



## **FACULTAD DE CIENCIAS AGRARIAS**

### **“DESARROLLO DE POBLACIONES MULTIPARENTALES COMO FUENTE DE NUEVOS RECURSOS GENÉTICOS PARA EL MEJORAMIENTO DE GIRASOL”**

Ing. Agr. Matías Domínguez

TESIS PARA OPTAR AL TÍTULO DE DOCTOR EN CIENCIAS AGRARIAS

**DIRECTORA:** Dra. Norma Beatriz Paniego

**CO-DIRECTOR:** Dr. Gerardo Domingo Lucio Cervigni

2022

# **DESARROLLO DE POBLACIONES MULTIPARENTALES COMO FUENTE DE NUEVOS RECURSOS GENÉTICOS PARA EL MEJORAMIENTO DE GIRASOL**

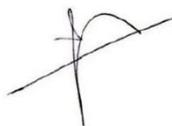
**Matías Domínguez**

**Ingeniero Agrónomo - Universidad Nacional de Luján**

Esta Tesis está presentada como parte de los requisitos para optar al grado académico de Doctor en Ciencias Agrarias, de la Universidad Nacional de Rosario y no ha sido previamente presentada para la obtención de otro título en esta u otra Universidad. La misma contiene los resultados obtenidos en investigaciones llevadas a cabo en el grupo de mejoramiento de girasol del Instituto Nacional de Tecnología Agropecuaria, en la EEA Pergamino, el Instituto de Biotecnología de INTA Castelar y la EEA Balcarce, durante el período comprendido entre el 1 de julio de 2017 y el 31 de marzo de 2022, bajo la dirección de la Dra. Norma Paniego y la codirección del Dr. Gerardo Cervigni.



Doctorando: Ing. Agr. Matías Domínguez



Directora Dra. Norma Beatriz Paniego



Co-director Dr. Gerardo D. Cervigni

## AGRADECIMIENTOS

**En las próximas líneas quiero expresar mi agradecimiento a todas las personas e instituciones que han hecho posible la realización de esta Tesis:**

Al CONICET y al INTA por la beca doctoral otorgada para la realización de esta Tesis.

A la Estación Experimental Agropecuaria Pergamino y al Instituto de Biotecnología de INTA Castelar y a sus directivos por permitir el desarrollo de este trabajo.

A mis directores de beca y Tesis Norma, Julio y Gerardo que me abrieron las puertas de esta institución para poder llevar adelante esta Tesis y que me permitió crecer tanto desde lo profesional como desde lo humano. Y a Gerardo también por tu acompañamiento en todo momento que te consultamos, tus ideas y tu buena predisposición ante nuestras consultas.

Gracias Norma por siempre estar presente en todo momento, por tus ideas, por darme la posibilidad de capacitarme en cursos en el exterior y abrireme paso en el mundo de la biotecnología y la genómica que jamás me hubiera imaginado que me iba a atrapar y apasionar tanto. Valoro mucho tu visión, paciencia y energía para poder encarar nuevos proyectos y nunca dejar que baje los brazos y enseñarme a que siempre hay que seguir intentando conseguir nuestros objetivos.

Gracias Julio por tu buena predisposición desde el primer momento que llegue a la EEA Pergamino, allá por octubre de 2017 a incursionar en el apasionante mundo del mejoramiento genético y en particular del cultivo de girasol. Y mas allá de lo profesional que pude adquirir en todo este tiempo quiero resaltar y agradecer tu calidad humana y las enseñanzas desde lo personal, que es sin duda lo mas importante que me queda después de estos años que tuvimos la suerte de compartir juntos.

A mis compañeros del grupo girasol de Pergamino, Carlitos Righetti y Gabriel Colacilli, por la integración al grupo desde el primer día y por estar siempre, incluso en los momentos más difíciles que vivimos estos años, de poder seguir adelante con el proyecto y con el sector girasol. Muchas gracias por hacerme sentir muy feliz cada día de trabajo, por las charlas, historias, comidas y momentos compartidos. Y sin dudas Carlitos que te vamos a extrañar mucho en esta nueva que esta por comenzar para vos. No quiero dejar de agradecer al gran hormiga Baroni, por su predisposición y ser siempre uno más del grupo que siempre nos salvó y nos salvará las papas!

A Juan Montecchia por su confianza plena para que pudiera encarar este proyecto, por su recibimiento en esos primeros tres meses de beca allá en el IB que fueron increíbles para mí y por todas las charlas, consultas y ni que hablar de aquella épica desmalezada del ensayo en Balcarce ¡¡¡Siempre te voy a estar agradecido!!!

A todos los integrantes del hermoso grupo de girasol de INTA: a Moni por sus capacidad para explicar y su paciencia en mis primeros días en el Laboratorio, a Carla y Sergio por su ayuda en los análisis, a Daniel Álvarez y todo el equipo de Manfredi por su excelente predisposición, a Caro Troglia, Silvio y Mauro por su tremendo laburo para que pudiéramos tener ese ensayo de Verticillium tan importante, a Salvador Nicosia por la posibilidad de realizar ese vuelo con el dron y siempre responder mis consultas, a Gabriel Rojo por ser un ladero increíble para esas castraciones eternas en Yuto y a Denis Colombo por las fotos del tan temido Cancro!!

A Emi Ben Guerrero, Mariana Conte y Ange Gutierrez por su buena onda en el lab y por ayudarme con esas extracciones que tantas idas y vueltas me hicieron dar!!!

A Andrea Puebla, Pablo Vera, Vero Nishi y Marianne de UGB por aguantarme y ayudarme con las librerías!!

A Ezequiel y Javier por su cálido recibimiento en la casa de INTA y por haberme hecho sentir como en casa en un momento de gran cambio al llegar a Pergamino. ¡¡¡Muchas gracias siempre!!!

A los chicos de campo y tareas de la EEA Pergamino por siempre estar dispuestos para coordinar trabajos, lotes y prestamos de herramientas.

A Stefania Herrera, Fátima Palifermo, Agustín Romero y Mirko Piazza por su aporte en este proyecto y en el sector girasol en distintos momentos de este período.

A mis amigos de General Las Heras que mas allá de que no nos veamos mucho siempre están y para aquellos que conocí en esta etapa (muchas gracias Dai Scaccia por tu buena onda, Agos y Andrés!!)

A mi mama y mi hermano por estar siempre, apoyarme y acompañarme en todo momento.

A mi novia Lu por haberme acompañado en todo momento durante esta etapa, enseñarme mucho, ayudarme siempre, tenerme mucha paciencia, brindarme todo su amor y volverse una persona muy especial que tuve la suerte de cruzarme en este camino.

Y a todos aquellos que olvidé de mencionar pero que de alguna manera contribuyeron a que podamos llevar adelante este hermoso proyecto... ¡¡¡**MUCHAS GRACIAS!!!**

## PUBLICACIONES Y PRESENTACIONES A CONGRESOS

Assessment of phenotypic variability among EEA INTA Pergamino sunflower lines: Its relationship with the grain yield and oil content. Domínguez, M., Herrera, S. & González, J. 2021. OCL 28 (33) <https://doi.org/10.1051/ocl/2021021>

Heredabilidad, correlaciones genéticas, análisis de sendero e índices de selección en una población de girasol (*Helianthus annuus*) de la EEA INTA Pergamino. Domínguez, M. & González, J. 2021. RIA 47 (2): 209-215.

Tratamientos de fertilización con zinc y boro mediante la impregnación de fuentes nitrogenadas en girasol. Ferraris, G., Domínguez, M. & González, J. 2019. RTA 10 (40): 16-20.

Aporte a resistencia a enfermedades, calidad industrial y resistencia a herbicidas de distintos orígenes de germoplasma de girasol en tres períodos de mejoramiento. González, J. & Dominguez, M. 2019. RTA 10 (39): 31-36.

Expositor en 5° Congreso Argentino de Fitopatología. Evaluación de respuesta al marchitamiento por *Verticillium dahliae* de familias F2 de una población de mapeo multiparental de girasol (*Helianthus annuus*). 22 y 23 de septiembre de 2021.

Expositor en 1° Simposio de Mejoramiento Genético Vegetal. Development of MAGIC populations as source of new genetic resources for Sunflower (*Helianthus annuus*) breeding. 13 y 14 de septiembre de 2021.

Expositor en 7° Congreso Argentino de Girasol: Desarrollo de poblaciones multiparentales de girasol como fuentes de nuevos recursos genéticos para el mejoramiento. 2019.

Expositor en Taller Pre-Congreso de ASAGIR: Desarrollo de poblaciones multiparentales de girasol como fuentes de nuevos recursos genéticos para el mejoramiento y Vigor híbrido hacia grupos heteróticos y selección genómica en Girasol (*Helianthus annuus* L.). 2019.

Expositor como coautor en Plant and Animal Genome Conference Integrating Phenotypic and Genetic diversity to breed for Sunflower *Verticillium* Wilt resistance. 2018.

---

---

## Contenido

---

---

RESUMEN.....	I
ABSTRACT .....	IV
INTRODUCCIÓN GENERAL .....	1
Taxonomía, origen, domesticación y mejoramiento	2
Importancia Económica	3
Cultivo del girasol en Argentina	4
Principales enfermedades del cultivo en Argentina	6
Genoma del girasol	8
Estrategias de mejoramiento asistido por marcadores moleculares	10
Ventajas y desventajas de las poblaciones MAGIC	16
Análisis de QTL en poblaciones multiparentales	18
El rol del fenotipado en los programas de mejoramiento genético	20
Programa de mejoramiento de girasol de INTA y Antecedentes del grupo	22
HIPÓTESIS .....	29
OBJETIVO GENERAL .....	31
OBJETIVOS ESPECÍFICOS.....	33
CAPITULO 1: Selección de las líneas parentales y construcción de las poblaciones multiparentales de girasol .....	35
INTRODUCCIÓN	36
MATERIALES Y MÉTODOS	39
Selección de las líneas parentales	39
Construcción de las poblaciones multiparentales	42
RESULTADOS	45
Selección de líneas parentales. Correlación entre las variables fenotípicas de resistencia a enfermedades fúngicas en líneas mantenedoras	45
Análisis de Componentes Principales	46
Diversidad genética de las líneas consideradas para la selección	48
Características de las líneas parentales seleccionadas	49
Construcción de las poblaciones multiparentales	54
DISCUSIÓN	57
CAPÍTULO 2: Caracterización genotípica de las líneas parentales y los híbridos formados en las primeras etapas de desarrollo de las poblaciones multiparentales .....	61
INTRODUCCIÓN	62

Análisis de parentesco	64
Reconstrucción de bloques haplotípicos	66
MATERIALES Y MÉTODOS	68
Material vegetal	68
Extracción de ADN	68
Secuenciación masiva de fragmentos de ADN obtenidos por doble restricción enzimática (ddRADSeq)	69
Identificación de Polimorfismos de Nucleótido Simple (SNP)	70
Anotación y efecto de los polimorfismos identificados	71
Análisis de parentesco para la evaluación de los cruzamientos	71
Análisis de diversidad genética de las líneas parentales de las poblaciones MAGIC	73
Reconstrucción de bloques genómicos ancestrales en las poblaciones MAGIC1 y MAGIC2 a través de RABBIT	74
RESULTADOS	77
Construcción de la librería ddRADseq	77
Secuenciación de la librería ddRADseq	79
Identificación de Polimorfismos de Nucleótido Simple (SNP)	80
Exploración funcional de los marcadores informativos obtenidos	83
Análisis de parentesco para la evaluación de los cruzamientos	84
Análisis de diversidad de las líneas parentales de las poblaciones MAGIC1 y MAGIC288	
Reconstrucción de bloques genómicos ancestrales en las poblaciones MAGIC1 y MAGIC2 a través de RABBIT	91
DISCUSIÓN	97
CAPÍTULO 3: Caracterización fenotípica de las familias F <sub>2</sub> de las poblaciones multiparentales	103
.....	
INTRODUCCION	104
1. Evaluación frente a Marchitez anticipada por <i>V. dahliae</i> en girasol	104
1.1 Importancia y descripción de la Marchitez anticipada por <i>V. dahliae</i> en girasol	104
1.2 Agente causal y taxonomía	104
1.3 Epidemiología y ciclo patogénico de <i>Verticillium dahliae</i>	105
1.4 Sintomatología de la MV en girasol	106
1.5 Evaluación a campo de la respuesta frente a MV en girasol	109
2. Impacto del daño por aves en el cultivo de girasol	110
3. Arquitectura de planta y su relación con la interceptación lumínica en el cultivo de girasol	111
4. Análisis multivariado	112

MATERIALES Y MÉTODOS	115
Diseño experimental	115
Manejo de los experimentos	115
Evaluaciones fenotípicas	116
Análisis estadístico del experimento de respuesta a MV por <i>V. dahliae</i>	120
Análisis estadístico del experimento de caracterización fenotípica de floración, altura de la punta de la hoja en relación con la inserción del pecíolo en el estrato superior de la planta, inclinación del capítulo y altura de planta	123
Análisis de Componentes Principales	124
Análisis de Conglomerados	124
Análisis Factorial Múltiple	125
RESULTADOS	126
Selección y Ajuste de Modelos Lineales Generales Mixtos, Generalizados Mixtos y Espaciales para la respuesta a MV	126
Estimación de la heredabilidad para cada variable fitopatométrica	129
Análisis multivariado de la respuesta de la población MAGIC2 frente a MV	130
Análisis Multivariados de Conglomerados	132
Selección y Ajuste de Modelos Lineales Generales Mixtos para las variables Altura de planta y DAF	134
Análisis multivariado integrando variables fitopatométricas, fenológicas y de arquitectura de planta en la población MAGIC2	137
DISCUSIÓN	145
CAPÍTULO 4: Implementación de imágenes de cámaras multispectrales para la el fenotipado de la marchitez prematura por <i>Verticillium dahliae</i> .....	151
INTRODUCCIÓN	152
El rol del fenotipado en el mejoramiento genético de plantas	152
Teledetección	153
Plataformas de fenotipado de alto rendimiento	155
Estrategias de análisis de imágenes espectrales y su implementación en los programas de mejoramiento	156
Antecedentes de fenotipado automático de enfermedades causadas por <i>V. dahliae</i>	158
MATERIALES Y MÉTODOS	160
Material vegetal	160
Diseño experimental	160
Evaluaciones fenotípicas manuales	161
Fenotipado automático: teledetección y procesamiento de imágenes	161

Adquisición de imágenes multiespectrales	161
Procesamiento de las imágenes	162
Cálculo de los índices de vegetación a partir de las imágenes obtenidas	163
Extracción de la información a nivel de parcela	163
Análisis de los datos obtenidos usando modelos predictivos de ML	164
RESULTADOS	167
Evaluaciones fenotípicas manuales	167
Adquisición y procesamiento de imágenes	167
Análisis de los resultados obtenidos con los diferentes modelos predictivos de ML implementados	171
DISCUSIÓN	173
CONSIDERACIONES FINALES Y PERSPECTIVAS.....	179
BIBLIOGRAFÍA.....	183
ANEXO.....	234

## INDICE DE TABLAS

<b>Tabla 1.1.</b> Ventajas y desventajas de las poblaciones utilizadas para el mapeo de QTLs.	17
<b>Tabla 1.1.</b> Marcadores EST-SSR utilizados para construir la matriz de distancias genéticas de la PMA.	41
<b>Tabla 1.2.</b> Varianza explicada por los primeros 5 componentes principales obtenidos del ACP.	46
<b>Tabla 1.3.</b> Contribuciones de las variables a los dos primeros CPs seleccionados.	46
<b>Tabla 1.4.</b> Características de las líneas parentales seleccionadas.	51
<b>Tabla 1.5.</b> Categorías fenotípicas de las líneas parentales seleccionadas frente a MV y PHC.	52
<b>Tabla 1.6.</b> Cruzamientos de 2 vías realizados en la primera etapa de desarrollo de las poblaciones MAGIC.	55
<b>Tabla 1.7.</b> Cruzamientos de 4 vías realizados en la segunda etapa de desarrollo de las poblaciones MAGIC.	55
<b>Tabla 1.8.</b> Cruzamientos de 8 vías realizados en la tercera etapa de desarrollo de las poblaciones MAGIC.	55
<b>Tabla 2.1.</b> Cantidad de SNPs por cromosoma identificados en la población MAGIC1.	80
<b>Tabla 2.2.</b> Cantidad de SNPs por cromosoma identificados en la población MAGIC2.	81
<b>Tabla 2.3.</b> Análisis de parentesco entre los híbridos obtenidos y las líneas parentales de la población MAGIC1.	85
<b>Tabla 2.4.</b> Análisis de parentesco entre los híbridos obtenidos y las líneas parentales de la población MAGIC2.	86
<b>Tabla 2.5.</b> Estimaciones de Heterocigosis esperada ( $H_e$ ) y observada ( $H_o$ ) para las líneas parentales de las poblaciones MAGIC1 y MAGIC2.	88
<b>Tabla 3.1.</b> Resultados de evaluación y selección de modelos para las 12 variables estudiadas.	126
<b>Tabla 3.2.</b> Valores mínimos, medios y máximos para las medias ajustadas de las variables fitopatométricas analizadas.	128
<b>Tabla 3.3.</b> Heredabilidad en sentido amplio para las 12 variables fitopatométricas estudiadas.	129
<b>Tabla 3.4.</b> Componentes principales obtenidos del análisis.	131
<b>Tabla 3.5.</b> Contribuciones de las variables a los dos primeros CPs seleccionados.	131
<b>Tabla 3.6.</b> Resultados de evaluación y selección de modelos para DAF y Altura de planta en la población MAGIC2.	135
<b>Tabla 3.7.</b> Resultados de evaluación y selección de modelos para DAF y Altura de planta en la población MAGIC1.	135
<b>Tabla 3.8.</b> Componentes principales obtenidos del análisis conjunto de las variables fitopatométricas, DAF y altura de planta.	138
<b>Tabla 3.9.</b> Contribuciones de las variables a los dos primeros CPs seleccionados.	138
<b>Tabla 4.1.</b> Especificaciones técnicas de la cámara multiespectral Parrot Sequioa.	162
<b>Tabla 4.2.</b> Ecuaciones para la obtención de los cuatro índices de vegetación utilizados.	163
<b>Tabla 4.3.</b> Correlaciones entre los índices de vegetación y las variables INC2, MSEV2, INT2 y AUDPC_INC.	170
<b>Tabla 4.4.</b> Precisión alcanzada por los 4 modelos predictivos evaluados.	172

## INDICE DE FIGURAS

<b>Figura I.1.</b> Área de cultivo de girasol, dividida en tres regiones: Norte (N), Centro (C) y Sur (S).	5
<b>Figura I.2.</b> Síntomas de las 5 enfermedades más relevantes del cultivo en Argentina.	8
<b>Figura I.3.</b> Comparación del número de cruzamientos y de individuos por cruzamientos entre las poblaciones biparentales, multiparentales y los paneles de asociación.	13
<b>Figura I.4.</b> Ejemplo de uno de los esquemas posibles para el desarrollo de una población MAGIC.	15
<b>Figura I.5.</b> Esquema de obtención de híbridos de girasol del programa de mejoramiento de INTA.	26
<b>Figura 1.1.</b> Esquema de cruzamientos para la formación de la PMP.	43
<b>Figura 1.2.</b> Métodos de emasculación utilizados en el desarrollo de las PMP.	43
<b>Figura 1.3.</b> Correlaciones de Pearson entre las medias ajustadas de las líneas mantenedoras para las variables fenotípicas consideradas.	45
<b>Figura 1.4.</b> Biplot conformado por las dos primeras componentes principales.	47
<b>Figura 1.5.</b> Árbol neighbor-joining construido en base a la distancia genética de las accesiones que componen la PMA.	49
<b>Figura 1.6.</b> Líneas parentales fundadoras de las poblaciones MAGIC 1 y 2.	53
<b>Figura 1.7.</b> Imágenes de los aquenios de cada una de las líneas parentales.	54
<b>Figura 1.8.</b> Imagen aérea tomada con dron del campo experimental de la EEA Pergamino donde se realizó la segunda generación de autofecundaciones de las poblaciones MAGIC1 y MAGIC2.	56
<b>Figura 2.1.</b> Diagrama de dispersión de todos los valores posibles de Gdijjk para una población dada.	73
<b>Figura 2.2.</b> Perfil de corrida electroforética del producto de ligación de la digestión del ADN de una de las líneas parentales en el equipo Fragment Analyzer.	77
<b>Figura 2.3.</b> Perfil de corrida electroforética del producto de la selección automatizada de fragmentos con adaptadores ligados del pool de muestras que conforman la librería en el equipo Fragment Analyzer.	78
<b>Figura 2.4.</b> Perfil de corrida electroforética del producto de PCR de la selección de tamaños del pool de muestras que conforman la librería en el equipo Fragment Analyzer.	78
<b>Figura 2.5.</b> Gráficos de cajas de valores de calidad a través de las bases secuenciadas.	80
<b>Figura 2.6.</b> Distribución de los SNPs obtenidos mapeados contra el genoma de referencia de girasol en la población MAGIC1.	82
<b>Figura 2.7.</b> Distribución de los SNPs obtenidos mapeados contra el genoma de referencia de girasol en la población MAGIC2.	83
<b>Figura 2.8.</b> Proporción de loci heterocigotas en las líneas parentales y en los híbridos de 2, 4 y 8 vías en la población MAGIC1.	87
<b>Figura 2.9.</b> Proporción de loci heterocigotas en las líneas parentales y en los híbridos de 2, 4 y 8 vías en la población MAGIC2.	88
<b>Figura 2.10.</b> Frecuencias alélicas para el alelo alternativo y de referencia en ambas poblaciones.	89
<b>Figura 2.11.</b> Filograma de Neighbor-Joining para las líneas parentales de la población MAGIC1.	90
<b>Figura 2.12.</b> Filograma de Neighbor-Joining para las líneas parentales de la población MAGIC2.	91

<b>Figura 2.13.</b> Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos de dos vías de la población MAGIC1.	93
<b>Figura 2.14.</b> Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos de 4 y 8 vías de la población MAGIC1.	94
<b>Figura 2.15.</b> Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos de 4 y 8 vías de la población MAGIC2.	95
<b>Figura 2.16.</b> Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos de 4 y 8 vías de la población MAGIC2.	96
<b>Figura 3.1.</b> Ciclo patogénico de la Marchitez causada por <i>V. dahliae</i> .	106
<b>Figura 3.2.</b> Sintomatología de la Marchitez causada por <i>V. dahliae</i> .	108
<b>Figura 3.3.</b> Escala fitopatológica de seis niveles usada para la evaluación de severidad.	117
<b>Figura 3.4.</b> Imágenes de la escala utilizada para la evaluación de la inclinación de capítulo.	119
<b>Figura 3.5.</b> Imágenes de la escala utilizada para la evaluación de la altura de la punta de la lámina en relación a la inserción del pecíolo (LTIP).	120
<b>Figura 3.6.</b> Función Densidad para las medias ajustadas de las familias F <sub>2</sub> para las variables INC2, MSEV_2, AUDPC_INC y AUDPC_INT.	128
<b>Figura 3.7.</b> Correlaciones fenotípicas entre las medias ajustadas para las variables registradas para la evaluación de la respuesta frente a MV de la población MAGIC2.	130
<b>Figura 3.8.</b> Biplot de los componentes principales 1 y 2.	132
<b>Figura 3.9.</b> Biplot del ACP particionado por los conglomerados definidos mediante K-Means (A) y Agrupamiento Jerárquico Average (B).	134
<b>Figura 3.10.</b> Valores medios y distribución de las medias ajustadas para DAF y Altura de planta para las poblaciones MAGIC1 y MAGIC2.	136
<b>Figura 3.11.</b> Distribución de las categorías obtenidas de las variables LTIP y Posición de Capítulo para las poblaciones MAGIC1 y MAGIC2.	137
<b>Figura 3.12.</b> Biplot de los componentes principales 1 y 2.	139
<b>Figura 3.13.</b> Representación gráfica en el plano principal de los dos grupos de variables en el AFM.	140
<b>Figura 3.14.</b> Representación gráfica de las variables cualitativas y cuantitativas utilizadas para el AFM.	142
<b>Figura 3.15.</b> Representación de los individuos en el plano principal del AFM y su caracterización para las variables LTIP y Posición de capítulo.	143
<b>Figura 4.1.</b> Espectro típico de luz reflejada por un cultivo.	155
<b>Figura 4.2.</b> Fotografía panorámica aérea del ensayo realizado en la EEA INTA Balcarce, Buenos Aires, Argentina.	160
<b>Figura 4.3.</b> Unidades experimentales correspondientes a parcelas de un surco.	161
<b>Figura 4.4.</b> A: Drone Parrot Disco-Pro Ag. B: Cámara multispectral Parrot Sequioa.	162
<b>Figura 4.5.</b> Esquema de trabajo para la obtención de modelos predictivos de la respuesta a la Marchitez por <i>V. dahliae</i> en girasol a partir de índices de vegetación.	166
<b>Figura 4.6.</b> Imagen del sitio de evaluación luego del procesamiento con los softwares PIX4d y QGIS para la obtención del índice NDVI.	168
<b>Figura 4.7.</b> Grilla de polígonos realizada para la definición de las parcelas del experimento.	169
<b>Figura 4.8.</b> Gráfico de violín de los 4 índices de vegetación obtenidos del análisis de las imágenes multispectrales.	170

**Figura 4.9.** Precisión obtenida en el entrenamiento de los distintos modelos por el método de validación cruzada K-Fold.

## ABREVIATURAS Y SÍMBOLOS

°C. *Grados Celsius*

A. *Líneas androestériles*

ACE. *Aptitud combinatoria específica*

ACG. *Aptitud combinatoria general*

ACM. *Análisis de correspondencia múltiple*

ACooP. *Análisis de Coordenadas Principales*

ACP. *Análisis de Componentes Principales*

ADN. *Ácido desoxirribonucleico*

ADNc. *ADN complementario*

AFLP. *Amplified fragment length polymorphism*

AFM. *Análisis factorial múltiple*

AIC. *Criterio de Akaike*

AM. *Análisis multivariado*

ANN. *Artificial neural networks*

AUDPC. *Area under disease progress curve*

AUDPC\_INC\_PHC. *Área por debajo de la curva de progreso de PHC para incidencia*

AUDPC\_SEV\_PHC. *Área por debajo de la curva de progreso de PHC para severidad*

B. *Líneas de citoplasma normal o mantenedor*

BAG-IM. *Banco Activo de Germoplasma de INTA Manfredi*

BC. *Backcrosses*

BIC. *Criterio de información bayesiana*

BLUPs. *Best linear unbiased predictors*

C. *Centro*

CART. *Classification and regression trees, Classification and Regression Training*

CC. *Complex Trait Consortium*

CDS. *Coding sequence*

CICVyA. *Centro de Investigaciones en Ciencias Veterinarias y Agropecuarias de INTA*

CIM. *Composite Interval Mapping*

cm. *centímetro*

cM. *centiMorgan*

CPS. *Componentes principales*

DAF. *Días a floración*

ddRADSeq. *Genotipado por secuenciación con doble digestión con enzimas de restricción*

DEM. *Digital elevation model*

DGC. *Prueba de comparaciones múltiples*

DH. *Doubled Haploids*

Dim. *Dimensión*

DM. *Downy mildew*

DO. *Diversity Outbred*

DSM. *Digital surface model*

EEA. *Estación Experimental Agropecuaria*

EECT. *Estación Experimental de Cultivos Tropicales*

ESTs. *Expressed sequence tag*

EVI2. *Enhanced vegetation index*

Fa. *Progenitor masculino*

FAO. *Organización de las Naciones Unidas para la Agricultura y la Alimentación*

FBPP. *Field based phenotypic platform*

g. *Gramo*

GA<sub>3</sub>. *Ácido giberélico*

Gb. *Gigabases*

GBS. *Genotyping by sequencing*

GD. *Coefficiente de similitud de Gower*

GGB-IABiMo. *Grupo de Genómica y Bioinformática del Instituto de Agrobiotecnología y Biología Molecular INTA-CONICET Hurlingham*

GL. *Grupo de ligamiento*

GLMM. *Modelos lineales mixtos generalizados*

GWAS. *Genome Wide Association Studies*

He. *Heterocigosis esperada*

HMM. *Hidden Markov Model*

Ho. *Heterocigosis observada*

HTPP. *High throughput phenotyping platform*

IAF. *Índice de área foliar*

IBD. *Identical by descent*

INC. *Incidencia*

INC\_PHC. *Incidencia de PHC*

INRA. *Institut National de la Recherche Agronomique*

INT. *Intensidad*

INTA. *Instituto Nacional de Tecnología Agropecuaria*

IR. *Intercepción de radiación solar por un cultivo*

IV. *Índices espectrales de vegetación*

k. *Coefficiente de extinción lumínica*

LCI. *Leaf chlorophyll index*

LDA. *Linear discriminant analysis*

LIDAR. *Light detection and ranging*

LTIP. *Altura de la punta de la lámina en relación al nivel de inserción del pecíolo en el estrato superior de la planta*

LTR-RT. *Long Terminal Repeat Retrotransposon*

m. *Metro*

m<sup>2</sup>. *Metro cuadrado, Metro cuadrado*

MA. *Mapeo de asociación*

MAF. *Frecuencia de alelo minoritario*

MAGIC. *Multi-parent Advanced Generation Inter-Crosses*

Mb. *Megabyte*

mg. *Miligramo*

MG. *Mapeo de ligamiento*

min. *Minuto*

ML. *Machine learning*

MLM. *Modelos lineales mixtos*

mM. *Milimolar*

Mo. *Progenitor femenino*

Mpb. *Mega par de bases*

MV. *Marchitez anticipada por Verticillium dahliae*

N. *Norte*

NAM. *Nested Association Mapping*

NDVI. *Normalized difference vegetation index*

NDWI. *Normalized difference water index*

ng. *Nanogramo*

NGS. *Next Generation Sequencing*

O. *Oeste*

Off. *Descendiente*

OSAVI. *Optimized soil-adjusted vegetation index*

Pa. *Progenitor*

pb. *Par de bases*

PCR. *Polymerase chain reaction*

PHC. *Podredumbre Húmeda del Capítulo*

PI\_PHC. *Período de incubación de PHC*

PMA. *Población de mapeo por asociación de girasol de INTA*

PMP. *Poblaciones multiparentales*

QTL. *Quantitative trait loci*

R. *Líneas restauradoras de la fertilidad*

RABBIT. *Reconstructing Ancestry Blocks BIT by Bit*

RAPD. *Random Amplification of Polymorphic DNA*

RFA. *Radiación fotosintéticamente activa*

RFLP. *Restriction fragment length polymorphism*

RFU. *Relative Fluorescence Units*

RILs. *Recombinant Inbred Lines*

s. *Segundo*

S. *Sur, Sur*

SEV. *Severidad*

SEV\_B. *Severidad binomial*

SEV\_PHC. *Severidad de PHC*

SR. *Simple ratio index*

SSR. *Simple sequence repeats*

SVM. *Support vector machine*

UAV-RSP. *Unmanned aerial vehicle remote sensing  
platforms*

UFC. *Unidades formadoras de colonias*

UV. *Ultravioleta*

vcf. *Variant Calling Format*

WGAIM. *Whole-Genome Average Interval Mapping*

µg. *Microgramo*

---

---

## RESUMEN

---

---

## DESARROLLO DE POBLACIONES MULTIPARENTALES COMO FUENTE DE NUEVOS RECURSOS GENETICOS PARA EL MEJORAMIENTO DE GIRASOL

El área de cultivo de girasol en Argentina, se extiende entre los 24 y 38 grados de latitud sur, abarcando una amplia gama de ambientes. La plasticidad del cultivo en su adaptación a diversas condiciones agroecológicas lo convierte en el segundo cultivo oleaginoso en importancia después de la soja. Desde fines de la década del 90' el cultivo de girasol fue desplazado a zonas con menor potencialidad productiva a causa de adversidades bióticas y abióticas que limitan los rendimientos obtenidos. Las principales enfermedades del girasol son la roya negra causada por *Puccinia helianthi*, downy mildew causado por *Plasmopara halstedii*, marchitez anticipada causada por *Verticillium dahliae*, el cancro del tallo y del capítulo causado por *Diaporthe helianthi* y la podredumbre húmeda del capítulo causada por *Sclerotinia sclerotiorum*. La diversidad de ambientes y de estreses a los que se ve expuesto el cultivo de girasol han impulsado los estudios de las bases genéticas de la resistencia a estreses bióticos y abióticos para poder lograr genotipos de mayor adaptabilidad, estabilidad y rendimiento.

La identificación de los factores genéticos responsables de la expresión de características de interés agronómico se basa en la disponibilidad de colecciones de germoplasma de base genética amplia. El desarrollo de la genómica en los últimos años favoreció la caracterización de la diversidad genética presente en estas colecciones y la implementación de diferentes estrategias de mejoramiento basado en poblaciones no estructuradas y más recientemente en poblaciones estructuradas a partir del cruzamiento de múltiples líneas parentales. Las poblaciones multiparentales de tipo MAGIC por sus siglas en inglés "*Multiparent Advanced Generation Inter-Cross populations*" han surgido como una alternativa en el estudio de QTLs ofreciendo características intermedias entre las poblaciones biparentales y los paneles de asociación en términos de diversidad, potencia y resolución.

El objetivo general de la presente Tesis fue diseñar, construir y caracterizar poblaciones multiparentales de girasol para complementar la plataforma actual de recursos genéticos del programa de mejoramiento de INTA para el estudio de regiones genómicas de interés agronómico. Se desarrollaron dos poblaciones MAGIC a partir de 8 líneas endocriadas mantenedoras de la fertilidad. Las líneas seleccionadas para la formación de las poblaciones presentan características diferenciales para distintos caracteres de interés para el mejoramiento del cultivo. Se priorizó la elección de materiales caracterizados en estudios previos del grupo girasol de INTA frente a las enfermedades marchitez anticipada por *V. dahliae* y podredumbre húmeda del capítulo por *S. sclerotiorum*. También, fueron

considerados otros caracteres de interés como la resistencia a Downy mildew, el contenido de aceite, la composición de ácidos grasos, la senescencia, el ciclo y la aptitud combinatoria. En la campaña 2020/21 se completó el segundo ciclo de autofecundaciones avanzando ambas poblaciones al estado de  $F_3$  y obteniéndose 1161 familias para la población MAGIC1 y 1497 familias para la población MAGIC2.

Fueron genotipadas las líneas parentales de ambas poblaciones y los híbridos de 2, 4 y 8 vías obtenidos en las etapas de desarrollo a través de la técnica de doble restricción enzimática para reducir la complejidad del genoma seguida por secuenciación de alto rendimiento. En esta instancia de desarrollo de las poblaciones, los datos genotípicos contribuyeron a testear la calidad de las líneas parentales altamente homocigotas, evaluar la diversidad genética entre ellas y la reconstrucción de haplotipos indicando la efectividad de los cruzamientos y el potencial de las poblaciones para el mapeo fino de QTLs.

Las familias  $F_2$  de ambas poblaciones MAGIC fueron caracterizadas fenotípicamente en INTA Pergamino en la campaña 2020/21 para caracteres de interés agronómico como floración, altura del ápice de la lámina en relación al nivel de inserción del pecíolo en el estrato superior de la planta, inclinación del capítulo y altura de planta. La población MAGIC2 fue caracterizada por su respuesta frente a la marchitez anticipada en el infectario natural de INTA Balcarce en la campaña 2020/21. Los resultados obtenidos confirmaron la variabilidad fenotípica de las poblaciones, permitieron identificar grupos de familias con respuesta contrastante a la marchitez anticipada y demostrar el potencial de ambas poblaciones para su utilización en futuros estudios de mapeo fino de QTLs. Asimismo, en este experimento fue evaluada la implementación de imágenes multiespectrales obtenidas por drones para el fenotipado de la marchitez anticipada en girasol. Los resultados obtenidos revelaron el potencial que tiene la implementación de índices de vegetación para el fenotipado de la marchitez anticipada en girasol y la utilización de modelos de aprendizaje que permiten, en base a la información aportada por los índices espectrales, poder clasificar genotipos de girasol como tolerantes o resistentes a la enfermedad. La construcción y caracterización de las poblaciones MAGIC presentadas en esta Tesis, constituyen recursos genéticos únicos hasta el momento para el mejoramiento del cultivo a nivel nacional e internacional.

---

---

## ABSTRACT

---

---

## **DEVELOPMENT OF MULTI-PARENT POPULATIONS AS A SOURCE OF NEW GENETIC RESOURCES FOR SUNFLOWER BREEDING**

The sunflower area in Argentina ranges from 24 to 38 °S and covers a wide range of environments. The diversity of stresses affecting sunflower has led to the study of the genetic basis of resistance to biotic and abiotic stresses in order to obtain genotypes with greater adaptability, stability and yield. Multiparent Advanced Generation Inter-Cross Populations (MAGIC) have emerged as an alternative in the study of QTLs offering intermediate characteristics in diversity, power and resolution between biparental populations and association panels. The objective of this work was to design, construct and characterize multiparental populations of sunflower to complement the genetic resources of the INTA breeding program for the study of genomic regions of agronomic interest. Two MAGIC populations were developed from 8 maintainer inbred lines. In the 2020/21 season, the second cycle of self-fertilization was completed, advancing both populations to F<sub>3</sub> status and a total of 1161 families were obtained for MAGIC1 population and 1497 families for MAGIC2 population. The parental lines of both populations and the 2-, 4- and 8-way hybrids were genotyped using Genotyping By Sequencing (GBS) technique. The genotypic data helped to test the quality of the highly homozygous parental lines, evaluate genetic diversity, and perform parentage analysis. The results demonstrated the effectiveness of the crosses and the potential of the populations for fine mapping of QTLs. The F<sub>2</sub> families of both MAGIC populations were phenotyped in the 2020/21 season at INTA Pergamino for agronomic traits such as flowering, petiole angle, head angle and plant height. The MAGIC2 population response to *Verticillium* wilt (VW) was characterized in infested fields of INTA Balcarce in the 2020/21 season. The results obtained confirmed the phenotypic variability of the populations, allowed the identification of groups of families with differential response to VW and demonstrated the potential of both populations for use in future fine mapping studies of QTLs. In addition, the use of multispectral images acquired from drones for VW phenotyping in sunflower demonstrated the potential of this technique for high-throughput phenotyping of fungal diseases in sunflower. The construction and characterization of the MAGIC populations presented in this Thesis represent unique genetic resources for sunflower breeding at national and international levels and will have a positive impact on the development of sunflower breeding as they can be use in research, development and transfer to industry.

---

---

## INTRODUCCIÓN GENERAL

---

---

## Taxonomía, origen, domesticación y mejoramiento

El girasol cultivado (*Helianthus annuus* L.) es una especie perteneciente a la familia *Asteraceae* (*Compositae*), subfamilia *Asteriodae*, tribu *Helianthae*, subtribu *Helianthinae*, género *Helianthus* (Gerald Seiler & Rieseberg, 1997). La familia *Compositae* es una de las familias más grandes y diversas de todas las plantas angiospermas (Funk et al., 2005; Kallunki & Heywood, 1994), conteniendo cerca de 2400 especies, un décimo de todas las especies con flor (Stevens, 2010). Las plantas pertenecientes a esta familia se caracterizan por su inflorescencia denominada capítulo, donde las flores se disponen en forma sésil, en círculos concéntricos, sobre un receptáculo ensanchado. En la República Argentina, es la familia vegetal más numerosa, contando con 227 géneros y cerca de 1400 especies, considerando taxones nativos y adventicios (Katinas et al., 2007).

El género *Helianthus* comprende 36 especies perennes y 12 herbáceas anuales (Heiser et al. 1981), entre las cuales se incluyen plantas diploides, tetraploides, y hexaploides, todas con un número cromosómico básico  $x = 17$  (Rieseberg, 1991). Su nombre deriva de las palabras griegas “*Helios*” que significa sol, y “*Anthus*” que significa flor; y le fue adjudicado por la característica que tiene la flor de mirar hacia el este hasta el período de antesis. La especie *H. annuus* está compuesta por tres subespecies: *H. annuus* ssp. *macrocarpus* (el girasol cultivado), *H. annuus* ssp. *lenticularis* y *H. annuus* ssp. *annuus*, ambos parientes silvestres del cultivo. El girasol es una especie diploide ( $2n = 2x = 34$ ), con un genoma haploide medio de 3600 Mb (Badouin et al., 2017).

La domesticación del girasol, comenzó aproximadamente en el centro-oeste de Norteamérica hace 3000-4000 años atrás gracias a las culturas pre-hispánicas (Crites, 1993; Harter et al., 2004; Smith & Yarnell, 2009), las cuales lo usaron como alimento, consumiendo la semilla directamente o moliéndola para harina e incluso, como colorantes para propósitos ceremoniales (Mandel et al., 2013). Desde entonces, el girasol ha sido usado también por sus propiedades medicinales, por sus características decorativas, para construcciones primitivas (Putt, 1997) y principalmente para uso alimenticio (aceite o directamente de las semillas), e industrial en la síntesis de polímeros, biocombustibles, emulsificantes y lubricantes (Dimitrijevic & Horn, 2018).

Luego de su introducción en Europa, el girasol se esparció rápidamente por el continente, donde fue cultivado principalmente con fines ornamentales. Su potencial como cultivo oleaginoso no fue reconocido hasta el siglo XVIII, en que se inició el mejoramiento para incrementar su producción de aceite en Rusia (Putt, 1978). La selección activa dirigida al contenido de aceite fue iniciada en 1860. Los Inmigrantes europeos fueron los responsables

de la reintroducción del girasol como cultivo mejorado a lo largo de toda América (Heiser et al., 1969).

El mejoramiento convencional ha sido exitoso en aumentar el potencial de rendimiento y la estabilidad del cultivo, así como en el control de la resistencia a algunas enfermedades fúngicas, plagas y malezas parasitarias (Sala et al., 2012). La aparición y el desarrollo de marcadores moleculares y mapas genéticos han facilitado la comprensión de la base genética de diferentes características agronómicas (Filippi, et al., 2020; Kiani et al., 2007; Montecchia, 2019; Talukder et al., 2020, 2021).

Diversas especies silvestres del género *Helianthus* han sido importantes fuentes de genes a lo largo de la domesticación y mejoramiento del girasol cultivado, contribuyendo con el aporte de genes de resistencia a enfermedades, tolerancia a sequía, esterilidad citoplasmática masculina y otras características de interés agronómico (Jan et al., 1998; Warburton et al., 2017). La explotación de la heterosis en girasol, descrita por primera vez en 1966 (Leclercq, 1966), sumada a la incorporación de la esterilidad citoplasmática masculina luego del cruzamiento inter-específico con *H. petiolaris* Nutt (Leclercq, 1969), y el desarrollo de líneas restauradoras de la fertilidad por Kinman (1970) permitieron el desarrollo de los híbridos de girasol, portadores de mayor rendimiento, calidad, estabilidad, alta homogeneidad, sincronización en el tiempo de maduración y mejor adaptación a las condiciones de cultivo.

### **Importancia Económica**

El girasol ha sido cultivado con diversos fines, destacándose principalmente la extracción de aceite comestible siendo este su principal uso, el consumo directo de sus semillas conocido como “girasol confitero”, el uso como especie ornamental, la utilización de subproductos de la industria aceitera para la alimentación animal, la utilización como pigmentos naturales de genotipos con alto nivel de antocianinas, el empleo de las proteínas en la industria alimenticia y del aceite como fuente para la producción de biodiesel (Park et al., 1997; Pilorgé, 2020).

La producción mundial de girasol alcanzada en el año 2018 fue de 52 millones de toneladas y la superficie sembrada fue de 27 millones de hectáreas en más de 50 países (Pilorgé, 2020). El girasol cuyo aceite se obtiene a partir de la extrusión de la semilla, es el tercer cultivo oleaginoso (sin considerar el aceite de palma por no obtenerse el aceite de la semilla) a nivel mundial con un promedio de 45 millones de toneladas por año. En el periodo 2014-18 su producción representó un 9 % de la producción global de aceite, ubicándose en primer lugar el aceite de soja con el 60 % y en segundo lugar la colza con 12 % (Pilorgé, 2020). El aceite de girasol es altamente valorado por el mercado mundial, en particular el europeo y con una

creciente demanda del asiático, debido a su alto contenido de ácidos grasos insaturados (oleico y linoleico) y bajo contenido de ácidos saturados (linolénico).

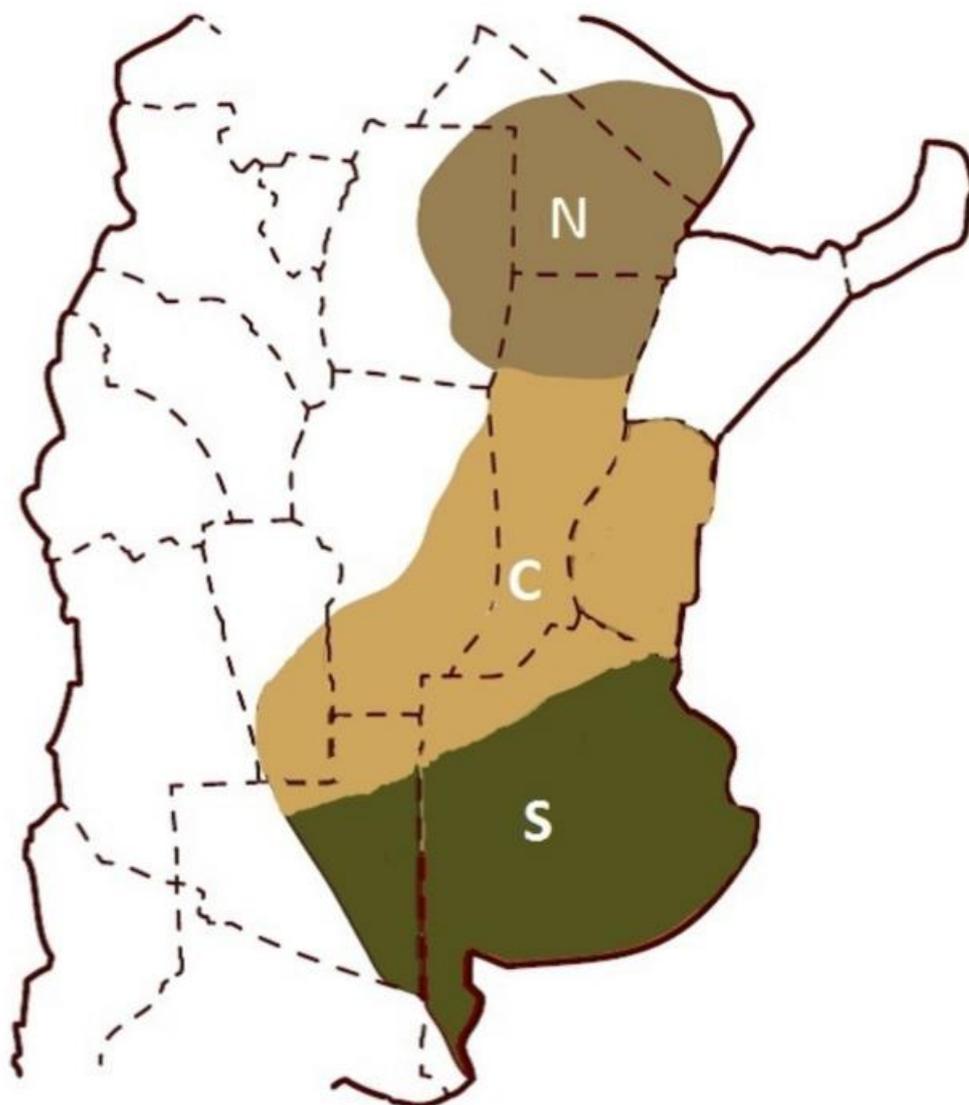
La producción del girasol a nivel mundial está concentrada principalmente por Ucrania, Rusia y la Unión Europea que en conjunto producen tres cuartos de la producción mundial total. Los demás países productores de importancia son Argentina, Estados Unidos y el sudeste de África. Actualmente Argentina, con una producción de 2,976 millones de toneladas (MT), ocupa el cuarto puesto del ranking mundial como productor (Pilorgé, 2020).

