0	0
HAR4	400	1	0	0	0	0	0
HAR4	Control	2	9	8,1	0,3705	12	0
HAR4	200	2	0	0,9	0,1048	3	25
HAR4	400	2	0	0	0	0	0
HAR4	Control	3	3	4,8	0,2282	7	0
HAR4	200	3	0	0,6	0,1036	2	28,5714286
HAR4	400	3	0	0	0	0	0
R420	Control	1	4	13,98	0,593	19	0
R420	200	1	3	0	0	0	0
R420	400	1	0	1,73	0,1817	6	31,579
R420	Control	2	9	15,5	0,7274	22	0
R420	200	2	2	0	0	0	0
R420	400	2	0	0	0	0	0
R420	Control	3	1	9	0,324	14	0
R420	200	3	1	0	0	0	0
R420	400	3	0	0	0	0	0
R415	Control	1	4	5,69	0,2855	13	0
R415	200	1	1	0	0	0	0
R415	400	1	0	0	0	0	0
R415	Control	2	15	6,4	0,2941	16	0
R415	200	2	4	0	0	0	0
R415	400	2	2	0	0	0	0
R415	Control	3	13	5	0,2972	14	0
R415	200	3	13	0,9	0,1506	6	42,8571429
R415	400	3	4	0,1	0,0304	1	7,14285714
430-C	Control	1	2	9,68	0,3808	12	0

430-C	200	1	0	0	0	0	0
430-C	400	1	0	0	0	0	0
430-C	Control	2	3	10,8	0,4917	16	0
430-C	200	2	0	0	0	0	0
430-C	400	2	0	0	0	0	0
430-C	Control	3	12	9,2	0,8791	15	0
430-C	200	3	10	1,8	0,2709	4	26,6666667
430-C	400	3	0	0	0	0	0
R427	Control	1	23	15,07	0,6349	24	0
R427	200	1	3	0	0	0	0
R427	400	1	0	0	0	0	0
R427	Control	2	19	14,3	0,7409	25	0
R427	200	2	3	0,7	0,1129	4	16
R427	400	2	4	0	0	0	0
R427	Control	3	2	11,3	0,5579	25	0
R427	200	3	0	0	0	0	0
R427	400	3	0	0,5	0,071	3	12
R435	Control	1	21	17,59	0,8041	23	0
R435	200	1	7	0	0	0	0
R435	400	1	0	1,8	0,1538	6	26,0870
R435	Control	2	25	18,9	1,0016	25	0
R435	200	2	0	1,5	0,1657	5	20
R435	400	2	0	1,3	0,3403	11	44
R435	Control	3	13	17,8	0,8226	24	0
R435	200	3	0	0,4	0,0674	2	8,33333333
R435	400	3	0	0	0	0	0
R422	Control	1	10	20,37	0,9266	23	0
R422	200	1	1	0	0	0	0
R422	400	1	0	0	0	0	0
R422	Control	2	23	17,7	0,8268	25	0
R422	200	2	4	0	0	0	0
R422	400	2	0	0	0	0	0
R422	Control	3	12	14,7	0,8338	25	0
R422	200	3	0	0	0	0	0
R422	400	3	0	0	0	0	0
RHA274	Control	1	10	11,45	0,571	18	0
RHA274	200	1	0	0	0	0	0
RHA274	400	1	0	0	0	0	0
RHA274	Control	2	18	12,4	0,6562	23	0
RHA274	200	2	5	1,8	0,1979	8	34,783
RHA274	400	2	6	0	0	0	0
RHA274	Control	3	11	10,4	0,4604	18	0
RHA274	200	3	1	0	0	0	0
RHA274	400	3	0	0	0	0	0
R423	Control	1	9	10,91	0,5599	18	0
R423	200	1	3	0	0	0	0
R423	400	1	0	1,27	0,0583	2	11,111
R423	Control	2	21	11,4	0,7032	24	0
R423	200	2	4	0	0	0	0
R423	400	2	3	0,1	0,0247	1	4,167
R423	Control	3	21	12,7	0,5914	24	0