### **Cultivo del girasol en Argentina**

La historia de introducción y adaptación del girasol en Argentina está íntimamente relacionada con la de los flujos migratorios humanos. El cultivo llegó al país por primera vez de la mano de inmigrantes provenientes del sur de Rusia, que trajeron consigo pequeñas cantidades de semillas y se instalaron en lo que se conoce como Colonia Mauricio de Hirsch en el Partido de Carlos Casares (Provincia de Buenos Aires) (Castaño, 2018). Luego, la introducción de materiales tempranos de Rusia, Canadá y Rumania, así como la introgresión con especies silvestres del género *Helianthus* permitieron la emergencia del germoplasma argentino, el cual tiene una constitución genética distintiva y está bien adaptado a las condiciones de crecimiento locales (Bertero de Romano & Vázquez, 2003; González et al., 2015; Moreno et al., 2013).

El área potencialmente cultivable en Argentina, se extiende entre los 24 y 38 grados de latitud sur, abarcando una amplia gama de ambientes. La plasticidad del cultivo en su adaptación a diversas condiciones agroecológicas lo convierte en el segundo cultivo oleaginoso en importancia después de la soja. La región de cultivo de girasol actualmente se extiende de norte a sur desde El Chaco hasta el sur de la provincia de Buenos Aires y desde el sur de San Luis y centro de La Pampa, hasta Entre Ríos. El área cultivada puede dividirse en tres grandes regiones, Norte, Centro y Sur (Figura I.1) (Castaño, 2018).

Las siembras comienzan con el incremento de la temperatura de suelo en la región norte desde fines de julio, hasta principios de diciembre en el sur bonaerense. El período de cosecha comienza a fines de octubre en la región norte y culmina a fines de abril en el sur de Buenos Aires.



**Figura I.1.** Área de cultivo de girasol, dividida en tres regiones: Norte (N), Centro (C) y Sur (S). Imagen extraída de (Castaño, 2018).

El crecimiento de la superficie sembrada con soja que se ha registrado desde la liberación de las variedades resistentes al glifosato en el año 1996 provocó un fuerte desplazamiento del área cultivada de girasol a zonas de menor calidad ambiental comprometiendo los rendimientos del cultivo y modificando los objetivos de los programas de mejoramiento. El mejoramiento del cultivo desde entonces se ha conducido a lograr genotipos estables y adaptados a estas regiones y a enfrentar las adversidades bióticas y abióticas propias de las mismas. La disponibilidad hídrica en estos ambientes puede estar limitada por los regímenes hídricos de los mismos, perfiles de suelo menos profundos, impedancias subsuperficiales de origen natural, períodos de cultivo más limitados por acortamiento del período libre de heladas, condiciones epidemiológicas más proclives a la aparición y desarrollo de

enfermedades, o de enfermedades emergentes o re-emergentes. Actualmente, la región Sur es la más relevante a nivel productivo desde el avance de la producción de soja, cubriendo la mayor proporción de la superficie sembrada total de girasol a nivel nacional, seguida por la región Norte y la región Centro (Castaño, 2018).

La brecha entre el rendimiento potencial y el rendimiento alcanzado en condiciones de producción es de aproximadamente 2-2,5 veces, y se atribuye principalmente a la presencia de estreses, tanto bióticos (malezas, enfermedades y plagas), como abióticos (reserva inicial de agua, precipitación, tipo de suelo), y a condiciones de manejo (fecha de siembra, fertilización, protección, labranza) (Hall et al., 2010).

En este contexto, los programas de mejoramiento de girasol cumplen un rol fundamental para poder acortar esta brecha. El mejoramiento tradicional ha logrado aumentar el rendimiento potencial del cultivo de girasol y obtener cultivares resistentes a estreses bióticos y abióticos (Sala et al., 2012). Asimismo, el advenimiento y desarrollo de marcadores moleculares y mapas genéticos posicionan al mejoramiento asistido por marcadores moleculares y la selección genómica como las herramientas más promisorias para discernir la arquitectura genética de diferentes caracteres agronómicos de interés y la obtención de materiales con mayor potencial de rendimiento, estabilidad y adaptabilidad a las diferentes zonas del área girasolera argentina.

### **Principales enfermedades del cultivo en Argentina**

El girasol es susceptible a enfermedades causadas por múltiples agentes, entre ellos virus, bacterias y hongos. La mayoría de las patologías reportadas en Argentina son de origen fúngico. Las enfermedades mayores del girasol, denominadas así por su prevalencia, capacidad de producir mermas en el rendimiento y por la ocurrencia reiterada de epifitias, son principalmente cinco: roya negra causada por *Puccinia helianthi*, Downy mildew o enanismo causado por *Plasmopara halstedii*, Marchitez anticipada o verticilosis causada por *Verticillium dahliae*, el cancro del tallo causado por *Diaporthe helianthi* y la podredumbre húmeda del capítulo causada por *Sclerotinia Sclerotiorum* (Figura I.2, A-E).



A



B



C



D



**Figura I.2.** Síntomas de las 5 enfermedades más relevantes del cultivo en Argentina. **A:** Roya negra; **B:** Downey mildew; **C:** Verticilosis; **D:** Cancro de tallo por *Phomopsis*; **E:** Podredumbre húmeda del capítulo.

### Genoma del girasol

El girasol es una especie diploide ( $2n = 2X = 34$ ), con un genoma haploide medio de 3,6 Gigabases (Gb) y un contenido de ADN (2C) aproximado de 7,23 pg (Baack et al., 2005, 2008; Badouin et al., 2017). Aproximadamente 81% del genoma está compuesto por elementos transponibles (TE), predominantemente del orden LTR-RT (de sus siglas en inglés *Long Terminal Repeat Retrotransposon*), superfamilia *Gypsy*, y repeticiones ribosomales. Se estima que la mayoría de las inserciones LTR-RT acontecieron concomitantemente al origen del girasol como especie, alrededor de un millón de años atrás (Staton et al., 2012). Más de 6000 transposones poseen fragmentos génicos en su estructura, muchos de ellos relacionados a procesos metabólicos y de defensa (Badouin et al., 2017).

Las primeras descripciones del genoma del girasol se realizaron a través de mapas de ligamiento empleando diversos tipos de marcadores moleculares, siendo los más comunes RFLP (Berry et al., 1995, 1996, 1997; Gentzbittel et al., 1995), RAPD (Rieseberg, 1998; Rieseberg et al., 1993), AFLP (Flores Berrios et al., 2000; Gedil et al., 2001; Peerbolte & Paleman, 1998; Rachid Al-Chaarani et al., 2004) y SSR (Al-Chaarani et al., 2004; Flores Berrios et al., 2000; Kiani et al., 2007; Paniego et al., 2002; Talia et al., 2010; Tang et al., 2002; Yu et al., 2003). Además, estos marcadores posibilitaron estudios de detección de caracteres cuantitativos (QTL de sus siglas en inglés *quantitative trait loci*) de importancia agronómica (Ebrahimi et al., 2008; Haddadi et al., 2010, 2012; Kiani, Talia, et al., 2007; Shunxue Tang et al., 2006), de diversidad y de filogenia del género *Helianthus* (Burke et al., 2002; Heesacker et al., 2008; Linder et al., 1998). La incorporación de marcadores tipo SNP (de sus siglas en

inglés *Single Nucleotide Polymorphism*) incrementó la densidad de los mapas genéticos y permitió la reconstrucción de haplotipos, elementos especialmente beneficiosos en estudios de caracteres complejos (Filippi et al., 2020; Filippi et al., 2015; Fusari et al., 2011; Montecchia, 2019; Talukder et al., 2021; Zubrzycki et al., 2017).

El advenimiento de nuevas tecnologías de secuenciación y plataformas de genotipificación de alta procesividad han permitido analizar simultáneamente miles de SNPs distribuidos a lo largo del genoma (Grover & Sharma, 2016). Se han desarrollado paneles de genotipificación para girasol, permitiendo la caracterización genotípica eficiente de colecciones de germoplasma y generando una herramienta importante para estudios genómicos de la especie (Bachlava et al., 2012; Filippi et al., 2015; Zubrzycki et al., 2017). En la última década las técnicas de GBS (*Genotyping by Sequencing*) se han convertido en una herramienta muy implementada para la genotipificación masiva en proyectos de mapeo de QTLs, selección asistida por marcadores moleculares y selección genómica (Scheben et al., 2017). Estas técnicas combinan estrategias de digestión del genoma con enzimas de restricción para reducir la complejidad del genoma con estrategias de secuenciación de nueva generación (Elshire et al., 2011; He et al., 2014). En comparación con metodologías anteriores, como las basadas en el polimorfismo de longitud de fragmentos de restricción (RFLP) y las repeticiones de secuencias simples (SSR), las técnicas de GBS proporcionan mayor cantidad de datos informativos con menor complejidad y costo (Scheben et al., 2017). Si bien los chips comerciales de SNP siguen proporcionando mayores densidades de marcadores y son más fáciles de analizar, son más costosos y menos flexibles que las técnicas de GBS (Bajgain et al., 2016). La gran ventaja de GBS es que no necesita de información previa del genoma para su aplicación lo cual expande su posibilidad de implementación en estudios de numerosas especies (Kim et al., 2016)

El grupo de Genómica y Bioinformática del Instituto de Agrobiotecnología y Biología Molecular de INTA Castelar, ha generado el banco de datos ATGC-SUR almacenando información genómica y transcriptómica de girasol ([www.atgc-sur.inta.gob.ar](http://www.atgc-sur.inta.gob.ar)) (Fernandez et al., 2012). También el *Institut National de la Recherche Agronomique* (INRA) en Francia, está activamente involucrado en la generación de secuencias de regiones genómicas y transcriptos (INRA *Sunflower Bioinformatics Resources*, [www.heliagene.org](http://www.heliagene.org)). Desde el 2017 se cuenta con la publicación del genoma de la línea XRQ de girasol como genoma de referencia (Badouin et al., 2017), dada la alta complejidad del genoma de girasol, transcurrió más de una década de investigación y el surgimiento de tecnologías de secuenciación de tercera generación (*Pacific Biosciences*, *Oxford Nanopore Technology*) para su publicación. El portal

(HanXRQ-SUNRISE: [www.heliagene.org/HanXRQ-SUNRISE](http://www.heliagene.org/HanXRQ-SUNRISE)) permite acceder a datos y herramientas para análisis genómico.

### **Estrategias de mejoramiento asistido por marcadores moleculares**

Gran parte de los caracteres de importancia agronómica están controlados por múltiples *loci* que determinan caracteres cuantitativos. El mapeo genético y la caracterización molecular de *loci* funcionales favorecen el desarrollo de programas de mejoramiento basados en marcadores moleculares, para caracteres de interés agronómico como el rendimiento, la longitud del ciclo, resistencia a estreses bióticos y abióticos entre otros. Las estrategias más utilizadas para la disección de caracteres complejos son el análisis de mapeo de ligamiento (MG) y el mapeo de asociación (MA) (Mackay, 2001; Risch & Merikangas, 1996; Timpson et al., 2018). Los análisis de mapeo de ligamiento exploran la ocurrencia de herencia compartida entre polimorfismos funcionales y marcadores adyacentes, dentro de familias o poblaciones de *pedigree* conocido. En plantas, este estudio ha sido conducido mayormente en poblaciones experimentales derivadas de cruzamientos biparentales. Los estudios de mapeo por asociación, también conocidos como mapeo por desequilibrio de ligamiento, se basan en el mismo principio fundamental de recombinación genética, aunque se diferencian del primero porque incluyen en su estudio colecciones de individuos, muchas veces con ascendencia desconocida, la cual puede remontarse a numerosas generaciones en el pasado (Yu & Buckler, 2006).

Los estudios de QTLs con marcadores genéticos tienen una larga historia. Sin embargo, la mayor parte de los avances se han alcanzado desde la década de 1980 con el desarrollo de marcadores de ADN, iniciando la era del mapeo de QTLs en poblaciones biparentales (Cockram & Mackay, 2018). Esto ha continuado hasta el presente, con nuevas y más accesibles metodologías para analizar marcadores genéticos más baratos y mejores métodos estadísticos y programas informáticos (Broman et al., 2019b; Paux et al., 2012). Los resultados que se obtienen de estudios de QTLs facilitan la introgresión de variantes de interés (por ejemplo, de germoplasma no adaptado) a líneas élite, mediante selección asistida por marcadores para apilar múltiples fuentes de resistencia a enfermedades (Collard & Mackill, 2008). Sin embargo, el mapeo de QTLs ha sido cuestionado por los riesgos de los posibles sesgos en la estimación de los efectos, de la falta de precisión en la localización de los QTLs, y de la relevancia práctica del germoplasma utilizado en los programas de mejoramiento (Cockram & Mackay, 2018; Kearsey & Farquhar, 1998). Bernardo (2008) reportó la escasa tasa de transferencia de los resultados de los experimentos de mapeo de QTLs a los programas de mejoramiento. Actualmente, los recursos poblacionales, genómicos y

estadísticos disponibles, están siendo dirigidos a estudios de mapeo de QTLs en intervalos más reducidos y en germoplasma que pueda ser directamente aplicado en los programas de mejoramiento (Cockram & Mackay, 2018).

La detección de QTLs se ha realizado en cruzamientos biparentales entre dos padres con fenotipos contrastantes como poblaciones de retrocruzas (BC de sus siglas en inglés *Backcrosses*), en familias  $F_2$  o en poblaciones de líneas endocriadas recombinantes (RILs de sus siglas en inglés *Recombinant Inbred Lines*) (Rakshit et al., 2012). Las poblaciones biparentales aprovechan principalmente la recombinación reciente que tiene lugar durante el experimento (Cavanagh et al., 2008). Como no hay suficientes eventos de recombinación para fragmentar el genoma en pequeños fragmentos, los QTLs se localizan en amplias regiones cromosómicas (Rakshit et al., 2012). Esto mantiene una gran extensión del desequilibrio de ligamiento y ayuda a mantener cierto poder de detección cuando la densidad de marcadores es baja. La principal limitación de las poblaciones biparentales es la reducida cantidad de diversidad genética que abordan, limitándose solamente a los alelos que segregan los dos progenitores (Garin, 2019).

Los paneles de asociación surgieron como una alternativa a los cruzamientos biparentales. Estos combinan un conjunto de individuos no emparentados, muestreados de una población hipotética con el fin de aumentar la diversidad genética (Myles et al., 2009). Los paneles de asociación aprovechan el gran número de recombinaciones históricas que han sufrido las poblaciones naturales (Cavanagh et al., 2008). Por lo tanto, el genoma se fragmenta en pequeños trozos, lo que puede mejorar la resolución de detección de QTLs si la cobertura de marcadores es lo suficientemente densa (Kessner & Novembre, 2015; Yu et al., 2008). Los paneles de asociación también sufren de debilidades como la falta de un mapa de consenso y sobre todo de una estructura poblacional desconocida (Astle & Balding, 2009; Cavanagh et al., 2008). La deriva génica, la selección, la mezcla de poblaciones, la estructura de las subpoblaciones y/o la mutación pueden crear desequilibrio de ligamiento entre loci distantes entre sí (por ejemplo, en diferentes cromosomas) (Flint-Garcia et al., 2003; Patterson et al., 2006). Esta asociación entre diferentes regiones del genoma puede causar la detección de QTLs falsos positivos (Garin, 2019).

Las poblaciones multiparentales (PMP) pueden considerarse una interesante alternativa para abordar las limitaciones de las poblaciones biparentales y los paneles de asociación. Hay diferentes tipos de PMP, una primera categoría se compone de individuos que provienen de varios cruzamientos con padres compartidos (al menos uno). La población de mapeo de asociación anidada (NAM de sus siglas en inglés *Nested Association Mapping*) (McMullen et

al., 2009), los diseños dialélicos y los diseños factoriales (Cockerham, 1963) son ejemplos de PMP. Existen PMP más complejas, como las poblaciones de tipo MAGIC (de sus siglas en inglés *Multi-parent Advanced Generation Inter-Crosses*), en estas poblaciones, el genoma de las líneas finales es una mezcla de los fundadores que se cruzaron durante varias generaciones (Cavanagh et al., 2008; Huang et al., 2011).

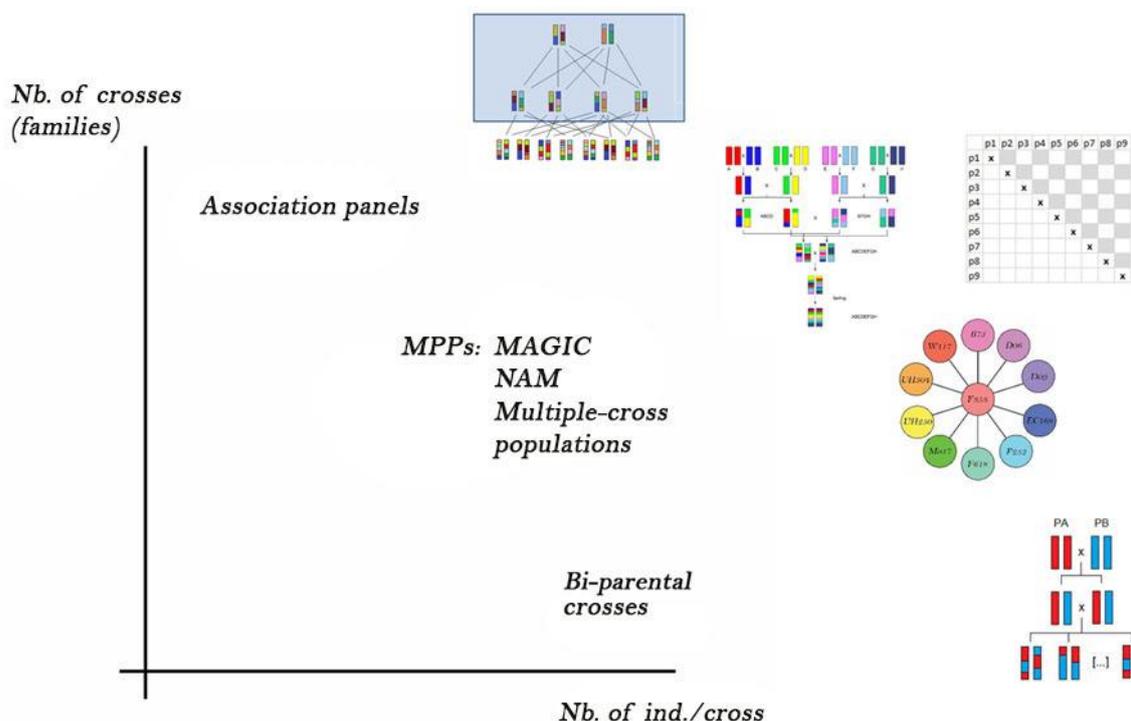
Con respecto a los cruzamientos biparentales, las PMP abordan una mayor diversidad genética (Blanc et al., 2006). El análisis conjunto de varios cruzamientos permite aumentar el nivel de polimorfismos y reducir la probabilidad de fijación en la posición del QTL (Muranty, 1996; S. Xu, 1998). El uso de las PMP también aumenta el espacio de inferencia estadística del efecto del QTL al probar el QTL en fondos genéticos diferentes (Xie et al., 1998).

Por lo tanto, el uso de las PMP permite detectar QTLs con efectos consistentes en varios fondos genéticos, lo que ayuda a encontrar efectos QTL que son más generalizables (Jansen et al., 2003; Verhoeven et al., 2006). Esta es una ventaja importante en comparación con los QTLs detectados en poblaciones biparentales que a menudo tienen un efecto específico acotado al fondo genético particular sobre el que se realizó el estudio (Bernardo, 2016).

En comparación con el análisis de cada cruzamiento por separado, el análisis de las PMP permite aumentar el tamaño total de la población y el tamaño de la muestra para medir el efecto del QTL cuando tiene un efecto consistente en los cruzamientos (Li et al., 2005; Rebai & Goffinet, 1993; Verhoeven et al., 2006). El uso de más de dos progenitores permite obtener una muestra más representativa de la variación del QTL presente en la hipotética población de referencia, especialmente para los QTLs multialélicos (Muranty, 1996; Wu & Jannink, 2004).

Con respecto a los paneles de asociación, las PMP ofrecen un mayor control de la tasa de falsos positivos debido a la información completa y disponible sobre el *pedigree* y la estructura de la población (Myles et al., 2009). Las recombinaciones de los genomas parentales que se producen durante el desarrollo de las PMPs permiten romper el parentesco encubierto existente entre posiciones no vecinas (Yu et al., 2008). Por lo tanto, se reducen las probabilidades de detectar asociaciones espurias entre marcadores y caracteres (Garin, 2019).

En la Figura 1.3 se muestra el trazado de una línea continua entre los cruces biparentales, por un lado, y los paneles de asociación. En las poblaciones biparentales, todos los individuos proceden de una única familia. En los paneles de asociación cada individuo representa a una familia de origen. Las PMP son una situación intermedia con varias familias de tamaño medio o pequeño.



**Figura I.3.** Comparación del número de cruzamientos y de individuos por cruzamientos entre las poblaciones biparentales, multiparentales y los paneles de asociación. Figura tomada de Garin (2019).

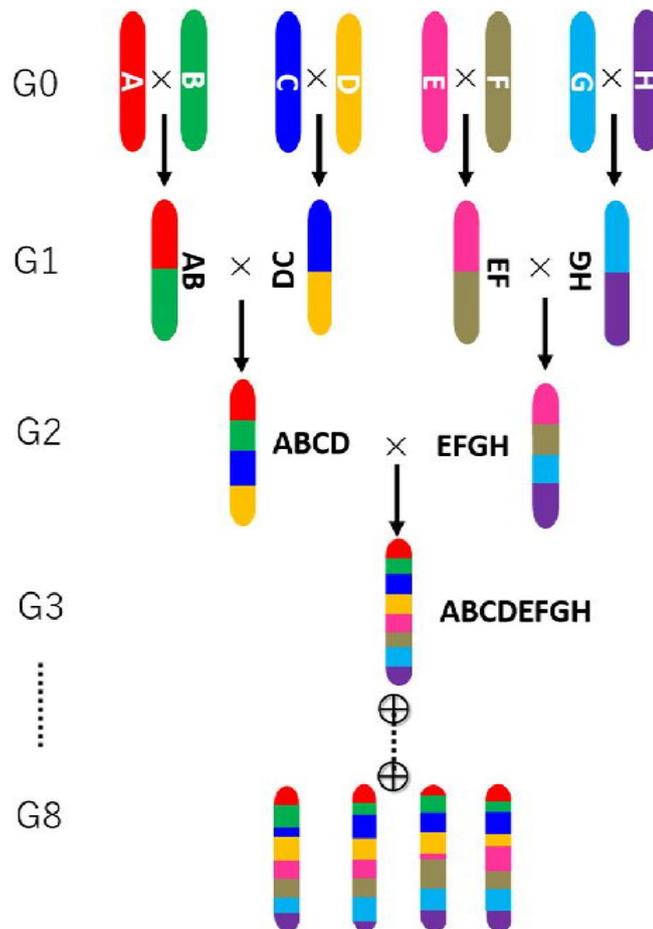
El mapeo fino es el proceso por el cual la localización de un QTL se reduce a un intervalo no mayor a 20 cM. Esta aproximación intenta la identificación de la variante genómica causante, que podría ser un polimorfismo de un solo nucleótido, SNP, u otra variante. Acotar la región del QTL o bien identificar el polimorfismo causante, permite avanzar hacia una caracterización molecular más profunda y a estudios funcionales dirigidos a inferir la relación entre el polimorfismo y el carácter. Asimismo, la reducción del intervalo del QTL a una región más pequeña sin identificar exactamente el polimorfismo o el gen funcional implicado es de gran importancia por sus implicancias para la introgresión de regiones cromosómicas de interés, para la predicción genómica, entre otras (Cockram & Mackay, 2018).

Los factores que afectan la precisión del mapeo fino son la fracción de recombinación entre el o los QTLs a mapear y los marcadores disponibles, la heredabilidad del QTL y el tamaño de la población de mapeo. Las poblaciones para el mapeo fino requieren de altos niveles de recombinación. Esto puede alcanzarse aumentando el tamaño de la población y aumentando el número de generaciones para desarrollar la población. El tamaño de la población tiene dos efectos en el mapeo, en primer lugar, influye directamente sobre la precisión ya que al aumentar el tamaño se incrementa el número de recombinaciones y también sobre el poder

de detección de un QTL. Finalmente, la heredabilidad del QTL será la que definirá su poder y precisión (Cockram & Mackay, 2018; Price, 2006)

En la mayoría de los estudios de mapeo de ligamiento de QTLs reportados para especies agronómicas, las poblaciones estructuradas de mapeo utilizadas incluyen poblaciones  $F_2$ , poblaciones de RILs, poblaciones de haploides duplicados (DH, *Doubled Haploids*), Retrocruzas, entre otras. A partir de la necesidad de recursos genéticos para alcanzar niveles de precisión más altos en la detección de QTLs se comenzaron los estudios en diferentes tipos de poblaciones experimentales con *pedigrees* más complejos que las conocidas biparentales (Cockram & Mackay, 2018).

Las poblaciones multiparentales de tipo MAGIC han surgido como una alternativa en el estudio de QTLs ofreciendo características intermedias entre las poblaciones biparentales y los paneles de asociación en términos de diversidad, potencia y resolución (Huang et al., 2015). Las MAGICs son desarrolladas a partir de múltiples cruzamientos entre un grupo de líneas parentales endocriadas. Luego de sucesivos ciclos de autofecundación de los híbridos se obtienen líneas recombinantes cuyos genomas son un mosaico de los genomas de las líneas fundadoras (Figura 1.4). Al igual que en las poblaciones biparentales, los alelos se presentan con frecuencias relativamente altas debido al número limitado de fundadores, pero la población captura una diversidad genética mucho mayor. Si bien una población MAGIC requiere una mayor inversión inicial en capacidad operativa y tiempo que una biparental, la cuidadosa selección de los fundadores permite su generalización para el estudio de diferentes caracteres de interés y garantiza su relevancia como panel de recursos genéticos a largo plazo (Arrones et al., 2020; Huang et al., 2015; Scott et al., 2020).



**Figura I.4.** Ejemplo de uno de los esquemas posibles para el desarrollo de una población MAGIC. En G0 se realizan los cruzamientos de 2 vías, en G1 los cruzamientos de 4 vías, en G2 los cruzamientos de 8 vías y a partir de G3 se desarrollan los ciclos de autofecundación para obtener líneas homocigotas recombinantes. Figura tomada de Yan et al. (2020).

La primera población multiparental reportada fue el cruzamiento colaborativo (CC, *Complex Trait Consortium* 2004; Churchill et al., 2004) en ratones, desde entonces el diseño ha tenido una gran aceptación en una amplia variedad de especies. El CC combinó los genomas de ocho razas endocriadas mediante múltiples cruzamientos y luego derivaron líneas endocriadas mediante el apareamiento entre hermanos. Al mismo tiempo, la población *Diversity Outbred* (DO) se desarrolló como una población heterogénea relacionada (HS) con los mismos ocho progenitores (*Collaborative Cross Consortium* 2012; Iraqi et al., 2012). Estas dos poblaciones constituyen un recurso genético abarcativo para estudios en ratones y se han utilizado para el mapeo y la identificación de genes candidatos para distintos caracteres, como el colesterol sérico y el color del pelaje (Ram et al., 2014; Svenson et al., 2012). Una estrategia de diseño similar se usó para crear una población multiparental de mapeo en *Drosophila* (King et al., 2012). En el caso de las plantas se han creado poblaciones MAGIC en *Arabidopsis thaliana* (Kover et al., 2009), en trigo (Huang et al., 2012; Mackay et al., 2014), en arroz

(Bandillo et al., 2013), en algodón (Islam et al., 2016a), en frutilla (Wada et al., 2017), en tomate (Pascual et al., 2015), sorgo (Ongom & Ejeta, 2018), cebada (Sannemann et al., 2015) y en una variedad de otros cultivos (Scott et al., 2020).

Debido a su amplia base genética, las poblaciones MAGIC son útiles para el mapeo de QTLs, el enriquecimiento de los recursos genéticos implementados en los programas de mejoramiento y el desarrollo de nuevos cultivares o variedades (Huynh et al., 2018). De esta manera, las poblaciones MAGIC proporcionan una plataforma ideal para el descubrimiento y el análisis de alta resolución de QTLs y genes responsables de caracteres agronómicos complejos para el mejoramiento genético de los cultivos (Cavanagh et al., 2008; Mackay & Powell, 2007).

### **Ventajas y desventajas de las poblaciones MAGIC**

La ventaja de las poblaciones MAGIC en la detección de QTLs radica en el número de recombinaciones acumuladas a lo largo de múltiples rondas de entrecruzamiento y autofecundación que mejoran la precisión del mapeo de QTLs, permitiendo identificar asociaciones genotipo-fenotipo con una mayor resolución (Arrones et al., 2020; Huang et al., 2012; Mackay & Powell, 2007; Pascual et al., 2015; Scott et al., 2020). El aumento de la tasa de recombinación también facilita una mayor reducción del desequilibrio de ligamiento y proporciona baja o nula estructura en la población reduciendo la ocurrencia de falsos positivos (Arrones et al., 2020; Cavanagh et al., 2008; Huang et al., 2012; Mackay & Powell, 2007; Sallam & Martsch, 2015; Valdar et al., 2006).

La combinación de múltiples parentales proporciona una mayor diversidad genética y fenotípica dentro una única población de mapeo, aumentando el número de QTLs que segregan y promoviendo nuevos reordenamientos de alelos (Kover et al., 2009). La elección óptima de las líneas fundadoras permitirá la presencia de múltiples caracteres en un fondo mixto de genomas parentales (Arrones et al., 2020; Huang et al., 2015; Kover et al., 2009). Por ejemplo, un QTL puede no expresarse en un fundador, pero si puede expresarse en la mezcla de los genomas parentales que dan lugar a las poblaciones MAGIC. Asimismo, estas proporcionan la oportunidad de evaluar simultáneamente el efecto de múltiples alelos en un locus y de estudiar las interacciones de los efectos del citoplasma, las introgresiones y la recombinación cromosómica que implica la diversidad alélica en el genoma (Arrones et al., 2020; Cavanagh et al., 2008; Huang et al., 2015).

Las RILs que constituyen la población MAGIC final, se consideran “inmortales” porque su genotipo ha sido fijado en cada marcador, preservando la identidad de la RIL en el tiempo, estas características simplifican en gran medida los experimentos repetidos en el tiempo, y

aseguran que la población pueda ser almacenada, difundida y utilizada para analizar diferentes caracteres fenotípicos en distintos ambientes. Además, proporcionan un germoplasma prometedor y útil para ser explotado en programas de mejoramiento para el desarrollo de líneas e híbridos mejorados, para ser liberados directamente como nuevos cultivares y para ser utilizados como parentales donadores de QTLs ya identificados (Arrones et al., 2020; Cockram & Mackay, 2018; Diouf et al., 2020; Ongom & Ejeta, 2018; Pascual et al., 2015; Scott et al., 2020).

Las principales desventajas de las poblaciones MAGIC radican en su desarrollo, dado que necesitan una mayor inversión de tiempo y esfuerzo para completar el esquema de cruzamientos de múltiples líneas fundadoras y un gran número de generaciones de autofecundación para obtener una población final MAGIC homocigótica estable (Diouf & Pascual, 2021; Mackay et al., 2014). Los cruzamientos de las líneas parentales seleccionadas y las siguientes generaciones de endocria llevan varios años debido a la limitante de una o dos generaciones por año en la mayoría de los cultivos, lo que ralentiza el desarrollo de las poblaciones MAGIC (Andrea Arrones et al., 2020). Metodologías como la obtención de haploides duplicados (DH) permiten reducir el tiempo de obtención de poblaciones totalmente homocigóticas, pero solo se incorporarán eventos de recombinación en las etapas de cruzamiento. Además, para muchos cultivos, esta técnica es fuertemente dependiente de la especie y del genotipo, y requiere el establecimiento previo de un protocolo optimizado y eficiente que suele ser muy laborioso (Murovec & Bohanec, 2012).

Por otra parte, si bien las poblaciones MAGIC son muy potentes para el mapeo fino requieren de un mayor tamaño poblacional por ejemplo que las biparentales, lo cual no solo limita el desarrollo de la población sino también el fenotipado de las mismas para distintos caracteres (Huang et al., 2015). Además, las poblaciones MAGIC debido a la compleja conformación de su estructura genómica demandan muchos marcadores moleculares para su análisis (Scott et al., 2020).

Una comparación entre las poblaciones biparentales, multiparentales y las colecciones de germoplasma para estudios de asociación, resaltando las ventajas y desventajas que presentan cada una, se muestran en la Tabla I.1.

**Tabla I.1.** Ventajas y desventajas de las poblaciones utilizadas para el mapeo de QTLs.

	<b>Ventajas</b>	<b>Desventajas</b>
Poblaciones biparentales	- Rápidas y fáciles de desarrollar	- Limitada por solo poder contrastar dos alelos. - Limitada recombinación.

	<ul style="list-style-type: none"> <li>- Útiles para el mapeo de alelos raros.</li> <li>- Análisis sencillo.</li> </ul>	<ul style="list-style-type: none"> <li>- Grandes intervalos para el mapeo de QTL.</li> </ul>
Poblaciones multiparentales	<ul style="list-style-type: none"> <li>- Mayor cantidad de alelos y QTLs segregando.</li> <li>- Mayor precisión que las poblaciones biparentales.</li> <li>- Rápido mapeo fino.</li> <li>- Útil para el escaneo de SNPs candidatos.</li> <li>- Apropiaada para selección de genotipos promisorios.</li> </ul>	<ul style="list-style-type: none"> <li>- Tiempo para desarrollarlas.</li> <li>- Requiere mayor cantidad de marcadores y mayor tamaño que las poblaciones biparentales.</li> <li>- Métodos de análisis en desarrollo o exploración.</li> </ul>
Paneles de asociación	<ul style="list-style-type: none"> <li>- Colecciones de germoplasma existentes.</li> <li>- Elevada diversidad.</li> <li>- Recombinación natural.</li> </ul>	<ul style="list-style-type: none"> <li>- Requiere gran cantidad de marcadores.</li> <li>- Posee estructura poblacional.</li> <li>- Si el desequilibrio de ligamiento es alto, el mapeo no es preciso.</li> <li>- Pobre identificación de alelos raros.</li> </ul>

### **Análisis de QTL en poblaciones multiparentales**

El análisis de QTLs para las poblaciones MAGIC fue discutido inicialmente por Xu (1996), quien utilizó un enfoque de mapeo por intervalos en base a la regresión de marcadores puntuales para un cruzamiento de cuatro vías (Haley & Knott, 1992). Sin embargo, Mott et al (2000) demuestran que la regresión simple utilizando las posiciones de los marcadores falla en la población CC y describen un método multipunto (conocido como HAPPY). Este método comienza construyendo las probabilidades de que un alelo haya sido heredado de los fundadores, lo cual es una alternativa al método anterior y resulta posible gracias a la estructura poblacional de la MAGIC (Verbyla et al., 2014). Esto puede proporcionar información valiosa, ya que los genotipos pueden no ser totalmente informativos con múltiples fundadores. Las probabilidades de que un marcador proceda de cada fundador se utilizan para identificar los QTLs y establecer sus tamaños para cada alelo fundador. Mott et al (2000) utilizan un modelo oculto de Markov (HMM de sus siglas en inglés *Hidden Markov Model*), también descrito por Broman (2006), para calcular las probabilidades del alelo fundador en cada intervalo.

Kover et al (2009) amplían los trabajos de otros autores como Mott et al (2000); Flint et al (2006); Valdar et al (2006), desarrollando tres métodos para el análisis de QTLs y aplicándolos a los datos de la población MAGIC de *Arabidopsis thaliana*. El primer método amplió la

propuesta de Mott et al (2000) utilizando un enfoque de selección hacia adelante después de un análisis inicial de un sólo marcador para permitir modelar múltiples QTLs considerando la posible estructura de la población. El análisis se repitió volviendo a muestrear los datos 500 veces e identificando QTLs en cada remuestreo. La confianza del QTL se determinó por la fracción de veces que se identificó en el total de los remuestreos. En el segundo enfoque de Kover et al (2009), se incluyó información de parentesco en las estimas de la estructura de la población, similar a la utilizada por Malosetti et al (2011) en una población de RILs en cebada derivadas de un cruzamiento de tres vías. El tercer enfoque utilizado fue un método jerárquico bayesiano. Los resultados alcanzados mediante los tres enfoques fueron muy similares.

El uso de estos métodos provoca un sesgo en los resultados obtenidos para poblaciones biparentales, lo que condujo al desarrollo del mapeo de intervalos compuesto (CIM de sus siglas en inglés *Composite Interval Mapping*) (Jansen, 1994; Zeng, 1994) y del mapeo de intervalos múltiples (Kao et al., 1999) para reducir los errores en la estimación de los QTLs. Un método que ha mostrado resultados superiores a los mencionados anteriormente es el mapeo de intervalos promedio de todo el genoma (WGAIM de sus siglas en inglés *Whole-Genome Average Interval Mapping*) (Verbyla et al., 2007, 2012, 2014; Verbyla & Cullis, 2012). Esta metodología incorpora simultáneamente toda la información de los marcadores en el análisis, superando la necesidad de repetir los escaneos del genoma. Además, utiliza un enfoque de selección hacia adelante que reduce el número de análisis, permite modelar la estructura de la población a través de la inclusión de la información del *pedigree* e incluir fácilmente cualquier efecto no genético (Verbyla et al., 2014).

A pesar de las dificultades encontradas por Mott et al (2000) utilizando directamente la información proporcionada por los marcadores, varios estudios han empleado métodos de mapeo por asociación (Bandillo et al., 2013; Mackay et al., 2014; Ongom & Ejeta, 2018) utilizando programas como TASSEL disponibles para los estudios conocidos como *Genome Wide Association Studies* (GWAS) (Bradbury et al., 2007). La ventaja de este enfoque es que la prueba es simple y computacionalmente sencilla, requiriendo menos grados de libertad que otros métodos. Estos métodos suelen coincidir con los basados en las probabilidades de los fundadores para QTLs mayores, pero pueden diferir para aquellos QTLs con efectos más pequeños (Huang et al., 2015).

Otro enfoque que se plantea para el análisis de QTLs en las poblaciones MAGIC es la reconstrucción de los mosaicos de haplotipos que representan la mezcla de los genomas parentales en las líneas endocriadas recombinantes. Los mosaicos se reconstruyen esencialmente en marcadores multialélicos a partir de las probabilidades de los fundadores.

Este enfoque puede realizarse gracias al conocimiento de la estructura de *pedigree* de la población. Uno de los problemas de este enfoque es que no tiene en cuenta la incertidumbre asociada al proceso de estimación. Además, en las regiones del genoma en las que los genotipos fundadores son muy similares o la densidad de los marcadores es baja, hay un menor contenido de información genómica que limita la capacidad de imputar alelos con gran certeza. Sin embargo, si se estiman con poco error, estos valores deberían ser los más cercanos a los verdaderos genotipos subyacentes, y pueden utilizarse para identificar puntos de recombinación en líneas individuales. El conocimiento de la estructura de los haplotipos de la población ofrece una estimación más informativa de los QTLs; sin embargo, debido al número de genomas fundadores, la observación directa de la contribución de cada progenitor a un QTL sigue siendo un reto (Broman et al., 2019a; Huang et al., 2015; Ogawa et al., 2018; Zheng et al., 2015).

Las diferentes etapas de este proceso de reconstrucción de haplotipos han dado lugar a tres representaciones de los datos genéticos de las poblaciones MAGIC: marcadores puntuales, probabilidades de los fundadores y los mosaicos haplotípicos mencionados. Cada uno de ellos puede utilizarse para investigar la estructura genómica y como datos de entrada para el análisis de QTLs (Huang et al., 2015).

Los principales retos que se plantean en el análisis de los datos de las poblaciones multiparentales son la reconstrucción de los haplotipos fundadores, la imputación de las variantes genéticas de todo el genoma, los métodos de análisis que pueden manejar los múltiples alelos fundadores y tener en cuenta la estructura de la población (Broman et al., 2019a).

### **El rol del fenotipado en los programas de mejoramiento genético**

El proceso de mejoramiento comprende tres principales clases de actividades: hibridación, fijación del genotipo y evaluación del rendimiento en ensayos multiambientales. En términos generales, la mejora genética se lleva a cabo mediante la selección de plantas que presentan características deseables, denominados caracteres (por ejemplo: altura de la planta, contenido de aceite del grano, días a floración, entre otros). El proceso de evaluación del rendimiento de un genotipo para un conjunto determinado de caracteres en un ambiente específico se denomina fenotipado (Serrano, 2020). En un programa de mejoramiento clásico, las actividades de fenotipado implican una operación logística a gran escala, que involucra de miles a cientos de miles de plantas al inicio del proceso de evaluación y selección, para terminar con un pequeño número de líneas seleccionadas que luego se liberan como cultivares o se utilizan como progenitoras de híbridos comerciales.

El fenotipado se define como el conjunto de métodos y protocolos aplicados para determinar con exactitud y precisión el crecimiento, la arquitectura y la composición de las plantas a diferentes escalas (Fiorani & Schurr, 2013). La cuantificación directa del fenotipo implica la medición de las características estructurales y funcionales, así como la composición química de las plantas o partes de las mismas. El método elegido para fenotipar depende del cultivo, del carácter a evaluar, de las fases de desarrollo y de los recursos disponibles. Como normalmente es necesario evaluar múltiples caracteres, resulta recomendable contar con un método que permita evaluar varios caracteres en paralelo para reducir los costos, el tiempo y los posibles errores relacionados con la revisión o la participación de muchos operadores en la adquisición de datos. Además, el tamaño de un ensayo típico de fenotipado varía entre cientos y miles de plantas/parcelas, lo que requiere la implementación de métodos y protocolos de medición eficientes (Chawade et al., 2019).

El fenotipado puede llevarse a cabo en diferentes condiciones: controladas, semicontroladas o naturales. Las condiciones controladas se consiguen en cámaras de crecimiento e invernaderos y pueden automatizarse utilizando cintas transportadoras o robots para transportar las plantas a la unidad de medición. En estos casos, se puede controlar un amplio conjunto de variables ambientales para simular las condiciones de crecimiento que son relevantes para el experimento de fenotipado. En el campo puede establecerse un ambiente semicontrolado que permite alterar o controlar parte de las condiciones ambientales para simular situaciones específicas (por ejemplo, utilizando refugios contra la lluvia para simular condiciones de sequía). Los ambientes naturales se corresponden con los datos que se pueden obtener en condiciones equivalentes a las de cultivo en producción, y tienen como ventaja que se pueden cultivar plantas del tamaño de un cultivo, pero están limitados a las condiciones de crecimiento estacional, lo que requiere evaluaciones repetidas en diferentes años o localidades (Fahlgren et al., 2015; Serrano, 2020).

El fenotipado de las plantas sigue dependiendo en gran medida de las mediciones visuales y en ciertos casos destructivas que demandan mucho tiempo, esfuerzo y recursos. Además, estas tienen como inconveniente la subjetividad y la falta de resolución temporal para describir procesos como el crecimiento de las plantas. Actualmente, las nuevas metodologías de fenotipado de alto caudal, *High Throughput Phenotyping* (HTP) complementan cada vez en mayor medida a las técnicas clásicas de fenotipado y permiten evaluar el comportamiento de las plantas de forma más objetiva y eficiente contribuyendo enormemente a aumentar la ganancia genética de los programas de mejoramiento (Reynolds et al., 2020; Serrano, 2020).

La ganancia genética es un concepto central en el mejoramiento genético y refleja el progreso del mismo. Se define como la magnitud del incremento anual en el desempeño alcanzado para un carácter determinado a través de la selección artificial (Fehr, 1991; Fischer & Edmeades, 2010; Xu et al., 2017). Comúnmente, la ganancia genética se define a través de la siguiente ecuación:

$$\text{Ganancia genética} = \frac{k * h * \sigma_A}{y}$$

Donde  $k$  refiere a la intensidad de selección,  $h$  a la heredabilidad en sentido estricto,  $\sigma_A$  a la varianza genética aditiva e  $y$  al número de años para completar el ciclo de mejoramiento de la especie. La ecuación de la ganancia genética traduce la eficiencia de la selección en los programas de mejoramiento de plantas en parámetros medibles que los mejoradores pueden manipular en sus líneas de trabajo (Zhou, 2021).

La ganancia genética puede incrementarse mediante: I) el aumento del tamaño del programa de mejoramiento permitiendo aplicar una mayor intensidad de selección, II) aumentando la precisión de la selección, (III) asegurando un *screening* adecuado y la explotación de la variabilidad genética, (IV) acelerando los ciclos de mejoramiento y (V) perfeccionando las herramientas de apoyo para la toma de decisiones (Araus et al., 2018; Masuka et al., 2017).

De lo expuesto anteriormente, se desprende que el rendimiento de un programa de mejoramiento genético puede volverse más eficiente mediante la ampliación y el aumento de la precisión de la evaluación de un gran número de plantas. Sin embargo, esto suele verse limitado por las incapacidades propias que conlleva el fenotipado de forma manual (Wang et al., 2019). Por lo tanto, un fenotipado de alto rendimiento puede contribuir directa o indirectamente a obtener mayores ganancias genéticas, ya que permite aumentar la escala de la evaluación, la precisión y objetividad de la medición. En combinación con la información genómica, se espera que las herramientas de fenotipado de precisión mejoren la predicción y aumenten la tasa de ganancia genética acortando los ciclos y/o logrando una mayor ganancia por ciclo de reproducción (Cobb et al., 2013; Crossa et al., 2017).

### **Programa de mejoramiento de girasol de INTA y Antecedentes del grupo**

A principios del siglo XX, las Estaciones Experimentales Pergamino (1939) y Manfredi (1947) iniciaron un programa de investigación y desarrollo para producir semillas de variedades identificadas y homogéneas en calidad y cantidad suficiente para el mercado interno de girasol, la principal oleaginosa de Argentina. La creación del INTA (1956) acentuó la actividad de mejoramiento genético del girasol en las Estaciones Experimentales con foco en el aumento del rendimiento (Alvarez et al., 2021).

La base para la producción de las primeras colecciones de trabajo para el mejoramiento del girasol en Argentina fue obtenida a partir de introducciones iniciales por inmigrantes que se adaptaron a las condiciones locales y que dieron lugar a poblaciones conocidas como el Gigante ruso, Gigante americano, Mamut, entre otras. Estas poblaciones eran altas, de ciclo largo y con amplia variabilidad en el color de grano, predominando el tipo estriado. Estas poblaciones, junto con nuevas introducciones, principalmente de Rusia, Estados Unidos, Canadá y Europa del Este, conformaron la colección de recursos genéticos de girasol del INTA (Alvarez et al., 2021; Bertero de Romano & Vázquez, 2003; Castaño, 2018).

Al principio, el programa desarrollaba variedades de polinización abierta utilizando las poblaciones introducidas por inmigrantes adaptadas a las condiciones locales, como así también los cruzamientos entre ellas y con otras especies del género *Helianthus*, desarrollando numerosas variedades de polinización abierta mejoradas, entre ellas Manfredi INTA, Impira INTA, Cordobés INTA, Guayacán INTA, Ñandubay INTA, Pehuén INTA, Aguaribay INTA, por citar algunas (Bertero de Romano & Vázquez, 2003). Con el desarrollo de la androesterilidad citoplasmática en los años 70 (Leclercq, 1969), los programas se centraron en la producción de líneas endocriadas para la obtención de híbridos. Ambos programas se han complementado para ampliar sus actividades, llegando actualmente a la evaluación en la fase precomercial de cohíbridos de germoplasma compartido entre ambos programas, ampliando el uso de la variabilidad genética del cultivo (González et al., 2015).

El Banco Activo de Germoplasma de INTA Manfredi (BAG-IM) es el único a nivel nacional que conserva germoplasma de girasol. Funciona desde hace más de 30 años y alberga cerca de 1200 entradas. El germoplasma 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, Argentina, Armenia, Australia, Bolivia, Brasil, Bulgaria, Canadá, Chile, China, España, Estados Unidos, Francia, Grecia, Hungría, Israel, Italia, Marruecos, Moldavia, Polonia, Rumania, Rusia, Serbia, Siria, Turquía, Uruguay, entre otros. Las colecciones fueron realizadas directamente en los lugares de cultivo y posteriormente recibidas, conservadas y multiplicadas por el BAG-IM (Moreno, 2010; Terzic et al., 2020).

El programa de mejoramiento de girasol conducido en EEA Pergamino desde 1939 ha desarrollado materiales de buenas características fitosanitarias y de calidad industrial (alto contenido de ácido oleico), además de buen rendimiento y contenido de aceite; contribuyendo al aumento de la tasa de ganancia genética del cultivo (Bertero de Romano & Vázquez, 2003; González et al., 2015).

Las características sanitarias, agronómicas, de calidad y rendimiento industrial de las líneas obtenidas ampliaron la base genética disponible para los programas de mejoramiento de girasol en la Argentina. A partir de la década del '90, se produjo un cambio significativo con la aparición híbridos estriados desarrollados por programas de mejoramiento local, que se tradujo en un incremento significativo de los rendimientos y en un mejor comportamiento sanitario; el cual estuvo principalmente sustentado por el aporte del germoplasma desarrollado por INTA (González et al., 2015).

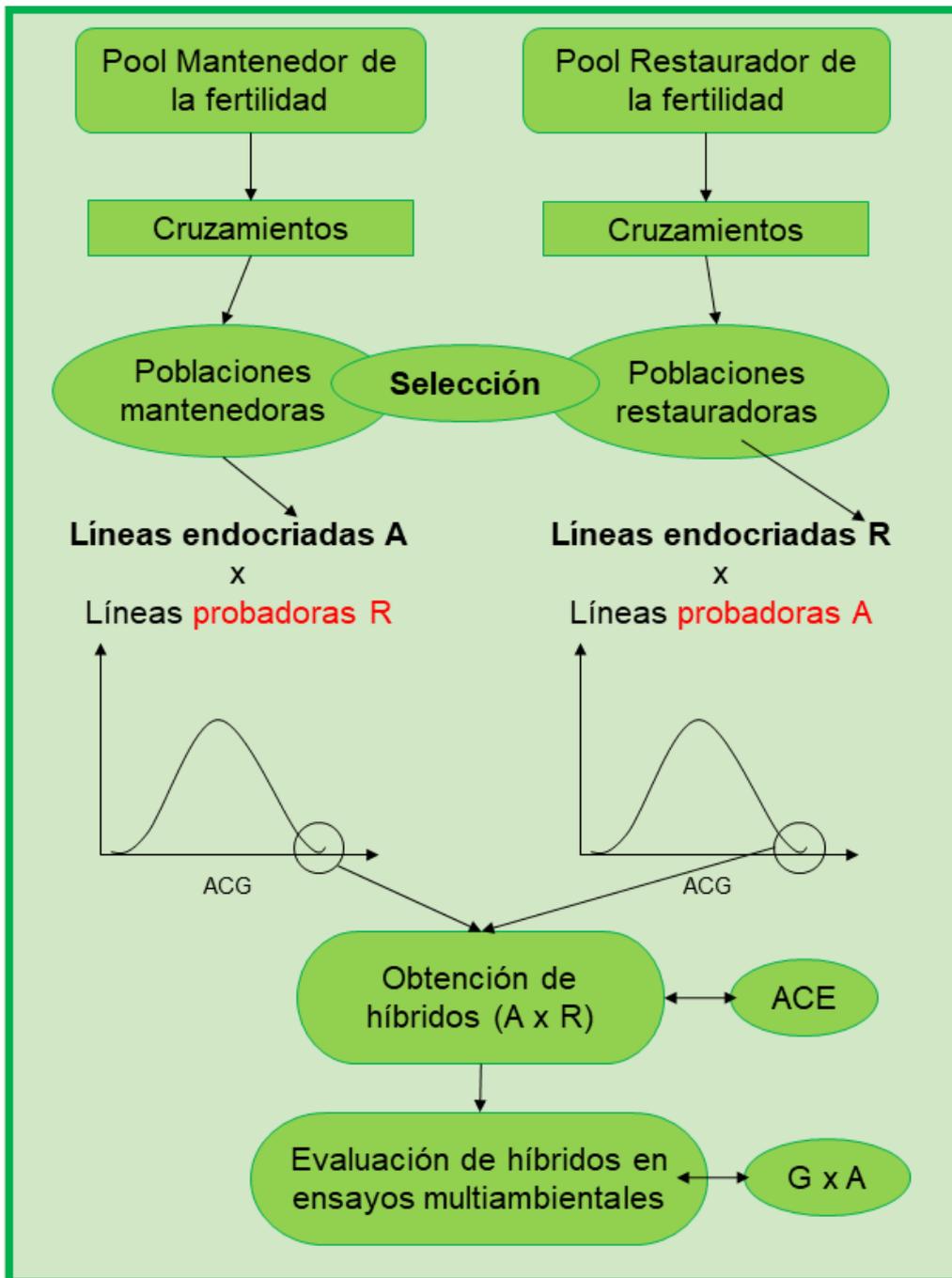
El germoplasma obtenido también aportó al mejoramiento del girasol en otros países. En Estados Unidos, se obtuvieron líneas y compuestos con resistencia a las principales enfermedades y buen comportamiento frente al ataque de pájaros, derivadas de germoplasma argentino (Gonzalez, 2016; Gulya, 1985). En Serbia, Skoric et al (1978) informaron que el cultivar RHA 93 con resistencia a Downy mildew se obtuvo a partir de la variedad Pehuén INTA. En Rusia, Pustovoit (1978) señalaron que líneas derivadas de Pehuén INTA mostraron un alto nivel de resistencia a *Sclerotium bataticola*. En Francia, Vear et al (2008) reportaron la utilización de materiales argentinos como fuente de resistencia a Downy mildew.

El Programa de Mejoramiento de Girasol de INTA actualmente aborda con un enfoque interdisciplinario la identificación y caracterización de nuevas fuentes de resistencia a estreses bióticos y abióticos en germoplasma propio o de origen público internacional. Tiene como objetivo obtener líneas endocriadas e híbridos de alta productividad, estabilidad y adaptabilidad a los diferentes ambientes de la región girasolera argentina. Los esfuerzos del programa se concentran en la obtención de genotipos competitivos en rendimiento, estabilidad, plasticidad y comportamiento agronómico para las distintas áreas de producción del país; con altos contenidos de aceite y perfiles diferenciales de ácidos grasos; con buen comportamiento frente a los principales estreses bióticos y abióticos que afectan al cultivo, entre ellos la Marchitez por *Verticillium dahliae*, la Podredumbre húmeda del capítulo por *Sclerotinia sclerotiorum*, el Enanismo del girasol causado por *Plasmopara halstedii*, el Cancro por *Phomopsis*, el daño por aves, el estrés hídrico, entre otros; y con tolerancia a herbicidas que aseguren sustentabilidad ambiental en diferentes sistemas productivos .

Para dar cumplimiento a estos objetivos el programa desarrolla actividades tendientes a la creación de variabilidad genética, conducción de poblaciones segregantes y selección por caracteres de interés, obtención de líneas endocriadas estabilizadas e híbridos y la evaluación de estos en múltiples ambientes.

Para la generación de variabilidad se realizan cruzamientos entre genotipos de diversos orígenes. Se trabaja en poblaciones genéticamente diversas empleando el método de

selección recurrente de progenies, y con cruzamientos entre líneas mantenedoras y entre líneas restauradoras con atributos destacados, y selección en las progenies segregantes. La obtención, desarrollo y evaluación de líneas e híbridos, se realiza utilizando la evaluación fenotípica convencional y se pretende incorporar la asistencia de marcadores moleculares en esquemas de selección asistida. La conducción del programa se divide en dos grupos heteróticos, el grupo de líneas mantenedoras de la fertilidad y el grupo de líneas restauradoras de la fertilidad. Las líneas mantenedoras con atributos destacados serán convertidas por un sistema de sucesivas retrocruzas a isolíneas androestériles para su utilización en la formación de híbridos. Una vez obtenidas las isolíneas androestériles (A) se las cruzan con líneas restauradoras de la fertilidad (R) para la evaluación de su aptitud combinatoria general (ACG) y específica (ACE). En la Figura I.5 se resume en un esquema general el procedimiento para la obtención de híbridos del programa de mejoramiento de girasol de INTA.



**Figura I.5.** Esquema de obtención de híbridos de girasol del programa de mejoramiento de INTA.

Este programa reúne además del esfuerzo de los grupos de mejoramiento (EEA Manfredi y EEA Pergamino), el de los grupos de fitopatología (EEA Balcarce, EEA Anguil y EEA Pergamino), biotecnología (Instituto de Biotecnología y EEA Manfredi) y el Banco Activo de Germoplasma (BAG-IM, EEA Manfredi).

La EEA Manfredi y la EEA Pergamino participan activamente en la creación y difusión de material mejorado de girasol a nivel local e internacional, y proveen los recursos genéticos que constituyen la base de los estudios sobre resistencia a estreses bióticos y abióticos realizados en INTA.

La EEA INTA Balcarce es la sede del estudio fenotípico de las enfermedades fúngicas producidas en girasol fundamentalmente por *Sclerotinia sclerotiorum* y *Verticillium dahliae*. Allí se han llevado a cabo estudios de monitoreo de enfermedades (Pereyra & Escande, 1994), se ha estudiado la patogenicidad del hongo *S. sclerotiorum* con establecimiento de un protocolo de fenotipificación para el estudio de la Podredumbre Húmeda del Capítulo (PHC) (Castaño et al., 1993; Maringolo, 2007; Prioretta et al., 1995) y se han realizado evaluaciones fenotípicas de la PHC de los materiales del programa de mejoramiento (Filippi et al., 2017, 2020; Fusari et al., 2012; Pedraza et al., 2004; Pereyra & Escande, 1994; Quiroz et al., 1998; Zubrzycki et al., 2017). También se han realizados estudios fenotípicos de los recursos genéticos del programa de mejoramiento de girasol de INTA, como la población de mapeo por asociación (PMA) y una población biparental frente a la Marchitez por *V. dahliae* en el infectario natural para identificar las regiones cromosómicas implicadas en la resistencia a esta enfermedad (Montecchia, 2019; Montecchia et al., 2021). Por otro lado, en la EEA Anguil se analizan las enfermedades producidas por *V. dahliae* y *Diaporthe helianthi*, mientras que en la EEA Pergamino los estudios involucran el comportamiento de materiales ante las infecciones con *V. dahliae*, *S. sclerotiorum* y *Plasmopara halstedii* (González et al., 2008, 2015).

El grupo de Genómica y Bioinformática del Instituto de Agrobiotecnología y Biología Molecular (GGB-IABiMo) desarrolló una amplia gama de herramientas moleculares para asistir a los programas de mejoramiento de girasol. Entre ellas se incluyen marcadores moleculares de tipo neutro (Paniego et al., 2002), marcadores funcionales derivados de secuencias que se expresan o ESTs (Fernandez et al., 2008; Filippi et al., 2014; Fusari et al., 2008; Zubrzycki et al., 2012), un banco de ESTs órgano-específicos aislados de colecciones de ADN complementario (ADNc) diferenciales (Fernández et al., 2003), mapas genéticos integrados en base a información local y pública (Kiani et al., 2007; Montecchia, 2019; Talia et al., 2010, 2011), una plataforma de análisis funcional que involucra análisis de metabolitos frente a infecciones con patógenos (Peluffo et al., 2010), la identificación de genes de referencia para cuantificación de genes candidato de girasol por PCR cuantitativa (Fernandez et al., 2011), el desarrollo de un microarreglo de expresión (Fernandez et al., 2012) y el desarrollo de plataformas de genotipificación de marcadores SNP (Filippi et al., 2015, 2020; Fusari et al., 2011; Montecchia, 2019; Zubrzycki et al., 2017).

Estas herramientas han contribuido al programa de mejoramiento de INTA a través de las identificaciones de regiones cromosómicas de interés alcanzada por los estudios de mapeo de QTL para resistencia a los patógenos *S. sclerotiorum* y *V. dahliae* en poblaciones biparentales y en la PMA (Filippi et al., 2015, 2020; Fusari et al., 2012; Maringolo, 2007; Montecchia, 2019; Zubrzycki et al., 2017). Se obtuvieron avances en el análisis de los perfiles de expresión de transcritos y de metabolitos en genotipos de respuesta contrastante frente a PHC (Fernandez et al., 2012; Peluffo et al., 2010). También se ha emprendido una estrategia de genética reversa para la búsqueda de genes candidatos para el comportamiento frente a PHC en poblaciones mutagenizadas, para comprobar su contribución al fenotipo resultante en girasol y *Arbidopsis thaliana* (Fass, 2018; Fass et al., 2020). Recientemente comenzó a estudiarse también la resistencia genética al Cancro del tallo causado por *D. helianthi*, una enfermedad emergente de gran impacto en el cultivo, que es parte de la Tesis doctoral del Ing. Agr. Denis Colombo.

El programa de mejoramiento de girasol de INTA también aborda estudios de resistencia a estreses abióticos, focalizando principalmente en la comprensión de la senescencia foliar, la respuesta al estrés hídrico y la asociación entre ambos procesos. Estos procesos se estudiaron a través de distintos niveles de organización: ecofisiológico, metabolómico y transcriptómico, culminando con la integración de los diversos enfoques ómicos (López Gialdi et al., 2016; Moschen et al., 2016, 2017, 2019).

La meta principal del trabajo publicado en esta Tesis es generar nuevos recursos genéticos de girasol que complementen la plataforma actual del programa de mejoramiento de INTA para contribuir al estudio y entendimiento de la arquitectura genética de diferentes atributos de interés. Hasta el momento no se cuenta a nivel internacional con poblaciones de este tipo en el cultivo de girasol, las poblaciones multiparentales que se desarrollan pretenden ser una herramienta de gran utilidad para el estudio de diferentes caracteres de interés agronómico.

---

---

## HIPÓTESIS

---

---

Dado que los parentales seleccionados para los cruzamientos son genéticamente diversos para una amplia gama de caracteres agronómicos, y muestran una amplia variabilidad genética; las poblaciones MAGIC resultantes mostrarán bajo desequilibrio de ligamiento, baja estructura poblacional y una mayor frecuencia de alelos minoritarios; aspectos que las convertirán en herramientas óptimas para el mapeo fino de QTLs, así como fuente de líneas endocriadas recombinantes adecuadas para el mejoramiento genético del girasol.

Dado que el tiempo requerido para la obtención de líneas recombinantes homocigotas excede el periodo de ejecución de la Tesis se plantean las siguientes hipótesis:

1. La información generada durante los últimos 15 años de estudio sobre los recursos genéticos del programa de mejoramiento de girasol de INTA permitirá la selección de parentales genéticamente diversos para la generación de poblaciones multiparentales que contengan una amplia diversidad genética para diversos caracteres de interés agronómico.
2. La homogeneidad de las líneas parentales y la calidad de los cruzamientos puede ser evaluada a partir de datos fenotípicos y genotípicos.
3. La implementación de herramientas de fenotipado de alto rendimiento facilitarán la evaluación de poblaciones de alto número de individuos para el mapeo fino de QTLs.

---

---

## **OBJETIVO GENERAL**

---

---

## Objetivo general

Este trabajo tiene como objetivo el diseño, la construcción y caracterización de poblaciones multiparentales para complementar la plataforma actual de recursos genéticos del programa de mejoramiento de girasol de INTA para el estudio de regiones genómicas de interés agronómico.

---

---

## **OBJETIVOS ESPECÍFICOS**

---

---

## Objetivos específicos

1. Seleccionar un grupo de líneas homocigotas del conjunto de accesiones que conforman la población de asociación de girasol, los parentales de poblaciones biparentales y las progenies derivadas como parentales para la construcción de poblaciones multiparentales de líneas mantenedoras.
2. Desarrollar marcadores basados en mutaciones puntuales para analizar la homogeneidad genética de las líneas parentales y la eficacia de los cruzamientos realizados para la construcción de las PMP.
3. Diseñar e implementar un esquema de cruzamientos para iniciar la construcción de poblaciones multiparentales.
4. Validar el potencial de la población a partir del análisis de segregación de características agronómicas conocidas en las líneas parentales y analizadas previamente en estudios de mapeo de asociación y poblaciones biparentales.
5. Evaluar el uso de índices de vegetación obtenidos a partir de imágenes multiespectrales en modelos predictivos para caracterizar la resistencia a la marchitez anticipada por *V. dahliae* en girasol.

---

---

**CAPITULO 1: Selección de las líneas parentales y construcción de las poblaciones multiparentales de girasol**

---

---

## INTRODUCCIÓN

La selección de las líneas parentales es un factor clave en el desarrollo de poblaciones multiparentales ya que determinará la variación genética que segregará en la población y definirá el potencial de uso de la población para estudios de mapeo de QTLs (Scott et al., 2020). Idealmente, los parentales a seleccionar deberían cubrir la mayor diversidad genética, fenotípica y geográfica para poder explotar al máximo el potencial de la población y lograr recursos genéticos altamente informativos (Andrea Arrones et al., 2020).

Las poblaciones multiparentales que se han desarrollado hasta el momento en cultivos difieren en las estrategias de selección para las líneas parentales, algunas poblaciones enfocaron la selección de parentales en genotipos muy diversos entre sí y provenientes de regiones geográficas diferentes, este es el caso de poblaciones desarrolladas en arroz y trigo (Bandillo et al., 2013; Mackay et al., 2014); otras poblaciones desarrolladas por ejemplo en cebada y *Arabidopsis* fueron fundadas incluyendo *landraces* y/o accesiones silvestres entre sus parentales con el objetivo de maximizar la diversidad genética (Kover et al., 2009; Maurer et al., 2015). En otros casos, la selección de los parentales fue en base a un carácter particular como roya del tallo en trigo pan y estreses bióticos y abióticos en arroz (Bajgain et al., 2016; Leung et al., 2015).

Las poblaciones multiparentales originadas a partir de parentales diversos, provenientes de regiones geográficas específicas y con características especiales son útiles en diferentes escenarios. Aquellas que parten de parentales muy diversos como *landraces* y accesiones de origen silvestre tienen como principal ventaja enriquecer notablemente la diversidad y lograr nuevas combinaciones alélicas de gran utilidad para el mapeo. La principal desventaja de estas poblaciones es el arrastre de alelos en los individuos que las componen lo cual limita su potencial uso como recursos genéticos para el mejoramiento. Otros inconvenientes de la inclusión de parentales silvestres pueden ser la incompatibilidad, el arrastre por ligamiento debido a la falta de recombinación en regiones particulares del genoma, dormición de las semillas y asincronía floral reduciendo el número de progenies (Huang et al., 2015). En cambio, las poblaciones que se desarrollaron con una estrategia más conservadora a partir de líneas o variedades mejoradas podrían ser rápidamente incorporadas en líneas o variedades superiores y resultan ser de gran interés para los mejoradores (Scott et al., 2020).

El diseño de las poblaciones multiparentales de tipo MAGIC requieren numerosas generaciones de cruzamientos proporcionales al número de parentales y estas generaciones de cruzamientos pueden llevarse a cabo de diferentes maneras. Los esquemas más utilizados son:

- A) Esquema de tipo **embudo**: consiste de varias generaciones de cruzamientos entre un número (n) de líneas parentales para obtener  $n/2$   $F_1$  híbridos, los cuales son seguidamente cruzados entre ellos en un diseño de cruzamientos específico para obtener múltiples combinaciones de los genomas de todos los parentales en la progenie.
- B) Esquema de tipo **dialélico**: los parentales son cruzados en un esquema de dialélicos completos o incompletos, generándose múltiples embudos de híbridos  $F_1$  que luego son cruzados nuevamente asegurando la representatividad de todos los parentales. El diseño dialélico completo o incompleto es similar al del embudo único pero su principal diferencia es que cubre un mayor número de combinaciones.

La elección del esquema de cruzamientos a realizar dependerá principalmente de los recursos disponibles para la formación de la población, si bien el esquema de dialélicos completos o incompletos logra un mayor número de combinaciones entre los parentales fundadores, el incremento de los recursos necesarios y el tiempo para formar la población se incrementarán significativamente. En cambio, la utilización de esquemas de tipo embudo reducen el número de combinaciones posibles pero tienen la ventaja de insumir menos recursos y tiempo (Huang et al., 2015). En el caso de poblaciones creadas en trigo y cebada (Sannemann et al., 2015; Stadlmeier et al., 2018) usando esquemas de tipo embudo, con un número reducido de combinaciones híbridas, sumaron una generación de cruzamientos con el propósito de compensar la reducción de combinaciones y aumentar la capacidad de resolución para el mapeo de QTLs. Otras modificaciones al esquema de cruzamientos fueron publicadas para los cultivos de frutilla, haba y algodón (Islam et al., 2016b; Sallam & Martsch, 2015; Wada et al., 2017) donde se realizaron numerosas generaciones de cruzamientos aleatorios libres para equilibrar la falta de recombinaciones.

Una vez alcanzadas las sucesivas generaciones de cruzamientos entre los parentales el proceso continúa hasta lograr individuos en homocigosis estable. Para alcanzar este objetivo, las poblaciones desarrolladas hasta el momento han sido obtenidas a través de sucesivos ciclos de autofecundación, al menos 5 generaciones, utilizando estrategias como la descendencia de única semilla (Brim, 1966) o el de *pedigree* (Acquaah, 2012). En otras poblaciones fue utilizada la técnica de haploides duplicados para alcanzar la homocigosis (Forster et al., 2007). Si bien la técnica de haploides duplicados es notablemente más rápida, las múltiples generaciones de endocria tienen la ventaja de introducir un mayor número de recombinaciones (Huang et al., 2015).

El tamaño de la población es también uno de los factores más importantes a tener en consideración durante el desarrollo de una población multiparental, para evitar posibles cuellos de botella que produzcan una pérdida de alelos por muestreo y para asegurar la capacidad de detección de QTLs de la población (Andrea Arrones et al., 2020). Un tamaño de población mínima entre 50 y 250 individuos generalmente se emplea para el mapeo grueso de QTLs (Collard et al., 2005; Jaganathan et al., 2020). El poder y la resolución del mapeo de QTLs se incrementa con el aumento del tamaño de la población (Broman, Gatti, Simecek, et al., 2019b; Valdar, Flint, et al., 2006b). Además, cuanto mayor sea el tamaño de la población, más débil será el desequilibrio de ligamiento entre y dentro de los bloques cromosómicos, y menor será el número de marcadores necesarios para definir un haplotipo (Visscher et al., 2012). Si bien incrementar el tamaño de la población confiere las ventajas mencionadas anteriormente, implica una mayor demanda de recursos para su mantenimiento y sus respectivas caracterizaciones fenotípicas y genotípicas. Por lo que resulta necesario buscar un tamaño poblacional equilibrado acorde a los recursos y estructura que se disponga tratando de resignar la menor capacidad de resolución posible (Andrea Arrones et al., 2020).

El tamaño óptimo de una población destinada para el mapeo de QTLs dependerá fuertemente del tamaño del genoma de la especie en cuestión. Especies con genomas de gran tamaño requerirán al menos unos 500 individuos para lograr una resolución en el rango del sub-centimorgan (Huang et al., 2015; Valdar et al., 2006). En cambio, en organismos con genomas de menor tamaño se podría emplear poblaciones más pequeñas sin afectar su resolución en el mapeo, por ejemplo en la población desarrollada en “Poroto tape” (*Vigna unguiculata*) cuyo genoma es de 620 Mb, 305 individuos fueron suficientes para obtener una alta resolución en todo el genoma (Huynh et al., 2018). Para obtener una resolución física del mismo tenor que la población anterior, en especies con genomas del orden de los 5 Gb como en cebada se necesitan unas 2500 RILs (Andrea Arrones et al., 2020; Huynh et al., 2018).

La selección de las líneas parentales de las poblaciones MAGIC que se presentarán en esta Tesis se realizó sobre un panel de líneas endocriadas de girasol que constituyen la PMA de INTA y que han sido extensamente caracterizadas genotípica y fenotípicamente para diferentes estreses bióticos, abióticos, atributos de calidad y rendimiento.

En este capítulo se presenta la metodología de selección de las líneas parentales de las poblaciones multiparentales de girasol en desarrollo, las características de las líneas parentales seleccionadas y el esquema de construcción implementado para las mismas.

## MATERIALES Y MÉTODOS

### Selección de las líneas parentales

Para la selección de las líneas parentales se analizaron los datos relevados en los últimos 15 años en el programa de mejoramiento de INTA sobre un grupo de 79 líneas mantenedoras de pre-mejoramiento que conforman la población de mapeo de asociación, los parentales de poblaciones biparentales y las progenies derivadas. Estos materiales han formado parte de estudios de identificación de QTLs y/o genes candidatos para resistencia a enfermedades como marchitez por *V. dahliae* (MV), podredumbre húmeda del capítulo por *S. sclerotiorum* (PHC), tolerancia a sequía, atributos relacionados con retraso de senescencia foliar y de calidad en cuanto a contenido de aceite y tocoferoles.

Las variables utilizadas para la selección de parentales fueron para PHC incidencia (INC\_PHC), severidad (SEV\_PHC), área por debajo de la curva de progreso de la enfermedad para incidencia y severidad (AUDPC\_INC\_PHC) y (AUDPC\_SEV\_PHC) y período de incubación de la enfermedad (PI\_PHC). Para MV las variables analizadas fueron incidencia en 4 fechas de evaluación, las primeras dos evaluaciones son en estadios prefloración (INC1\_MV y INC2\_MV), la tercera fecha es en plena floración (INC3\_MV) y la cuarta fecha es a final de floración y principios de llenado (INC4\_MV). También se incluyeron el área por debajo de la curva de progreso de la enfermedad para incidencia (AUDPC\_INC\_MV) y para intensidad (es una media ponderada a nivel parcela en función de las frecuencias registradas para cada nivel de severidad) (AUDPC\_INT\_MV) y la intensidad proporcional en la fecha de evaluación 4 (IP\_MV). Se incluyó también en el análisis los días a floración (DAF). A fin de definir si la información que aportan las variables citadas anteriormente podía resumirse efectivamente en un menor número de variables a través del ACP se analizaron las correlaciones de Pearson entre las variables. Con esos datos se realizó un análisis de componentes principales (ACP) usando el paquete FactoMineR de R (Lê, S., Josse, J. & Husson, 2008) y el biplot fue obtenido con el paquete factoextra de R (Kassambara & Mundt, 2020).

Finalmente, los genotipos fundadores fueron seleccionados a partir de un ACP usando las medias ajustadas de las variables fenotípicas analizadas para MV y PHC obtenidas en estudios previos del grupo (Filippi et al., 2015; 2017; 2020; Fusari et al., 2012; Montecchia, 2019; Zubrzycki et al., 2017). La matriz de datos utilizada se encuentra en el Anexo digital "Datos\_fenotípicos\_PMA".

Se tuvo en cuenta el nivel de diversidad y la distancia genética estimada en las líneas que conforman la población de mapeo de asociación de INTA. Las medidas de distancia genética

entre las 159 líneas que conforman la PMA fueron estimadas utilizando la proporción de alelos compartidos empleando el software GenAlEx (Peakall & Smouse, 2012). La matriz de datos genotípicos utilizada para la estimación de la distancia entre accesiones está compuesta por 14 marcadores EST-SSR (SSR derivados de secuencias que se expresan o *expressed sequence tag*) seleccionados por Filippi et al (2015) a partir de los trabajos publicados por (Chapman et al., 2008; Mandel et al., 2011) y que poseen información alélica para todas las líneas que componen la PMA. Si bien en el grupo de Genómica de Girasol de INTA se dispone de un mayor número de marcadores para la PMA, se utilizó este número reducido con el objetivo de incluir en el análisis un grupo de líneas que no pudo ser genotificado con un mayor número de marcadores debido a una baja eficiencia en la digestión del ADN con las enzimas de corte poco frecuente utilizadas en el protocolo original (Filippi et al., 2020). En la Tabla 1.1 se presenta la información de los marcadores utilizados.

Se construyó un árbol *neighbor-joining* en base a las distancias genéticas calculadas entre pares de accesiones. Los análisis de agrupamiento y re-muestreos (1000 pseudo-réplicas) fueron realizados utilizando el software MEGA X (Kumar et al., 2018). El árbol resultante fue editado y visualizado con el software iTOL (Letunic & Bork, 2019).

**Tabla 1.1.** Marcadores EST-SSR utilizados para construir la matriz de distancias genéticas de la PMA. GL: Grupo de ligamiento.

Alias	GL	Motivo	Mín.	Máx.	Iniciador Forward	Iniciador reverse
HT0974	1	GT	173	195	HEX-AAAGGCGATTGAGTTTTTCGAT	GCAAGAATGAAAGGCCAACTTTA
HT0761	2	TG	126	128	NED-GGTGGGTGTTTGATTTTGTGTG	TCTCCTCTCTCCCTTTGTCTGC
HT0572	2	TA	300	321	FAM-CACGCCATATGAAAGCAACTCT	TCATGGATGTTTCATCACCAGAC
HT0292	3	TC	311	334	HEX-ACTGTCAACACCTCCTTCGACT	GGCTTACACTTCTCTCCATCTCAA
HT0734	3	TA	294	300	NED-GTTTCGCAGCTCTAGCATTGA	CCCCCTACAAGGCAACAAAATA
HT1045	4	NA	312	325	NED-GGATGAGAGACCGGAGAAATA	AAAGGAAGTATTGATTTATGTATGGA
HT0375	4	TA	398	418	NED-GCTGAAACGGGTCAATAAAGTC	TGAAGAAATGCTTCCAATCTGA
HT0440	5	AGG	236	248	NED-GAAGTTGGGAGGGTTGTTCAAG	CCTCCTGTTGGAACACCAAAT
HT0519	8	AT	252	274	HEX-TTTCGTGACCGAGAAAGGTAT	TGTGGCATACATAGAAATGATCTAA
c5783	11	CTA	204	207	FAM-GCCATCACAACATCAAATGG	AAGCTTGCACGTCCATACAC
HT0426	11	TA	312	324	HEX-AACCGTAAATGAAATCGGTGTG	GAGGGCAAAGTTGGGATACTCT
HT0271	13	TA	242	257	HEX-CCCATTGTTGCCTAATTCAAGAT	TTGGAGTGCATAACACGTATCAG
c2070	16	ATGT	235	261	FAM-GGGTAATGCAAAGTACTAAGATGTG	GCATCATCCAACAACTAGAAGG
HT0723	16	CAA/CCG	279	303	FAM-TCAAGCAATCAGACACCACATC	TGAACACAACCAAGAAATCCAA

*Mín:* Mínimo tamaño esperado. *Máx:* Máximo valor esperado.

### **Construcción de las poblaciones multiparentales**

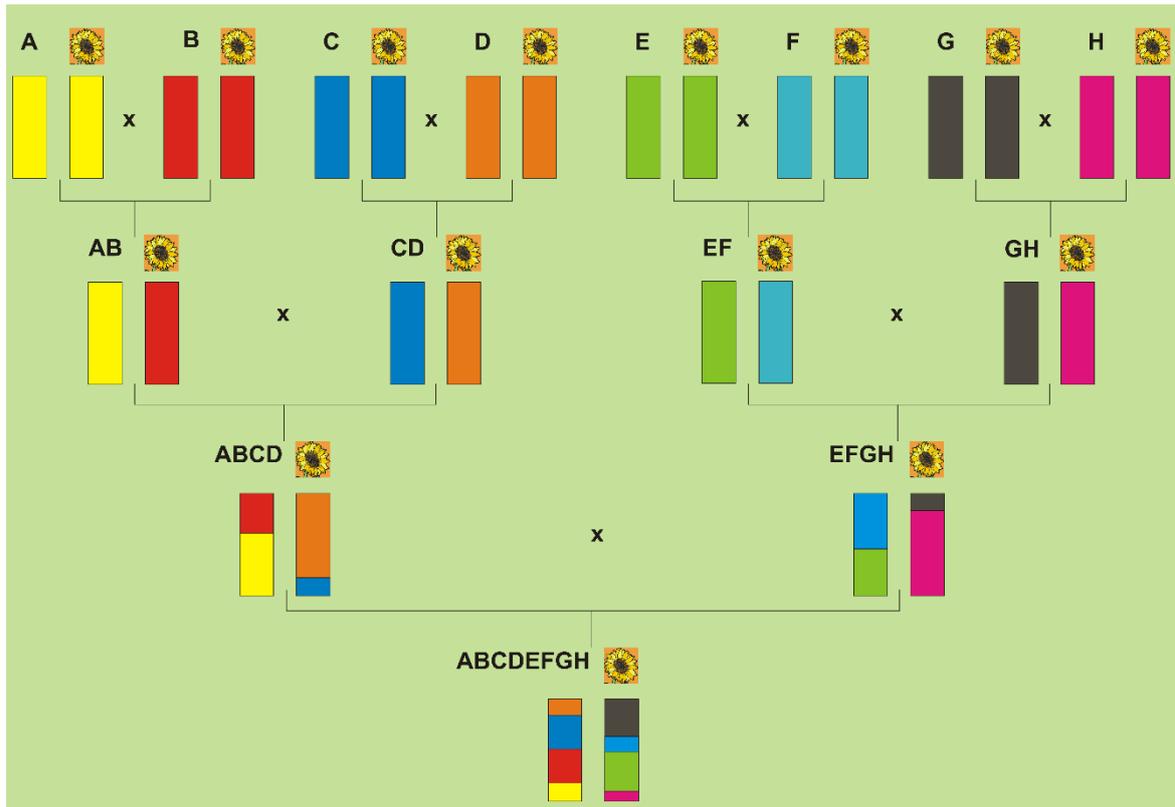
La construcción de la PMP se inició con líneas de citoplasma Normal (B) para posteriormente mediante endocría y conversión poder generar líneas androestériles (A) que actúen como parentales femeninos en el desarrollo de híbridos de alta productividad con los atributos mencionados anteriormente.

Para iniciar las poblaciones multiparentales se realizaron cruzamientos de las líneas homocigotas A; B; C; D; E; F; G; H, seleccionadas en base a un esquema de cruzamientos simples replicados para la generación de dialelos incompletos que confluyan en único embudo (Ladejobi et al., 2016).

El primer ciclo de cruzamientos fue realizado en la campaña 2017/18 en el campo experimental de la EEA INTA Pergamino (33° 57' 02" S, 60° 34' 08" O). En esta etapa se realizaron los cruzamientos entre las líneas endocriadas seleccionadas para obtener los híbridos simples. Para la población MAGIC1 los cruzamientos fueron: (2071/GP 06/04), (GP 03/03/B10), (GP 08/05/GP 07/15) y (AO 01/04/B477-1). Para la población MAGIC2 los cruzamientos realizados fueron: (2071/HA853), (GP 03/03/B481-6), (GP 08/05/GP 07/15) y (GP 14/06/B10). Se realizaron 4 réplicas por cruzamiento para cada población obteniéndose un total de 16 plantas cruzadas por población.

El segundo ciclo de cruzamientos fue realizado en 2018 en contra-estación en el campo de invierno de la EECT INTA Yuto (23°35' 1" S, 64° 30' 25" O), en esta etapa se efectuaron dos cruzamientos de cuatro vías. Para la población MAGIC1 los cruzamientos fueron [(2071/GP 06/04)//(GP 03/03/B10)] y [(GP 08/05/GP 07/15)//(AO 01/04/B477-1)]. Para cada cruzamiento se cruzaron 42 plantas obteniéndose un total de 84 réplicas. Para la población MAGIC2 los cruzamientos fueron [(2071/HA853)//(GP 03/03/B481-6)] y [(GP 08/05/GP 07/15)//(GP 14/06/B10)]. Para cada cruzamiento se cruzaron 38 plantas obteniéndose un total de 76 réplicas.

En la campaña 2018/19 en la EEA INTA Pergamino se continuó con el tercer ciclo de cruzamientos de ocho vías (A·B·C·D) x (E·F·G·H) maximizando el número de réplicas de este cruzamiento para evitar la pérdida de alelos por muestreo. Para la población MAGIC1 el cruzamiento fue [(2071/GP 06/04)//(GP 03/03/B10)//(GP 08/05/GP 07/15)//(AO 01/04/B477-1)] y su recíproco lográndose un total de 135 réplicas. Para la población MAGIC2 el cruzamiento fue [(2071/HA853)//(GP 03/03/B481-6)//(GP 08/05/GP 07/15)//(GP 14/06/B10)] y su recíproco lográndose un total de 171 réplicas. En la Figura 1.1 se representa el esquema de cruzamientos realizados para el desarrollo de las poblaciones MAGIC.



**Figura 1.1.** Esquema de cruces para la formación de la PMP.

La esterilización masculina de las plantas a ser utilizadas como parental femenino en cada cruzamiento fueron realizadas por emasculación manual mayoritariamente y en menor medida a través de la aplicación de ácido giberélico ( $GA_3$ ) en el estado de botón floral no mayor a 2 cm de diámetro siguiendo el protocolo propuesto por (Miller & Fick, 1978) (Figura 1.2). La emasculación se realizó en un lapso muy corto de tiempo entre las 6 AM y 9 AM aproximadamente, para evitar la apertura de las anteras y la contaminación con polen de la planta.



**Figura 1.2.** Métodos de emasculación utilizados en el desarrollo de las PMP. A: Emasculación manual. B: Emasculación química con aplicación de  $GA_3$ .

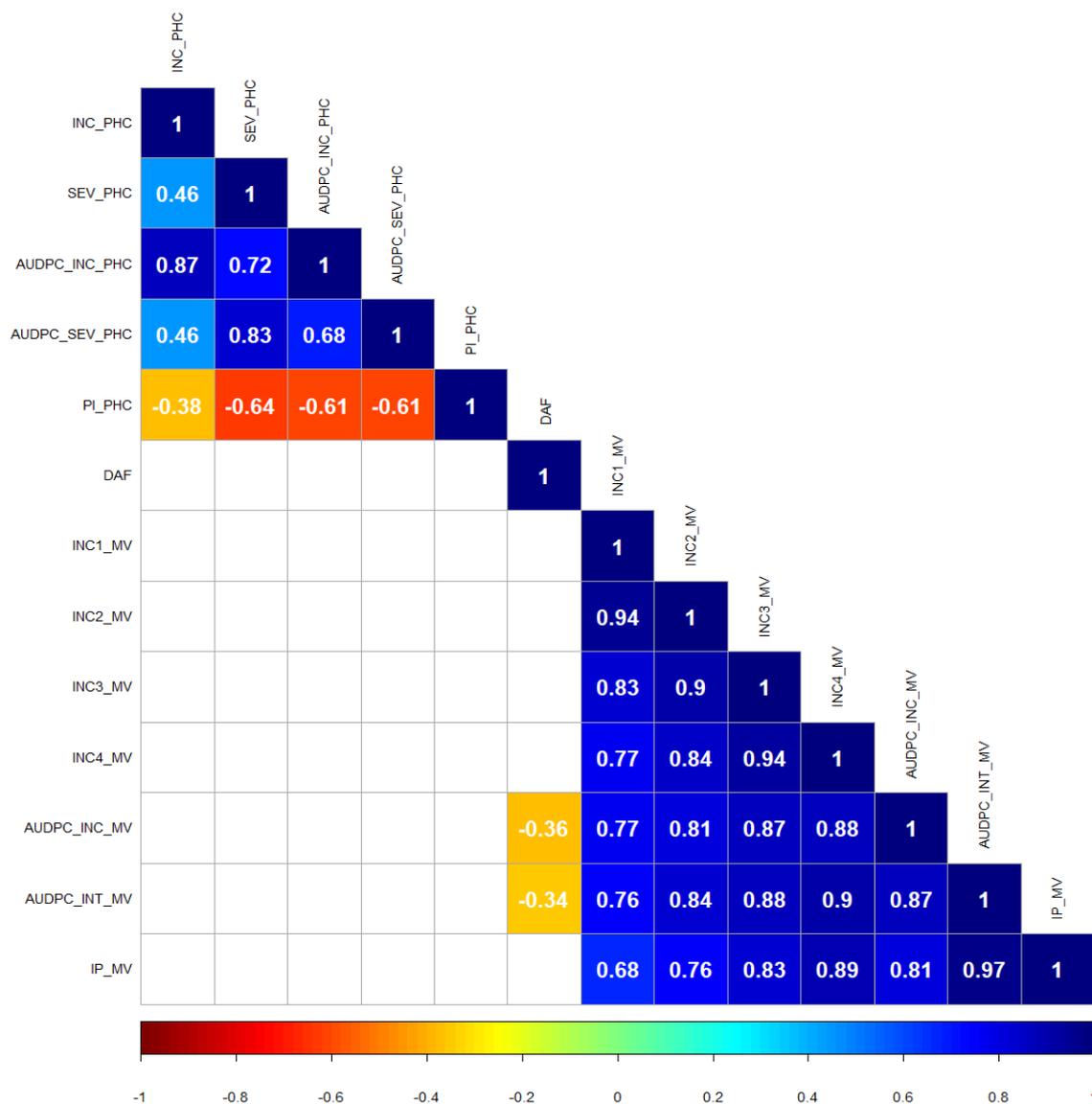
Los cruzamientos fueron realizados entre las plantas previamente esterilizadas y las plantas dadoras de polen. Cada inflorescencia fue tapada con bolsas de polinización para evitar la contaminación por insectos polinizadores antes de la apertura de la inflorescencia (R4) (Schneiter & Miller, 1981). En el caso de que las plantas a cruzar estuvieran ubicadas en surcos contiguos los cruzamientos se realizaron frotando ambos capítulos. Para aquellas plantas que estuvieran a una distancia que no permitiera el método anterior se optó por recolectar el polen de la planta macho en la cara abaxial de una hoja de la misma y llevarlo cuidadosamente hasta la planta hembra seleccionada para el cruzamiento.

En la campaña 2019/20 en la EEA INTA Pergamino se comenzó con el primer ciclo de autofecundaciones para la formación de las familias  $F_2$  utilizando el método de descendencia por *pedigree*. Por cada híbrido de 8 vías obtenido en el tercer ciclo de cruzamientos se sembró en esta campaña un surco y se autofecundaron 3 plantas por cada surco. Las familias  $F_2$  obtenidas en la campaña 2019/20 fueron en la campaña 2020/21 autofecundadas nuevamente, el número de plantas autofecundadas fue de 3 por familia. Las poblaciones PMP actualmente se encuentran en estado  $F_3$ .

## RESULTADOS

### Selección de líneas parentales. Correlación entre las variables fenotípicas de resistencia a enfermedades fúngicas en líneas mantenedoras

Las correlaciones de las medias ajustadas para las 79 líneas mantenedoras pertenecientes a la PMA y consideradas para la selección de las líneas parentales de las PMP se muestran en la Figura 1.3.



**Figura 1.3.** Correlaciones de Pearson entre las medias ajustadas de las líneas mantenedoras para las variables fenotípicas consideradas. Los cuadros vacíos corresponden a correlaciones que no fueron significativas ( $p$ -valor < 0.01).

Como puede observarse en la Figura 1.3, las correlaciones entre las variables que caracterizan la respuesta a MV fueron positivas, significativas y en la mayoría de los casos

superiores al 0,70. Las correlaciones entre las variables AUDPC\_INC\_MV y AUDPC\_INT\_MV fueron negativas con DAF. Las correlaciones entre las variables de PHC también fueron significativas, siendo altas y positivas para las variables INC, SEV, AUDPC\_INC y AUDPC\_SEV, y negativas para la variable PI\_PHC y las variables anteriores.

### Análisis de Componentes Principales

Con el propósito de resumir e integrar la información de todas las variables consideradas se realizó un ACP utilizando las funciones descritas en Materiales y Métodos. El total de la varianza multivariada fue explicado con 13 Componentes Principales (CPs), alcanzando explicar más del 90 % de la varianza multivariada total con sólo cinco componentes (Tabla 1.2).

**Tabla 1.2.** Varianza explicada por los primeros 5 componentes principales obtenidos del ACP.

Componentes principales	Autovalor	% Varianza explicada	% Varianza explicada acumulada
CP1	6,24	48,04	48,04
CP2	3,46	26,62	74,66
CP3	1,02	7,9	82,56
CP4	0,69	5,34	87,9
CP5	0,46	3,57	91,46

Se seleccionaron los primeros dos ya que conjuntamente explican el 74,66 % de la varianza multivariada total, alcanzando el objetivo de resumir la información disponible en un número acotado de variables. El CP1 explica el 48,04 % y el CP2 el 26,62% de la varianza total. Las contribuciones de las variables a los dos primeros CPs seleccionados se detallan en la Tabla 1.3.