Apéndice

R423	200	3	16	0,2	0,0385	2	0
R423	400	3	3	0	0	0	0
R049	Control	1	14	16,58	0,7391	22	0
R049	200	1	1	0	0	0	0
R049	400	1	1	0	0	0	0
R049	Control	2	20	14,7	0,6901	24	0
R049	200	2	9	0	0	0	0
R049	400	2	0	0	0	0	0
R049	Control	3	11	11,2	0,6688	22	0
R049	200	3	8	1,4	0,2028	6	27,2727273
R049	400	3	2	0	0	0	0
HA89	Control	1	8	23,9	1,1123	25	0
HA89	200	1	1	3,03	0,226	4	16
HA89	400	1	0	0	0	0	0
HA89	Control	2	20	24,7	1,4947	24	0
HA89	200	2	0	3,3	0,371	7	29,167
HA89	400	2	0	0	0	0	0
HA89	Control	3	18	23,9	1,4282	23	0
HA89	200	3	2	0	0	0	0
HA89	400	3	0	0	0	0	0
R431	Control	1	9	8,96	0,3855	17	0
R431	200	1	1	0	0	0	0
R431	400	1	0	1,24	0,038	4	23,529
R431	Control	2	10	9,2	0,3864	19	0
R431	200	2	5	1	0,1442	5	26,316
R431	400	2	0	0,2	0,0599	4	21,053
R431	Control	3	9	6,2	0,3799	17	0
R431	200	3	11	0,4	0,0792	3	17,6470588
R431	400	3	3	0	0	0	0
R433	Control	1	19	14,34	0,9047	20	0
R433	200	1	2	0	0	0	0
R433	400	1	0	2,84	0,3749	9	45
R433	Control	2	23	17,3	0,8656	24	0
R433	200	2	0	1,8	0,1413	6	25
R433	400	2	0	0,6	0,1455	5	20,83
R433	Control	3	22	19,1	1,0514	24	0
R433	200	3	7	0	0	0	0
R433	400	3	0	0	0	0	0
1063-C	Control	1	6	11,1	0,513	13	0
1063-C	200	1	1	0	0	0	0
1063-C	400	1	0	0	0	0	0
1063-C	Control	2	10	7,9	0,4637	12	0
1063-C	200	2	3	0	0	0	0
1063-C	400	2	0	0	0	0	0
1063-C	Control	3	6	7,2	0,4126	11	0
1063-C	200	3	3	0	0	0	0
1063-C	400	3	0	0	0	0	0
964-C	Control	1	1	14,28	0,5922	13	0
964-C	200	1	0	0	0	0	0
964-C	400	1	0	0	0	0	0
964-C	Control	2	1	9,2	0,4371	9	0

964-C	200	2	1	0	0	0	0
964-C	400	2	0	0	0	0	0
964-C	Control	3	14	23,5	1,0803	20	0
964-C	200	3	0	0	0	0	0
964-C	400	3	0	0	0	0	0
R043	Control	1	1	4,96	0,2245	8	0
R043	200	1	0	0	0	0	0
R043	400	1	0	0	0	0	0
R043	Control	2	0	2,7	0,1491	5	0
R043	200	2	0	0	0	0	0
R043	400	2	0	0	0	0	0
R043	Control	3	1	2	0,2578	4	0
R043	200	3	0	0	0	0	0
R043	400	3	0	0	0	0	0
B10	Control	1	7	22,15	1,3457	20	0
B10	200	1	0	0	0	0	0
B10	400	1	0	0	0	0	0
B10	Control	2	15	31,9	1,6655	25	0
B10	200	2	0	5,6	0,6795	11	44
B10	400	2	0	0,9	0,2103	3	12
B10	Control	3	20	30,8	1,4991	25	0
B10	200	3	0	0,8	0,1355	2	8
B10	400	3	0	0	0	0	0
B71	Control	1	3	14,32	0,5597	18	0
B71	200	1	0	0	0	0	0
B71	400	1	0	0	0	0	0
B71	Control	2	0	12,5	0,571	19	0
B71	200	2	0	1,2	0,2097	6	31,579
B71	400	2	0	0	0	0	0
B71	Control	3	5	13,4	0,5436	20	0
B71	200	3	0	0,2	0,0345	1	5
B71	400	3	0	0	0	0	0
R429	Control	1	20	18,53	0,7284	25	0
R429	200	1	3	0	0	0	0
R429	400	1	0	1,41	0,1123	4	16
R429	Control	2	16	19,1	0,6504	25	0
R429	200	2	0	1,4	0,1665	4	16
R429	400	2	0	0	0	0	0
R429	Control	3	25	16,9	0,7186	23	0
R429	200	3	6	1,2	0,1447	3	13,0434783
R429	400	3	0	0	0	0	0
A59	Control	1	6	19,48	1,0434	18	0
A59	200	1	0	0	0	0	0
A59	400	1	0	0	0	0	0
A59	Control	2	19	26,2	1,3279	25	0
A59	200	2	3	0,9	0,071	2	8
A59	400	2	0	0,2	0,0373	2	8
A59	Control	3	11	23,8	1,3078	25	0
A59	200	3	0	0	0	0	0
A59	400	3	0	0	0	0	0
R432	Control	1	1	14,68	0,5555	12	0