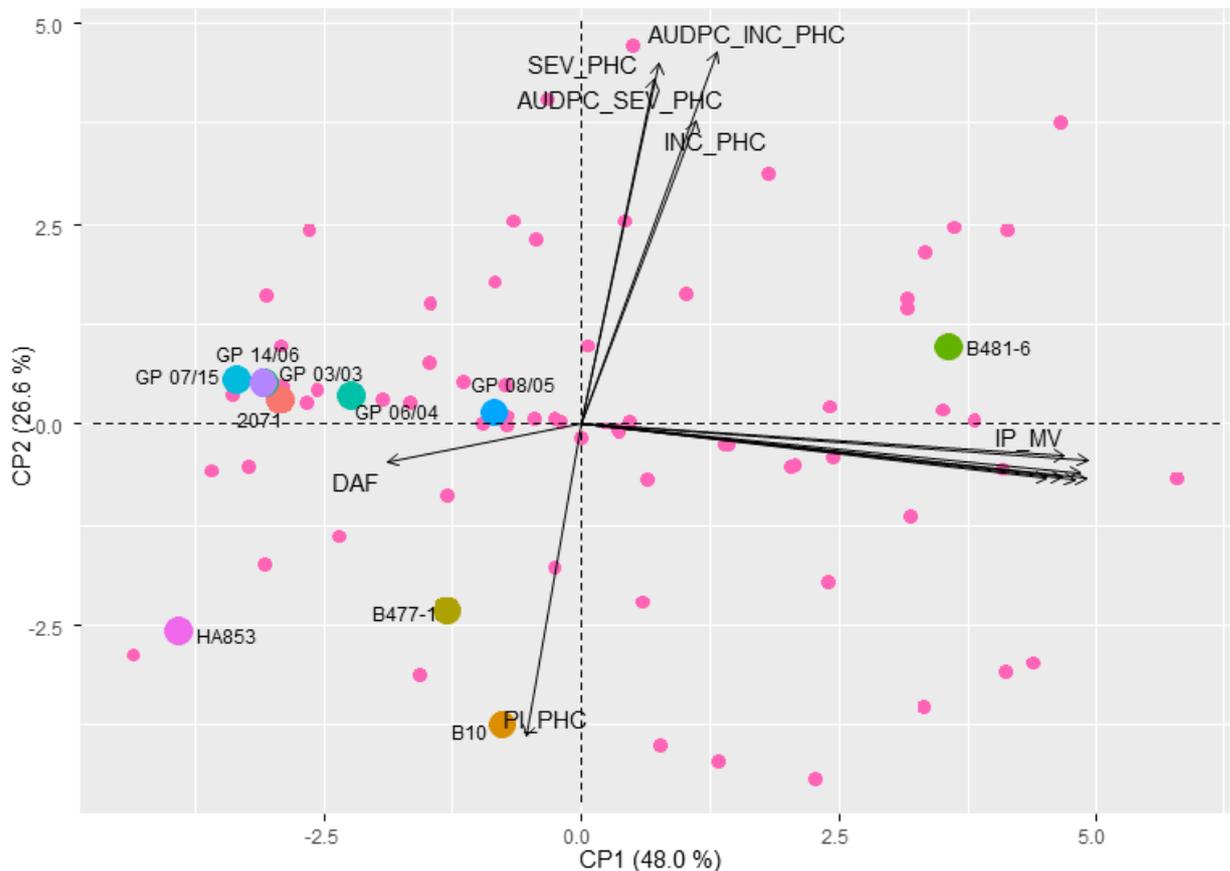
**Tabla 1.3.** Contribuciones de las variables a los dos primeros CPs seleccionados.

Variable	CP1	CP2
INC_PHC	0,75	15,49
SEV_PHC	0,34	21,84
AUDPC_INC_PHC	1,03	23,27
AUDPC_SEV_PHC	0,3	20,08
PI_PHC	0,17	16,36
DAF	2,12	0,22
INC1_MV	12,22	0,48
INC2_MV	13,79	0,51
INC3_MV	14,43	0,5
INC4_MV	14,08	0,39
AUDPC_INC_MV	13,07	0,47
AUDPC_INT_MV	14,49	0,21

IP\_MV | 13,22 0,16

Las variables que caracterizan la respuesta a MV son las que más contribuyeron al CP1, en cambio las variables que caracterizan la respuesta a PHC tuvieron mayor contribución al CP2.

En la Figura 1.4 se presenta el biplot de los dos primeros CPs utilizados para evaluar el comportamiento de las líneas a seleccionar frente a las enfermedades PHC y MV para la formación de la PMP. Los vectores de las variables indican hacia donde se incrementa el valor de las mismas. La ortogonalidad entre vectores implica mayor independencia entre variables. La proyección de los mismos sobre los ejes indica la proporción de correlación con cada CP. El CP1 se define principalmente por las variables de MV, por lo que los genotipos que se posicionaron a la derecha del biplot son los más susceptibles a MV y los que se ubicaron a la izquierda son los más resistentes.



**Figura 1.4.** Biplot conformado por las dos primeras componentes principales. Las líneas seleccionadas se diferencian del resto por su color y etiqueta. Los demás puntos representan las líneas mantenedoras no seleccionadas y los vectores las variables.

El CP2 en cambio quedó definido por las variables que caracterizan la PHC. Los genotipos con mayor resistencia a la enfermedad se ubicaron en la parte inferior del biplot y los más

susceptibles en la parte superior del mismo. El gráfico también muestra una correlación negativa entre la variable DAF y las variables de MV, lo cual indica que aquellos genotipos de ciclo más largo tendrán una mayor resistencia a la enfermedad.

#### **Diversidad genética de las líneas consideradas para la selección**

Para evaluar la diversidad genética de las líneas se construyó una matriz de distancia basada en el porcentaje de alelos compartidos para todos los pares de líneas que conforman la PMA usando la matriz de datos genotípicos construída a partir de 14 marcadores EST-SSR (Filippi et al., 2015). La misma se encuentra en el Anexo digital “EST-SSR”.

Las relaciones entre las 159 líneas endocriadas que conforman la PMA incluyendo líneas mantenedoras y restauradoras de la fertilidad fueron graficadas utilizando el método de *Neighbor-joining* (Figura 1.5).



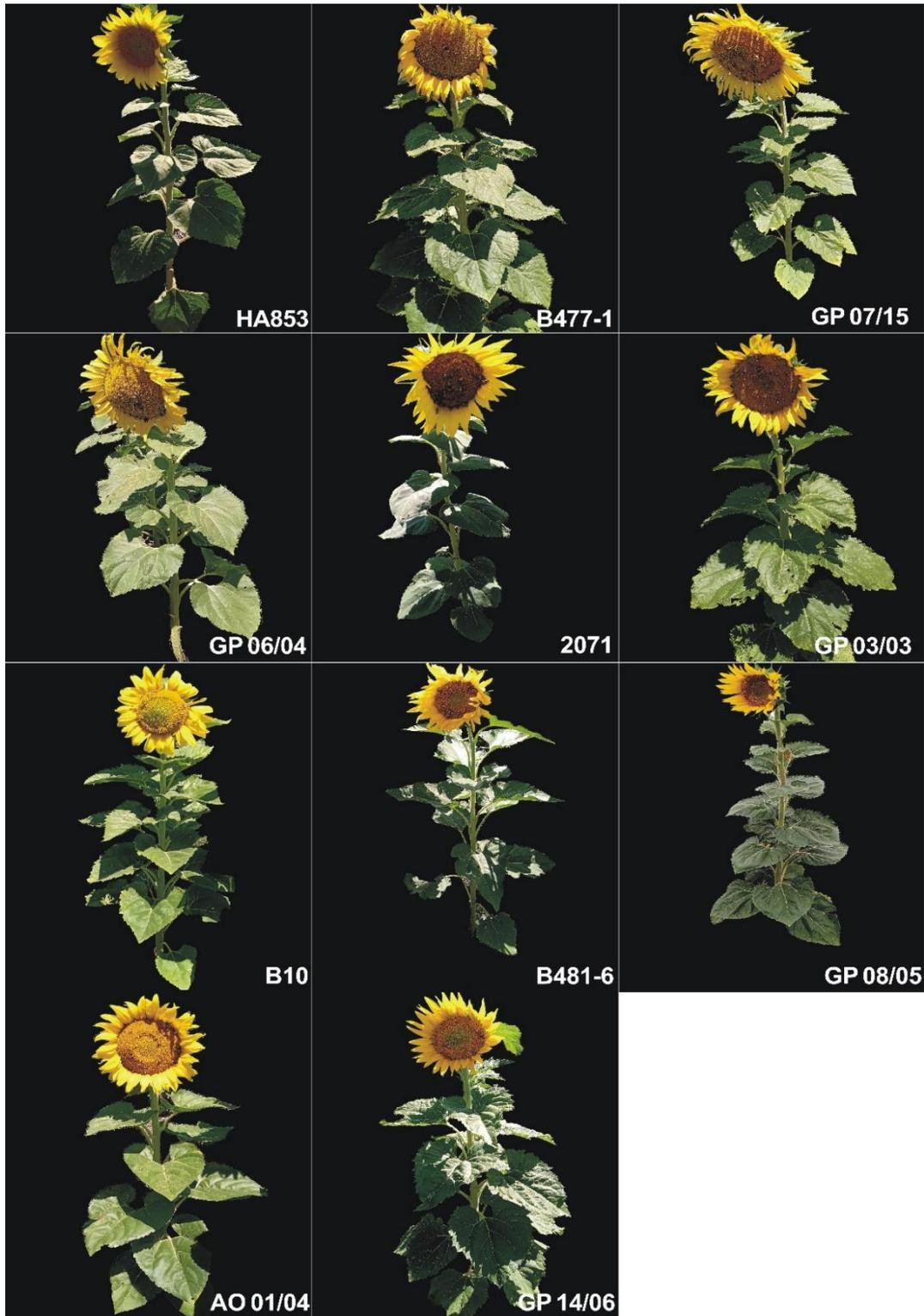
En las Figuras 1.6 y 1.7 se muestran imágenes de las líneas parentales seleccionadas. En la Figura 1.6 se destacan la variabilidad de las líneas parentales seleccionadas en cuanto a su arquitectura de planta, presentando diferentes ángulos de inclinación del capítulo y niveles de inserción del pecíolo con respecto al ápice de la lámina de la hoja. Asimismo, se notan diferencias en la morfología de las hojas, láminas triangulares y cordadas con bordes aserrados nulos hasta bordes con aserrado marcado. En la Figura 1.7 se representan las diferencias de las líneas seleccionadas en cuanto a la morfología de los aquenios. Pueden identificarse aquenios de color blanco marcado, aquenios con estrías blancas, grises y marrones, como aquenios de color negro oscuro. También, se notan diferencias en el tamaño de los aquenios y la forma

**Tabla 1.4.** Características de las líneas parentales seleccionadas.

<b>Línea</b>	<b>Pedigree</b>	<b>Características</b>	<b>DAF</b>	<b>% Aceite</b>
HA853	Derivada del compuesto de alto rendimiento 1975	Resistente a MV, PHC y EH	64	50,8
B477-1	Derivada de BHA64/B71_1	Resistente a MV, PHC y tolerante a EH	66	50,6
GP 07/15	SB-04/01/BZ-3669-2-5-3-2-2-1-1-2-4 SB	Resistente a MV	67	44,5
GP 06/04	LUCH-1-1-1-1-1/AO-01/03 -4-1-1	Moderadamente resistente a MV, resistente a PHC y alto contenido de ácido oleico	68	44,7
2071	Derivada del cruzamiento HA335/A69	Resistente a MV, DM y roya negra	67	46,3
GP 03/03	AxB-3479-2-2-3/BxC-3495-2-1-1-1-3-3	Moderadamente resistente a MV, PHC y vuelco	70	52,1
B10	Derivada del cruzamiento PxS81-162	Resistente a PHC y buena aptitud combinatoria general	70	41,1
B481-6	Derivada del cruzamiento B71/R423_6	Retrasa senescencia y tolerante a EH	68	52,0
GP 08/05	Derivada del cruzamiento de 2341/23	Moderadamente resistente a MV y a PHC y resistente a DM	66	46,3
AO 01/04	ND-01-1397-3-1-3-4 compuesto ND 01	Moderadamente susceptible a MV, moderadamente resistente a PHC y alto contenido de ácido oleico	65	50,9
GP 14/06	2309-1-5-1-3-3-3/AO-03/15-2-4-2-3-3	Resistente a DM (raza 710) y susceptible a MV	60	47,8

**Tabla 1.5.** Categorías fenotípicas de las líneas parentales seleccionadas frente a MV y PHC.

<b>Línea</b>	<b>MV</b>	<b>PHC</b>
HA853	Resistente	Resistente
B477-1	Resistente	Resistente
GP 07/15	Resistente	Mod. susceptible
GP 06/04	Mod. resistente	Resistente
2071	Resistente	Mod. susceptible
GP 03/03	Mod. resistente	Mod. resistente
B10	Mod. resistente	Resistente
B481-6	Mod. resistente	Susceptible
GP 08/05	Mod. resistente	Mod. resistente
AO 01/04	Mod. susceptible	Mod. resistente
GP 14/06	Susceptible	Mod. susceptible



**Figura 1.6.** Líneas parentales fundadoras de las poblaciones MAGIC 1 y 2.



**Figura 1.7.** Imágenes de los aquenios de cada una de las líneas parentales.

### **Construcción de las poblaciones multiparentales**

Como se describió en la sección anterior, los cruzamientos entre las líneas parentales fueron realizados en un único embudo. Las líneas parentales fueron cruzadas para crear cuatro

híbridos simples diferentes para cada población, obteniéndose un total de 4 réplicas por cruzamiento (Tabla 1.6).

**Tabla 1.6.** Cruzamientos de 2 vías realizados en la primera etapa de desarrollo de las poblaciones MAGIC.

Población	Cruzamientos de 2 vías	N° Plantas cruzadas	Total
MAGIC1	2071/GP 06/04	4	16
	GP 03/03/B10	4	
	GP 08/05/GP 07/15	4	
	AO 01/04/B477-1	4	
MAGIC2	2071/HA853	4	16
	GP 03/03/B481-6	4	
	GP 08/05/GP 07/15	4	
	GP 14/06/B10	4	

En la segunda etapa de cruzamientos, se realizaron cruzamientos para obtener dos híbridos de 4 vías, para la población MAGIC1 se obtuvieron 42 réplicas por cruzamiento y para la población MAGIC2 se obtuvieron 38 réplicas por cruzamiento (Tabla 1.7).

**Tabla 1.7.** Cruzamientos de 4 vías realizados en la segunda etapa de desarrollo de las poblaciones MAGIC.

Población	Cruzamientos de 4 vías	N° réplicas	Total
MAGIC1	(2071/GP 06/04)//(GP 03/03/B10)	42	84
	(GP 08/05/GP 07/15)//(AO 01/04/B477-1)	42	
MAGIC2	(2071/HA853)//(GP 03/03/B481-6)	38	76
	(GP 08/05/GP 07/15)//(GP 14/06/B10)	38	

En la tercera etapa de cruzamientos, se obtuvo un híbrido de 8 vías por población que fue cruzado recíprocamente (ambos híbridos de 4 vías actuando como parentales femeninos y masculinos). Este cruzamiento de 8 vías fue replicado 135 veces para la población MAGIC1 y 171 veces para la población MAGIC2 (Tabla 1.8).

**Tabla 1.8.** Cruzamientos de 8 vías realizados en la tercera etapa de desarrollo de las poblaciones MAGIC.

Población	Cruzamiento de 8 vías	N° replicas	Total
MAGIC1	[(2071/GP 06/04)//(GP 03/03/B10)]//[[(GP 08/05/GP 07/15)//(AO 01/04/B477-1)]]	74	135
	[(GP 08/05/GP 07/15)//(AO 01/04/B477-1)]//[[(2071/GP 06/04)//(GP 03/03/B10)]]	61	
MAGIC2	[(2071/HA853)//(GP 03/03/B481-6)]//[[(GP 08/05/GP 07/15)//(GP 14/06/B10)]]	100	171
	[(GP 08/05/GP 07/15)//(GP 14/06/B10)]//[[(2071/HA853)//(GP 03/03/B481-6)]]	71	

En la campaña 2019/20 se completó el primer ciclo de autofecundaciones y formación de familias  $F_2$ , se obtuvieron 382 y 470 familias para las poblaciones MAGIC1 y MAGIC2 respectivamente. En la campaña 2020/21 se completó el segundo ciclo de autofecundaciones avanzando ambas poblaciones al estado de  $F_3$ , se obtuvieron 1161 y 1497 familias para las poblaciones MAGIC1 y MAGIC2. En la Figura 1.8 se muestra la conducción de las autofecundaciones en la última campaña.



**Figura 1.8.** Imagen aérea tomada con dron del campo experimental de la EEA Pergamino donde se realizó la segunda generación de autofecundaciones de las poblaciones MAGIC1 y MAGIC2.

## DISCUSIÓN

La selección criteriosa de las líneas parentales es muy importante para definir el potencial que tendrá la población multiparental que se pretende desarrollar. Existe una gran variedad de criterios y estrategias publicadas para llevar a cabo este proceso. Como se mencionó anteriormente, se reportaron poblaciones que priorizan la diversidad de los parentales incluyendo accesiones de origen silvestre o *landraces* (Campanelli et al., 2019; Pascual et al., 2015), en detrimento de la utilidad que puedan tener las progenies para su incorporación en programas de mejoramiento genético. En cambio, en otras poblaciones publicadas se priorizaron la inclusión de líneas parentales pertenecientes a programas de mejoramiento o variedades cultivadas (Dell'Acqua et al., 2015; Ongom & Ejeta, 2018) resignando diversidad y posibles nuevas combinaciones alélicas, pero incrementando la potencialidad de la población para su uso directo como nuevas fuentes de germoplasma para el mejoramiento.

Las poblaciones que se presentan en esta Tesis fueron construidas con el objetivo de complementar, ampliar y potenciar los recursos genéticos del programa de mejoramiento de girasol de INTA. La selección de las líneas parentales fue realizada sobre un conjunto de accesiones pertenecientes a la PMA, poblaciones biparentales y líneas endocriadas desarrolladas por los programas de mejoramiento de girasol en sus unidades en Manfredi y Pergamino. La PMA fue conformada a partir de la selección de 164 líneas endocriadas albergadas en el BAG-IM y desarrolladas por los programas de mejoramiento de girasol de INTA en las EEA Pergamino y Manfredi. La PMA es esencialmente una colección núcleo (Brown, 1989), con el objetivo de reunir la mayor variabilidad posible de los recursos genéticos de girasol conservados y desarrollados por INTA, en un número reducido de accesiones, permitiendo una mayor practicidad para estudios genéticos. Para la construcción de la PMA, se incluyeron principalmente líneas endocriadas de origen argentino, provenientes del programa de mejoramiento de INTA para diferentes caracteres de interés, entre ellos resistencia a estreses bióticos y abióticos y líneas públicas de origen internacional con el objetivo de maximizar la diversidad genética de la misma. Esta población en un estudio realizado por Filippi et al (2020), fue comparada con otras poblaciones internacionales de girasol, como la UBC-USDA y la de INRA. Los autores revelaron diferencias genéticas significativas entre las poblaciones, aunque éstas sólo explicaron una pequeña proporción de la variación genética total. Además, identificaron patrones similares de desequilibrio de ligamiento, por lo que sus resultados mostraron cierta homogeneidad entre las tres poblaciones internacionales analizadas y una estrecha base genética del mejoramiento actual de girasol.

El perfil de las líneas seleccionadas para la construcción de la PMP fue de líneas con comportamientos diferenciales frente a distintos atributos de interés para el mejoramiento de girasol. Se priorizó la elección de materiales caracterizados en estudios previos del grupo frente a las enfermedades marchitez prematura causada por *V. dahliae* y podredumbre húmeda del capítulo por *S. sclerotiorum*. También, se tuvo en consideración otros caracteres de interés como la resistencia a downy mildew provocado por *P. halstedii*, el contenido de aceite, la composición de ácidos grasos, la senescencia, el ciclo y la habilidad combinatoria.

Si bien las poblaciones en desarrollo no incluyen materiales silvestres o *landraces* que incrementen la diversidad como se mencionó anteriormente, la selección de los genotipos fundadores fue realizada considerando lograr la mayor diversidad posible y tratando de obtener una elevada representación de la variabilidad genética del germoplasma de INTA en los parentales. La inclusión en este caso de líneas con un fondo genético más estrecho, adaptado a las condiciones locales del área girasolera argentina y con atributos deseables para distintos caracteres permitirá la implementación directa de las progenies de estas poblaciones en los programas de mejoramiento, casos similares que implementaron materiales de pre-mejoramiento o *elite* fueron reportados en poblaciones MAGIC de arroz por Descalsota et al., (2018) y Zaw et al., (2019) en las cuales las MAGIC *RILS* fueron empleadas directamente en los programas de mejoramiento de arroz.

Las líneas parentales seleccionadas son líneas de citoplasma normal, mantenedoras de la fertilidad, presentan ciclos de siembra a floración que varían desde los 60 hasta los 70 días, varían en morfología y arquitectura foliar con diferentes ángulos de inclinación de las hojas, en la morfología de los aquenios, incluyéndose líneas con aquenios de pericarpio blanco, y con estrías de distintas tonalidades de grises y marrones. Las líneas también presentan distintos grados de inclinación y curvatura del capítulo. Aportan diversos grados de resistencia frente a las enfermedades MV y a PHC (Filippi et al., 2017; Montecchia et al., 2021), resistencia a downy mildew, resistencia a estreses abióticos como la sequía (Moreno, 2010) y el vuelco. También, fueron incluidas líneas con respuesta contrastante a la senescencia foliar (Moschen et al., 2019). El contenido de aceite de las líneas varía entre el 41 al 52 % y fueron incluidas líneas de alto contenido de ácido oleico. El origen de las líneas proviene de materiales desarrollados por INTA, a partir de diferentes compuestos formados por cruzamientos entre germoplasma adaptado a las condiciones locales y germoplasma proveniente de Rusia, Estados Unidos, Rumania y Francia principalmente.

Las estrategias de desarrollo de las poblaciones MAGIC resultan clave para maximizar el potencial de la población para los estudios de mapeo fino de QTLs. En esta etapa del proceso,

resulta desafiante y muy importante ser eficiente en la utilización de los recursos disponibles y el tiempo para llevar adelante el proyecto. La premisa de estas poblaciones es aumentar los niveles de diversidad genética con respecto a la estrecha variabilidad de las poblaciones biparentales clásicas. Por lo que resulta necesario lograr un balance entre los recursos disponibles para desarrollar la población y el potencial que tendrá la misma.

Dado el elevado número de combinaciones que pueden lograrse en función del número de parentales que inicie las poblaciones y considerando las complicaciones prácticas y de recursos para llevar adelante los múltiples cruzamientos, se han reportado poblaciones con diferentes estrategias que redujeron el número de embudos, disminuyeron el número de parentales que confluyen en cada embudo, realizaron cruzamientos libres entre los parentales seleccionados e implementaron el uso de líneas androestériles para reducir los esfuerzos en emasculación (Huang et al., 2011; Islam et al., 2016b; Ongom & Ejeta, 2018; Pascual et al., 2015; Sallam & Martsch, 2015; Stadlmeier et al., 2018).

La estrategia de desarrollo implementada en las poblaciones presentadas en esta Tesis redujo el número de embudos y el número de fundadores por embudos de manera similar a la población MAGIC de tomate publicada por Pascual et al., (2015) generando dialelos incompletos lo cual puede asimilarse al esquema básico de acuerdo a la definición dada por Yang et al., (2021). La elección de esta estrategia fue con el objetivo de reducir los costos y el tiempo que hubiera insumido la realización de todas las posibles combinaciones. A partir de su aplicación se lograron efectivamente los cruzamientos de dos, cuatro y ocho vías, obteniéndose un elevado número de cruzamientos exitosos en cada una de las etapas, logrando semillas viables que pudieron ser continuadas en el avance de generaciones de desarrollo de las poblaciones multiparentales. Asimismo, la elección de genotipos provenientes de programas de mejoramiento para la formación de las poblaciones, evita posibles incompatibilidades en los cruzamientos como si puede darse en poblaciones que incluyen genotipos silvestres dentro de sus parentales, permitiendo evitar una reducción en el número de descendientes de cruzamientos (Arrones, 2019; Huang et al., 2015). Las generaciones de cruzamientos que dan lugar a este tipo de poblaciones favorecen la recombinación de los haplotipos parentales y permitirán el mapeo preciso de QTLs (Diouf & Pascual, 2021). En estas etapas es fundamental maximizar el tamaño poblacional dentro de los recursos y límites operativos, de lo contrario, si en cualquier generación el número de individuos se reduce, los individuos derivados de ellos estarán relacionados entre sí, reduciendo la diversidad genética. Esta relación también puede crear una estructura poblacional que puede sesgar los análisis, afectando a la construcción de los mapas genéticos y al mapeo de QTLs (Huang et al., 2015).

Estudios de simulación realizados por Ladejobi et al., (2016) demostraron que la implementación de esquemas de cruzamientos que confluyen en único embudo redujeron el número de haplotipos recombinantes en comparación a esquemas de múltiples embudos, pero tuvieron valores de recombinación mayores que los de poblaciones biparentales y multiparentales de tipo NAM. Asimismo, la realización de un elevado número de réplicas de los cruzamientos de 4 y 8 vías resultan fundamentales para evitar la pérdida de alelos por muestreo y favorecer la captura de haplotipos recombinantes (Scott et al., 2020). En este trabajo, se introdujeron modificaciones al diseño de cruzamientos a fin de reforzar el número de recombinaciones y acortar el desequilibrio de ligamiento realizando cruzamientos recíprocos en la etapa de obtención de híbridos de ocho vías. Asimismo, el cruzamiento fue replicado 135 y 171 veces en las poblaciones MAGIC1 y MAGIC2 respectivamente, para favorecer el contenido de haplotipos recombinantes en la población final.

El proceso de obtención de las líneas homocigotas derivadas de estas poblaciones está siendo conducido a través de la endocría de los híbridos de 8 vías obtenidos. Esta metodología es muy frecuente en los programas de mejoramiento convencionales y particularmente en girasol, en el cual hasta el momento no se cuenta con una técnica de obtención de haploides duplicados efectiva. Las ventajas que brindaría poder contar con la técnica de haploides duplicados sería la reducción en el tiempo para obtener líneas recombinantes homocigotas y poder alcanzar la homocigosis total de las líneas. En cambio, el proceso de autofecundación conlleva un mínimo de 5 generaciones que extiende considerablemente el tiempo de desarrollo y no genera líneas completamente homocigotas como el sistema de haploides duplicados (Huang et al., 2015). Una de las principales ventajas de la endocría es la generación de recombinaciones adicionales debido a las sucesivas rondas de meiosis hasta alcanzar la homocigosis (Tripodi, 2021).

La formación de las poblaciones multiparentales descritas en este capítulo, desarrolladas a partir de 11 líneas fundadoras provenientes del programa de mejoramiento genético de girasol de INTA y caracterizadas previamente para distintos caracteres de interés, las convierte en recursos genéticos con alto potencial para descifrar la arquitectura genética de distintos caracteres de interés para el mejoramiento. Asimismo, serán fuentes de recursos genéticos de gran interés para su aplicación directa o indirecta en los programas de mejoramiento públicos y privados de girasol.

--

---

---

**CAPÍTULO 2: Caracterización genotípica de las líneas parentales y los híbridos formados en las primeras etapas de desarrollo de las poblaciones multiparentales**

---

---

## INTRODUCCIÓN

La caracterización genotípica de las líneas parentales es una actividad fundamental durante la formación y el desarrollo de las poblaciones multiparentales. Conocer la diversidad genética entre las líneas parentales seleccionadas es de gran interés para determinar el potencial de las poblaciones de mapeo. En este capítulo se describirán las metodologías implementadas para la caracterización genotípica de las líneas parentales y los híbridos de dos, cuatro y ocho vías producidos para la formación de las poblaciones. El estudio de la diversidad presente entre las líneas parentales. La estrategia utilizada para el análisis de parentesco entre los híbridos obtenidos en cada etapa y los parentales y la implementación de un análisis de reconstrucción de bloques haplotípicos ancestrales en las progenies obtenidas con el fin de evaluar esta técnica para su utilización futura en el mapeo de QTLs.

En girasol se han implementado diferentes tipos de marcadores moleculares, siendo los microsatélites (SSR) y los Polimorfismos de Nucleótido Simple (SNP, *single nucleotide polymorphisms*) los más reportados (Brunel, 1994; Chapman et al., 2008; Dehmer & Friedt, 1998; Filippi et al., 2020; Heesacker et al., 2008; Kiani et al., 2007; Montecchia, 2019; Paniego et al., 2002; Tang et al., 2002; Yu et al., 2003). Los avances en la automatización del genotipado de SNP y la genotipificación por secuenciación han marcado la preferencia de los marcadores SNPs por sobre otros tipos de marcadores moleculares (Dimitrijevic & Horn, 2018).

Un SNP se define como una diferencia a nivel de base nucleotídica entre dos secuencias de ADN. En teoría, las variantes de tipo SNP pueden involucrar las cuatro bases nitrogenadas posibles A, T, G o C, no obstante, se consideran dos alelos para un determinado locus, a causa de que la probabilidad de hallar dos mutaciones puntuales en una misma posición de un genoma de 3,6 Gb resulta extremadamente baja. De este modo, la información que aportan estos marcadores es de tipo bialélica. Esto, que en principio es una desventaja, se ve contrarrestado por la gran abundancia y estabilidad de los SNPs respecto de los SSRs (Oraguzie & Wilcox, 2007).

Diversos estudios han estimado la diversidad genética presente en distintas poblaciones usando SNPs en girasol (Bachlava et al., 2012; Cadic et al., 2013; Filippi et al., 2015; Fusari et al., 2008; Kolkman et al., 2007; Mandel et al., 2011). La frecuencia de polimorfismos de tipo SNP, su estabilidad en el genoma, y la versatilidad que ofrecen, para diseñar diferentes métodos automatizados para su detección a distintos niveles de procesividad, los convierten en marcadores ideales para estudios de análisis de diversidad genética, de asociación, de selección genómica, entre otros (He et al., 2014).

El grupo de girasol de INTA desarrolló un arreglo de SNP a partir de genes candidatos para tolerancia a estrés bióticos y abióticos que fue usado para saturar el mapa genético descrito por Talia et al (2010) y optimizar la búsqueda de QTLs para resistencia a podredumbre húmeda de capítulo producida por *S. sclerotiorum* en girasol (Zubrzycki et al., 2017). Asimismo, se aplicó a estudios de estructura genética y diversidad poblacional (Filippi et al. 2015), y mapeo de QTL mediante asociaciones genotipo-fenotipo (Filippi et al., 2020).

Las estrategias de genotipado masivo han incorporado los avances de la secuenciación de nueva generación para generar marcadores SNPs de amplia cobertura genómica aplicando la estrategia de reducción de complejidad del genoma en base a la restricción enzimática acoplada a la secuenciación masiva (Baird et al., 2008; Elshire et al., 2011; Peterson et al., 2012; Poland & Rife, 2012a). El genotipado por secuenciación (GBS, del inglés *Genotyping by Sequencing*) utiliza enzimas de restricción para reducir la complejidad del genoma en combinación con el uso de adaptadores identificados por “*barcodes*” (códigos de barras) para componer librerías multiplexadas de muestras listas para ser secuenciadas con tecnologías NGS. Esta metodología ha demostrado ser robusta a través de diversas especies y de ser capaz de producir grandes volúmenes de marcadores moleculares (Elshire et al., 2011; Poland & Rife, 2012a). La flexibilidad de aplicación de GBS en consideración de especies, poblaciones, y objetivos de investigación, la convierte en una herramienta ideal para estudios de genética vegetal (Poland & Rife, 2012a).

Una de las metodologías más difundidas es la de secuenciación de fragmentos obtenidos por doble digestión con enzimas de restricción (ddRADSeq) descrita por Peterson et al., (2012). Esta tecnología basada en la reducción de complejidad del genoma a partir de combinaciones de enzimas de restricción aporta ventajas vinculadas a la exploración de nuevas variantes alélicas y a la reducción de costos por muestra (Rasheed et al., 2017). La reducción del costo de secuenciación por muestra se debe a las posibilidades de multiplexado que permiten las combinaciones de adaptadores con *barcodes* e iniciadores indexados de PCR para secuenciación. Esta ventaja de las técnicas de secuenciación masiva sumada a la simplificación del genoma por selección de fragmentos de restricción, permiten dirigir la secuenciación a regiones aleatorias distribuidas regularmente a lo largo de todo el genoma y a través de distintos genotipos (Peterson et al., 2012). Además, estas herramientas permiten realizar diferentes estudios a nivel poblacional sin la necesidad de contar con información genómica previa de la especie, pudiendo ser aplicada en organismos no modelo, con un costo menor que empleando microarreglos ya que no necesita del desarrollo de un chip (Andrews et al., 2016; Davey et al., 2011).

Entre las desventajas de las metodologías de GBS, se encuentran los errores de genotipado por baja profundidad de lecturas, errores de identificación de heterocigotas, la pérdida de alelos por polimorfismos en los sitios de corte de las enzimas, entre otros. Estas limitantes resultan en un incremento de datos faltantes por SNP y desviaciones en la frecuencia de alelos por amplificación diferencial de algunas regiones del genoma durante el armado de librerías (Davey et al., 2011; He et al., 2014; Rasheed et al., 2017). Para resolver este inconveniente, es posible utilizar sólo los *loci* con un porcentaje mínimo de individuos con datos faltantes o bien, aumentar la profundidad de secuenciación y/o imputar los datos perdidos a partir de la matriz de datos completa obtenida complementada con la información de relación genética entre genotipos de la población (Andrews et al., 2016; Bajgain et al., 2016).

Gracias a los avances en las diversas tecnologías de secuenciación y más allá de las limitaciones anteriormente mencionadas, GBS se ha convertido en una de las técnicas más utilizadas en estudios genéticos en plantas y es ampliamente aplicada en el mejoramiento genético de cultivos. Además, la reciente disponibilidad de plataformas de secuenciación de moléculas únicas condujo al enriquecimiento de la información de genomas de referencia resultando en una mayor confiabilidad y practicidad de la implementación de éstas metodologías de genotipado (Rasheed et al., 2017).

### **Análisis de parentesco**

Las relaciones padre-progenie, ya sea entre individuos de poblaciones naturales *in situ* o de colecciones de recursos genéticos *ex situ*, son de interés fundamental para los mejoradores de plantas y animales, los ecólogos moleculares y los genetistas de poblaciones. El análisis de parentesco implica el análisis del ADN de la progenie y de los posibles progenitores o progenitores candidatos, comparando sus alelos para identificar y confirmar las relaciones existentes. En esta técnica, los perfiles de ADN de la progenie y de los progenitores maternos y paternos se verifican asignando cada alelo de la progenie analizada de acuerdo a los alelos de sus progenitores. Se basa en el principio de que los padres pueden asignarse a sus respectivas progenies con un cierto nivel de confianza, basándose en la significancia de la compatibilidad genética entre generaciones. Las leyes mendelianas de la herencia permiten inferir las relaciones genealógicas, siempre que se disponga de un conjunto suficientemente informativo de marcadores genéticos que se transmitan de forma estable de padres a hijos (Jones et al., 2010; Melo & Hale, 2019).

Históricamente, el análisis del parentesco en plantas se realizó mediante el análisis de caracteres fenotípicos discriminantes. Esta técnica tiene ciertas desventajas como ser

limitada, laboriosa, ciertos caracteres pueden verse influidos por el ambiente y carece de precisión cuando los fenotipos de los individuos a distinguir son muy similares. Estas limitaciones del análisis de parentesco fenotípico sumadas al avance de la biología molecular llevaron a la aparición de nuevas técnicas superiores de análisis de parentesco (Norman et al., 2018). Las primeras técnicas de análisis de parentesco utilizando aloenzimas fueron muy ineficientes. En este sentido, los marcadores moleculares tipo microsatélites primero y SNPs más recientemente, junto con paquete estadísticos robustos facilitaron su implementación en áreas de mejoramiento genético, evolución, ecología entre otras disciplinas (Jones et al., 2010; Jones & Ardren, 2003; Norman et al., 2018).

La teoría relativa al análisis de parentesco no ha cambiado sustancialmente en la última década, y los marcadores de nueva generación, no cambian las reglas en relación a las técnicas analíticas necesarias para un análisis de parentesco exitoso. Los principales métodos de análisis de parentesco se clasifican de acuerdo a su naturaleza en cuatro categorías: de exclusión, de asignación, de reconstrucción parental y de análisis bayesiano (Flanagan & Jones, 2018).

El análisis de parentesco utilizando marcadores moleculares está siendo ampliamente implementado en los programas de mejoramiento debido a sus contribuciones en el control parental, reduciendo los costos en tiempo y mantenimiento de parentales, en la estimación de los valores de mejora, en la detección de contaminaciones y errores en las labores fitotécnicas, entre otras (Norman et al., 2018). En el mejoramiento de especies de explotación forestal, dadas sus particulares características biológicas de polinización y los extensos períodos de cultivo, es donde el análisis de parentesco ha sido mayormente implementado y cumple un rol clave en los protocolos de trabajo de mejoramiento de diversas especies forestales (Wang et al., 2010).

La mayoría de las herramientas de análisis de parentesco disponibles actualmente se desarrollaron originalmente para datos SSR, siendo el software más empleado CERVUS (Kalinowski et al., 2007), utilizando bases de datos relativamente pequeñas (de docenas a cientos de puntos de datos) (Jones et al., 2010; Jones & Ardren, 2003). Con la creciente disponibilidad de recursos genómicos y métodos multiplexados para obtener numerosos SNPs simultáneamente, numerosos investigadores pasaron a utilizar SNPs para llevar a cabo el análisis de parentesco (Flanagan & Jones, 2018). La adopción de los polimorfismos de tipo SNP en algunas especies fue promovido por el bajo polimorfismo de los microsatélites (Cramer et al., 2011). No obstante, las ventajas que proporcionan estos marcadores, como una automatización y puntuación más fáciles (Anderson & Garza, 2006) y las tasas de

mutación más bajas (Amorim & Pereira, 2005; Fisher et al., 2009) favorecen su uso por sobre el de los microsatelites. La difusión de los SNP promovió el desarrollo de herramientas bioinformáticas como “Apparent”, “ParentOffspring”, “SOLOMON”, “Hsphase” entre otros, capaces de manipular el volumen de datos producidos por las nuevas plataformas de genotipado (Abdel-Haleem et al., 2013; Christie et al., 2013; Ferdosi et al., 2014; Hadfield et al., 2006; Melo & Hale, 2019). El éxito del análisis de parentesco depende del diseño de muestreo utilizado, la elección de los marcadores moleculares, la estructura familiar de los individuos a evaluar y la biología reproductiva de la especie (Ashley, 2010; Jones et al., 2010; Jones & Ardren, 2003; Norman et al., 2018).

### **Reconstrucción de bloques haplotípicos**

El genoma es un mosaico de bloques de haplotipos que capturan la historia evolutiva o de reproducción de un individuo. Reconstruir los mapas de haplotipos ancestrales de las poblaciones es importante para imputar regiones no tipificadas del genoma (Davies et al., 2016), mapear QTLs (Mott et al., 2000), analizar el paisaje de recombinación (Morgan et al., 2017) e inferir la historia evolutiva y la estructura de los haplotipos (Aylor et al., 2011; Gabriel et al., 2002). Asimismo, la reconstrucción de los bloques haplotípicos descendientes de cromosomas homólogos, puede ayudar a caracterizar el patrón de recombinación, a dilucidar las relaciones genotipo-fenotipo y a mejorar las predicciones genómicas (Manching & Wisser, 2021).

Los individuos que componen las poblaciones multiparentales, dadas las sucesivas generaciones de cruzamientos, presentan genomas que son un mosaico aleatorio de bloques haplotípicos, cada uno de ellos heredado alternativamente de un fundador homocigota. La reconstrucción de estos bloques de ancestría es un paso necesario para el mapeo posterior de QTL (Zheng, 2015a). Cuando hay múltiples haplotipos y/o fundadores segregando en las poblaciones de mapeo, los marcadores bialélicos ya no son completamente informativos como en las poblaciones biparentales. En estos casos se utiliza el contexto haplotípico dado por los marcadores circundantes para inferir el fundador que contribuyó con cada locus genómico dentro de cada RIL (Mott et al., 2000). Los enfoques basados en el *pedigree*, como la metodología MERLIN (Abecasis et al., 2002), son utilizados a menudo para inferir los bloques ancestrales en genética humana. En cambio, en mejoramiento animal y vegetal estos algoritmos se vuelven computacionalmente intensos debido a la extensión de los *pedigrees*, la ausencia de datos genotípicos en las generaciones intermedias y la densidad de marcadores en la última generación (Zheng, 2015a). A partir de esto, se han desarrollado diferentes herramientas basadas en enfoques bayesianos que pueden ser utilizadas en las

áreas de mejoramiento animal y vegetal, entre ellas se puede mencionar GAIN, HAPPY, RABBIT, R/qtl2, entre otros (Broman., et al., 2019b; Liu et al., 2010; Mott et al., 2000; Zheng, 2015a).

La reconstrucción de los haplotipos es valiosa no sólo para el análisis de QTL, sino también para evaluar la calidad de las matrices de datos genotípicos que son implementadas para dichos estudios (Broman et al., 2019b). Por ejemplo, el número inferido de puntos de recombinación es un diagnóstico útil para evaluar la calidad de los individuos genotipificados que componen la población. Además, los genotipos reconstruidos pueden ser utilizados para derivar genotipos de SNP predichos y la comparación de estos con los genotipos de SNP observados permite identificar problemas tanto en las muestras como en los SNPs (Broman, 2021).

En este capítulo se presenta la metodología de genotipificación implementada para las líneas parentales y los híbridos obtenidos en cada generación de entrecruzamientos. El análisis de parentesco para comprobar los cruzamientos obtenidos se realizó utilizando la herramienta “Apparent” y la implementación del paquete RABBIT para la reconstrucción de los bloques ancestrales en los híbridos obtenidos. Las actividades desarrolladas en este capítulo se enfocan en la evaluación y entrenamiento de herramientas de análisis para su incorporación en futuros estudios de QTLs utilizando las poblaciones multiparentales que se presentan en esta tesis.

## MATERIALES Y MÉTODOS

### Material vegetal

Las líneas parentales de las dos poblaciones y un individuo de cada híbrido formado en las distintas etapas de desarrollo de las poblaciones multiparentales fueron genotipados utilizando la metodología de ddRAD. En total se procesaron 26 muestras correspondiendo a las 11 líneas parentales, 8 híbridos de 2 vías, 4 híbridos de 4 vías y 4 híbridos de 8 vías. Se genotiparon 7 híbridos de 2 vías ya que ambas poblaciones comparten un híbrido simple.

### Extracción de ADN

Para la extracción de ADN se recolectaron de 3 a 5 hojas de plantas jóvenes en estado vegetativo (estadio fenológico V4-10) (Schneiter & Miller, 1981). El material fresco fue congelado inmediatamente en atmósfera de nitrógeno líquido y fue liofilizado para su conservación a largo plazo. La extracción de ADN se realizó a partir de 20 a 25 mg de material liofilizado al cual se le incorporaron 3 bolillas de tungsteno (QIAGEN, CAT 69989) para realizar la disrupción del tejido con el sistema *Tissue Lyser* (QIAGEN). El ADN se extrajo mediante el kit Easy Pure Plant Genomic DNA (TransGen Biotech, China) siguiendo las recomendaciones del fabricante.

La calidad del ADN extraído se evaluó mediante corridas electroforéticas en gel de agarosa 0,8 % (p/v) (TAE 1X: Tris-Acetato 20 mM, pH 8,5, EDTA 1mM), de acuerdo a lo descrito por Sambrook et al., (1989). La concentración de bromuro de etidio (BrEt) utilizada fue de 0,3 µg/ml y el volumen de muestra se mezcló con solución de siembra (1/6) bromophenolblue 0,25%. Para la corrida electroforética se aplicó voltaje constante entre 1 y 5 V/cm. Para la visualización del ADN se empleó un sistema de adquisición de imágenes provisto con un transiluminador de luz UV (320 nm) con cámara incorporada (GEL DOC XR System, BioRad).

La cuantificación de ADN se realizó por medio de 3 metodologías:

- Espectrofotometría de luz UV (NanoDrop ND-100 spectrophotometer, NanoDrop Technologies). Este equipo estima la concentración a partir de la absorbancia (A) a 260 nm, asumiendo que  $A(260\text{ nm}) = 1$  equivale a 40 µg/ml de ADN; además, calcula la relación  $A_{260\text{ nm}}/A_{280\text{ nm}}$ , teniendo en cuenta que el pico máximo de absorción de las proteínas y otros contaminantes es 280 nm y la relación  $A_{260\text{ nm}}/A_{230\text{ nm}}$ , siendo la absorbancia a 230 nm, la correspondiente a distintos polifenoles; de este modo una relación cercana a 2 indica una buena calidad y pureza de la muestra de ADN.

- Fluorimetría usando el Fluorómetro Qubit® 2.0 con el protocolo Qubit™ dsDNA BR Assay Kit. El kit provee los reactivos fluorescentes, los estándares de peso molecular a ser empleados para la calibración y el buffer de dilución. Una alícuota de 2  $\mu\text{L}$  de la solución de ADN se diluye a un volumen final de 200  $\mu\text{L}$  con el buffer de dilución y el fluoróforo Qubit™ dsDNA BR. El ensayo es específico para medir ADN de doble cadena en un rango de concentraciones entre 100  $\text{pg}/\mu\text{L}$  y 1000  $\text{ng}/\mu\text{L}$ .
- Para el protocolo de Genotipado por Secuenciación, la cuantificación de ácidos nucleicos en los pasos intermedios de la construcción de las librerías ddRADSeq se realizan con un fluorómetro FluorStar Optima (BMG-LABTECH) que permite el uso de placas en el formato de 96 y 384 pocillos.

### **Secuenciación masiva de fragmentos de ADN obtenidos por doble restricción enzimática (ddRADSeq)**

La caracterización genotípica de las 11 líneas parentales y de una selección 15 híbridos que representen los cruzamientos realizados fue mediante ddRADSeq según el protocolo publicado por Aguirre et al (2019).

Previo a la restricción, la concentración de los ADN debe ser ajustada a 10  $\text{ng}/\mu\text{L}$  para todas las muestras realizando cuantificaciones de las diluciones preparadas con el Fluorómetro Qubit. El ADN de alta calidad se digirió con dos enzimas endonucleasas con distinta frecuencia de corte, una de ellas sensible a metilación (Mbol) y la otra no (SphI-HF). Entre los pasos de restricción, ligación, agrupamiento, selección de fragmentos y PCR se realizan purificaciones con micro esferas magnéticas Agencourt Ampure XP para recuperar los fragmentos deseados e ir eliminando buffers empleados en pasos previos, fragmentos de tamaños menores a los deseados, dNTPs, Adaptadores no ligados, restos de enzimas desnaturalizadas (Endonucleasas, Ligasas, Polimerasas), sales, entre otros.