R432	200	1	0	0	0	0	0
R432	400	1	0	1,74	0,1602	3	25
R432	Control	2	14	25,1	1,1221	24	0
R432	200	2	6	0,4	0,075	1	4,167
R432	400	2	0	0	0	0	0
R432	Control	3	13	21,1	1,0413	22	0
R432	200	3	8	3,9	0,5047	10	45,4545455
R432	400	3	0	0	0	0	0

Tabla A.6. Valores predichos de probabilidad de germinación para cada línea en el control y bajo condición de déficit moderado y severo. Predicho-200: valor predictivo bajo tratamiento con solución de manitol 200mM. Predicho-400: valor predictivo bajo tratamiento con solución de manitol 400mM. Predicho-C: valor predictivo en control irrigado.

Genotipo	Repetición	Predicho-200	Predicho-400	Predicho-C
1063-C	1	4,20259E-10	4,20259E-10	0,394671334
1063-C	2	9,17682E-10	9,17682E-10	0,587408435
1063-C	3	5,44502E-10	5,44501E-10	0,457920231
430-C	1	0,036573236	4,20259E-10	0,489392428
430-C	2	0,076548106	9,17682E-10	0,676677337
430-C	3	0,046878658	5,44501E-10	0,553930235
962-C	1	0,027268945	4,20259E-10	0,929226018
962-C	2	0,05768303	9,17682E-10	0,966295586
962-C	3	0,035048025	5,44501E-10	0,944478396
964-C	1	4,20259E-10	4,20259E-10	0,475489674
964-C	2	9,17682E-10	9,17682E-10	0,664376827
964-C	3	5,44501E-10	5,44501E-10	0,540133499
A59	1	0,018072441	0,018072441	0,877438853
A59	2	0,038636731	0,038636731	0,939878191
A59	3	0,023290828	0,023290828	0,902682956
B10	1	0,125270548	0,027268945	0,91184017
B10	2	0,238220862	0,05768303	0,957600485
B10	3	0,156508591	0,035048025	0,930559345
B59	1	0,018072441	0,036573236	0,877438853
B59	2	0,038636731	0,076548106	0,939878191
B59	3	0,023290828	0,046878658	0,902682956
B67	1	4,20259E-10	4,20259E-10	0,517570432
B67	2	9,17682E-10	9,17682E-10	0,700837912
B67	3	5,44501E-10	5,44501E-10	0,581591656
B71	1	0,065139587	4,20259E-10	0,697102651
B71	2	0,132057924	9,17682E-10	0,834037864
B71	3	0,082802489	5,44501E-10	0,748859484
B99	1	0,065139587	0,008983028	0,946735105
B99	2	0,132057924	0,019409083	0,974881769
B99	3	0,082802489	0,011607889	0,958383126
HA89	1	0,104776985	4,20259E-10	0,946735105
HA89	2	0,203549088	9,17682E-10	0,974881769
HA89	3	0,131673927	5,44501E-10	0,958383126
HAR4	1	0,045985996	4,20259E-10	0,223195365
HAR4	2	0,095232131	9,17682E-10	0,385525287
HAR4	3	0,058781873	5,44501E-10	0,271279347
R023	1	0,027268945	0,036573236	0,728771037
R023	2	0,05768303	0,076548106	0,854380139

R023	3	0,035048025	0,046878658	0,776848824
R043	1	4,20259E-10	4,20259E-10	0,167630123
R043	2	9,17682E-10	9,17682E-10	0,305437585
R043	3	5,44501E-10	5,44501E-10	0,206932291
R049	1	0,055507894	4,20259E-10	0,877438853
R049	2	0,113735319	9,17682E-10	0,939878191
R049	3	0,070756787	5,44501E-10	0,902682956
R073	1	4,18866E-10	0,128602515	0,821985333
R073	2	9,14641E-10	0,243720041	0,909770589
R073	3	5,42697E-10	0,160518961	0,856787041
R307	1	0,008983028	4,20259E-10	0,793602869
R307	2	0,019409083	9,17682E-10	0,893572309
R307	3	0,011607889	5,44501E-10	0,832824821
R413	1	0,027268945	4,20259E-10	0,964367256
R413	2	0,05768303	9,17682E-10	0,983360356
R413	3	0,035048025	5,44501E-10	0,972272388
R415	1	0,055507894	0,008983028	0,489392428
R415	2	0,113735319	0,019409083	0,676677337
R415	3	0,070756787	0,011607889	0,553930235
R417	1	0,027268945	4,20259E-10	0,546245848
R417	2	0,05768303	9,17682E-10	0,724420134
R417	3	0,035048025	5,44501E-10	0,609334017
R419	1	0,027268945	0,027268945	0,877438853
R419	2	0,05768303	0,05768303	0,939878191
R419	3	0,035048025	0,035048025	0,902682956
R420	1	4,20259E-10	0,055507894	0,665933177
R420	2	9,17682E-10	0,113735319	0,813183314
R420	3	5,44501E-10	0,070756787	0,720883509
R421	1	0,084734921	4,20259E-10	0,777208205
R421	2	0,168162618	9,17682E-10	0,88395735
R421	3	0,107102461	5,44501E-10	0,818834445
R422	1	4,20259E-10	4,20259E-10	0,964367256
R422	2	9,17682E-10	9,17682E-10	0,983360356
R422	3	5,44501E-10	5,44501E-10	0,972272388
R423	1	0,018072441	0,027268945	0,843532386
R423	2	0,038636731	0,05768303	0,921704149
R423	3	0,023290828	0,035048025	0,874763465
R426-1	1	0,065139587	0,055507894	0,63526282
R426-1	2	0,132057924	0,113735319	0,79180536
R426-1	3	0,082802489	0,070756787	0,692931821
R427	1	0,036573236	0,027268945	0,982122285
R427	2	0,076548106	0,05768303	0,991732664
R427	3	0,046878658	0,035048025	0,98614505
R429	1	0,065139587	0,036573236	0,964367256
R429	2	0,132057924	0,076548106	0,983360356
R429	3	0,082802489	0,046878658	0,972272388
R431	1	0,074881719	0,074881719	0,63526282
R431	2	0,150200244	0,150200244	0,79180536
R431	3	0,094918037	0,094918037	0,692931821
R432	1	0,104776985	0,027268945	0,7128745
R432	2	0,203549088	0,05768303	0,844272341
R432	3	0,131673927	0,035048025	0,762853159
R433	1	0,055507894	0,13568807	0,877438853
R433	2	0,113735319	0,255290014	0,939878191

R433	3	0,070756787	0,169021916	0,902682956
R435	1	0,065139587	0,167630123	0,946735105
R435	2	0,132057924	0,305437585	0,974881769
R435	3	0,082802489	0,206932291	0,958383126
R437	1	0,027268945	0,074881719	0,810121858
R437	2	0,05768303	0,150200244	0,903067514
R437	3	0,035048025	0,094918037	0,846810629
RHA274	1	0,074881719	4,20259E-10	0,728771037
RHA274	2	0,150200244	9,17682E-10	0,854380139
RHA274	3	0,094918037	5,44501E-10	0,776848824

Tabla A.7. Genotipificación con SSR de 34 líneas pertenecientes al programa de mejoramiento para tolerancia a sequía del BAG-IM. a. La identificación individual dentro de cada línea se indica después del guión.