Luego de la restricción y purificación se realiza la ligación de adaptadores. El adaptador portador del barcode fue el complementario a la enzima SphI. A los fines de multiplexado de muestras se diseñaron 26 adaptadores con barcodes específicos para el sitio de corte de la enzima SphI-HF y 2 adaptadores para el sitio de corte de Mbol.

Entre los pasos de ligación de adaptadores y combinación de muestras en el pool, se realizó una cuantificación de ADN en fluorómetro de placa FluorStar Optima (BMG-LABTECH), para realizar mezclas equimolares de las distintas muestras y evitar sub o sobre representación de material genético.

La ligación de adaptadores se llevó a cabo usando la enzima T4 DNA-ligasa (Invitrogen). El proceso de selección de tamaños fue realizado de forma automatizada con el equipo SageELF™ (Sage Science, USA) utilizando cassettes de 2% agarosa especificados para cubrir el rango de tamaños buscado.

Como control de calidad de algunos pasos del protocolo, luego de las ligaciones y al concluir el paso de selección de fragmentos, se realizaron corridas electroforéticas automatizadas con el equipo Fragment Analyzer (Agilent), y el kit High Sensitivity NGS Fragment Analysis Kit (DNF-486-0500, DNF-486-1000 35 – 6000 pb).

Una vez obtenidos los fragmentos purificados en la concentración adecuada, se ligan adaptadores asociados a diferentes *barcodes*. En el protocolo implementado se usan *barcodes* en sentido *forward* y *reverse* para aumentar el número de combinaciones y por ende de muestras a multiplexar por corrida de secuenciación. En este paso se toma el recaudo de no repetir combinaciones de *barcodes* a los fines de reconocer cada muestra mediante el uso programas bioinformáticos.

Finalmente se realiza un paso de amplificación con iniciadores Illumina que también contienen índices de reconocimiento de muestras que en este caso se combinan para identificar *pools*. Estos índices son incorporados mediante PCR no son muestra específicos, sino *pools* específicos. Al final de la corrida, el programa que controla este proceso reconoce los índices y envía los datos de secuenciación a archivos específicos para cada *pool* de muestras. El programa de ciclado utilizado fue el siguiente: una desnaturalización inicial de 1 min a 95°C, 10 ciclos de (a) 30 s a 95°C, (b) 30 s a 60°C y (c) 45 s a 72°C, seguido por una extensión final de 2 min a 72°C, Stand by de 8°C.

La secuenciación de la librería se realizó en un secuenciador NextSeq 550 (Illumina). Se utilizó un kit de secuenciación de mediana productividad, Nextseq 500 Mid Output v2 PE 2x74.

### **Identificación de Polimorfismos de Nucleótido Simple (SNP)**

El análisis de datos de GBS requiere de múltiples fases desde el procesamiento de las lecturas crudas que salen del secuenciador empezando por la separación de las lecturas por muestra utilizando los *barcodes* identificadores (demultiplexado), el filtrado de lecturas de baja calidad, el alineamiento de lecturas solapantes y finalmente la identificación de SNPs y el relevamiento de los mismos entre los individuos que componen la muestra (Catchen et al., 2013; Rochette et al., 2019). Para el procesamiento de las salidas del secuenciador y todas las demás etapas fue utilizado el software Stacks (Catchen et al., 2013; Rochette et al., 2019). Se empleó

también Bowtie2 (Langmead & Salzberg, 2012), un programa que permite alinear de manera eficiente las lecturas a genomas de referencia para asignar la localización de los *loci* informativos definidos por Stacks. En el caso del girasol las lecturas se mapearon a la referencia genómica (NCBI GenBank assembly accession: GCA\_002127325.2) publicada por Badouin et al. (2017).

La asignación de alelos se realizó según los pasos recomendados por Stacks para el caso de usar una referencia genómica. El panel final de marcadores SNP resultó de aplicar un filtrado según los parámetros siguientes: profundidad mínima de 3 lecturas por locus, una frecuencia de alelo minoritario (MAF) menor a 5 % y una representación mínima de cada marcador en las muestras del 20 %.

### **Anotación y efecto de los polimorfismos identificados**

Las posiciones genómicas de los SNPs obtenidas para cada muestra generadas por Stacks fueron analizadas contra la anotación del genoma de referencia de girasol (Badouin et al., 2017), para definir la funcionalidad de las regiones en las que se encontraron polimorfismos. La herramienta bioinformática utilizada fue el programa SnpEff (Cingolani et al., 2012).

### **Análisis de parentesco para la evaluación de los cruzamientos**

A partir de la matriz de datos genotípicos generados para las líneas parentales y para los representantes de los híbridos generados en cada etapa del desarrollo de las poblaciones se realizó un análisis de parentesco con el objetivo evaluar el desarrollo de los cruzamientos.

Para este análisis se utilizó el paquete de R “apparent” (Melo & Hale, 2019), el cual emplea una matriz de SNPs con el conjunto de loci dispuesto en las columnas y los individuos en las filas. En la columna 2 de la matriz se le asigna a cada individuo una de las cinco clases que el paquete ofrece para el análisis: *Mo* (considerado exclusivamente como madre potencial, o progenitor femenino), *Fa* (considerado exclusivamente como padre potencial, o progenitor masculino), *Off* (considerado exclusivamente como descendiente), *Pa* (considerado exclusivamente como progenitor, tanto femenino como masculino), o *All* (considerado como progenitor femenino potencial, progenitor masculino y descendiente dentro de la población).

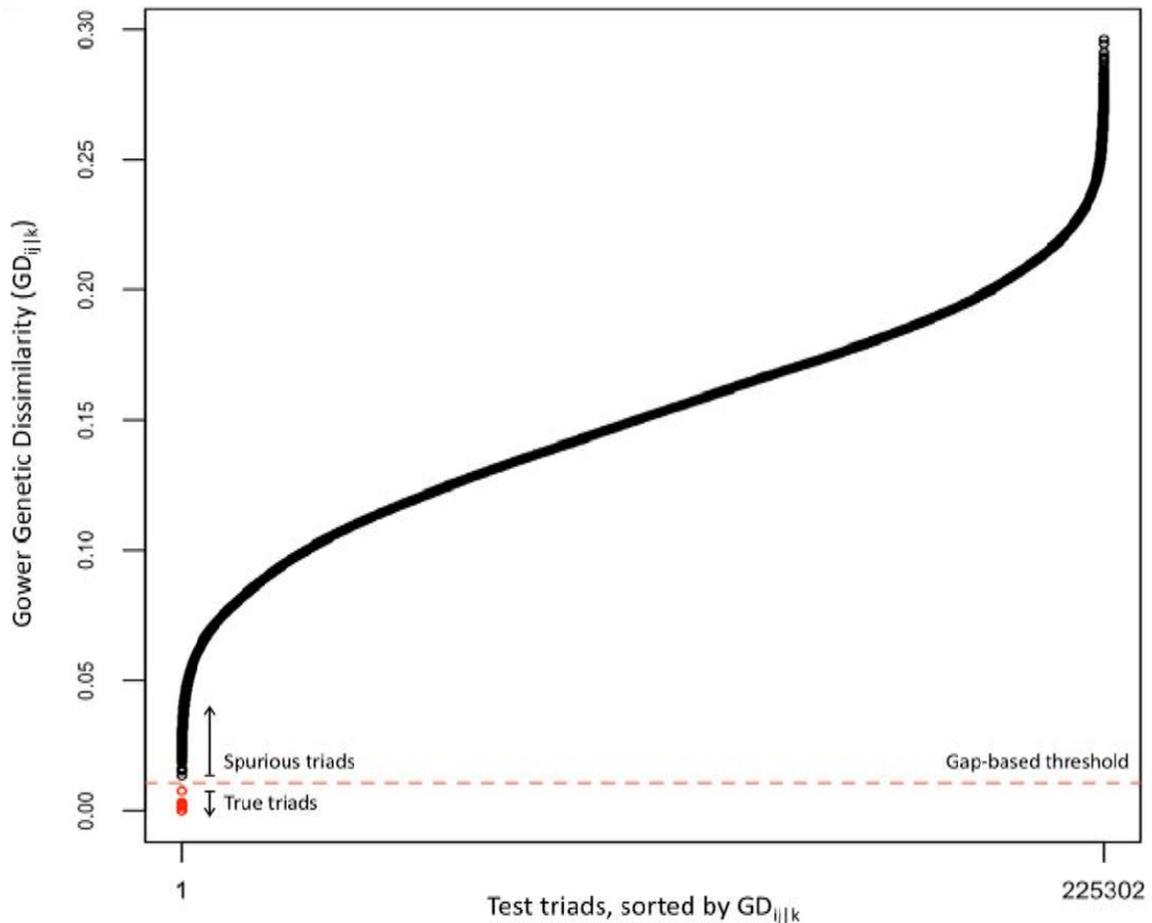
Para cada uno de los posibles pares de *i* progenitores femeninos (*Mo*, *Pa* y *All*) y *j* progenitores masculinos (*Fa*, *Pa* y *All*), se construye el genotipo de la Progenie Esperada (*Ep<sub>ij</sub>*) basándose únicamente en los marcadores que son homocigotos en ambos progenitores. A continuación, se calcula el coeficiente de similitud de Gower (GD) (Gower, 1971), entre cada *Ep<sub>ij</sub>* y todos los *k* descendientes potenciales (*Pok*) de la población (*Off* y *All*). Con un rango de 0 (identidad

perfecta) a 1 (disimilitud perfecta), el GD capta el grado de parentesco genético entre dos individuos cuantificando la identidad por estado de todos los  $n$  SNPs, según:

$$GD_{ij|k}(EP_{ij}|PO_k) = 1 - \left( \frac{\sum_{l=1}^n S_l W_l}{\sum_{l=1}^n W_l} \right)$$

Donde, para cada SNPI,  $S_l = 1$  si los estados genotípicos son los mismos;  $S_l = 0,5$  si los estados genotípicos difieren en un alelo (es decir, heterocigoto frente a homocigoto);  $S_l = 0$  si los estados genotípicos difieren en ambos alelos (es decir, homocigoto primario frente a homocigoto secundario);  $W_l = 1$  si ambos individuos están genotipados; y  $W_l = 0$  si alguno de los individuos carece de un genotipo asignado ((por ejemplo, si faltan datos debido a baja cobertura).

Por lo tanto, si  $Moi$  y  $Faj$  son los verdaderos padres de  $Pok$ ,  $Ep_{ij}$  y  $Pok$  serán genéticamente idénticos en todos los *loci* parentales homocigóticos, lo que dará como resultado una GD por pares igual a cero. Sin embargo, debido a posibles errores de secuenciación y genotipado, en la práctica el valor de GD calculado para una tríada verdadera ( $Moi$ ,  $Faj$ ,  $Pok$ ) será mayor que cero; pero será significativamente menor que los valores de GD obtenidos entre  $Ep_{ij}$  y todos los falsos descendientes. Para una población dada de individuos, un diagrama de dispersión de todos los valores posibles de  $G_{dij|k}$  muestra una brecha significativa que separa las tríadas verdaderas de las asociaciones espurias. Esta brecha se localiza escaneando el conjunto ordenado de valores  $G_{dij|k}$  y detectando el lugar de máxima diferencia entre dos valores adyacentes; y el punto medio de esta brecha se toma como un umbral (Figura 2.1).



**Figura 2.1.** Diagrama de dispersión de todos los valores posibles de  $G_{dij|k}$  para una población dada. Figura extraída de Melo & Hale, (2019).

Una vez identificada la brecha, la importancia de su magnitud se evalúa mediante una prueba de Dixon (Dixon, 1950, 1951). Si el tamaño de la brecha se declara significativo, la importancia individual de cada tríada por debajo de la brecha (es decir, aquellas tríadas declaradas como potenciales asociaciones reales entre padres e hijos) se comprueba entonces contra una muestra de los valores de  $G_{dij|k}$  más estrechamente relacionados por encima de la brecha (es decir, aquellas tríadas declaradas como espurias). Si esta segunda prueba de Dixon también resulta ser estadísticamente significativa, la tríada implicada se declara como verdadera y se informa su valor de probabilidad asociada.

Por otro lado, se evaluó el nivel de heterocigosis en las líneas parentales y en los híbridos de 2, 4 y 8 vías con el objetivo de comprobar el grado de homocigosis de las líneas parentales utilizadas y poder detectar si aumenta el nivel de heterocigosis en los híbridos obtenidos.

### **Análisis de diversidad genética de las líneas parentales de las poblaciones MAGIC**

Se estimaron las medidas de heterocigosis esperada ( $H_e$ ) (Nei, 1978), heterocigosis observada ( $H_o$ ) y las frecuencias alélicas para las líneas parentales que conforman cada

población. Las estimaciones fueron realizadas con los paquetes de R hierfstat, adegenet, VCFr y poppr (Goudet, 2005; Jombart, 2008; Kamvar et al., 2014; Knaus & Grünwald, 2017)

Se construyó un árbol *neighbor-joining* en base a las distancias genéticas de Nei (Nei, 1972) calculadas entre pares de líneas parentales para cada población utilizando el paquete de R StAMPP (Pembleton et al., 2013). La construcción de los árboles fue realizada con la función *nj* del paquete APE de R (Paradis & Schliep, 2019). Los árboles resultantes fueron editados con el software iTOL (Letunic & Bork, 2019).

### **Reconstrucción de bloques genómicos ancestrales en las poblaciones MAGIC1 y MAGIC2 a través de RABBIT**

RABBIT por sus siglas en inglés *Reconstructing Ancestry Blocks BIT by Bit*, es una plataforma bayesiana que permite reconstruir haplotipos, imputar genotipos y construir mapas genéticos para poblaciones multiparentales. Para la función de reconstrucción de haplotipos requiere de datos genotípicos, modelo y diseño de la población. Además, incorpora a partir de un hiperparámetro  $\Omega$ , datos sobre nivel de endocria y densidad de sitios de recombinación en cromosomas homólogos (Zheng, 2014, 2015b). En una primera etapa, los datos genotípicos se analizan mediante modelos Ocultos de Markov *Hidden Markov Model* (HMM) para generar dos modelos. El modelo de proceso que describe los procesos de origen ancestral a lo largo de dos cromosomas homólogos, y el modelo de observación que describe la probabilidad de los genotipos dados los estados latentes de origen ancestral. El modelo de observación contempla la existencia de datos perdidos y errores de secuenciación presentes en la matriz de datos genotípicos tanto de los individuos fundadores como de los híbridos.

El análisis se realiza independientemente para cada grupo de ligamiento de cada individuo muestreado. Cada individuo es genotipado en  $T$  SNPs bialélicos de cada grupo de ligamiento y las distancias genéticas  $d_t$  ( $t = 1 \dots T - 1$ ) entre la ubicación de los marcadores consecutivos  $t$  y  $t + 1$  son medidas en Morgans. El término  $\{Y_\tau\}_{\tau=1}^T$  denota los genotipos a lo largo de dos cromosomas homólogos de un individuo muestreado y  $H$  representa el haplotipo fundador, siendo el elemento de la matriz  $H_{ti}$  el alelo observado en el locus  $t = 1 \dots T$  del fundador  $i = 1 \dots L$ .

El término  $\{O_t^m\}_{t=1}^T$  denota el origen ancestral materno y el término  $\{O_t^p\}_{t=1}^T$  denota el origen ancestral paterno. El término  $O_t = (O_t^m, O_t^p)$  determina el estado ordenado de origen ancestral, y es idéntico por descendencia (*identical by descent* IBD) si  $O_t^m = O_t^p$  y de no ser iguales no son IBD. El número de posibles orígenes ancestrales queda definido por el número de fundadores  $n$  de la población en estudio.

El **modelo de proceso** utilizado fue el *indepmodel* el cual considera que dos cromosomas homólogos tienen *a priori* orígenes ancestrales completamente independientes. La distribución inicial está dada por:

$$P[(O_1^m, O_1^p)] = P(O_1^m)P(O_1^p) = \frac{1}{n^2}$$

y, por tanto, la probabilidad *f* de IBD *a priori* se establece implícitamente en  $1/n$ . La matriz de probabilidad de transición para el *indepModel* viene dada por:

$$\begin{aligned} P[(O_{t+1}^m, O_{t+1}^p)|(O_t^m, O_t^p)] &= P(O_{t+1}^m|O_t^m)P(O_{t+1}^p|O_t^p), \\ P(O_{t+1}^x|O_t^x)|(O_t^m, O_t^p) &= \delta(O_{t+1}^x = O_t^x)e^{-R^x D_t} + \delta(O_{t+1}^x \neq O_t^x) \\ &X(1 - e^{-R^x d_t})\frac{1}{n-1}, x \in \{m, p\}, \end{aligned}$$

Donde  $\delta$  es una función indicadora que será igual a 1 si el argumento es verdadero y 0 si no lo es, y  $R^m$  ( $R^p$ ) es el largo del mapa o la densidad de sitios de recombinación total en los cromosomas derivados de los parentales. Por lo que el hiperparámetro  $\Omega$  se define por  $\{R^m, R^p\}$ .

En el **modelo de observación** los genotipos  $\{Y_t\}_{t=1}^T$  están dados condicionalmente por los estados latentes de origen ancestral  $\{O_t\}_{t=1}^T$ . Por lo tanto, nos centramos en la probabilidad en un locus y dejamos caer el locus  $t$ . Dado  $\mathbf{D}(\mathbf{H}, \mathbf{O}) = (H_{O^m}, H_{O^p})$  como el genotipo en fase derivados de los haplotipos fundadores  $\mathbf{H}$  y el estado de origen ancestral  $\mathbf{O}$  en cada locus. Siendo  $\epsilon$  y  $\epsilon_F$  las probabilidades de error de tipificación alélica para los individuos muestreados y las líneas fundadoras respectivamente. El objetivo es calcular la probabilidad  $l = P(Y|D(H, O), O, \epsilon, \epsilon_F)$  en cada locus.

Siendo  $\mathbf{Z}$  el genotipo verdadero en fase en el locus del individuo muestreado. La probabilidad  $l$  se calcula integrando el genotipo verdadero desconocido  $\mathbf{Z}$ :

$$l = \sum_{\mathbf{Z}} P(Y|\mathbf{Z}, \epsilon)P(\mathbf{Z}|D, O, \epsilon_F),$$

Donde  $P(\mathbf{Z}|D, O, \epsilon_F)$  es la probabilidad posterior dado el genotipo derivado  $\mathbf{D}$  y el origen ancestral  $\mathbf{O}$ . De acuerdo al teorema de Bayes se tiene:

$$P(\mathbf{Z}|D, O, \epsilon_F) = \frac{P(D|\mathbf{Z}, O, \epsilon_F)P(\mathbf{Z}|O)}{P(D|O, \epsilon_F)}$$

Donde  $P(\mathbf{Z}|O)$  es la probabilidad *a priori* del verdadero genotipo  $\mathbf{Z}$ , y la probabilidad marginal es  $P(D|O, \epsilon_F) = \sum_{\mathbf{Z}} P(D|\mathbf{Z}, O, \epsilon_F)P(\mathbf{Z}|O)$  de acuerdo a ley de probabilidades.

La reconstrucción de los bloques de ancestría se realiza muestreando en numerosas ocasiones la distribución posterior conjunta de  $\{O_t\}_{t=1}^T$  condicionada por el hiperparámetro  $\Omega$  y la probabilidad de error de tipificación  $\epsilon_F$  y  $\epsilon$ . Para cada muestra posterior, se calcula  $\alpha(O_t) = P(\{Y_\tau\}_{\tau=1}^t, O_t | \Omega, \epsilon, \epsilon_F)$  iterativamente para  $t = 1 \dots T$  mediante el algoritmo de Rabiner (1989). Luego la muestra  $O_T$  de acuerdo a la distribución  $\alpha(O_T)$ , y posteriormente se muestrea  $O_t$  de acuerdo a:

$$P(O_t | \{O_\tau\}_{\tau=1}^T, \{Y_\tau\}_{\tau=1}^T, \Omega, \epsilon, \epsilon_F) \propto \alpha(O_t) P(O_{t+1} | O_t, \Omega)$$

Las muestras posteriores contienen información completa de los bloques de ancestría a lo largo de dos cromosomas homólogos. La secuencia óptima de los bloques de ancestría se obtiene seleccionando la muestra posterior con la máxima probabilidad marginal  $P(\Omega, \epsilon, \epsilon_F) = \sum_{O_T} \alpha(O_T)$ .

Las distancias físicas entre los marcadores utilizados fueron transformadas a distancias genéticas considerando una relación de recombinación de 0,5 cM/Mpb siguiendo lo recomendado por Zheng, (2015a).

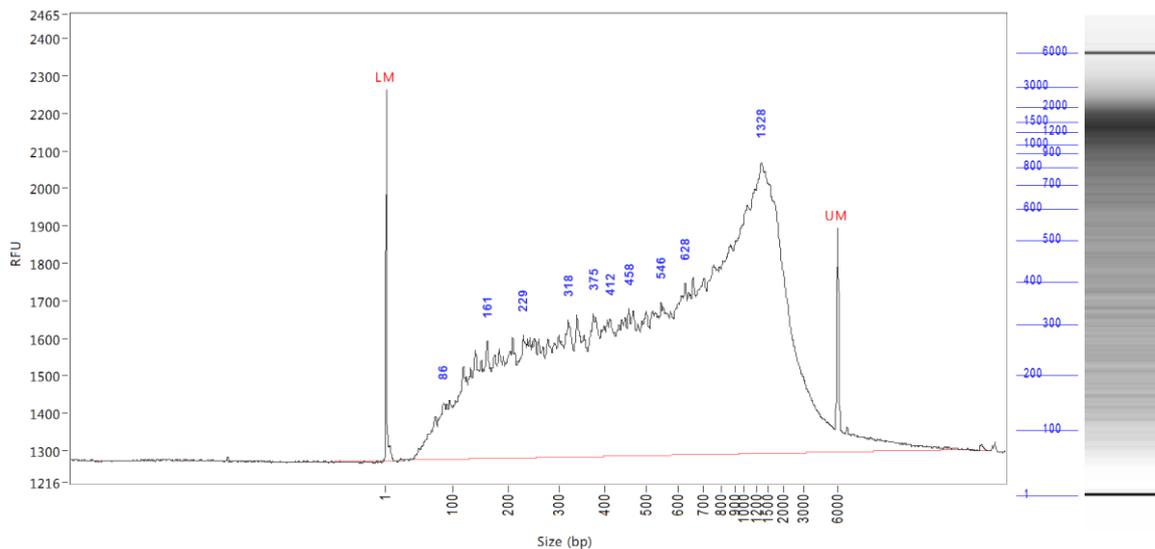
La reconstrucción de los bloques genómicos ancestrales se realizó considerando a las 8 líneas parentales de cada población como fundadoras y la reconstrucción de los bloques fue sobre 4 híbridos simples, 2 híbridos de cuatro vías y 2 híbridos de ocho vías para ambas poblaciones.

El paquete RABBIT está implementado en lenguaje de programación Mathematica 9.0 (Wolfram Research, 2012) y se encuentra disponible en <https://github.com/chaozhi/RABBIT.git>.

## RESULTADOS

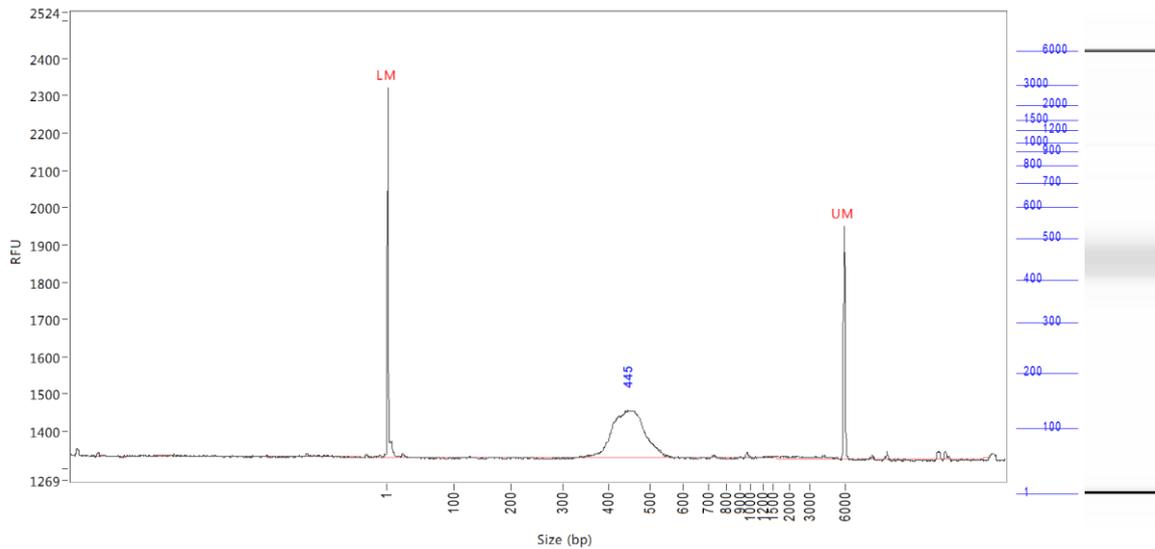
### Construcción de la librería ddRADseq

El patrón de restricción observado por la acción conjunta de las enzimas SphI-HF y MboI fue acorde a lo predicho y permitió obtener una distribución homogénea de fragmentos que exploraron un rango continuo de 300 a 3000 pb (adaptadores incluidos). La Figura 2.2 muestra el perfil de una corrida electroforética automatizada del producto de ligación de adaptadores de una de las líneas parentales evaluada al azar.



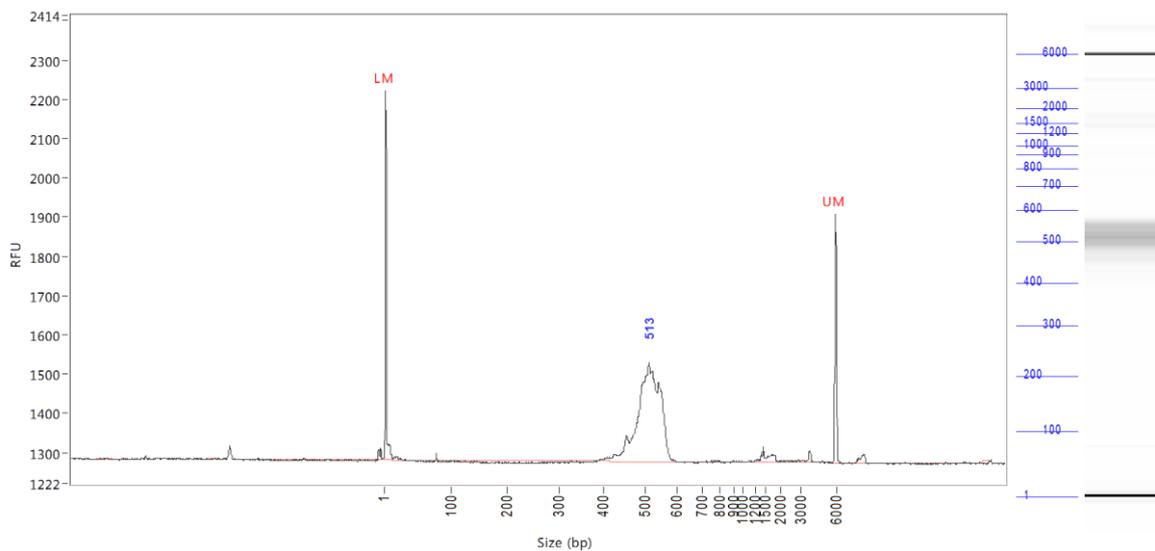
**Figura 2.2.** Perfil de corrida electroforética del producto de ligación de la digestión del ADN de una de las líneas parentales en el equipo Fragment Analyzer. A la izquierda se observa la curva de tamaños de fragmentos en función de su concentración (Relative Fluorescence Units, RFU). A la derecha se grafica la misma información, pero como si observáramos el perfil de una corrida electroforética de un gel de agarosa.

La Figura 2.3 muestra el perfil electroforético de la fracción de ADN (fragmentos más adaptadores) recuperada a partir de la corrida del equipo SageELF. Según lo indicado por el fabricante la dispersión esperada es de 70 pb respecto del tamaño seleccionado, los resultados fueron consistentes con esta dispersión.



**Figura 2.3.** Perfil de corrida electroforética del producto de la selección automatizada de fragmentos con adaptadores ligados del pool de muestras que conforman la librería en el equipo Fragment Analyzer. A la izquierda se observa la curva de tamaños de fragmentos en función de su concentración (Relative Fluorescence Units, RFU). A la derecha se grafica la misma información, pero como si observáramos el perfil de una corrida electroforética de un gel de agarosa.

Luego de la obtención de los fragmentos de tamaños deseados, se procedió a la purificación con el kit Ampure XP y se llevó a cabo la reacción de PCR final con la inclusión de índices Illumina. Los iniciadores indexados suman a la secuencia con adaptadores, unos 60 nucleótidos en cada extremo, por lo que el tamaño promedio de salida de la selección de rangos de tamaños se incrementó en unas 120 pb (Figura 2.4).



**Figura 2.4.** Perfil de corrida electroforética del producto de PCR de la selección de tamaños del pool de muestras que conforman la librería en el equipo Fragment Analyzer. A la izquierda

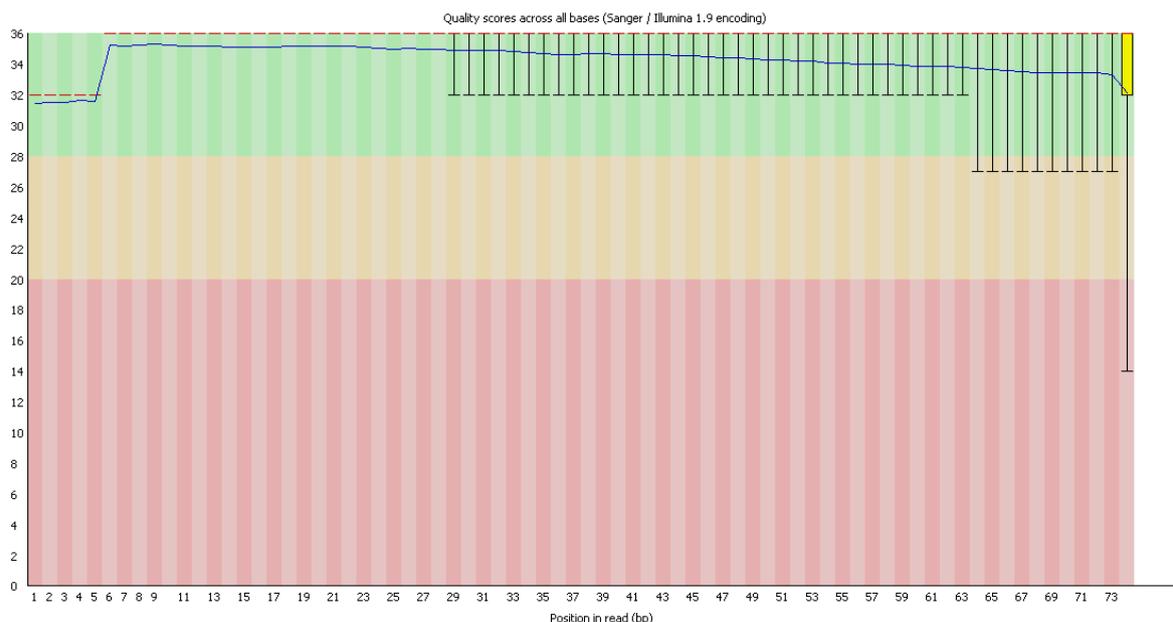
se observa la curva de tamaños de fragmentos en función de su concentración (Relative Fluorescence Units, RFU). A la derecha se grafica la misma información, pero como si observáramos el perfil de una corrida electroforética de un gel de agarosa.

### **Secuenciación de la librería ddRADseq**

La secuenciación de la librería ddRADseq tuvo un rendimiento de 39 millones de lecturas. El 94 % de las posiciones secuenciadas obtuvo un valor de calidad  $QV > 30$ . El 89,4 % de las lecturas pasaron los filtros de calidad del secuenciador y de la función *process\_radtags* del software Stacks y pudieron ser mapeadas al genoma de referencia (Badouin et al., 2017).

Las lecturas obtenidas para las 26 muestras fueron procesadas mediante el programa Stacks. Este programa como fue mencionado en Materiales y Métodos, filtra las lecturas obtenidas en función del índice de calidad de Phred (Phred score), elimina adaptadores en la secuencia, e identifica los sitios de corte de las enzimas. La caída de la calidad de las secuencias hacia los ciclos finales de secuenciación, permite definir el umbral de calidad de las secuencias con el que se va a trabajar y el largo de las secuencias para la búsqueda de polimorfismos. En este caso se definió un punto de corte en función del *barcode* más largo (9 bases), a fin de producir secuencias del mismo tamaño que es un requisito de Stacks para poder alinearlas. El *trimming*, o punto de corte, fue establecido entonces en 66 bases.

En la Figura 2.5 se puede apreciar que la calidad de secuenciación por posición hasta la base 75, la secuenciada en el último ciclo de la lectura paired-end. El percentil 75 de las muestras, e incluso un número mayor, se mantienen dentro de los Phred scores aceptables para el *process\_radtags* (el cual establece un umbral de Phred = 33).



**Figura 2.5.** Gráficos de cajas de valores de calidad a través de las bases secuenciadas. En el eje de las ordenadas se grafica el índice de calidad de secuencia (Phred score) y en el de las abscisas la posición dentro de la secuencia de 75 nucleótidos.

### Identificación de Polimorfismos de Nucleótido Simple (SNP)

Las matrices de SNPs obtenidas a partir del análisis de 26 individuos habiendo realizados los procesos de filtrado descritos anteriormente fueron de 4191 SNPs y de 4674 SNPs para las poblaciones MAGIC1 y MAGIC2, respectivamente. La cantidad de SNPs obtenidos por cromosoma para ambas poblaciones se presenta en las Tablas 2.1 y 2.2. La distribución de los marcadores obtenidos en el genoma de girasol se representa en las Figuras 2.6 y 2.7. La proporción de datos faltantes para las matrices de SNPs obtenidas fue de 11,48 % para la población MAGIC1 y de 12,66 % para la población MAGIC2.

**Tabla 2.1.** Cantidad de SNPs por cromosoma identificados en la población MAGIC1.

Cromosoma	Largo Mb	Nro. SNPs	Kb/SNP	SNP/Mb
1	154,00	274	562,04	1,78
2	180,80	241	750,21	1,33
3	151,60	273	555,31	1,80
4	161,00	214	752,34	1,33
5	197,20	378	521,69	1,92
6	93,50	160	584,38	1,71
7	93,50	124	754,03	1,33
8	137,90	276	499,64	2,00
9	188,90	332	568,98	1,76

10	221,60	296	748,65	1,34
11	151,60	230	659,13	1,52
12	149,80	164	913,41	1,09
13	177,50	277	640,79	1,56
14	157,00	180	872,22	1,15
15	154,20	173	891,33	1,12
16	169,80	253	671,15	1,49
17	193,30	346	558,67	1,79
<b>Total</b>		<b>4.191</b>		

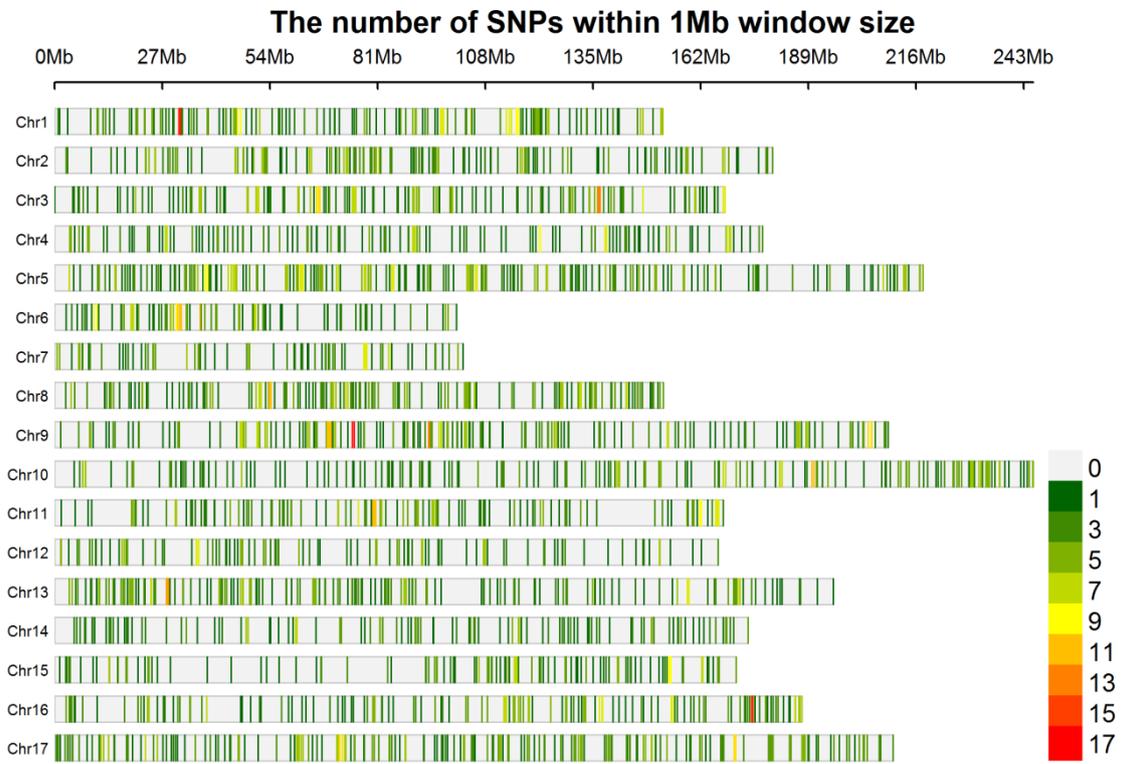
Para la población MAGIC1, el promedio de SNPs por cromosoma fue de 246,5 SNPs, con una densidad de 1,53 SNP por Mb. El cromosoma más densamente cubierto fue el Cromosoma 8 con 2,00 SNPs/Mb. El menos cubierto fue el Cromosoma 12 con 1,09 SNPs/Mb.

**Tabla 2.2.** Cantidad de SNPs por cromosoma identificados en la población MAGIC2.

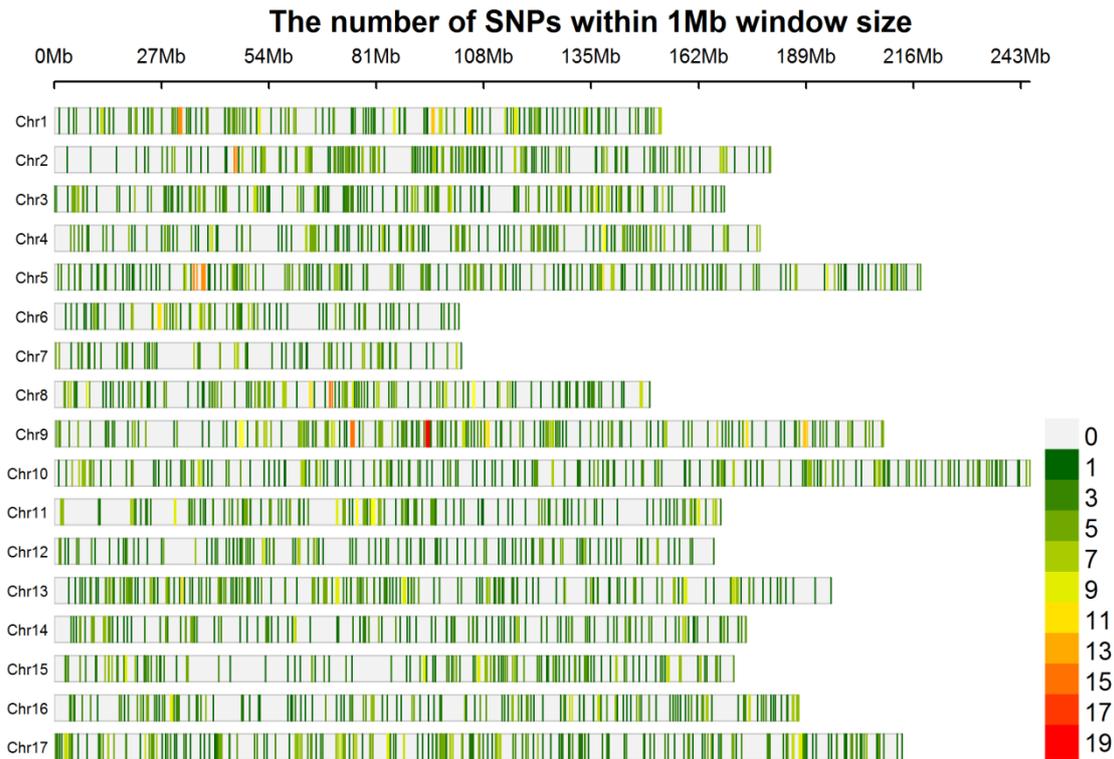
<b>Cromosoma</b>	<b>Largo Mb</b>	<b>Nro. SNPs</b>	<b>Kb/SNP</b>	<b>SNP/Mb</b>
1	154,00	289	532,87	1,88
2	180,80	286	632,17	1,58
3	151,60	303	500,33	2,00
4	161,00	269	598,51	1,67
5	197,20	378	521,69	1,92
6	93,50	153	611,11	1,64
7	93,50	128	730,47	1,37
8	137,90	277	497,83	2,01
9	188,90	392	481,89	2,08
10	221,60	336	659,52	1,52
11	151,60	242	626,45	1,60
12	149,80	175	856,00	1,17
13	177,50	320	554,69	1,80
14	157,00	216	726,85	1,38
15	154,20	249	619,28	1,61
16	169,80	270	628,89	1,59
17	193,30	391	494,37	2,02
<b>Total</b>		<b>4674</b>		

Para la población MAGIC2, el promedio de SNPs por cromosoma fue de 274,9 SNPs, con una densidad de 1,71 SNP por Mb. El cromosoma más densamente cubierto fue el Cromosoma 9 con 2,08 SNPs/Mb. El menos cubierto fue el Cromosoma 12 con 1,17 SNPs/Mb.

La distribución de los SNPs obtenidos en los 17 cromosomas del genoma del girasol para ambas poblaciones se grafican en las Figuras 2.6 y 2.7.



**Figura 2.6.** Distribución de los SNPs obtenidos mapeados contra el genoma de referencia de girasol en la población MAGIC1. La leyenda indica el código de color definido para el número de SNPs presentes por ventana de 1Mb.



**Figura 2.7.** Distribución de los SNPs obtenidos mapeados contra el genoma de referencia de girasol en la población MAGIC2. La leyenda indica el código de color definido para el número de SNPs presentes por ventana de 1Mb.

### Exploración funcional de los marcadores informativos obtenidos

El análisis de los 4.191 SNPs polimórficos entre los parentales de la población MAGIC1, usando el programa SnpEff, ubicó a la mayoría de los polimorfismos en regiones intergénicas (67,89 %), siguiéndole en representatividad regiones codificantes corriente abajo (3') (11,26 %) y corriente arriba (5') (9,43 %). Las regiones exónicas e intrónicas representaron el 3,87 y 6,07 % respectivamente. Los SNPs en las regiones codificantes (CDS) condujeron en un 34,03 % a mutaciones con cambio de sentido (MisSense) y en un 65,97 % a mutaciones silentes. Las 4.191 variantes pueden potencialmente causar 5.302 efectos putativos, de los cuales 5.095 tendrían un impacto modificador, 157 impacto bajo, 49 impacto moderado y un solo efecto alto impacto, según la clasificación propuesta por Cingolani et al (2012).

En la población MAGIC2 el análisis de los 4.674 SNPs realizado con el programa SnpEff ubicó a la mayoría de los polimorfismos en regiones intergénicas (66,60 %), siguiéndole en representatividad regiones codificantes corriente abajo (3') (11,31 %) y corriente arriba (5') (10,17 %). Las regiones exónicas e intrónicas representaron el 4,44 y 5,83 % respectivamente. Los SNPs en las regiones codificantes (CDS) condujeron en un 30,36 % a mutaciones con cambio de sentido (MisSense), en un 0,52 % a mutaciones sinónimas (NonSense) y en un

69,11 % a mutaciones silentes. Las 4674 variantes pueden potencialmente causar 5967 efectos putativos, de los cuales 5690 tendrían un efecto modificador, 211 impacto bajo, 58 impacto moderado y un solo efecto alto impacto, según la clasificación propuesta por Cingolani et al (2012).

### **Análisis de parentesco para la evaluación de los cruzamientos**

El análisis de parentesco para la población MAGIC1 mostró los valores más bajos de GD para los cruzamientos efectivamente realizados. Si bien los resultados no fueron significativos se puede apreciar la diferencia de las GD entre los cruzamientos verdaderos y los espurios. En la Tabla 2.3 se muestran los resultados obtenidos.

En la población MAGIC2 las diferencias fueron estadísticamente significativas para el caso de los híbridos simples, siendo significativos aquellos híbridos que efectivamente se habían realizado. Para los cruzamientos de 4 y 8 vías los resultados no fueron significativos, pero también se puede observar la diferencia de las GD entre las tríadas verdaderas y las espurias. En la Tabla 2.4 se muestran los resultados obtenidos

**Tabla 2.3.** Análisis de parentesco entre los híbridos obtenidos y las líneas parentales de la población MAGIC1.

MAGIC1 - Híbridos simples			
Madre	Padre	Descendencia	SNPs GD
<b>AO 01/04</b>	<b>B477-1</b>	<b>AO 01/04 x B477-1</b>	<b>3658 0,011</b>
<b>GP 08/05</b>	<b>GP 07/15</b>	<b>GP 08/05 x GP 07/15</b>	<b>3624 0,011</b>
<b>2071</b>	<b>GP 06/04</b>	<b>2071 x GP 06/04</b>	<b>2083 0,019</b>
<b>GP 03/03</b>	<b>B10</b>	<b>GP 03/03 x B10</b>	<b>1854 0,041</b>
B477-1	GP 06/04	2071 x GP 06/04	3644 0,079
AO 01/04	2071	AO 01/04 x B477-1	2059 0,083
GP 08/05	GP 06/04	2071 x GP 06/04	3577 0,095
AO 01/04	GP 08/05	GP 08/05 x GP 07/15	3658 0,096
GP 07/15	B477-1	AO 01/04 x B477-1	3610 0,098
GP 07/15	2071	GP 08/05 x GP 07/15	2103 0,108
MAGIC1 - Híbridos de cuatro vías			
<b>GP 08/05 x GP 07/15</b>	<b>AO 01/04 x B477-1</b>	<b>((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))</b>	<b>1985 0,011</b>
2071 x GP 06/04	GP 08/05 x GP 07/15	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	1894 0,087
GP 08/05 x GP 07/15	GP 08/05 x GP 07/15.1	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	2577 0,101
AO 01/04 x B477-1	AO 01/04 x B477-1.1	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	2768 0,108
AO 01/04 x B477-1	2071 x GP 06/04	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	2135 0,120
<b>2071 x GP 06/04</b>	<b>GP 03/03 x B10</b>	<b>((2071 x GP 06/04)x(GP 03/03 x B10))</b>	<b>996 0,146</b>
AO 01/04 x B477-1	GP 03/03 x B10	((2071 x GP 06/04)x(GP 03/03 x B10))	945 0,194
2071 x GP 06/04	2071 x GP 06/04.1	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	2849 0,210
GP 03/03 x B10	GP 08/05 x GP 07/15	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	1344 0,216
AO 01/04 x B477-1	GP 03/03 x B10	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	1579 0,231
MAGIC1 - Híbridos de ocho vías			
<b>((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))</b>	<b>((2071 x GP 06/04)x(GP 03/03 x B10))</b>	<b>((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))x((2071 x GP 06/04)x(GP 03/03 x B10))</b>	<b>705 0,139</b>
((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))x((2071 x GP 06/04)x(GP 03/03 x B10))	2362 0,153
((2071 x GP 06/04)x(GP 03/03 x B10))	((2071 x GP 06/04)x(GP 03/03 x B10))	((2071 x GP 06/04)x(GP 03/03 x B10))x((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	1255 0,218
<b>((2071 x GP 06/04)x(GP 03/03 x B10))</b>	<b>((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))</b>	<b>((2071 x GP 06/04)x(GP 03/03 x B10))x((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))</b>	<b>645 0,218</b>
((2071 x GP 06/04)x(GP 03/03 x B10))	((2071 x GP 06/04)x(GP 03/03 x B10))	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))x((2071 x GP 06/04)x(GP 03/03 x B10))	1363 0,228
((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	((2071 x GP 06/04)x(GP 03/03 x B10))x((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))	2225 0,229

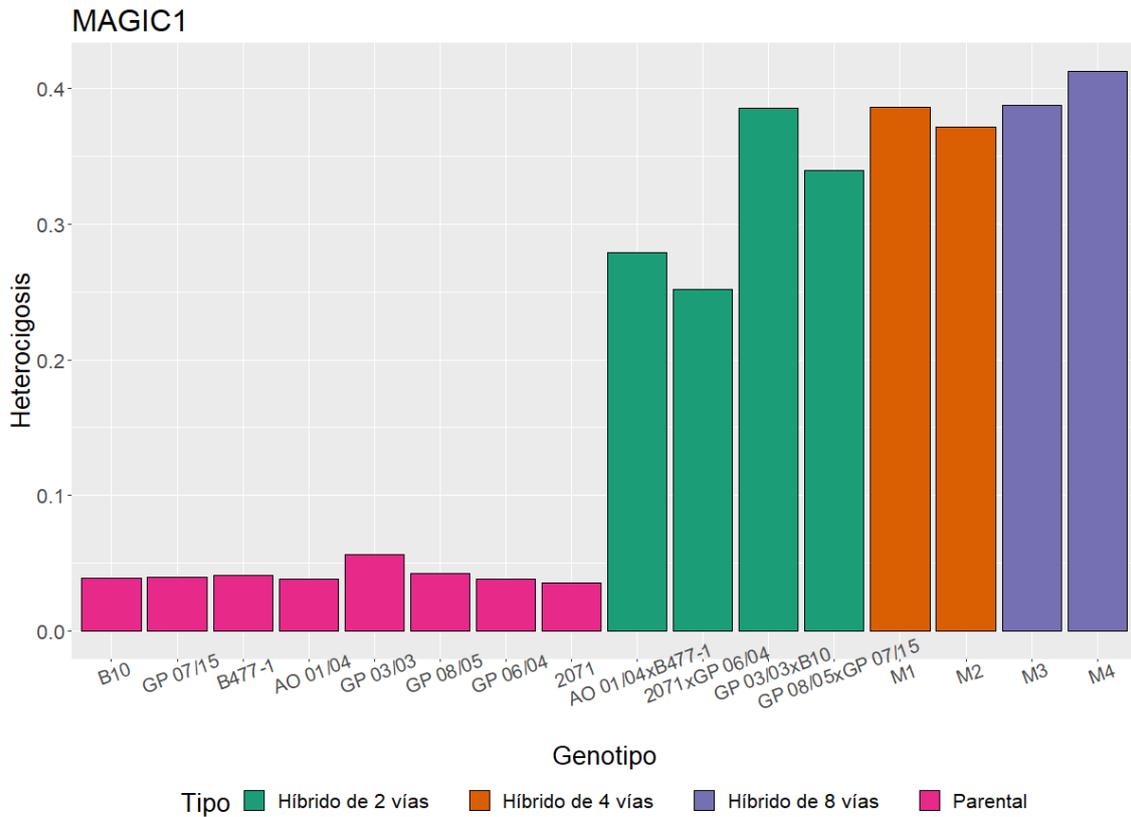
En rojo se indican las tríadas verdaderas, en color negro se muestran algunas de las relaciones espurias que fueron analizadas. La columna SNP indica el número de marcadores utilizados por el análisis.

**Tabla 2.4.** Análisis de parentesco entre los híbridos obtenidos y las líneas parentales de la población MAGIC2.

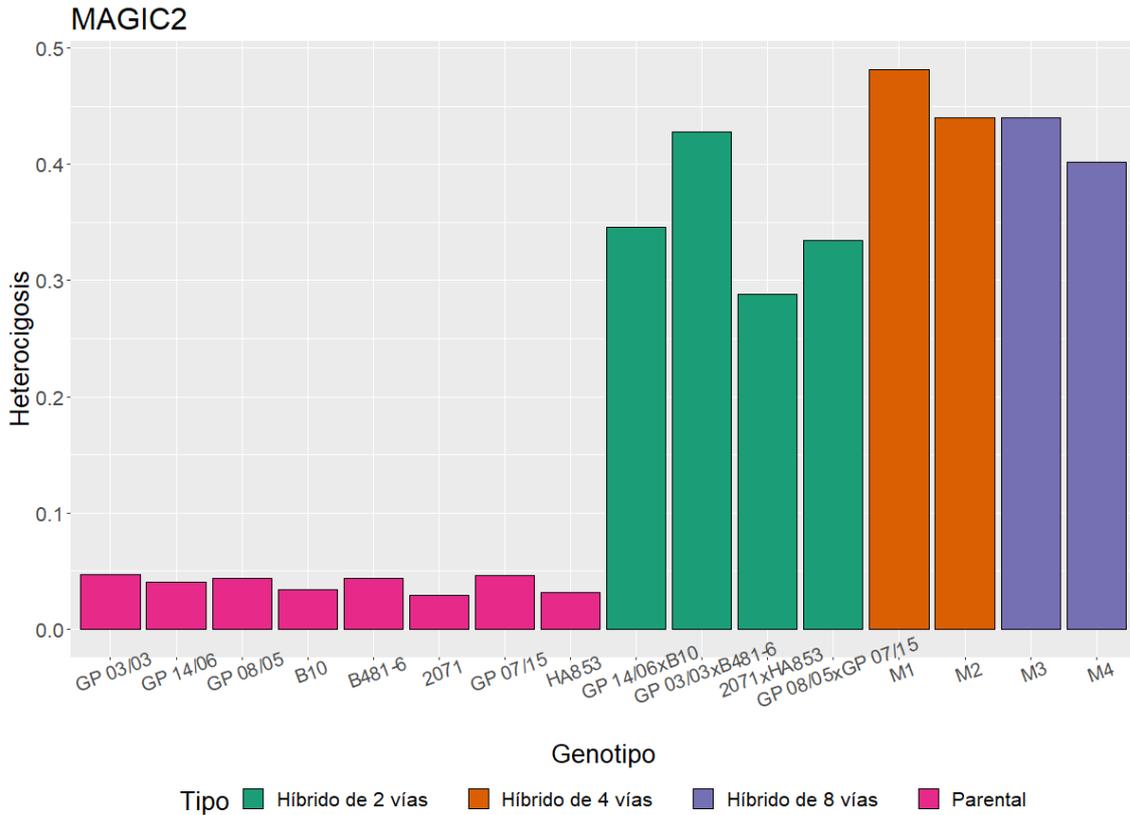
MAGIC2 - Híbridos simples				
Madre	Padre	Descendencia	SNPs	GD
GP 03/03	B481-6	GP 03/03 x B481-6*	3974	0,010
GP 08/05	GP 07/15	GP 08/05 x GP 07/15*	3797	0,010
2071	HA853	2071 x HA853*	1444	0,014
GP 14/06	B10	GP 14/06 x B10*	2068	0,032
MAGIC2 - Híbridos de cuatro vías				
2071 x HA853	GP 03/03 x B481-6	((2071 x HA853)x(GP 03/03 x B481-6))	1817	0,018
GP 08/05 x GP 07/15	GP 14/06 x B10	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	1802	0,023
GP 14/06 x B10	2071 x HA853	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	2090	0,117
2071 x HA853	GP 08/05 x GP 07/15	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	1980	0,135
GP 14/06 x B10	GP 14/06 x B10.1	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	2837	0,141
GP 14/06 x B10	GP 03/03 x B481-6	((2071 x HA853)x(GP 03/03 x B481-6))	1546	0,148
GP 03/03 x B481-6	GP 08/05 x GP 07/15	((2071 x HA853)x(GP 03/03 x B481-6))	1353	0,148
GP 08/05 x GP 07/15	GP 08/05 x GP 07/15.1	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	2708	0,155
GP 03/03 x B481-6	GP 03/03 x B481-6.1	((2071 x HA853)x(GP 03/03 x B481-6))	2513	0,189
2071 x HA853	2071 x HA853.1	((2071 x HA853)x(GP 03/03 x B481-6))	3159	0,195
MAGIC 2 - Híbridos de 8 vías				
((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	((2071 x HA853)x(GP 03/03 x B481-6))	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))x((2071 x HA853)x(GP 03/03 x B481-6))	1010	0,082
((2071 x HA853)x(GP 03/03 x B481-6))	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	((2071 x HA853)x(GP 03/03 x B481-6))x((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	1093	0,091
((2071 x HA853)x(GP 03/03 x B481-6))	((2071 x HA853)x(GP 03/03 x B481-6)).1	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))x((2071 x HA853)x(GP 03/03 x B481-6))	1922	0,148
((2071 x HA853)x(GP 03/03 x B481-6))	((2071 x HA853)x(GP 03/03 x B481-6)).1	((2071 x HA853)x(GP 03/03 x B481-6))x((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	2067	0,163
((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	((GP 08/05 x GP 07/15)x(GP 14/06 x B10)).1	((GP 08/05 x GP 07/15)x(GP 14/06 x B10))x((2071 x HA853)x(GP 03/03 x B481-6))	2014	0,167
((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	((GP 08/05 x GP 07/15)x(GP 14/06 x B10)).1	((2071 x HA853)x(GP 03/03 x B481-6))x((GP 08/05 x GP 07/15)x(GP 14/06 x B10))	2258	0,187

En rojo se indican las tríadas verdaderas, en color negro se muestran algunas de las relaciones espurias que fueron analizadas. La columna SNP indica el número de marcadores utilizados por el análisis. El asterisco en la descendencia denota aquellas tríadas que resultaron estadísticamente significativas p-valor < 0,01.

La proporción de *loci* heterocigotas en las líneas parentales y en los híbridos obtenidos de las poblaciones MAGIC1 y MAGIC2 se muestra en las Figura 2.8 y 2.9.



**Figura 2.8.** Proporción de *loci* heterocigotas en las líneas parentales y en los híbridos de 2, 4 y 8 vías en la población MAGIC1. Referencias: M1: ((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1)); M2: ((2071 x GP 06/04)x(GP 03/03 x B10)); M3: ((2071 x GP 06/04)x(GP 03/03 x B10))x((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1)); M4: ((GP 08/05 x GP 07/15)x(AO 01/04 x B477-1))x((2071 x GP 06/04)x(GP 03/03 x B10)).



**Figura 2.9.** Proporción de loci heterocigotas en las líneas parentales y en los híbridos de 2, 4 y 8 vías en la población MAGIC2. Referencias: M1: ((2071 x HA853)x(GP 03/03 x B481-6)); M2: ((GP 08/05 x GP 07/15)x(GP 14/06 x B10)); M3: ((2071 x HA853)x(GP 03/03 x B481-6))x((GP 08/05 x GP 07/15)x(GP 14/06 x B10)); M4: ((GP 08/05 x GP 07/15)x(GP 14/06 x B10))x((2071 x HA853)x(GP 03/03 x B481-6)).

La heterocigosis residual que presentaron las líneas parentales es del orden 3 al 5 % y corresponden a valores similares obtenidos por estudios previos del grupo. El aumento de la heterocigosis en los híbridos confirma también el desarrollo exitoso de los cruzamientos realizados para el armado de las poblaciones.

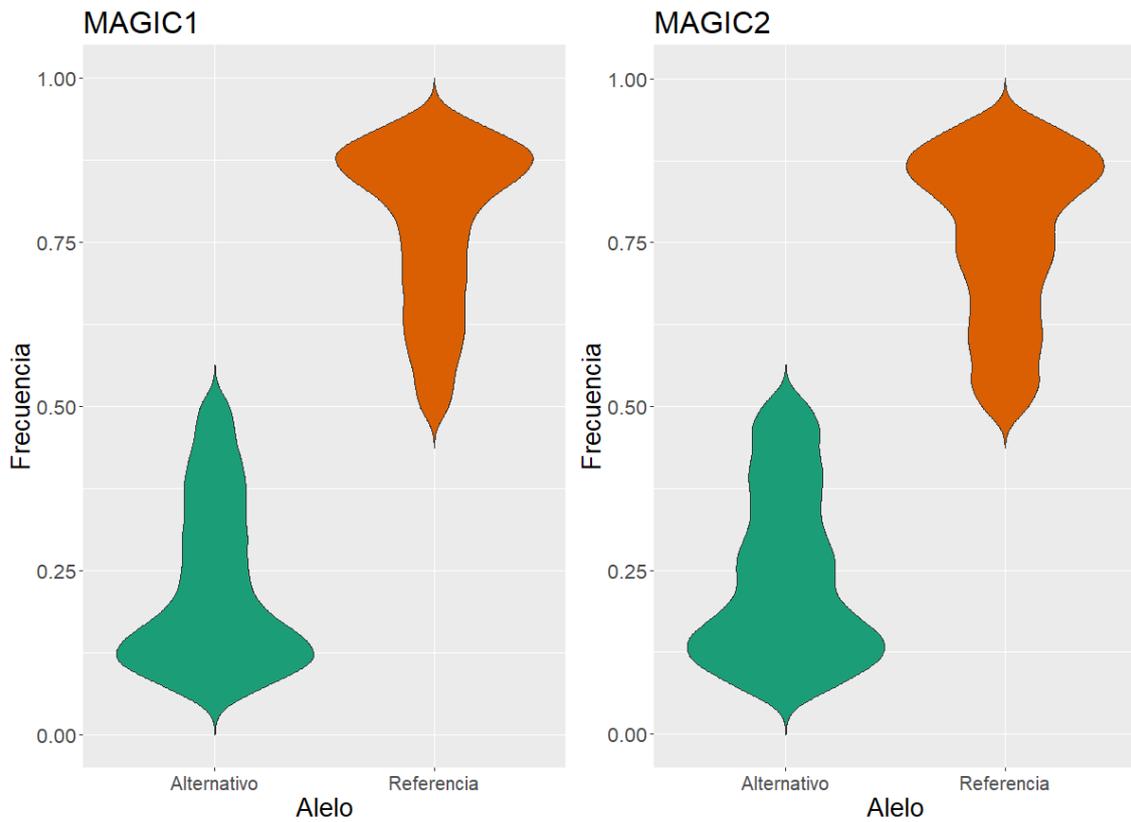
**Análisis de diversidad de las líneas parentales de las poblaciones MAGIC1 y MAGIC2**

Las estimaciones de diversidad genética entre las líneas parentales se muestran en la Tabla 2.5.

**Tabla 2.5.** Estimaciones de Heterocigosis esperada (He) y observada (Ho) para las líneas parentales de las poblaciones MAGIC1 y MAGIC2.

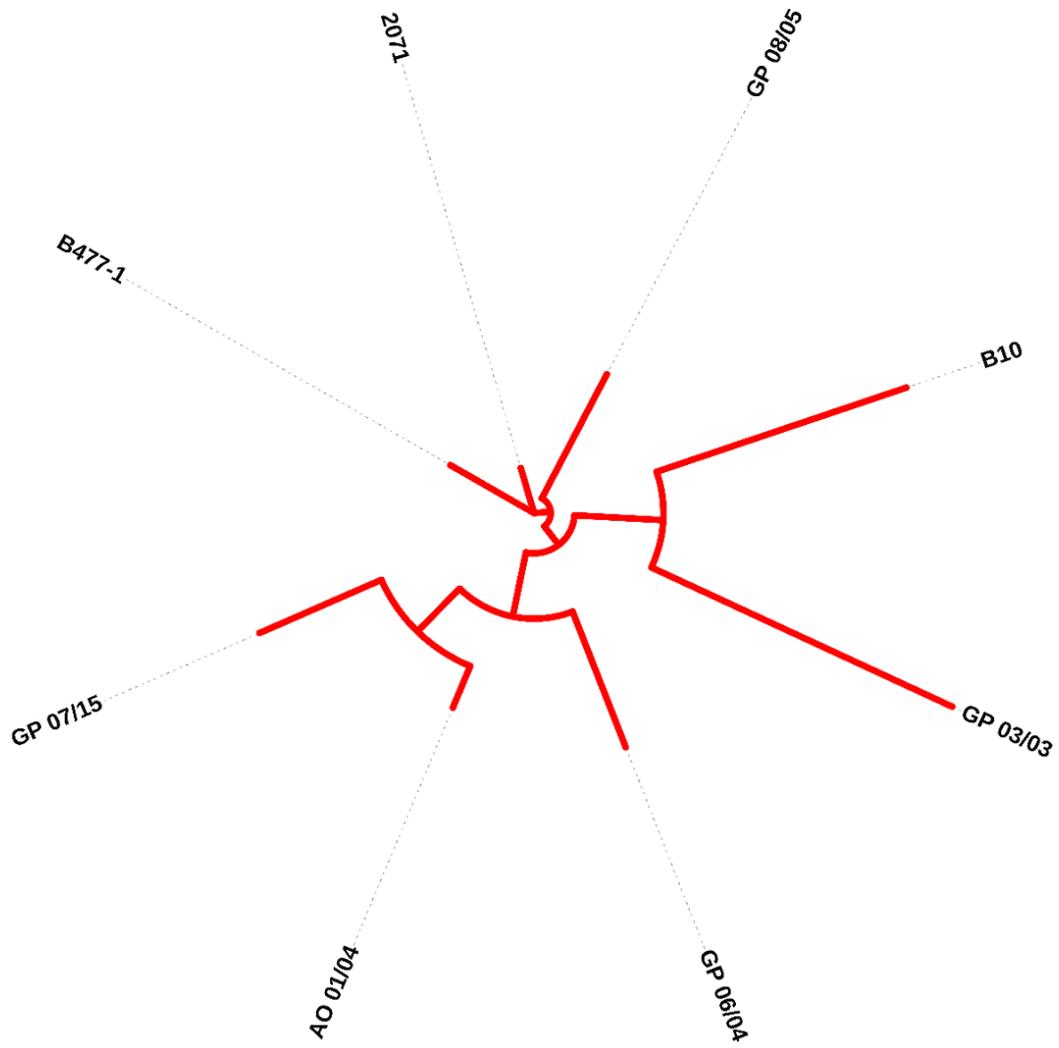
Población	He min	He media	He máx	Ho
MAGIC1	0,000	0,319	0,500	0,044
MAGIC2	0,000	0,338	0,500	0,043

Las frecuencias alélicas para el alelo de referencia y alternativo para cada población se grafican en la Figura 2.10.

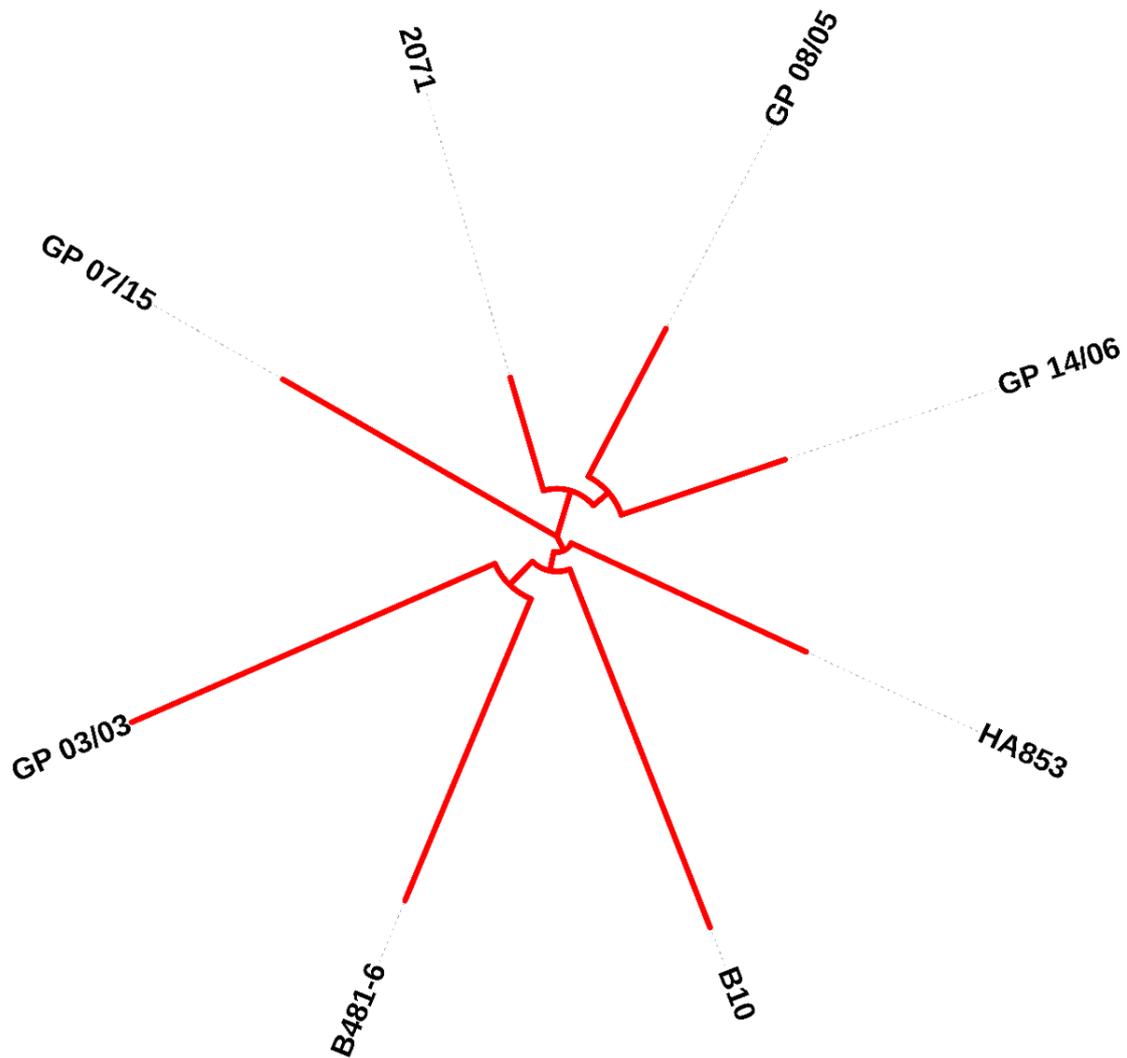


**Figura 2.10.** Frecuencias alélicas para el alelo alternativo y de referencia en ambas poblaciones.

Las relaciones entre las líneas parentales de las poblaciones MAGIC1 y MAGIC2 fueron graficadas utilizando el método *Neighbor-Joining* Figuras 2.11 y 2.12.



**Figura 2.11.** Filograma de Neighbor-Joining para las líneas parentales de la población MAGIC1.

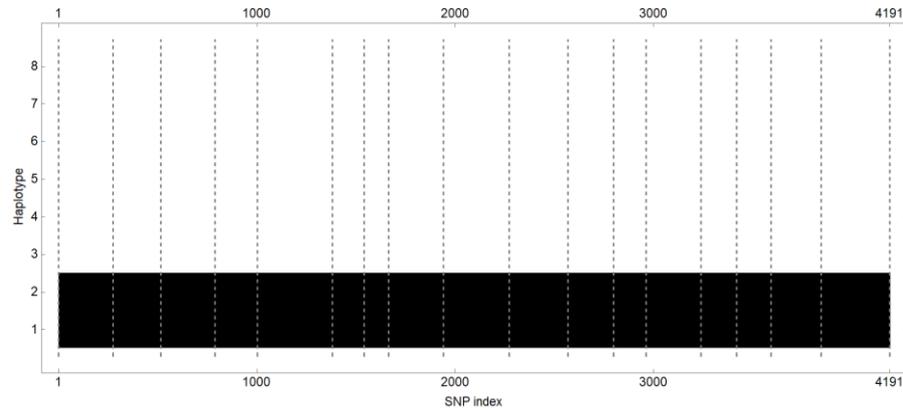


**Figura 2.12.** Filograma de Neighbor-Joining para las líneas parentales de la población MAGIC2.

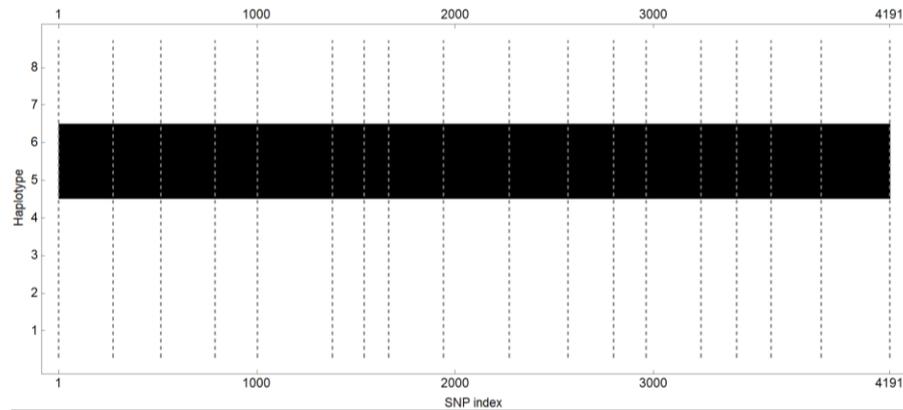
### Reconstrucción de bloques genómicos ancestrales en las poblaciones MAGIC1 y MAGIC2 a través de RABBIT

La reconstrucción de los bloques genómicos ancestrales para las poblaciones MAGIC1 y MAGIC2 pudo ser realizado con la implementación del paquete RABBIT. En las Figuras 2.13, 2.14, 2.15 y 2.16 se muestran las probabilidades posteriores de los 8 orígenes ancestrales para cada híbrido. Las Figuras 2.13 y 2.14 corresponden a los híbridos de la población MAGIC1, en el eje de las ordenadas se representan los haplotipos parentales y en el eje de las abscisas se representan las posiciones de los marcadores SNPs a lo largo del genoma. Las líneas discontinuas verticales delimitan el largo de cada uno de los 17 cromosomas. Los haplotipos parentales están codificados del 1 al 8 y corresponden a las líneas parentales AO 01/04, B477-1, 2071, GP 06/04, GP 03/03, B10, GP 08/05 y GP 07/15 respectivamente. Las Figuras 2.15 y 2.16 corresponden a los híbridos de la población MAGIC2, los haplotipos parentales también se representan en el eje de las ordenadas, codificados del 1 al 8 y

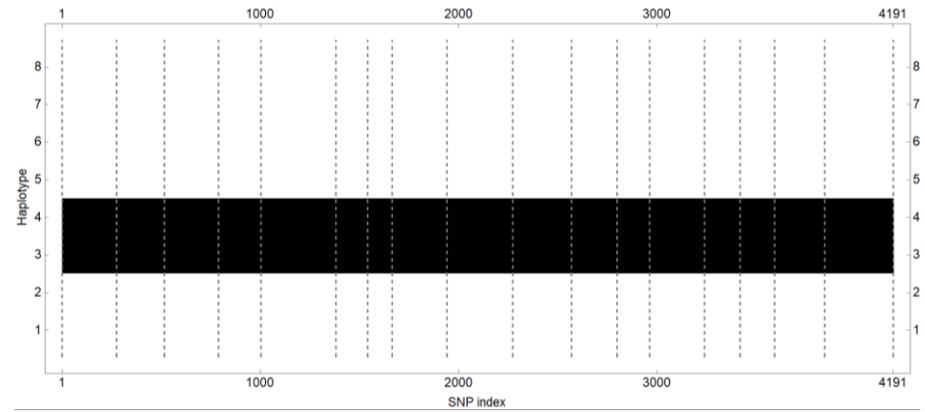
corresponden a las líneas parentales GP 14/06, B10, GP 03/03, B481-6, 2071, HA853, GP 08/05 y GP 07/15 respectivamente. Las probabilidades posteriores de los orígenes ancestrales se representan en una escala de grises, tomando valores entre 0 (color blanco) y 1 (color negro). Las distintas tonalidades de grises refieren a valores intermedios entre 0 y 1. La presencia de diferentes tonalidades de grises en ciertas regiones del genoma indica una menor precisión en la asignación del origen ancestral en esa zona. En las Figuras 2.14 y 2.16 se nota un menor tamaño de los bloques ancestrales debido a la mayor recombinación que se obtiene en los híbridos de 8 vías.



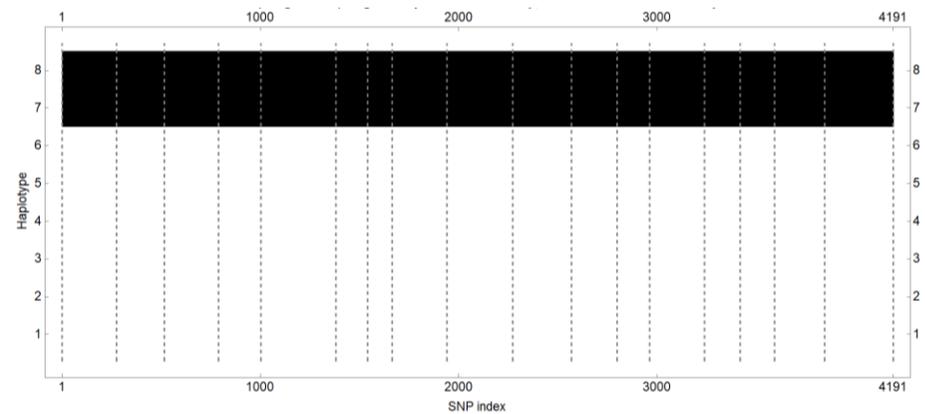
AO 01/04 x B477-1



GP 03/03 x B10

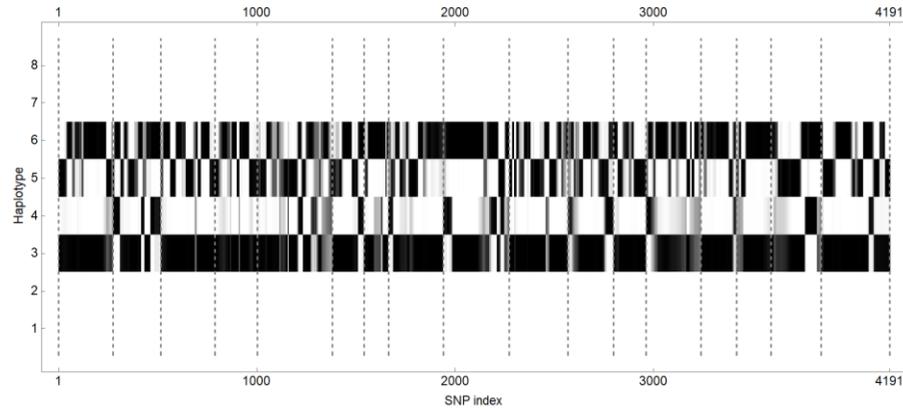


2071 x GP 06/04

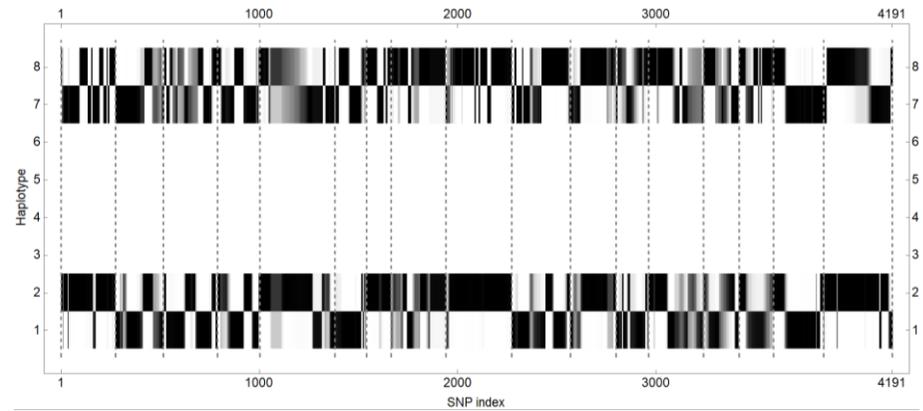


GP 08/05 x GP 07/15

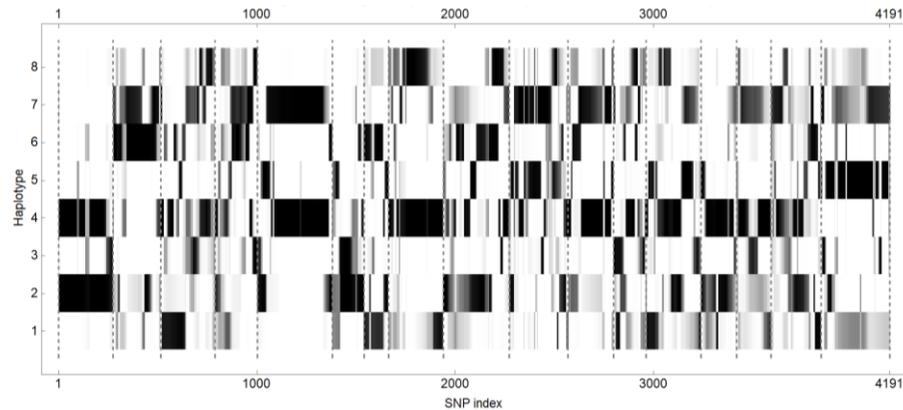
**Figura 2.13.** Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos de dos vías de la población MAGIC1. Las líneas verticales punteadas indican los límites de los 17 cromosomas. Las probabilidades se grafican en una escala de grises siendo el 0 = blanco y 1 = negro.



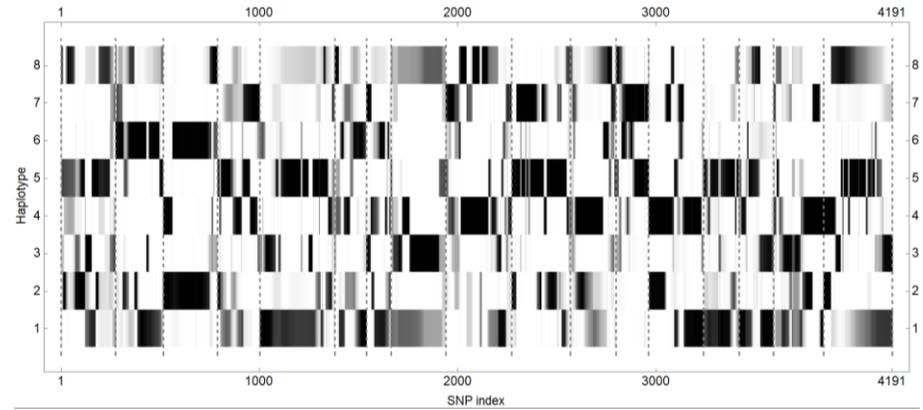
(2071 x GP 06/04) x (GP 03/03 x B10)



(GP 08/05 x GP 07/15) x (AO 01/04 x B477-1)

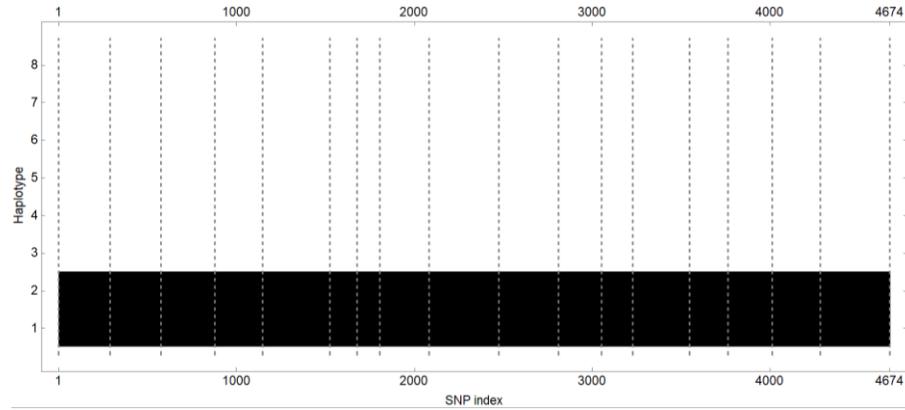


[(2071 x GP 06/04) x (GP 03/03 x B10)] x [(GP 08/05 x GP 07/15) x (AO 01/04 x B477-1)]

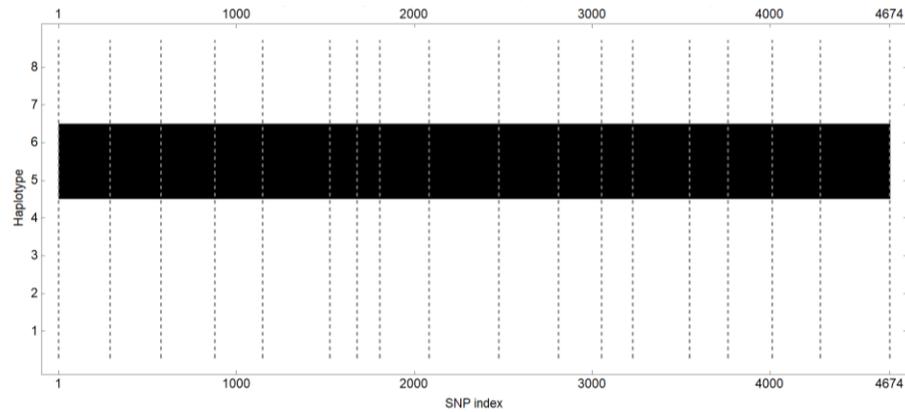


[(GP 08/05 x GP 07/15) x (AO 01/04 x B477-1)] x [(2071 x GP 06/04) x (GP 03/03 x B10)]

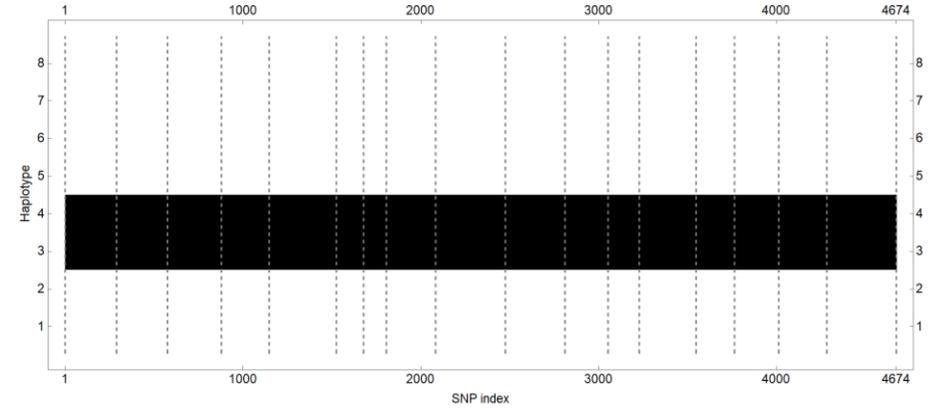
**Figura 2.14.** Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos de 4 y 8 vías de la población MAGIC1. Las líneas verticales punteadas indican los límites de los 17 cromosomas. Las probabilidades se grafican en una escala de grises siendo el 0 = blanco y 1 = negro.



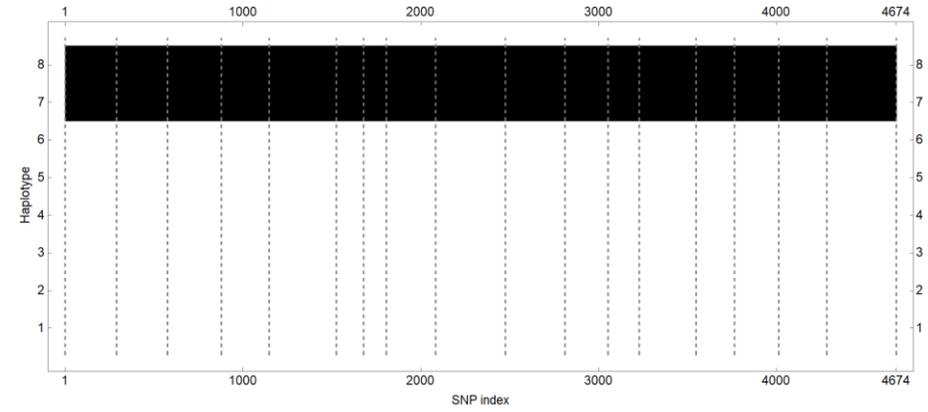
GP 14/06 x B10



2071 x HA853

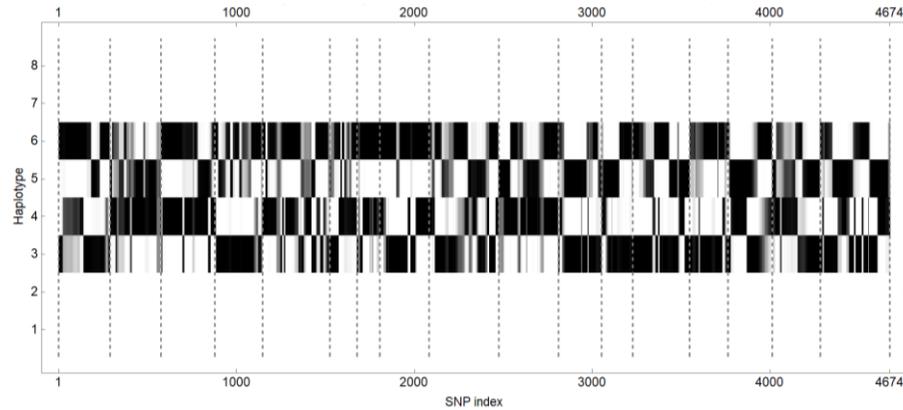


GP 03/03 x B481-6

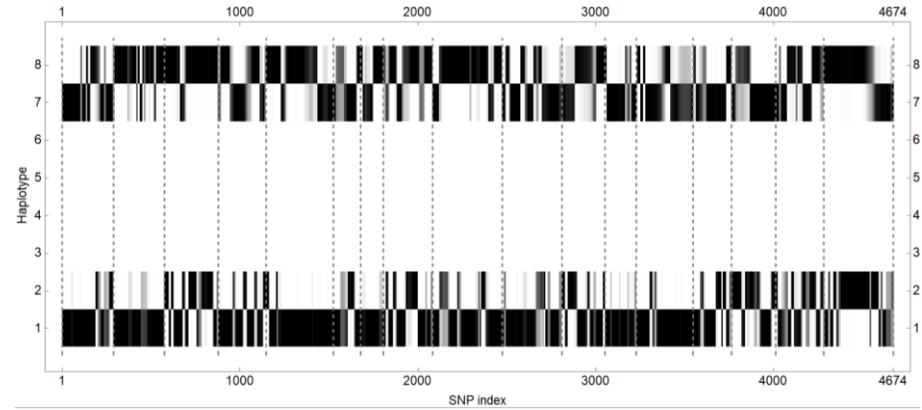


GP 08/05 x GP 07/15

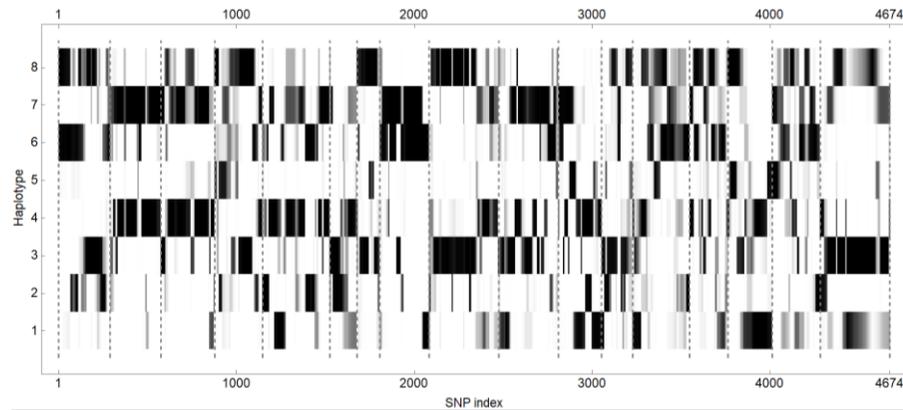
**Figura 2.15.** Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos d vías de la población MAGIC2. Las líneas verticales punteadas indican los límites de los 17 cromosomas. Las probabilidades se grafican en una escala de grises siendo el 0 = blanco y 1 = negro.



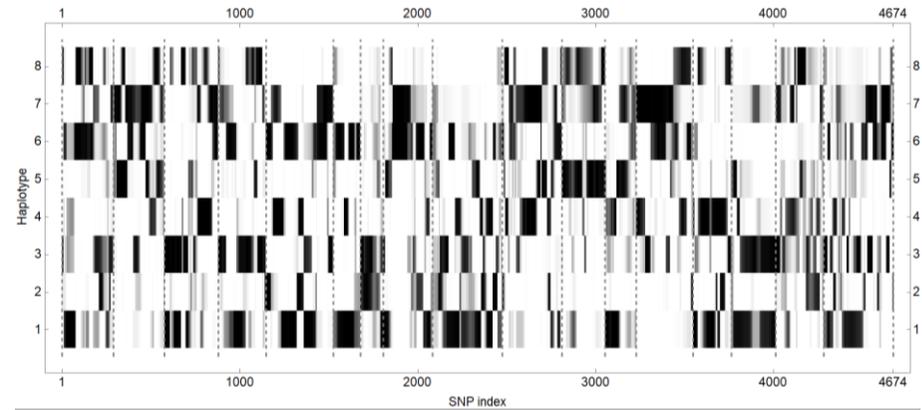
(2071 x HA853) x (GP 03/03 x B481-6)



(GP 08/05 x GP 07/15) x (GP 14/06 x B10)



[(2071 x HA853) x (GP 03/03 x B481-6)] x [(GP 08/05 x GP 07/15) x (GP 14/06 x B10)]



[(GP 08/05 x GP 07/15) x (GP 14/06 x B10)] x [(2071 x HA853) x (GP 03/03 x B481-6)]

**Figura 2.16.** Probabilidades posteriores de los ocho orígenes ancestrales para los híbridos de 4 y 8 vías de la población MAGIC2. Las líneas verticales punteadas indican los límites de los 17 cromosomas. Las probabilidades se grafican en una escala de grises siendo el 0 = blanco y 1 = negro.