Líneas ^a	HA77		HA293		HA928		HA2063		HA2920		HA3239		HA4103	
B10-1	259	259	0	0	217	217	178	178	173	173	144	144	150	150
B10-2	0	0	87	87	217	217	178	178	173	173	144	144	150	150
B10-3	247	259	87	87	217	217	178	178	173	173	144	144	150	150
A59-1	259	259	113	121	217	217	170	170	170	173	137	137	141	150
A59-2	259	259	0	0	0	0	170	170	170	170	137	137	150	150
A59-3	259	259	121	121	227	227	170	170	170	170	137	137	150	150
R073-1	259	259	113	113	221	221	180	180	173	173	134	134	150	150
R073-2	259	259	113	113	221	221	180	180	173	173	0	0	150	150
R422-1	259	259	119	119	221	221	180	180	173	173	134	134	141	141
R422-2	0	0	113	113	221	221	171	180	173	173	134	134	141	141
R422-3	259	259	113	113	221	221	180	180	173	173	134	134	141	141
R435-1	259	259	121	121	227	227	178	178	175	175	144	144	141	141
R435-2	259	259	121	121	227	227	178	178	175	175	144	144	141	141
R435-3	259	259	121	121	227	227	178	178	175	175	144	144	141	141
B59-1	259	259	121	121	227	227	170	170	170	170	137	137	150	150
B59-2	259	259	121	121	0	0	170	170	170	170	137	137	0	0
B59-3	259	259	121	121	227	227	170	170	170	170	0	0	150	150
430C-1	259	259	119	119	227	227	170	170	170	170	137	137	150	150
430C-2	259	259	121	121	0	0	170	170	175	175	137	137	150	150
430C-3	259	259	121	121	217	217	170	170	175	175	137	137	150	150
R307-1	259	259	121	121	221	221	170	170	173	173	137	137	150	150
R307-2	259	259	121	121	221	221	170	170	173	173	137	137	141	141
R307-3	259	259	121	121	221	221	170	170	164	164	137	137	141	141
R423-1	259	259	113	113	221	221	171	171	175	175	134	134	141	141
R423-2	259	259	113	113	221	221	171	171	175	175	134	134	141	141
R437-1	259	259	121	121	221	221	180	180	175	175	144	144	141	141
R437-2	259	259	113	113	221	221	180	180	175	175	144	144	141	141
B67-1	247	247	121	121	217	217	178	178	173	173	144	144	150	150
B67-2	247	247	121	121	217	217	178	178	173	173	144	144	148	148
B67-3	247	247	121	121	217	217	178	178	173	173	144	144	150	150
962C-1	259	259	121	121	217	217	171	171	175	175	134	134	141	141
962C-2	259	259	121	121	217	217	171	171	175	175	134	134	141	141
962C-3	259	259	121	121	217	217	171	171	175	175	134	134	141	141
R413-1	259	259	121	121	221	221	171	171	173	173	134	134	150	150
R413-2	259	259	121	121	221	221	171	171	173	173	134	134	150	150
R4261-1	259	259	113	113	221	221	171	171	173	173	137	137	150	150
R4261-2	259	259	113	113	221	221	171	171	173	173	137	137	150	150
R4261-3	259	259	113	113	221	221	171	171	173	173	137	137	150	150
B71-1	259	259	119	119	217	217	180	180	173	173	134	134	141	150
B71-2	259	259	119	119	0	0	180	180	173	173	134	134	141	150
B71-3	259	259	119	119	217	217	180	180	173	173	134	134	141	150
964C-1	259	259	121	121	217	217	178	178	175	175	134	134	141	141
964C-2	259	259	121	121	0	0	178	178	175	175	134	134	141	141
964C-3	259	259	121	121	217	217	178	178	175	175	134	134	141	141
R415-1	247	259	119	119	221	221	180	180	173	173	134	134	150	150
R415-2	247	247	119	119	217	217	170	170	0	0	137	137	150	150
R415-3	259	259	119	119	221	221	180	180	173	173	134	134	150	150