## DISCUSIÓN

Los avances en las tecnologías de secuenciación de nueva generación (NGS) y la disminución de los costos de secuenciación automática han proporcionado nuevas oportunidades a los programas de mejoramiento (Kim et al., 2016). El genotipado por secuenciación (GBS) es una herramienta rápida y de bajo costo para la genotipificación de las poblaciones de mejoramiento, posibilitando el descubrimiento de marcadores moleculares, los estudios de diversidad genética, el análisis de ligamiento, los estudios de GWAS y la selección genómica en programas a gran escala (He et al., 2014).

La implementación de GBS como metodología para el genotipado ha sido ampliamente reportada en la caracterización de poblaciones multiparentales de arroz (Raghavan et al., 2017), sorgo (Ongom & Ejeta, 2018), algodón (Islam et al., 2016b), cebada (Hemshrot et al., 2019), poroto (Diaz et al., 2020), maíz (Anderson et al., 2018) entre otras, tanto para la caracterización de las líneas parentales y las progenies o solamente las progenies.

Filippi et al (2020) y Montecchia (2019) emplearon la metodología de ddRADseq para el genotipado de la PMA de girasol de INTA con resultados muy satisfactorios, obteniendo una matriz final de 34.309 SNPs. El mismo protocolo se aplicó en este trabajo reproduciendo los valores de calidad esperados tanto en la preparación de la librería (patrones de corte enzimático homogéneos, concentraciones adecuadas en los rangos “blanco” de selección de fragmentos, rendimiento adecuado de la amplificación de reacción de PCR) como así también en las pruebas de secuenciación realizadas en el equipo Illumina MySeq (resultados no presentados).

Una de las principales desventajas de la implementación de metodologías de GBS para el genotipado de poblaciones es la generación de datos faltantes. Para la población MAGIC1 la proporción de datos faltantes para la matriz de SNPs obtenida fue de 11,48 % y para la población MAGIC2 fue de 12,66 %. Entre las causas principales de la ocurrencia de datos faltantes están el nivel de cobertura y la profundidad de secuenciación alcanzada en cada región secuenciada. Los datos faltantes por baja cobertura pueden interpretarse como regiones que no han sido muestreadas en el proceso de GBS, aunque también pueden estar reflejando variabilidad real relacionada con polimorfismos de presencia/ausencia, polimorfismos en los sitios de corte de las enzimas y/o diferencias en la metilación. Otras causas están relacionadas con la complejidad de la librería lo que está ligado a las características del genoma de la especie, el girasol posee un genoma de gran tamaño (3,6 Gb), tres veces mas grande que el genoma de la colza y 8 veces mayor que el genoma del arroz (Badouin et al., 2017; Dimitrijevic & Horn, 2018) y además está constituido en un 85 %

por regiones repetitivas lo cual condiciona el tipo de enzimas de restricción a seleccionar, el tipo de enzimas seleccionadas para conseguir la reducción de esa complejidad (enzimas con sitios de reconocimiento más largos producirán menos cortes) y la sensibilidad de esas enzimas a metilación. (Poland & Rife, 2012b; Torkamaneh et al., 2018). El problema de los datos faltantes puede ser corregido en parte mediante el incremento de la profundidad de secuenciación y/o mediante metodologías de imputación de datos. La primera alternativa puede alcanzarse aumentando la profundidad de secuenciación o reduciendo el nivel de multiplexado, lo que resultaría en un incremento del costo por muestra. La alternativa de imputación resulta confiable dada la eficiencia que han demostrado las metodologías implementadas en distintos trabajos (Merino, 2018; Poland & Rife, 2012b; Torkamaneh et al., 2018).

En este capítulo, la cobertura del genoma y la distribución de marcadores por unidad de distancia física mostró regiones amplias desprovistas de marcadores, siendo el cromosoma 12 para ambas poblaciones el que tuvo menor densidad de marcadores, observándose 1,09 SNP/Mb para la población MAGIC1 y 1,17 SNP/Mb para la población MAGIC2. Estos resultados coinciden con lo observado en otros estudios que aplican GBS en poblaciones biparentales de girasol (Talukder et al., 2014; Celik et al., 2016; Qi et al., 2016; Montecchia, 2019). En las poblaciones multiparentales se esperaría una mayor densidad de marcadores por el mayor número de parentales en comparación con las poblaciones biparentales citadas anteriormente. Cabe mencionar, que este número puede aumentar a instancias de imputar datos faltantes. En esta instancia del trabajo, y dado el número inicial de líneas genotipadas no se aplicaron métodos de imputación y se utilizaron únicamente los marcadores con una representación mínima del 20 %. En estas condiciones, el menor número de SNPs obtenido puede atribuirse al elevado porcentaje de datos faltantes que tienen algunas de las líneas parentales (ver Anexo digital “N°\_Lecturas” y “Prop\_datos\_faltantes\_xmuestra”).

La utilización de herramientas bioinformáticas como SnpEff (Cingolani et al., 2012) resultan de gran utilidad para poder caracterizar de forma más acabada las regiones genómicas exploradas por los marcadores moleculares identificados. Para ambas poblaciones, los polimorfismos obtenidos fueron ubicados mayoritariamente en regiones intergénicas con valores por encima del 65 %. Similares resultados fueron obtenidos por Malmberg et al (2018) en un estudio de diversidad genética en 633 variedades mejoradas de colza provenientes de bancos de germoplasma de los principales países productores. Los SNPs en regiones codificantes obtenidos para ambas poblaciones condujeron a un mayor porcentaje de mutaciones no sinónimas que sinónimas, similares resultados fueron obtenidos en otras

especies de plantas cultivadas como en palma aceitera (Pootakham et al., 2015), sorgo (Zheng et al., 2011) y arroz (Jeong et al., 2013; Subbaiyan et al., 2012). Por su parte, Aballay et al (2021) desarrollaron un panel de marcadores SNPs por GBS para una colección de variedades de duraznero y reportaron resultados similares con respecto a la mayor proporción de SNPs de efecto modificador según la clasificación propuesta por Cingolani et al (2012). Montecchia (2019) en una caracterización genotípica de la PMA de girasol de INTA reportó un porcentaje superior al 75 % de SNPs en regiones intergénicas y con una relación de mutaciones no sinónimas y silenciosas similar a la presentada en este capítulo para las poblaciones MAGIC. La aproximación del posible efecto de los polimorfismos mapeados al genoma de referencia y la inferencia de su efecto en relación a las funciones biológicas asociadas a las estructuras génicas adyacentes, resulta un recurso valioso para la formulación de hipótesis una vez obtenidas asociaciones genotipo-fenotipo o QTLs identificados por mapeo de ligamiento.

Los conjuntos de marcadores SNPs obtenidos fue aplicado al análisis de parentesco de los cruzamientos realizados para la formación de las poblaciones multiparentales. Este tipo de análisis resulta de gran interés para la comprobación de los cruzamientos y no ha sido reportado en otras poblaciones multiparentales publicadas hasta el momento. Asimismo, tiene interesantes prestaciones para su aplicación en los programas de mejoramiento por sus contribuciones en el control parental, en la reducción de recursos para mantenimiento de parentales, en la estimación de los valores de mejora, en la detección de contaminaciones y errores en las labores fitotécnicas (Norman et al., 2018). Asimismo, es de utilidad para analizar los efectos y maximizar la ganancia genética en los programas de mejoramiento (Norman et al., 2020). La utilización del análisis de parentesco en los programas de mejoramiento genético vegetal se presentó por primera vez en pino taeda (*Pinus taeda* L.) (Lambeth et al., 2001). Posteriormente, también se aplicó en otras especies forestales, como en *Eucalyptus* sp. (Cupertino et al., 2009), en pino silvestre (*Pinus sylvestris* L.) (Wang et al., 2010), e industriales como olivo (Baruca Arbeiter et al., 2014), en vid (Mihaljevic et al., 2020) y para identificar la tasa de contaminación del polen en pino taeda (Vidal et al., 2015). Asimismo, este análisis fue utilizado con éxito en especies forrajeras como trébol rojo (Riday, 2011) y alfalfa (Riday et al., 2013) y en cultivos hortícolas como el repollo, implementándose como pieza clave de un protocolo innovador de mejoramiento genético para el incremento de las combinaciones híbridas a evaluar reduciendo el número de operaciones fitotécnicas (Rudolf-Pilih et al., 2019). En este capítulo el análisis de parentesco de las PMP se realizó utilizando el paquete “Apparent” (Melo & Hale, 2019). La utilización de este paquete se describió recientemente en un análisis de parentesco en pino radiata (*Pinus radiata* L.) (Graham et al.,

2019). Los resultados obtenidos en las PMP de girasol mostraron los valores más bajos de distancia genética para los cruzamientos efectivamente realizados. Si bien para la población MAGIC1 los resultados no fueron significativos, si lo fueron para los cruzamientos de dos vías de la población MAGIC2. La falta de significancia de las tríadas puede deberse, según los resultados descritos por Graham et al (2019), al número reducido de individuos analizados lo cual influye en el tamaño de la brecha entre las tríadas espurias y verdaderas. Por otro lado, el éxito de los cruzamientos realizados en las etapas de desarrollo de las poblaciones MAGIC1 y 2 puede inferirse a partir del aumento del nivel de heterocigosis detectado mediante la caracterización genotípica en las progenies híbridas.

El estudio de la diversidad entre las líneas fundadoras es clave para conocer el potencial de las poblaciones para su utilización en el mapeo fino de QTLs y para la utilización de las progenies derivadas en fondos genéticos de elite para su aplicación en los programas de mejoramiento. Las estimaciones de diversidad genética (heterocigosis esperada media) obtenidas entre las líneas parentales (0,319 y 0,338 para la MAGIC1 y MAGIC2 respectivamente) mostraron valores superiores a los reportados por Mackay et al., (2014) en una población multiparental de trigo a partir de líneas fundadoras *elite*, quienes obtuvieron valores de 0,186. La diversidad presente entre las líneas parentales, considerando que los parentales son líneas pertenecientes al programa de mejoramiento de INTA y no a *landraces* o silvestres emparentados como en otras poblaciones multiparentales publicadas (Maurer et al., 2015), resulta muy interesante por su potencial para el mapeo de QTL como para su implementación directa en el programa de mejoramiento.

Los niveles de heterocigosis observada en las líneas parentales fueron entre 3 a 5 %, siendo similares a los obtenidos en trabajos previos del grupo (Filippi, et al., 2020; Montecchia, 2019). Estos bajos valores obtenidos de heterocigosis en las líneas parentales utilizadas confirman la pureza genética del material de partida para la formación de las poblaciones.

La reconstrucción de los bloques genómicos ancestrales en las poblaciones multiparentales es un paso previo para la realización de estudios de mapeo de QTLs. Los resultados de la reconstrucción, como las probabilidades posteriores de los orígenes ancestrales en los QTLs putativos, se utilizan como predictores genéticos en un modelo mixto de mapeo de QTL (Huang et al., 2011). Una metodología confiable para obtener valores precisos de IBD es fundamental para inferir los orígenes parentales y realizar el mapeo de QTLs basado en las probabilidades de IBD, sin embargo, hasta el momento solo se dispone de un reducido número de metodologías que se limitan a diseños específicos de PMP (Broman et al., 2019a; Verbyla et al., 2014). En este trabajo se empleó el paquete RABBIT publicado por Zheng (2015a) para

la reconstrucción de los bloques ancestrales utilizando la información obtenida en la caracterización genotípica. Una de las ventajas de la utilización de este paquete es que permite la evaluación de individuos en etapas intermedias de desarrollo de las poblaciones, permitiendo la inclusión de individuos con alto grado de heterocigosis. Esta característica, dado el extenso período de tiempo necesario para la obtención de líneas recombinantes en homocigosis, la convierte en una herramienta de gran utilidad para estudios de mapeo de QTLs en generaciones intermedias en poblaciones multiparentales. RABBIT ha sido utilizado en poblaciones multiparentales de papa diploide (Korontzis et al., 2020), papa tetraploide (Amadeu et al., 2020) y en cebada (Liller et al., 2017) con resultados satisfactorios. Por su parte, Li et al (2021) en un estudio de caso sobre diferentes diseños de poblaciones PMP simuladas y reales, demostraron que los orígenes ancestrales pueden ser ambiguos en los *loci* no segregados, faltantes o mal genotipados basados en la información de las probabilidades de idéntico por estado (IBS), mientras que las probabilidades de IBD inferidas por la información genómica y de *pedigree* redujeron la incertidumbre. Además, los autores reportaron que el método de modelo mixto multi-QTL basado en IBD, es robusto a través de una variedad de diseños de PMP y patrones de segregación de alelos en comparación con el método de mapeo por asociación, y en la mayoría de los casos, superó o se comportó al menos tan bien como otras herramientas desarrolladas para diseños PMP específicos en términos de potencia de mapeo y resolución. Asimismo, Bhat et al (2021) en una revisión reciente sobre la implementación de los haplotipos para estudios de QTLs y selección genómica resaltaron las ventajas y la mayor precisión que tiene su utilización en comparación con los SNPs. Estas ventajas de los haplotipos radican en el mayor contenido de información polimórfica que ofrecen debido a su naturaleza multialélica, la mayor información que proveen sobre posibles mutaciones recientes y el aporte al conocimiento de efectos genéticos epistáticos. Esta última ventaja, surge de que el rango heredable de las combinaciones alélicas de todas las variantes dentro de una región del QTL está determinado por la descendencia local de los haplotipos, proporcionando una mayor comprensión de la variación genética y, por lo tanto, permite una mayor potencia para la detección de los QTLs (Zhang et al., 2014). Además, Bhat et al (2021) plantean que la implementación de los haplotipos en estudios de mapeo de QTLs y en modelos de selección genómica será cada vez más frecuente, no sólo por su mayor precisión sino también gracias a el advenimiento de las nuevas tecnologías de secuenciación de fragmentos largos, que facilitan el descubrimiento de los haplotipos. En el mismo sentido, Zhang et al (2014) sostienen que el uso de los haplotipos inferidos en lugar de los genotipos observados para estudios de QTLs en PMP confiere otras importantes ventajas, entre las que se incluyen: una mayor solidez frente a los errores de

genotipado debido a la complementación de información entre los marcadores y el modelado implícito de la caída del desequilibrio de ligamiento, permitiendo una mayor resolución en el mapeo de los QTLs.

Por otro lado, una aplicación muy interesante de la reconstrucción de los bloques ancestrales es que el genotipo de los fundadores puede proyectarse en las progenies, siendo importante destacar que las variantes pueden ser imputadas incluso cuando no están directamente genotipadas en ninguna de las *RILs*. Por lo cual, los datos de genotipado de baja cobertura de las *RILs* pueden utilizarse para inferir los mosaicos y, a continuación, los datos genómicos densos de los fundadores pueden proyectarse en ellos para imputar todas las variantes (Zheng, 2015a; Zheng et al., 2014). Por otra parte, las probabilidades de IBD no solo resultan útiles como predictores genéticos para el mapeo de QTL, sino también valiosas en la construcción de mapas de consenso. El proceso tradicional de construcción de mapas de consenso puede ser tedioso en el contexto de las poblaciones multiparentales, ya que incluye la depuración de marcadores, la agrupación, el ordenamiento, y la integración de mapas (Endelman & Plomion, 2014; W. Li et al., 2021; Taylor, 2018; Y. Wu et al., 2008).

La caracterización genotípica de las líneas parentales de las poblaciones MAGIC presentada en este capítulo resulta de suma importancia para conocer el potencial para su utilización en estudios de mapeo de QTLs. En esta instancia de desarrollo de las poblaciones, los datos genotípicos contribuyeron a testear la calidad de las líneas parentales altamente homocigotas y la reconstrucción de haplotipos indicando la efectividad de los cruzamientos y el potencial de las poblaciones para el mapeo fino de QTLs. La implementación de una estrategia para el chequeo de los cruzamientos y la corroboración de los mismos permiten el control del proceso de desarrollo de las poblaciones multiparentales y la posibilidad de aplicación del chequeo de cruzamientos específicos como método de rutina en el programa de mejoramiento. El entrenamiento obtenido en la utilización de paquete RABBIT para la reconstrucción de los bloques genómicos ancestrales es un paso importante y necesario para los estudios futuros de QTLs que se pretenden realizar con el desarrollo de estas poblaciones multiparentales.

--

---

---

## **CAPÍTULO 3: Caracterización fenotípica de las familias F<sub>2</sub> de las poblaciones multiparentales**

---

---

## INTRODUCCION

En este capítulo se presenta la caracterización fenotípica realizada a las familias F<sub>2</sub> de las poblaciones MAGIC en desarrollo. La población MAGIC2 fue analizada frente a la Marchitez anticipada causada por el hongo de suelo *Verticillium dahliae*. Ambas poblaciones fueron fenotipificadas para caracteres de interés agronómico como floración, altura de planta, inclinación de la lámina de las hojas y el capítulo.

### 1. Evaluación frente a Marchitez anticipada por *V. dahliae* en girasol

#### 1.1 Importancia y descripción de la Marchitez anticipada por *V. dahliae* en girasol

La marchitez anticipada y quebrado del tallo causada por *V. dahliae* (MV), también denominada Verticilosis, es una de las principales enfermedades del cultivo de girasol en la Argentina (Pereyra & Escande, 1994), causando pérdidas de rendimiento de alrededor del 30% en híbridos comerciales susceptibles y pudiendo llegar hasta 73% en campos altamente infestados (Pereyra et al., 1999). El quebrado de tallos no tan frecuente en los cultivares actuales, pero cuando ocurre causa pérdidas de gran magnitud. La marchitez y el secado anticipado de las plantas provocan una disminución del rendimiento de menor importancia, pero más difundida y frecuente (Pereyra & Escande, 1994).

La MV es una enfermedad monocíclica, cuya fuente de inóculo son los microesclerocios formados por el hongo como estructura de resistencia, que persisten viables en suelo por varios años (entre 10 y 15), pudiendo infectar un amplio rango de hospedantes alternativos. El área endémica de la enfermedad abarca aproximadamente 1.200.000 hectáreas, cubriendo el sur de Buenos Aires y el centro-Este de La Pampa. El inóculo está presente en más del 50% de la superficie total cultivada con girasol en la República Argentina, siendo esta distribución y prevalencia una condición excepcional a nivel mundial (Montecchia, 2019; Montecchia et al., 2021).

#### 1.2 Agente causal y taxonomía

El agente causal de la MV es el hongo de suelo *V. dahliae*, perteneciente al Filo: *Ascomycota*, Sub-filo: *Pezizomycotina*, Clase: *Sordariomycetes*, Orden: *Phyllachorales*, género: *Verticillium*.

El género *Verticillium* representa uno de los más patogénicos a nivel mundial afectando un amplio espectro de especies cultivadas y silvestres. Es preponderante en regiones cálidas y templadas, aunque existen especies patogénicas reportadas en áreas tropicales y subtropicales. Se pueden distinguir siete especies patogénicas preponderantes afectando árboles, plantas herbáceas, especies cultivadas y hongos: *V. dahliae* Kleb., *V. albo-atrum*

(Reinke et Berth.), *V. nigrescens* (Pethybr.), *V. nubilum* (Pethybr.), *V. tricorpus* (Isaac.), *V. theobromae* (Turc.) y *V. fungicola* (Preuss). De estos, las especies polífagas causales de marchitamientos más destacables resultan *V. dahliae* Kleb. y *V. albo-atrum*, dada su amplio rango de hospedantes, aproximadamente 200 especies dicotiledóneas, su importancia en la agricultura y la bibliografía científica generada sobre ellas (Pegg & Brady, 2002).

*V. dahliae* es un hongo hemibiótrfico de suelo con un amplio espectro de especies dicotiledóneas hospedantes. Causa enfermedades vasculares asociadas a marchitez en más de 200 especies dicotiledóneas. El micelio del hongo es hialino, simple o ramificado, septado y multinucleado. Se reproduce de forma asexual y la producción de conidios ovoides se genera a través de conidióforos verticilados, que son el rasgo morfológico característico del género. Se propaga por medio de microesclerocios melanizados que constituyen la fuente de inóculo del patógeno y su estructura de resistencia, pudiendo permanecer viables en suelo por períodos de entre 10 y 15 años (Fradin & Thomma, 2006).

### **1.3 Epidemiología y ciclo patogénico de *Verticillium dahliae***

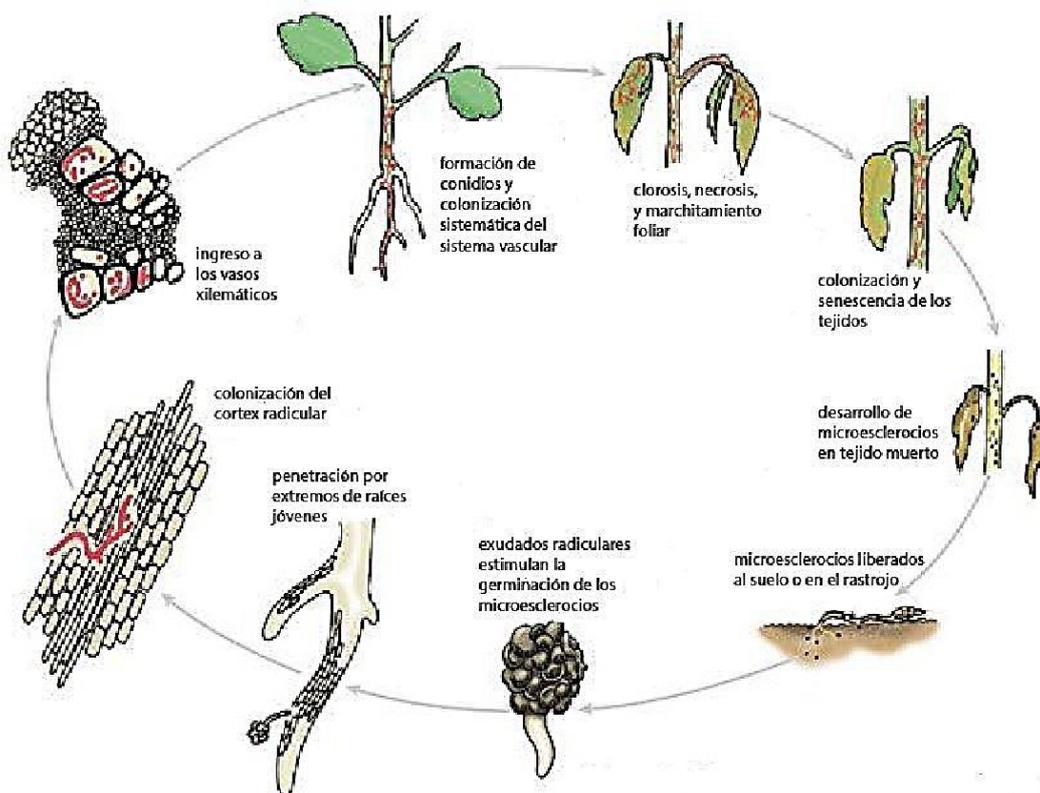
La MV es favorecida por suelos húmedos y un rango de temperaturas entre 21 a 27°C. *V. dahliae* es un patógeno de suelo que coloniza a las plantas por la raíz, y causa una enfermedad monocíclica. El ciclo de vida del hongo puede dividirse en tres fases: (a) fase de dormición, (b) fase parasítica y (c) fase saprofítica (Figura 3.1; Fradin & Thomma, 2006).

Durante la fase de dormición los microesclerocios del hongo permanecen en el suelo sin germinar por estar inhibidos. La dormición es interrumpida por el incremento de las concentraciones de carbono en el entorno del microesclerocio y/o por la liberación de exudados radiculares de la planta (Huisman, 1982; Mol et al., 1995; Olsson & Nordbring-Hertz, 1985; Schreiber & Green, 1963). Una vez rota la dormición, todas las células del microesclerocio pueden germinar, lo que incrementa las chances del patógeno de alcanzar un hospedante. El rango de alcance del microesclerocio es de 300 µm por lo que la planta es la que tiende a alcanzar al inóculo (Huisman, 1982).

Para iniciar la fase parasítica el hongo debe alcanzar las raíces de una planta y penetrar a través de las puntas de raíces secundarias o raíces jóvenes, o por el punto de unión de raíces secundarias. La endodermis representa una barrera preformada para el avance del hongo hacia los vasos xilemáticos, por lo que requiere entrar por tejidos jóvenes en los que todavía no fue formada, o por heridas (Bowers et al., 1996; Huisman, 1982; Pegg & Brady, 2002; Schnathorst, 1981).

Luego de superar la endodermis, el micelio del hongo rápidamente reinicia un ciclo de esporulación y liberación de conidios, este proceso es conocido como gemación y es la base del proceso infeccioso del hongo. Los conidios liberados son transportados mediante la corriente xilemática y encuentran puntos de detención en los extremos de los vasos, o en las perforaciones laterales de los mismos, llamados “sitios de detención”. Desde estos puntos, el conidio germina, desarrolla micelio intentando colonizar vasos adyacentes y reinicia el ciclo de esporulación (Bishop & Cooper, 1983).

La tercera etapa del ciclo se cumple durante la necrosis de tejidos o durante la senescencia de la planta, en donde el hongo entra en estado saprofítico. Para este momento, no sólo los vasos sino también las raíces y los brotes son alcanzados por el hongo. *V. dahliae* produce una enorme cantidad de microesclerocios que son liberados al suelo con los restos vegetales. El hongo puede también pasar el invierno como micelio en hospedantes alternativos perennes, o en semillas en algunos casos de infección severa.



**Figura 3.1.** Ciclo patogénico de la Marchitez causada por *V. dahliae*. Adaptado de American Phytopathology Society (<http://www.apsnet.org>).

#### 1.4 Sintomatología de la MV en girasol

La sintomatología de la enfermedad tiene un sentido acrópeto, comenzando a observarse en los estratos foliares inferiores. El primer síntoma se manifiesta como una mancha clorótica

internerval entre las venas principales de la hoja, generalmente de uno de los lados de la vena media (unilateral), el avance de la enfermedad por los haces xilemáticos tiende a hacerse visible a través de la sintomatología progresiva de las hojas alimentadas por un mismo haz vascular. Las manchas cloróticas tienden a un amarillamiento progresivo y coalescente que concluye en tejido necrótico con un halo clorótico. Los tejidos afectados adquieren un color castaño y mueren, y dependiendo de la susceptibilidad del cultivo, la hoja morirá o no (Figura 3.2.A y 3.2.B). En cultivos medianamente tolerantes la enfermedad puede quedar contenida por los haces vasculares observándose síntomas unilateralmente a nivel planta (Figura 3.2.C y 3.2.D), en cultivares susceptibles la contención del haz es vulnerada rápidamente y se observan síntomas en una mayor proporción de hojas, en todos los ortósticos del tallo (Figura 3.2.A). La tolerancia al avance de la sintomatología se manifiesta en el grado de contención de los síntomas, a nivel hoja mediante la superficie necrosada y a nivel planta mediante la cobertura de ortósticos y estratos. En base a este conjunto de síntomas se define la severidad de la enfermedad en el cultivar observado. En líneas susceptibles puede observarse una banda negra en la epidermis del tallo del lado afectado (Pereyra & Escande, 1994).





**Figura 3.2.** Sintomatología de la Marchitez causada por *V. dahliae*. **A:** Planta susceptible afectada hasta el estrato superior. **B; C; D:** Plantas medianamente tolerantes con síntomas en distintos estratos.

El taponamiento de los tejidos conductivos y la liberación de toxinas del hongo generan un impacto sobre el área foliar, la planta manifiesta un achaparramiento general, un menor diámetro de capítulo, adelanto de la senescencia, marchitamiento generalizado y, en casos de susceptibilidad elevada, quebrado del tallo y/o muerte de la planta. El desprendimiento de la médula y el recubrimiento de ésta por microesclerocios representan síntomas diagnósticos de la enfermedad a la madurez del cultivo (Montecchia, 2019).

La marchitez, la reducción del índice de área foliar y el secado anticipado comprometen íntegramente los componentes de rendimiento del cultivo de girasol. Esta enfermedad provoca una reducción de la interceptación de radiación fotosintéticamente activa (RFA), de la eficiencia en el uso de la misma y acorta el período de llenado de granos. El nivel de manifestación de síntomas se relaciona directamente con las pérdidas de rendimiento y de contenido de aceite en grano, siendo el peso del grano la variable más afectada (Pereyra et al., 1999; Pereyra & Escande, 1994).

### 1.5 Evaluación a campo de la respuesta frente a MV en girasol

La respuesta frente a la MV en girasol ha sido estudiada usando las variables incidencia y severidad. La primera se evalúa a través del conteo de plantas sintomáticas sobre el total de plantas evaluadas en la parcela. La segunda evalúa el nivel de sintomatología expresado por cada planta a la fecha de evaluación. La Incidencia refleja la capacidad de la planta de evitar la infección del patógeno, definiéndose como la resistencia a la penetración y colonización del patógeno. Esta variable presenta, por su naturaleza binaria, una distribución de probabilidades binomial. La severidad, refleja el grado de resistencia a la dispersión, o avance, del patógeno una vez dentro de los tejidos conductivos y es estimada en base a una escala fitopatológica de seis niveles (0= planta sin síntomas, 5= planta marchita, o con síntomas en todos sus estratos foliares). De estas dos variables y de las distintas fechas de evaluación, se desprenden otras que se detallan en los materiales y métodos de este Capítulo.

Una particularidad de los hongos fitopatógenos de suelo es que interactúan directamente con la variación ambiental y con el ambiente rizosférico en donde son condicionados por la actividad metabólica de las plantas y de otros microorganismos (Turner et al., 2013). Por lo que el solapamiento en el espacio de las poblaciones del patógeno y del cultivo son definidos por la interacción de diversos factores y, a la vez, críticos en la determinación de la ocurrencia de la enfermedad y su dinámica de desarrollo (Huisman, 1982).

Un problema del estudio de la resistencia a patógenos de suelo es su distribución discontinua y no aleatoria. Esta distribución en parches que causa los típicos manchones observados a campo, es propia de la subsistencia del inóculo primario; y de las limitaciones de dispersión que generalmente presentan los hongos de suelo (Nelson, 1973; Thrall & Burton, 1997). En el caso de *Verticillium dahliae* la ausencia de un ciclo sexual limita la dispersión del hongo a sus estructuras de resistencia. Bajas cargas de inóculo en regiones particulares pueden conducir a la interpretación errónea de comportamiento resistente de cultivares susceptibles (Missonnier et al., 2017). Por ello el fenotipado de enfermedades fúngicas causadas por hongos de suelo se realiza mediante estudios a campo sobre lotes con historia de la enfermedad o mediante inoculación artificial (Postman et al., 2010).

Una manera de aproximarse a la variabilidad espacial local es mediante el uso de parcelas control. Estas son formadas por genotipos cuya respuesta frente a la enfermedad es conocida y tiene un comportamiento estable, a menudo son materiales con una susceptibilidad media y permiten relevar la presencia o la ausencia de condiciones favorables (ambientales y epidemiológicas/patogénicas) para el desarrollo de la enfermedad (Missonnier et al., 2017; Montecchia, 2019; Montecchia et al., 2021). La combinación del diseño experimental y la

implementación de modelos mixtos, con los factores de diseño (Repeticiones, Bloques, Filas, Columnas) como aleatorios, permite relevar las covarianzas de las mediciones no independientes dentro de los niveles de los factores de diseño y obtener estimaciones más precisas de las fuentes de variación a lo largo del espacio en un experimento (Barreto et al., 1997; Bolker, 2008).

El modelado de datos experimentales de las medidas fenotípicas en el marco teórico de los modelos lineales generalizados mixtos permite realizar el análisis de datos con estructuras de dependencia, desbalances y falta de normalidad, como los que se presentan en este trabajo. El objetivo de estos modelos es despejar la variabilidad ambiental (error experimental) de las mediciones realizadas sobre las unidades experimentales, a fin de poder estimar precisamente la contribución de las fuentes de variación que se desea estudiar (genotipo y factores de diseño experimental). La varianza ambiental, propia de cualquier experimento realizado en condiciones de cultivo, se manifiesta a través de plantas o parcelas experimentales que presentan aleatoriedad para la variable respuesta evaluada, aún en condiciones de homogeneidad genética.

## **2. Impacto del daño por aves en el cultivo de girasol**

El cultivo de girasol es uno de los más vulnerables al ataque de aves en todo el mundo, entre las principales especies depredadoras en Sudamérica y en Argentina podemos mencionar a la “Cotorra argentina” *Myiopsitta monachus*, a la “Paloma torcaza” *Zenaida auriculata*, a la “Paloma manchada” *Columba maculosa* y a la “Paloma picazuró” *Columba picazuro*. El daño producido por la cotorra se da a partir de la etapa de floración y en la etapa de madurez del cultivo. Por su parte, las palomas inciden en el período de emergencia del cultivo y también provocan daños a final de ciclo, cuando los aquenios están próximos al contenido de humedad comercial para cosecha (Sausse & Lévy, 2021).

Los daños provocados por las aves generan altas pérdidas económicas. En Argentina, un estudio realizado por la FAO en 1980 reportó pérdidas de 36 millones de dólares (Bruggers et al., 1998). Estudios realizados en el norte de la región girasolera argentina reportaron lotes con niveles de daño por aves que oscilaron entre el 5 al 40 % en el 20 % del área sembrada en la región (Vitti & Zuil, 2012). En La Pampa se registraron áreas aún mayores alcanzando el 35 % del área sembrada en la región oeste (Bernardos & Farrel, 2012). Los daños fueron atribuidos principalmente a la cotorra en la región norte y a la paloma en La Pampa.

Desde el enfoque del mejoramiento genético del cultivo, se buscan nuevos ideotipos de girasol que se adapten al daño provocado por las aves y puedan contribuir al manejo integrado de la plaga como una estrategia defensiva. Entre los caracteres de la planta que influyen

directamente en la vulnerabilidad del cultivo frente al ataque de las aves se incluyen: la concavidad e inclinación del capítulo, la distancia entre el capítulo y el tallo, la fijación de los aquenios al capítulo, la densidad de aquenios en el capítulo, el tamaño, color, estrías y composición de los aquenios, brácteas involucrales largas orientadas hacia la parte delantera del capítulo, menor altura de planta y ciclos de cultivo cortos (Gross & Hanzel, 1991; Khaleghizadeh, 2011; Mah et al., 1990; Seiler & Rogers, 1987; Silva et al., 2019).

### **3. Arquitectura de planta y su relación con la intercepción lumínica en el cultivo de girasol**

Uno de los principales objetivos en cultivos de alta producción es lograr que las hojas intercepten la mayor parte de la radiación solar incidente, ya que es la fuente de energía utilizada para la producción de materia seca en las plantas. Así, cuanto mayor radiación solar intercepta un cultivo, mayor es su tasa de crecimiento. La producción de biomasa de un cultivo puede ser expresada como el producto entre la cantidad de radiación incidente, la capacidad del cultivo para interceptar dicha radiación, y la eficiencia con la que el cultivo convierte la radiación interceptada en biomasa (Andrade et al., 2002).

La intercepción de radiación solar por un cultivo (IR), relativa a la radiación incidente, está dada por su índice de área foliar (IAF) ( $\text{m}^2$  de hojas por  $\text{m}^2$  de suelo) y su coeficiente de extinción (k) (Gardner et al., 1985).

La capacidad del cultivo de interceptar luz frente a distintas condiciones ambientales está regulada fundamentalmente por el IAF. Modificaciones en el área foliar por planta son el mecanismo principal por el cual el cultivo ajusta su crecimiento y consumo de agua a las condiciones ambientales (Trapani et al., 1992).

Las hojas del girasol cultivado constan de una lámina de forma generalmente acorazonada y un pecíolo. Los primeros tres a cinco pares de hojas son opuestos mientras que las hojas superiores se disponen de manera alterna. El área de hojas de la planta es función del número de hojas, que depende del genotipo y del fotoperíodo, y del tamaño final de las mismas (Seiler, 1997).

Las hojas del girasol se disponen en forma horizontal denominándose canopia planófila, a causa de esta característica es que el girasol posee un coeficiente de extinción lumínica mayor que otras especies cultivadas como por ejemplo el maíz. El coeficiente de extinción del girasol es superior a 1,5 a comienzos de su ciclo y disminuye hasta valores cercanos a 1 al aumentar su área foliar (Rawson et al., 1984; Zaffaroni & Schneiter, 1989). Las variaciones en el coeficiente de extinción a lo largo del ciclo pueden ser explicadas por diferencias en la inclinación foliar predominante en hojas insertas en diferentes estratos (y que aparecen por lo

tanto en diferentes etapas del ciclo de la planta). Esto se evidencia en girasol, donde las hojas inferiores (las más cercanas a los cotiledones) poseen una inclinación más próxima a la horizontal que las superiores (Sinoquet & Andrieu, 1993).

Las canopias planófilas permiten interceptar más luz con un menor índice de área foliar, sin embargo, la distribución de la luz en diferentes estratos del canopeo es menos favorable que en canopias erectófilas. En girasol, los dos estratos superiores de hojas (0,6 a 1 de altura relativa de la planta) interceptan una alta proporción de la radiación incidente; mientras que en el maíz la distribución de la radiación en el perfil resulta ser más homogénea (Andrade et al., 2002). Una distribución más homogénea de la radiación en el perfil del canopeo permite una mayor eficiencia de conversión de la radiación interceptada en biomasa, ya que disminuye la saturación lumínica en estratos foliares superiores y por lo tanto aumenta la iluminación y la fotosíntesis en los inferiores (Aguirrezábal et al., 1996). Además, el llenado de los frutos se vería beneficiado por el aporte de fotoasimilados provenientes también de hojas que están más abajo del tercio superior del canopeo (Hernández & Orioli, 1994).

A través del mejoramiento genético, se han obtenido cambios en el coeficiente de extinción de los cultivos, a través de modificaciones de la inclinación foliar, destacándose los obtenidos en arroz (Donald & Hamblin, 1983). En el cultivo de girasol se ha reportado la existencia de variación genética para este carácter, Hernández & Orioli (1983) mostraron que genotipos de girasol con hojas superiores erectas presentaron una mejor distribución de luz dentro de su canopeo que los genotipos con hojas superiores planófilas.

#### **4. Análisis multivariado**

El Análisis Multivariado (AM) es la rama de la estadística y del análisis de datos que estudia, analiza, representa e interpreta los datos que resultan de observar más de una variable estadística sobre una muestra de individuos. Puede definirse como el conjunto de técnicas cuyo objetivo es el análisis descriptivo y/o la realización de inferencias a partir de datos en los que cada observación está constituida por los valores de un grupo de variables homogéneas y correlacionadas, sin que alguna predomine sobre las demás. La información estadística en AM es de carácter multidimensional, por lo tanto la geometría, el cálculo matricial y las distribuciones multivariantes juegan un papel fundamental (Cuadras, 2019; Defacio, 2016; Franco & Hidalgo, 2003; Palacio et al., 2020).

Entre los métodos clásicos de análisis multivariado se encuentran métodos de ordenación como el Análisis de Componentes Principales (ACP) y el Análisis de Coordenadas Principales (ACooP) y métodos de clasificación como el conjunto de técnicas para la obtención de árboles, clusters o dendrogramas. En estos métodos de análisis, la información es presentada

mediante una “*matriz de dos vías*” de orden “ $n \times p$ ”, cuyas filas y columnas recogen los datos de “ $n$ ” individuos o poblaciones y “ $p$ ” variables respectivamente. Los métodos de ordenación tienen como objetivos ordenar y representar gráficamente el material de estudio en un número de dimensiones reducidas. Los métodos de clasificación se focalizan en la búsqueda de grupos similares, lo más homogéneos posible, en los cuales se clasifican los individuos en estudio (Bramardi, 2000).

El objetivo del ACP es reducir el número de variables cuantitativas de respuesta correlacionadas entre sí, en un número menor de nuevas variables no correlacionadas (Johnson, 2000). Estas nuevas variables son una combinación lineal de las variables originales y explican la máxima variabilidad total, las que se denominan Componentes Principales (CP) (Bramardi, 2000). La contribución de las variables originales a cada CP se expresa por autovalores (variabilidad asociada con el CP) y autovectores (coeficientes de las combinaciones lineales de cada una de las variables originales) (Franco & Hidalgo, 2003). Esta metodología resulta de gran interés y aplicación en múltiples áreas de estudio y en particular en la caracterización de recursos genéticos porque permite conocer:

- La relación existente entre las variables cuantitativas, pudiéndose determinar cuáles se encuentran relacionadas y cuales caracterizan en el mismo sentido o en sentido contrario
- La relación entre los individuos, pudiendo agruparlos en función de sus similitudes y diferencias. Además, permite seleccionar aquellas variables que más discriminan entre las poblaciones, pudiendo limitarse el número de variables a registrar en evaluaciones posteriores (Defacio, 2016; Franco & Hidalgo, 2003).

El ACooP es similar al ACP pero aplicado al análisis de variables cualitativas. Esta técnica permite agrupar a los individuos en base a su similitud (Palacio et al., 2020). Por su parte, el análisis de clusters o agrupamiento intenta clasificar a los individuos en base a sus semejanzas fenotípicas y/o genotípicas cuantificadas a partir de un coeficiente de similaridad determinado (Bramardi, 2000).

El Análisis Factorial Múltiple (AFM) es una técnica de análisis multivariado que permite un estudio comparativo de un grupo de datos referidos a los mismos individuos. Su objetivo es estudiar la estructura común de las distintas tablas de datos que constituyen el estudio, proporcionando información sobre los individuos homogéneos, así como también acerca de los elementos heterogéneos. El AFM busca identificar los principales factores causales de la variabilidad de los individuos, descriptos estos, de manera equilibrada, por los diferentes grupos de variables que definen la estructura de partición de la información (Escofier & Pagès,

1992; Moncada Herrera, 2007; Pagès, 2014). El equilibrio de la influencia de los distintos grupos en el análisis se logra mediante una ponderación de los distintos grupos, y de esa manera se posibilita el análisis de:

- Relaciones entre los distintos grupos de variables, además de medir el grado de semejanza entre tales grupos.
- Relaciones entre las variables que conforman un determinado grupo y las del resto de los grupos.
- Semejanzas entre los individuos, observados desde las ópticas de los diferentes grupos de variables (Moncada Herrera, 2007).

En el presente Capítulo se exponen las metodologías empleadas para caracterizar fenotípicamente la resistencia de una de las poblaciones MAGIC en desarrollo frente a la Marchitez prematura por *V. dahliae*. La caracterización fenotípica de ambas poblaciones para caracteres de arquitectura de planta como la inclinación del capítulo y la inclinación de las láminas de las hojas del estrato superior de la planta, los días a floración y la altura de la planta. Finalmente, se presenta el análisis integrado de todos los caracteres evaluados mediante la utilización de técnicas de análisis multivariado.

## MATERIALES Y MÉTODOS

Las familias  $F_2$  de las poblaciones MAGIC ya descritas en el capítulo II fueron caracterizadas fenotípicamente en la EEA INTA Pergamino ( $33^{\circ} 57' 02''$  S,  $60^{\circ} 34' 08''$  O) para caracteres de interés agronómico como floración, altura de la punta de la lámina en relación al nivel de inserción del pecíolo en el estrato superior de la planta (LTIP), inclinación del capítulo y altura de planta. La población MAGIC2 fue a su vez caracterizada por su respuesta frente a la Marchitez anticipada causada por el hongo de suelo *V. dahliae* en el infectario natural de la EEA INTA Balcarce ( $37^{\circ}50' 0''$  S,  $58^{\circ}15' 33''$  O, Buenos Aires, Argentina).

### Diseño experimental

Los experimentos a campo realizados en la EEA INTA Pergamino fueron sembrados bajo un diseño de tipo alpha lattice de bloques incompletos aleatorizados en dos repeticiones para la población MAGIC1 y bajo un diseño experimental de tipo aumentado (Federer & Raghavarao, 1975) para la población MAGIC2. Para el experimento realizado en el infectario de la EEA INTA Balcarce el diseño experimental fue de tipo alpha lattice con dos repeticiones. Una particularidad del diseño alpha lattice es que, en caso de no controlar las fuentes de variación adecuadamente mediante bloques y sub-bloques, permite ser evaluado como un diseño Fila/Columna. En el cual pueden estimarse con una mayor resolución la variabilidad ambiental debida a la distribución de las unidades experimentales en el espacio, siendo las filas los sub-bloques del diseño alpha y las columnas los surcos de siembra.

En todos los experimentos realizados se incluyeron junto a las familias las líneas parentales fundadoras de cada población. En el caso del experimento con diseño aumentado las mismas actuaron como testigos repetidos o “Checks”. En el experimento realizado en la EEA INTA Balcarce se utilizó un control interno susceptible para poder relevar la carga patogénica y emplear la información reportada por los controles para ajustar la variación ambiental propia de cada región del infectario.

### Manejo de los experimentos

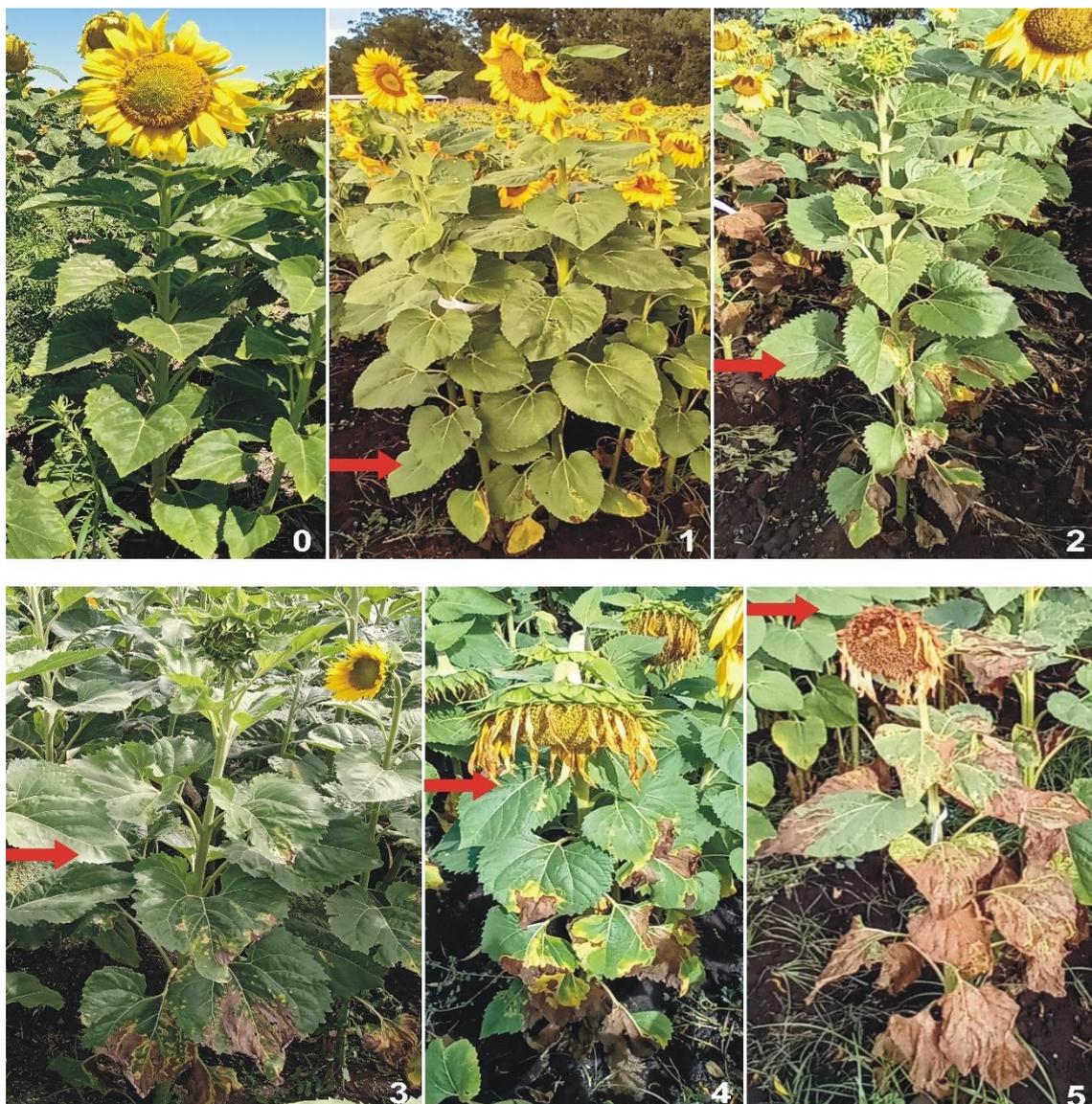
El lote en el que se realizó la caracterización fenotípica en la EEA INTA Pergamino fue trabajado en siembra directa proviniendo de un cultivo de cobertura como antecesor al girasol y siendo interrumpido el ciclo del mismo un mes antes de la siembra. La fecha de siembra fue el 28 de octubre de 2020 y fue realizada con máquina. Las malezas y plagas fueron controladas adecuadamente durante todo el ciclo del cultivo. Los experimentos no fueron regados. El tamaño de las parcelas fue de un surco con una distancia entre surcos de 0,52 m, una longitud de 6 m de largo y una densidad de 5 plantas por metro lineal.