R427-1	259	259	121	121	221	221	171	171	173	173	134	134	150	150
R427-2	259	259	121	121	221	221	171	171	173	173	144	144	150	150
R427-3	259	259	121	121	221	221	171	171	173	173	144	144	150	150
B99-1	247	247	119	119	217	217	170	170	173	173	137	137	150	150
B99-2	247	247	119	119	217	217	170	170	173	173	137	137	150	150
B99-3	247	247	119	119	217	217	170	170	173	173	137	137	150	150
1063C-1	259	259	121	121	217	217	178	178	173	173	144	144	150	150
1063C-2	259	259	121	121	217	217	178	178	173	173	144	144	141	150
1063C-3	259	259	121	121	217	217	178	178	173	173	144	144	141	150
R417-1	247	247	121	121	217	217	180	180	173	173	137	137	150	150
R417-2	247	247	121	121	217	217	180	180	173	173	137	137	150	150
R417-3	247	247	121	121	217	217	180	180	173	173	0	0	0	0
R429-1	259	259	121	121	221	221	180	180	170	170	144	144	150	150
R429-2	259	259	119	119	221	221	171	171	170	170	0	0	150	150
R429-3	0	0	0	0	221	221	171	171	173	173	137	137	150	150
RHA274-1	259	259	121	121	221	221	171	171	170	170	134	134	150	150
RHA274-2	259	259	121	121	221	221	180	180	170	170	0	0	0	0
RHA274-3	259	259	121	121	221	221	180	180	170	170	134	134	150	150
R023-1	259	259	113	113	221	221	171	171	173	173	137	137	141	141
R023-2	259	259	113	113	221	221	171	171	173	173	137	137	141	141
R419-1	259	259	113	113	221	221	171	171	173	173	134	134	141	141
R419-2	259	259	113	113	221	221	171	171	173	173	134	134	141	141
R419-3	259	259	113	113	221	221	171	171	173	173	134	134	141	141
R431-1	253	259	119	119	221	221	171	171	173	173	144	144	150	150
R431-2	253	253	119	119	0	0	171	171	173	173	0	0	150	150
R431-3	253	259	119	119	221	221	171	171	173	173	144	144	150	150
HAR4-1	259	259	119	119	225	225	170	170	175	175	144	144	141	141
HAR4-2	259	259	113	113	225	225	170	170	175	175	144	144	141	141
HAR4-3	259	259	0	0	225	225	170	170	175	175	144	144	141	141
R043-1	259	259	113	113	221	221	174	174	175	175	144	144	150	150
R043-2	259	259	113	113	221	221	174	174	175	175	144	144	150	150
R420-1	259	259	119	119	221	221	176	176	173	173	134	134	141	141
R420-2	259	259	113	113	0	0	176	176	170	170	134	134	141	141
R420-3	259	259	113	113	221	221	176	176	170	170	0	0	141	141
R432-1	259	259	119	119	227	227	178	178	173	173	144	144	150	150
R432-2	259	259	119	119	221	221	178	178	0	0	144	144	141	141
R432-3	259	259	119	119	221	221	178	178	173	173	144	144	141	141
HA89-1	247	247	119	119	217	217	170	170	173	173	134	137	141	141
HA89-2	259	259	121	121	217	217	178	178	173	173	134	134	141	150
HA89-3	259	259	121	121	217	217	178	178	173	173	134	134	141	150
R049-1	253	253	119	119	221	221	180	180	173	173	144	144	150	150
R049-2	253	253	119	119	221	221	180	180	173	173	0	0	150	150
R049-3	253	253	119	119	221	221	180	180	173	173	144	144	150	150
R421-1	259	259	113	113	221	221	171	171	173	173	134	134	141	150
R421-2	259	259	113	113	221	221	171	171	173	173	134	134	141	150
R433-1	259	259	121	121	227	227	178	178	173	173	144	144	150	150
R433-2	259	259	121	121	227	227	178	178	173	173	144	144	141	150
R433-3	259	259	121	121	0	0	178	178	173	173	0	0	141	150

FIGURAS

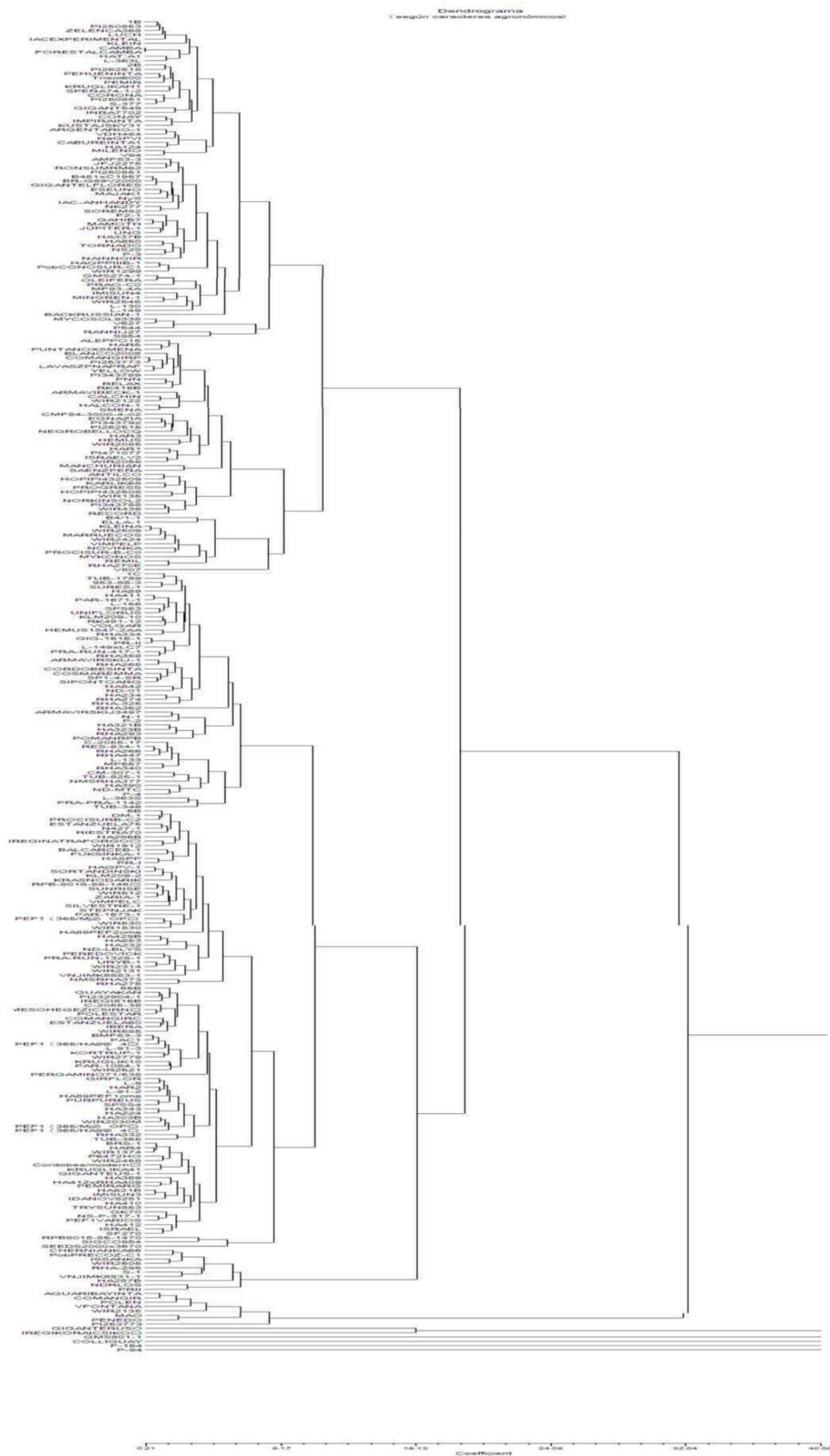


Figura A.1. Árbol obtenido en base a distancia taxonómica promedio. Método de agrupamiento UPGMA en función de caracteres agronómicos y morfológicos.

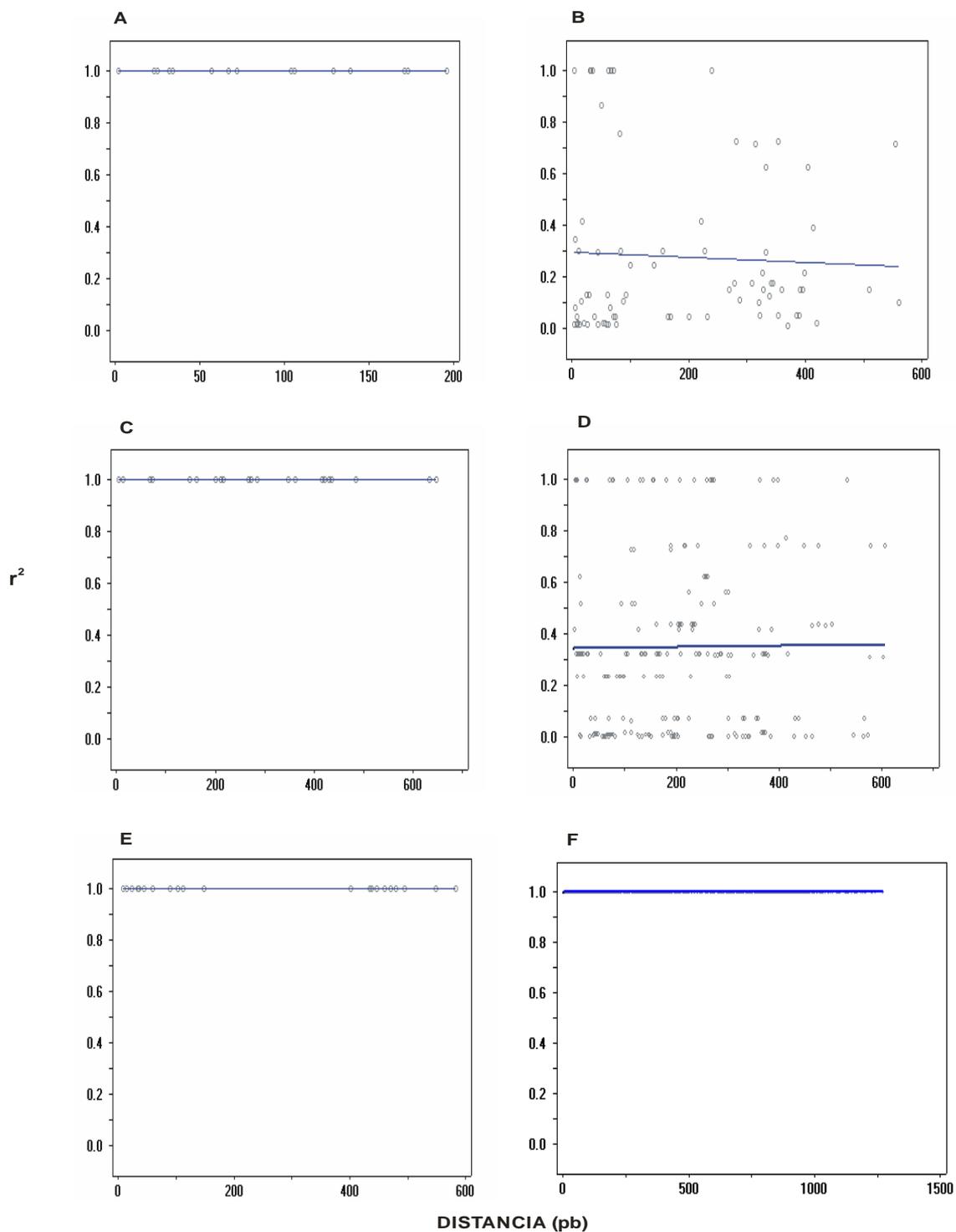


Figura A.2. Análisis del desequilibrio de ligamiento en girasol. Gráfico de dispersión del DL para las regiones estudiadas. El DL se expresó como r^2 y la distancia en pb. A: CK. B: HaDhn1. C: FB. D: *Suntip*. E: HaL1L. F: Hahb4 (región promotora y transcripto).