El lote en el que se ubica el infectario de la EEA INTA Balcarce fue laboreado convencionalmente con arado, rastra de discos y rolo para el afinado de los agregados. Se realizaron aplicaciones de herbicida en pre-emergencia para el control del *Cyperus* (Cebollín), *Tagetes* (Chinchilla) y *Xanthium* (Abrojo). La fecha de siembra fue el 9 de noviembre de 2020. La siembra se realizó de forma manual sobre los surcos previamente marcados a 0.7 m con una sembradora, mediante el uso de “escopeta” y golpes de 3 semillas (aprox.) cada 20 cm para alcanzar una densidad de 5 plantas por metro lineal. Las unidades experimentales (parcelas) consistieron en surcos de 5 m de largo.

Las parcelas con menos de 5 plantas fueron descartadas, por no poder obtener una estimación confiable del comportamiento de cada genotipo en esas condiciones. Se realizaron desmalezamientos manuales entre fechas de evaluación para controlar malezas (papa, cebollín, abrojo, chinchilla, entre otras) y conseguir una buena visibilidad del stand de plantas.

### **Evaluaciones fenotípicas**

La evaluación de la MV se realizó en 2 fechas (R4 y R6) escala de Schneiter & Miller (1981) a nivel planta individual, registrándose dos variables fitopatológicas, la Incidencia (INC) y la Severidad (SEV). Cada fecha de evaluación es definida por un número que se adjunta a la variable para identificar el momento en que esa variable fue registrada (por ejemplo: INC1 en R4 e INC2 en R6). La INC representa la sumatoria de plantas sintomáticas sobre el total de plantas por parcela. La SEV se estimó usando una escala ordinal de 6 niveles de (0 a 5) siendo el valor “0” = Planta sana, “1” = Estrato basal sintomático (~20% de la parte aérea), “2” = Síntomas hasta por debajo del estrato medio de la planta (20 - 40% de la planta), “3” = Síntomas en la zona media de la planta (40-60%), “4” = Síntomas en estratos superiores y el “5” = Planta marchita o sintomática en todos sus estratos (Figura 3.3). La INC, por ser una variable discreta de conteos sobre un número definido de plantas, tiene una distribución binomial, mientras que la SEV, al ser una escala ordinal podría analizarse con modelos multinomiales. En este trabajo se resolvió transformarla por Normal Scores (Gianola & Norton, 1981) y modelarla transformada como una variable de distribución normal (MSEV).



**Figura 3.3.** Escala fitopatológica de seis niveles usada para la evaluación de severidad. Niveles de sintomatología crecientes de 0 (planta asintomática) a 5 (Todos los tejidos afectados). Flechas indican el último estrato sintomático.

De estas dos variables se desprenden otras como por ejemplo la Intensidad, que es una media ponderada a nivel parcela en función de las frecuencias registradas para cada nivel de severidad (INT). El Área Bajo la Curva de Progreso de la Enfermedad (AUDPC, de sus siglas en inglés *Area Under Disease Progress Curve*) pudiendo estimarse sobre cualquiera de las otras variables. El AUDPC integra las mediciones realizadas a través del tiempo en un único valor final mediante la fórmula descrita por (Shaner & Finney, 1977), para obtener un solo dato por línea que describa la enfermedad:

$$AUDPC = \sum \left( \frac{(X_{i+1} + X_i)}{2} \right) (T_{i+1} - T_i)$$

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

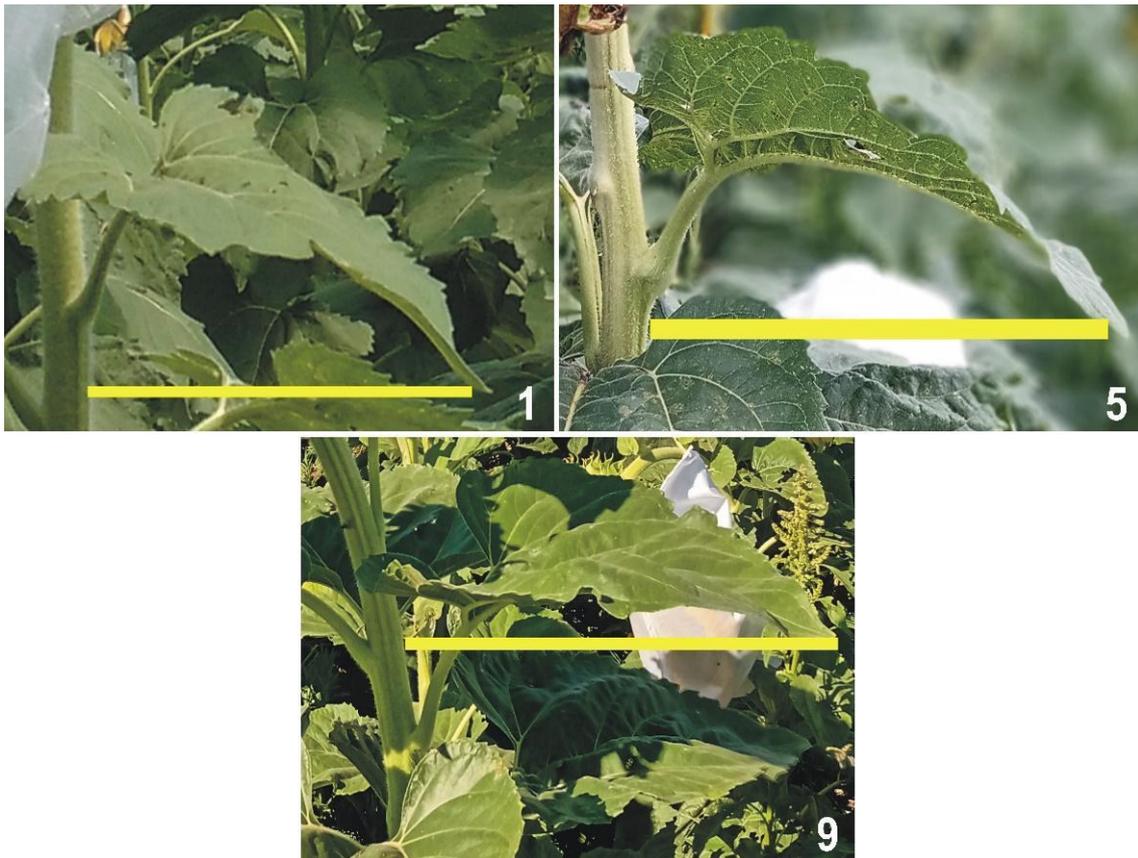
Otra variable generada, evaluada durante la exploración de metodologías de análisis de este conjunto de datos e implementada para el análisis multivariado, fue la Severidad binomial, la cual consiste en conteos a ambos lados de una partición simétrica de la escala ordinal de seis niveles, considerando tolerantes a las plantas con scores de severidad de 0 a 2, y susceptibles a aquellas con scores de 2 a 5 en la variable SEV\_B, y considerando valores de severidad de 0 a 3 como plantas tolerantes y de 3 a 5 como plantas susceptibles en la variable denominada en este trabajo SEV\_B3.

Las evaluaciones de altura de planta, y altura de la punta de la hoja en relación a la inserción del pecíolo fueron realizadas en el estado fenológico R6. La floración fue registrada como los días desde la emergencia hasta plena floración (50 % del capítulo florecido). La inclinación del capítulo a madurez fue registrada usando una escala de 3 niveles siendo "1" = ángulo de inclinación entre el capítulo y el tallo entre 0° y 90°, "5" = ángulo de inclinación entre el capítulo y el tallo entre 90° y 180° y "9" = ángulo de inclinación entre el capítulo y el tallo mayor a 180° (Figura 3.4).



**Figura 3.4.** Imágenes de la escala utilizada para la evaluación de la inclinación de capitulo.

La altura de la punta de la lámina en relación a la inserción del pecíolo fue registrada usando una escala de 3 valores siendo “1” = cuando la punta de la lámina se ubica por debajo del nivel de inserción del pecíolo, “5” = cuando la punta de la lámina se ubica al mismo nivel que la inserción del pecíolo y “9” = cuando la punta de la lámina se ubica por encima del nivel de inserción del pecíolo (Figura 3.5).



**Figura 3.5.** Imágenes de la escala utilizada para la evaluación de la altura de la punta de la lámina en relación a la inserción del pecíolo (LTIP). Las barras amarillas hacen referencia al nivel alcanzado por la punta de la lámina.

#### **Análisis estadístico del experimento de respuesta a MV por *V. dahliae***

El análisis estadístico de los datos fenotípicos se realizó utilizando modelos lineales mixtos (MLM) y modelos lineales mixtos generalizados (GLMM) según se detalla a continuación. Para ambos casos se empleó el software estadístico libre R (R Core Team, 2020) y el paquete lme4, que permite ajustar modelos lineales mixtos y modelos lineales mixtos generalizados (Bates et al., 2015). También se evaluaron modelos que tengan en cuenta la variabilidad espacial del terreno, como el modelo P-splines con la utilización del paquete SpATS (Rodríguez-Álvarez et al., 2018; Velazco et al., 2017).

Se evaluó la Incidencia como el número de plantas enfermas a floración sobre el número total de plantas por surco (parcela). Dado que la distribución de probabilidades de esta variable se ajusta a una distribución de tipo binomial, conteo sobre un número definido de experimentos, el análisis se realizó utilizando modelos lineales mixtos generalizados, indicando la familia de distribución “Binomial” y empleando la función de enlace “*logit*” para vincular la esperanza de la variable respuesta al predictor lineal. Para definir el número de casos totales por parcela,

se empleó como covariable el número total de plantas en la misma al momento de la evaluación.

Con el propósito de definir el mejor modelo para ajustar a los datos obtenidos se ajustaron tres tipos de modelos: Alpha Lattice [1.1], Fila-Columna [1.2] y el modelo espacial P-splines [1.3]. Entre los modelos [1.1] y [1.2] se seleccionó al más parsimonioso en función de los criterios Bayesianos de comparación de modelos usando para ello al Criterio de Akaike (AIC) y al Criterio de Información Bayesiana (BIC). Para comparar los modelos Alpha Lattice y Fila-Columna con el modelo espacial propuesto, se utilizó la metodología propuesta por Rodríguez-Álvarez et al (2018) en base a la comparación de la heredabilidad generalizada propuesta por Cullis et al (2006).

El primer modelo propuesto representa al diseño Alpha Lattice (Bloques incompletos aleatorizados) implementa el factor bloque (o "Fila") el cuál se encuentra anidado a la repetición [1.1]. El segundo contempla al diseño Fila-Columna, además del factor bloque, incluye el factor "Columna" siendo constituido por el surco de siembra [1.2]. Todos modelos fueron comparados considerando al factor Genotipo (Familia  $F_2$ ) como factor de efecto fijo y a los factores de diseño experimental (Repetición, Fila y Columna) como de efectos aleatorios.

El modelo [1.1] se define como:

$$Y_{ijk} = \mu + F_i + R_j + B(R)k(j) + \varepsilon_{ijk}$$

Donde  $Y_{ijk}$  es la incidencia (proporción de plantas enfermas) de la familia  $F_2$  (genotipo)  $i$ ;  $\mu$  es la media general;  $F_i$  es el efecto fijo del genotipo con  $i = 1-358$ ;  $R_j$  es el efecto aleatorio de la repetición;  $B(R)k(j)$  es el efecto aleatorio del bloque  $k$  dentro de la repetición  $j$ ; y  $\varepsilon_{ijk}$  es el término de error aleatorio asociado a la observación  $Y_{ijk}$ .

El modelo [1.2] se define como:

$$Y_{ijkm} = \mu + F_i + R_j + B(R)k(j) + C(R)m(k) + \varepsilon_{ijkm}$$

Donde  $Y_{ijkm}$  es la incidencia (proporción de plantas enfermas) de la familia  $F_2$  (genotipo)  $i$ ;  $\mu$  es la media general;  $F_i$  es el efecto fijo del genotipo con  $i = 1-358$ ;  $R_j$  es el efecto aleatorio de la repetición;  $B(R)k(j)$  es el efecto aleatorio del bloque  $k$  dentro de la repetición  $j$ ;  $C(R)m(k)$  es el efecto aleatorio de la columna (surco)  $m$  dentro de la repetición  $k$  y  $\varepsilon_{ijkm}$  es el término de error aleatorio asociado a la observación  $Y_{ijkm}$ .

El modelo [1.3] se define como:

$$Y_{ijkm} = \mu + f(f, c) + F_i + R_j + B(R)k(j) + C(R)m(k) + \varepsilon_{ijkm}$$

Donde el término fijo (no penalizado)  $ss$  y el componente aleatorio (penalizado)  $ss$  forman la expresión del modelo mixto de la superficie espacial lisa,  $f(f, c) = Xs\beta s + ZsS$ , donde el vector de efectos espaciales aleatorios  $s$  tiene una matriz de covarianza  $S$ . El vector  $u$  comprende los subvectores mutuamente independientes de los efectos aleatorios de fila y de columna aleatorios, que representan la variación espacial, con la matriz de diseño  $Zu = [Zf | Zc]$  y la matriz de covarianza  $U = \text{diag}(\sigma_f^2 I_f, \sigma_c^2 I_c)$ . El detalle de la descomposición de los términos fijos y aleatorios puede verse en el Anexo I. Los demás términos del modelo son coincidentes con el modelo [1.2] descrito anteriormente.

Como se describió anteriormente, la SEV de la enfermedad hace referencia a una escala ordinal de 6 niveles que define la proporción de tejido sintomático de la planta. Promediando los scores obtenidos por las plantas enfermas del surco a la fecha final de evaluación, se obtiene el valor de SEV para cada parcela. Por lo que la SEV igual a cero no está definida, de modo que para el análisis de este carácter se contó con menos datos que para el carácter INC, dado que no se tuvieron en cuenta para el análisis aquellas familias que no presentaron enfermedad. Considerando que los datos no presentan continuidad se realizó una transformación mediante la metodología Normal Scores (Gianola & Norton, 1981), para poder analizarlos como normales. Nuevamente, se probaron para este carácter los mismos tres modelos presentados anteriormente en 1.1, 1.2, 1.3, implementando la función lmer del paquete lme4 para los modelos 1.1 y 1.2 que permite ajustar Modelos Lineales Mixtos a variables con distribución de probabilidades de ajuste normal. Se corroboró el cumplimiento de los supuestos de homogeneidad de varianzas del error y de normalidad de los residuos.

Para todas las variables estudiadas, las comparaciones múltiples entre medias se hicieron aplicando la prueba de comparaciones múltiples (DGC) (Di Rienzo et al., 2001), utilizando un nivel de significación del 0,10.

Los mismos modelos fueron ajustados con el factor Genotipo como efecto aleatorio para obtener las estimaciones de los efectos aleatorios de cada genotipo. Estos últimos predictores se denominan Mejores Predictores Lineales Insesgados (BLUPs, por sus siglas del inglés *Best Linear Unbiased Predictors*). Asimismo, ajustando el factor Genotipo como efecto aleatorio se pudo estimar las varianzas genotípicas y poder realizar las estimaciones de las heredabilidades para cada variable y para cada modelo.

Las estimaciones de la heredabilidad por variable fueron realizadas a través de la aproximación tradicional de la heredabilidad en sentido amplio  $H^2 = \frac{\sigma_{Genotípica}^2}{\sigma_{Fenotípica}^2}$ , la heredabilidad generalizada fue solo utilizada como se mencionó anteriormente para la

evaluación del modelo espacial 1.3. En el caso particular de las variables que presentan una distribución de tipo binomial, la varianza residual de los modelos empleados fue fijada por convención en cada caso a un valor de  $\frac{\pi^2}{3}$  tal como se describe en Snijders & Bosker (2012).

**Análisis estadístico del experimento de caracterización fenotípica de floración, altura de la punta de la hoja en relación con la inserción del pecíolo en el estrato superior de la planta, inclinación del capítulo y altura de planta**

El análisis estadístico de los datos fenotípicos de floración y altura de planta fue realizado mediante modelos lineales mixtos con el paquete lme4 de R (Bates et al., 2015).

Para el análisis de la población MAGIC1 dado que el diseño experimental empleado fue el mismo que en la evaluación frente a MV de la población MAGIC2 se utilizaron los mismos modelos 1.1 y 1.2 descritos anteriormente.

Para el análisis de la población MAGIC2 en un diseño aumentado se ajustaron 4 modelos diferentes incluyendo al genotipo (familias MAGIC2  $F_2$ ) como factor fijo con el objetivo de obtener las medias ajustadas de cada familia:

El primer modelo 1.4 se define como:

$$Y_{ij} = \mu + F_i + R_j + \varepsilon_{ij}$$

Donde  $Y_{ij}$  es el fenotipo (Altura de planta y DAF) de la familia  $F_2$  (genotipo)  $i$ ;  $\mu$  es la media general;  $F_i$  es el efecto fijo del genotipo con  $i= 1-512$ ;  $R_j$  es el efecto aleatorio del bloque y  $\varepsilon_{ijk}$  es el término de error aleatorio asociado a la observación  $Y_{ij}$ .

El segundo modelo 1.5 contempla un diseño fila-columna incluyendo estos factores como aleatorios:

$$Y_{ijk} = \mu + F_i + R_j + S_k + \varepsilon_{ijk}$$

Donde  $Y_{ijk}$  es el fenotipo (Altura de planta y DAF) de la familia  $F_2$  (genotipo)  $i$ ;  $\mu$  es la media general;  $F_i$  es el efecto fijo del genotipo con  $i= 1-512$ ;  $R_j$  es el efecto aleatorio de la fila o bloque;  $S_k$  es el efecto aleatorio de la columna (surco)  $k$  y  $\varepsilon_{ijk}$  es el término de error aleatorio asociado a la observación  $Y_{ijk}$ .

El tercer modelo 1.6 incluye al factor *check* como efecto fijo y se define como:

$$Y_{ijkm} = \mu + C_i + F(C)j(i) + R_k + S_m + \varepsilon_{ijk}$$

Donde  $Y_{ijkm}$  es el fenotipo (Altura de planta y DAF) de la familia  $F_2$  (genotipo)  $i$ ;  $\mu$  es la media general;  $C_i$  es el efecto fijo del check con  $i= 1-8$ ;  $F(C)j(i)$  es el efecto fijo de las familias  $j= 1-504$  dentro del check  $i$ ;  $R_k$  es el efecto aleatorio de la fila o bloque;  $S_m$  es el efecto aleatorio

de la columna (surco)  $m$  y  $\varepsilon_{ijklm}$  es el término de error aleatorio asociado a la observación  $Y_{ijklm}$ .

El cuarto modelo 1.7 es similar al modelo 1.6 con la única diferencia que el *check* fue considerado como factor aleatorio en el modelo.

Se corroboró el cumplimiento de los supuestos de homogeneidad de varianzas del error y de normalidad de los residuos.

Los mismos modelos fueron ajustados con el factor Genotipo como efecto aleatorio para obtener las estimaciones de los efectos aleatorios de cada genotipo y las varianzas genotípicas y poder realizar las estimaciones de las heredabilidades para cada variable y para cada modelo. Las estimaciones de la heredabilidad por variable fueron realizadas a través de la aproximación tradicional de la heredabilidad en sentido amplio.

### **Análisis de Componentes Principales**

El análisis de componentes principales fue realizado mediante el paquete FactoMineR de R (Lê, S., Josse, J. & Husson, 2008) y el biplot fue obtenido con el paquete factoextra de R (Kassambara & Mundt, 2020). Se utilizaron las medias ajustadas por los modelos detallados previamente, aplicados según la distribución de probabilidades de cada variable: Incidencia (Binomial), Intensidad (Gamma), Severidad Media (Normal, transformada N.S.), Severidad Binomial (Binomial) y las AUDPC para incidencia e intensidad a tiempo final (Gamma), conformando un conjunto de 12 variables fenotípicas en total.

### **Análisis de Conglomerados**

Sobre estos mismos conjuntos de datos se realizaron Análisis de Conglomerados Jerárquicos y No Jerárquicos con los paquetes de R stats, ape y factoextra (Kassambara & Mundt, 2020; Paradis & Schliep, 2019). Se estimó una matriz de distancia Euclídea para las medias ajustadas de los genotipos usando la función "*dist*" del paquete ape. Se estimaron los números de grupos más probables para cada método de agrupamiento con el paquete Nbclust (Charrad et al., 2014) y a partir de ello se conformaron los grupos con el método No-Jerárquico K-Means mediante la función "*kmeans*" del paquete stats y con la función "*hclust*" del paquete stats para el desarrollo de los métodos jerárquicos. Finalmente se realizaron comparaciones de los coeficientes de correlación cofenética de distintos métodos de agrupamiento jerárquico para definir el más adecuado (correlación más alta) y así asignar los genotipos a los distintos clusters.

**Análisis Factorial Múltiple**

Con el objetivo de caracterizar la población MAGIC2 integrando la información aportada por las variables de respuesta a MV, los días a floración, la altura de la planta, la posición del capítulo y la altura de la punta de la hoja en relación a la inserción del pecíolo se empleó la técnica multivariada de Análisis Factorial Múltiple (AFM) (Escofier & Pagès, 1992; Pagès, 2014). Esta metodología permite analizar simultáneamente dos grupos de variables de distinta naturaleza (cuantitativa y cualitativa), evaluados sobre el mismo conjunto de individuos (Del Medico et al., 2021; Moncada Herrera, 2007).

## RESULTADOS

### Selección y Ajuste de Modelos Lineales Generales Mixtos, Generalizados Mixtos y Espaciales para la respuesta a MV

Los resultados de la evaluación y selección de los diferentes modelos ajustados para el análisis de las 12 variables fitopatométricas descritas en materiales y métodos se representa en la Tabla 3.1.

**Tabla 3.1.** Resultados de evaluación y selección de modelos para las 12 variables estudiadas. En rojo se identifican los modelos elegidos.

INC1	AIC	BIC	logLik	deviance
Modelo 1.1	3375,40	5054,10	-1328,70	2657,40
Modelo 1.2	3377,40	5060,70	-1328,70	2657,40
Modelo 1.1 genotipo aleatorio	3598,50	3617,20	-1795,20	3590,50
Modelo 1.2 genotipo aleatorio	3590,50	3613,90	-1790,20	3580,50
INC2	AIC	BIC	logLik	deviance
Modelo 1.1	3788,00	5466,70	-1535,00	3070,00
Modelo 1.2	3775,70	5459,00	-1527,80	3055,70
Modelo 1.1 genotipo aleatorio	3986,20	4004,90	-1989,10	3978,20
Modelo 1.2 genotipo aleatorio	3936,70	3960,10	-1963,30	3926,70
MSEV_1	AIC	BIC	logLik	deviance
Modelo 1.1	1894,70	3472,90	-601,37	1202,70
Modelo 1.2	1895,20	3477,80	-600,58	1201,20
Modelo 1.1 genotipo aleatorio	1926,80	1949,60	-958,40	1916,80
Modelo 1.2 genotipo aleatorio	1927,80	1955,20	-957,92	1915,80
MSEV_2	AIC	BIC	logLik	deviance
Modelo 1.1	1918,70	3598,80	-599,33	1198,70
Modelo 1.2	1916,30	3601,10	-597,16	1194,30
Modelo 1.1 genotipo aleatorio	2002,50	2025,90	-996,26	1992,50
Modelo 1.2 genotipo aleatorio	2002,90	2030,90	-995,44	1990,90
AUDPC_INC	AIC	BIC	logLik	deviance
Modelo 1.1	-19,12	1664,20	369,56	-739,12
Modelo 1.2	-17,12	1670,80	369,56	-739,12
Modelo 1.1 genotipo aleatorio	-264,86	-241,48	137,43	-274,86
Modelo 1.2 genotipo aleatorio	-263,37	-235,31	137,69	-275,37
INT_1	AIC	BIC	logLik	deviance
Modelo 1.1	761,10	2339,70	-34,55	69,10
Modelo 1.2	761,65	2344,80	-33,83	67,65
Modelo 1.1 genotipo aleatorio	835,41	858,22	-412,71	825,41
Modelo 1.2 genotipo aleatorio	830,40	857,77	-409,20	818,40
INT_2	AIC	BIC	logLik	deviance
Modelo 1.1	1768,30	3448,40	-524,16	1048,30
Modelo 1.2	1762,20	3447,00	-520,12	1040,20
Modelo 1.1 genotipo aleatorio	1743,90	1767,20	-866,94	1733,90

<b>Modelo 1.2 genotipo aleatorio</b>	1733,00	1761,00	-860,50	1721,00
<b>AUDPC_INT</b>	<b>AIC</b>	<b>BIC</b>	<b>logLik</b>	<b>deviance</b>
<b>Modelo 1.1</b>	-691,94	984,48	705,97	-1411,90
Modelo 1.2	-689,94	991,14	705,97	-1411,90
Modelo 1.1 genotipo aleatorio	-808,73	-785,44	409,36	-818,73
<b>Modelo 1.2 genotipo aleatorio</b>	-807,99	-780,05	409,99	-819,99
<b>SEV_B_1</b>	<b>AIC</b>	<b>BIC</b>	<b>logLik</b>	<b>deviance</b>
<b>Modelo 1.1</b>	2485,40	4164,10	-883,72	1767,50
Modelo 1.2	2486,80	4170,10	-883,39	1766,80
Modelo 1.1 genotipo aleatorio	2752,50	2771,20	-1372,20	2744,50
<b>Modelo 1.2 genotipo aleatorio</b>	2730,80	2754,20	-1360,40	2720,80
<b>SEV_B_2</b>	<b>AIC</b>	<b>BIC</b>	<b>logLik</b>	<b>deviance</b>
Modelo 1.1	3511,10	5189,70	-1396,50	2793,10
<b>Modelo 1.2</b>	3487,80	5171,10	-1383,90	2767,80
Modelo 1.1 genotipo aleatorio	3888,60	3907,30	-1940,30	3880,60
<b>Modelo 1.2 genotipo aleatorio</b>	3832,50	3855,90	-1911,30	3822,50
<b>SEV_B3_1</b>	<b>AIC</b>	<b>BIC</b>	<b>logLik</b>	<b>deviance</b>
<b>Modelo 1.1</b>	1758,40	3437,00	-520,21	1040,40
Modelo 1.2	1759,00	3442,20	-519,47	1039,00
Modelo 1.1 genotipo aleatorio	1812,80	1831,50	-902,41	1804,80
<b>Modelo 1.2 genotipo aleatorio</b>	1789,00	1812,40	-889,50	1779,00
<b>SEV_B3_2</b>	<b>AIC</b>	<b>BIC</b>	<b>logLik</b>	<b>deviance</b>
Modelo 1.1	2920,10	4598,80	-1101,10	2202,10
<b>Modelo 1.2</b>	2917,20	4600,50	-1098,60	2197,20
Modelo 1.1 genotipo aleatorio	3206,60	3225,30	-1599,30	3198,60
<b>Modelo 1.2 genotipo aleatorio</b>	3176,70	3200,10	-1583,30	3166,70

Las medias fenotípicas ajustadas a cada familia para cada una de las variables fitopatométricas estudiadas, se detallan en el Anexo digital "MAjustadas\_VD".

Asimismo, la comparación de estos modelos con el modelo espacial 1.3 se realizó en base a la heredabilidad generalizada (Cullis et al., 2006). Para esto se indicó al factor Genotipo como aleatorio. La heredabilidad generalizada fue menor para el modelo 1.3 por lo que se optó por no utilizar este modelo. Por otra parte, se estimaron las correlaciones entre las medias ajustadas de los modelos 1.2 y 1.3 y las mismas fueron significativas y superiores a 0,9, demostrando que la implementación de los modelos que incluyen a los factores fila y columna como aleatorios son suficientes para modelar la variabilidad espacial del experimento.

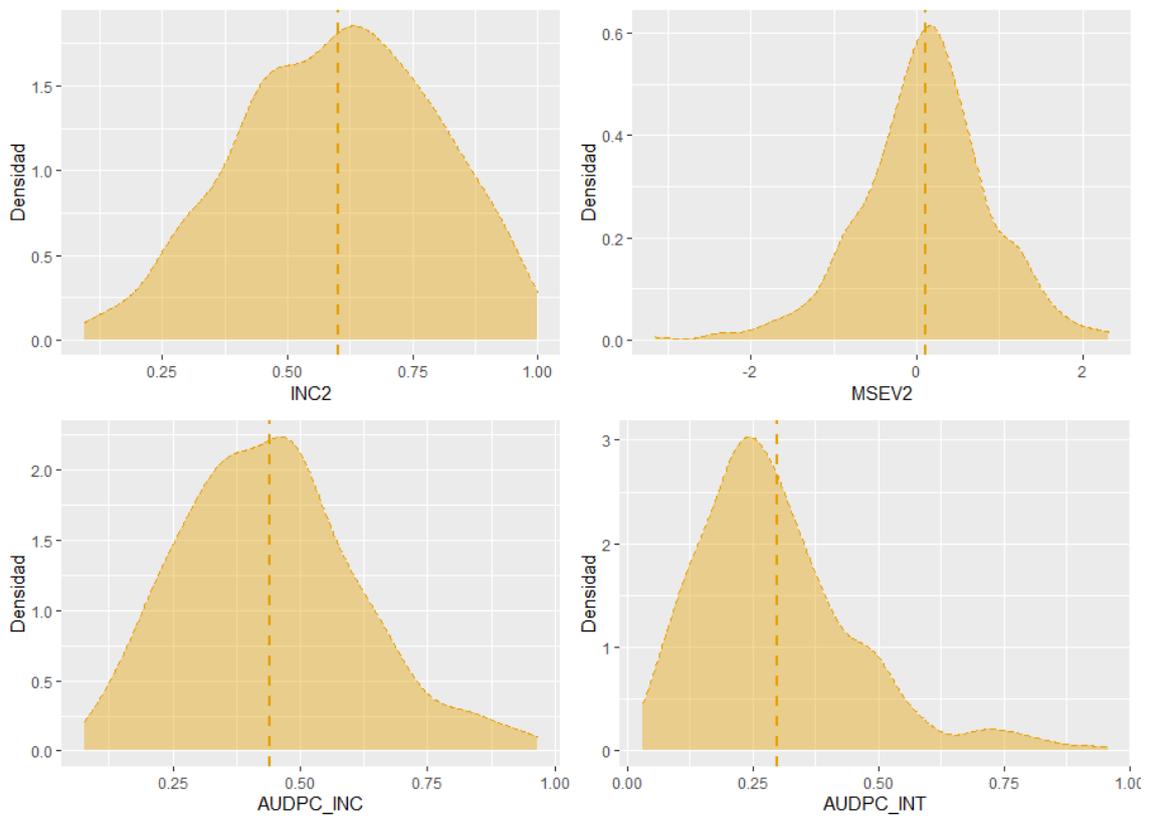
En la Tabla 3.2 se muestran los valores mínimos, medios y máximos de las medias ajustadas para cada una de las variables evaluadas.

**Tabla 3.2.** Valores mínimos, medios y máximos para las medias ajustadas de las variables fitopatométricas analizadas.

Variable	Mínimo	Media	Máximo
INC1	0,02	0,29	1,00
INC2	0,09	0,60	1,00
MSEV_1	-2,78	-0,08	2,14
MSEV2	-3,16	0,10	2,32
AUDPC_INC	0,08	0,44	0,97
INT1	0,05	0,79	3,85
INT2	0,23	1,78	9,10
AUDPC_INT	0,03	0,30	0,96
SEV_B_1	0,00	0,15	1,00
SEV_B_2	0,00	0,37	1,00
SEV_B3_1	0,00	0,07	0,75
SEV_B3_2	0,00	0,19	0,97

En la Tabla 3.2 se destaca el incremento que mostró la enfermedad entre la primera y la segunda fecha de evaluación.

En la Figura 3.6 se observa la distribución de las medias ajustadas obtenidas para las familias  $F_2$  para las variables INC2, MSEV\_2, AUDPC\_INC y AUDPC\_INT.



**Figura 3.6.** Función Densidad para las medias ajustadas de las familias  $F_2$  para las variables INC2, MSEV\_2, AUDPC\_INC y AUDPC\_INT.

Se realizó la comparación múltiple de medias mediante la metodología DGC (Di Rienzo et al., 2001) para definir las clases significativamente distintas de cada variable fitopatométrica. Este análisis permitió identificar diferencias significativas entre las medias estimadas de las distintas familias con un nivel de significancia de 0,10 solamente para las variables INC1, AUDPC\_INC, INT\_2, AUDPC\_INT y SEV\_B3\_2. Para estas variables fueron obtenidas dos clases significativamente distintas a excepción de INT\_2 la cual presentó 3 clases.

Para la variable INC1 la clase “A” con la media más alta (mayor probabilidad de incidencia de la enfermedad) es donde se encuentran las familias más susceptibles y la “B” representa a las familias con mejor comportamiento frente a MV. La clase “A” agrupo al 20 % de los genotipos mientras que la clase “B” fue conformada por el 80 % restante. Para la variable AUDPC\_INC la representación entre las clases estuvo aún más desequilibrada que INC1, la clase “A” agrupo sólo al 3,64 % de los genotipos mientras que la clase “B” agrupo al resto. Para la variable INT\_2, la única que presentó 3 clases DGC, mostró una distribución muy heterogénea entre ellas, la clase “A” estuvo conformada por sólo 2 genotipos, la “B” casi la totalidad de los genotipos (354) y la “C” un sólo genotipo. Para la variable AUDPC\_INT la clase “A” agrupo al 20,45 % de los genotipos y la clase “B” a la porción restante. Por último, la variable SEV\_B3\_2 presente solamente 3 genotipos en la clase “A” y los restantes genotipos en la clase “B”. En el Anexo digital “MAjustadas\_VD” se detallan las clases DGC a la que pertenece cada uno de los genotipos.

### **Estimación de la heredabilidad para cada variable fitopatométrica**

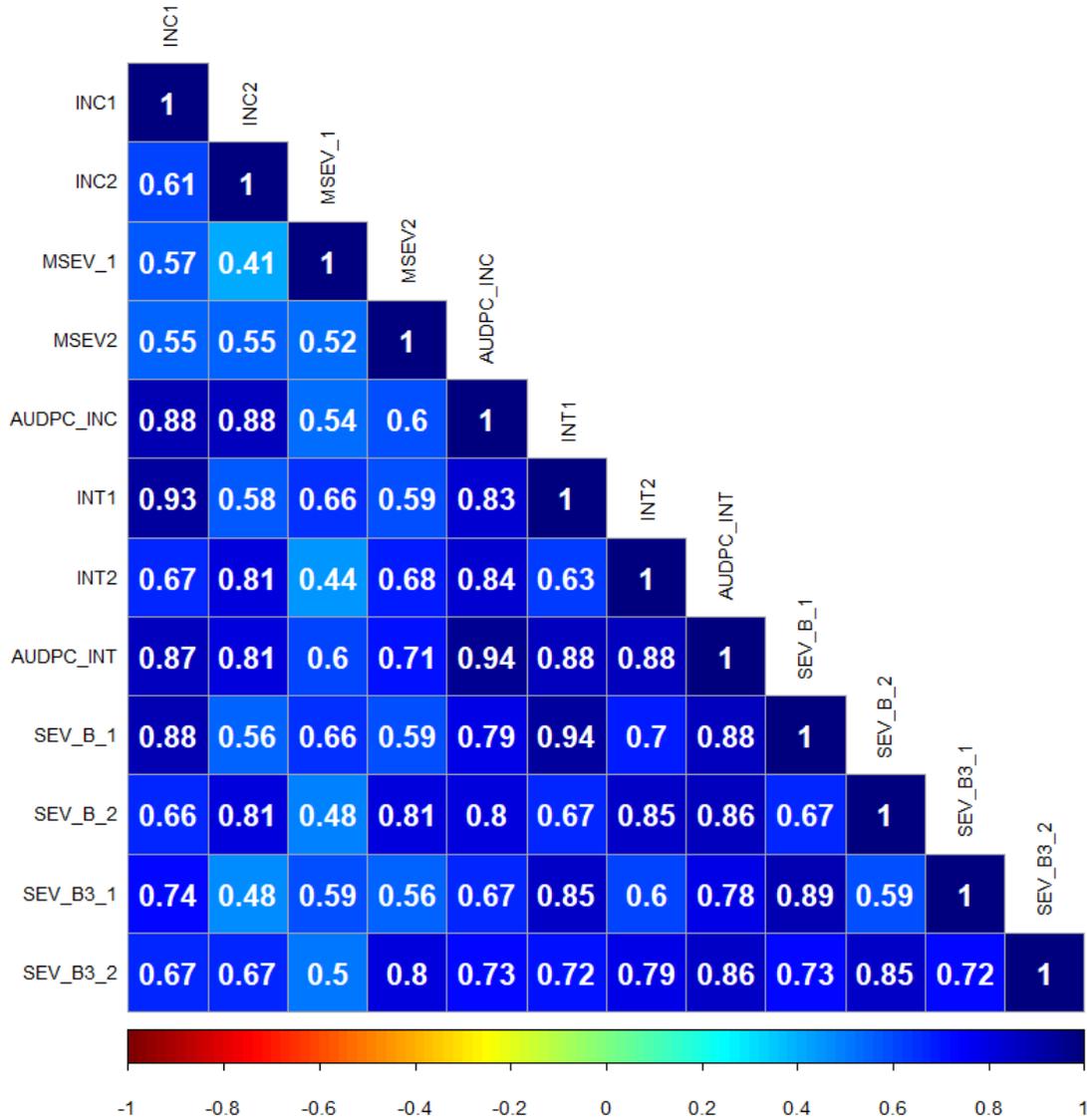
Empleando los modelos que mejor ajustaron pudieron definirse las contribuciones de cada factor a la varianza fenotípica total y así poder estimar para cada variable la heredabilidad en sentido amplio. Los resultados se detallan en la Tabla 3.3.

**Tabla 3.3.** Heredabilidad en sentido amplio para las 12 variables fitopatométricas estudiadas.

<b>Variable</b>	<b>Heredabilidad</b>
INC1	0,35
INC2	0,28
MSEV1	0,53
MSEV2	0,60
AUDPC_INC	0,70
INT_1	0,72
INT_2	0,27
AUDPC_INT	0,90
SEV_B1	0,57
SEV_B2	0,45
SEV_B3_1	0,63
SEV_B3_2	0,49

**Análisis multivariado de la respuesta de la población MAGIC2 frente a MV**

Las correlaciones fenotípicas obtenidas entre las variables consideradas para evaluar la población frente a MV se muestran en la Figura 3.7.



**Figura 3.7.** Correlaciones fenotípicas entre las medias ajustadas para las variables registradas para la evaluación de la respuesta frente a MV de la población MAGIC2.

Como puede observarse en la Figura 3.5 las correlaciones entre las 12 variables fitopatológicas fueron positivas, significativas y en la mayoría de los casos superiores al 0,5 por lo que podemos considerar que la correlación entre variables es alta y que existe, dada su colinealidad, oportunidad de resumir tal información en componentes principales multivariados.

Se realizó un ACP utilizando las funciones descriptas en materiales y métodos. El total de la varianza multivariada fue explicado con 12 CPs, alcanzando explicar más del 95% de la varianza multivariada total con los primeros 5, Tabla 3.4.

**Tabla 3.4.** Componentes principales obtenidos del análisis.

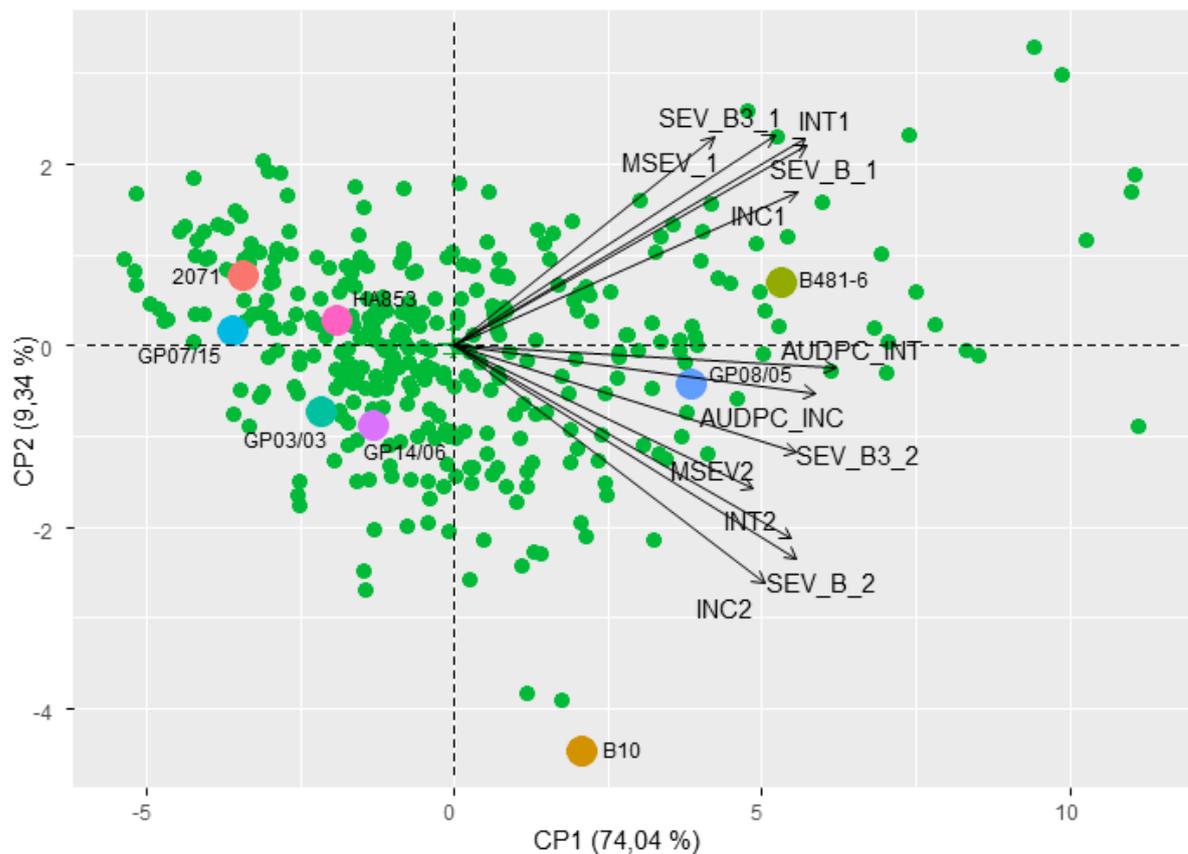
<b>Componentes principales</b>	<b>CP1</b>	<b>CP2</b>	<b>CP3</b>	<b>CP4</b>	<b>CP5</b>	<b>CP6</b>	<b>CP7</b>	<b>CP8</b>	<b>CP9</b>	<b>CP10</b>	<b>CP11</b>	<b>CP12</b>
<b>Varianza</b>	8,89	1,12	0,69	0,49	0,28	0,17	0,13	0,1	0,05	0,04	0,03	0,01
<b>Propor. de la Varianza (%)</b>	74,04	9,34	5,78	4,1	2,35	1,42	1,06	0,82	0,43	0,29	0,26	0,11
<b>Acum. de la Var. (%)</b>	74,04	83,38	89,16	93,26	95,61	97,03	98,09	98,91	99,34	99,63	99,89	100

Se seleccionaron los primeros dos CPs, que conjuntamente explicaron el 83,38 % de la varianza multivariada total, alcanzando el objetivo de resumir la información disponible en un número acotado de variables. El CP1 explica el 74,04 % y el CP2 el 9,34 % de la varianza total. Las contribuciones de las variables a los dos primeros CPs seleccionados se detallan en la Tabla 3.5.

**Tabla 3.5.** Contribuciones de las variables a los dos primeros CPs seleccionados.

<b>Variables</b>	<b>CP1</b>	<b>CP2</b>
<b>INC1</b>	8,78	6,1
<b>INC2</b>	7,14	15,87
<b>MSEV_1</b>	4,98	12,3
<b>MSEV2</b>	6,64	5,01
<b>AUDPC_INC</b>	9,7	0,76
<b>INT1</b>	9,18	11,35
<b>INT2</b>	8,44	10,22
<b>AUDPC_INT</b>	10,85	0,15
<b>SEV_B_1</b>	9,2	10,85
<b>SEV_B_2</b>	8,72	12,18
<b>SEV_B3_1</b>	7,61	12,32
<b>SEV_B3_2</b>	8,73	2,88

En la Figura 3.8 se presenta el biplot de los dos primeros CPs utilizados para evaluar el comportamiento de las familias  $F_2$  de la población MAGIC2 frente a MV. Los vectores de las variables indican hacia donde se incrementa el valor de las mismas. La ortogonalidad entre vectores implica mayor independencia entre variables. La proyección de los mismos sobre los ejes indica la proporción de correlación con cada CP.



**Figura 3.8.** Biplot de los componentes principales 1 y 2. Las líneas parentales se identifican con distintos colores y con etiquetas respecto de las familias  $F_2$  de la población MAGIC2.

En los cuadrantes de la izquierda del Biplot se ubican los genotipos con mejor respuesta frente a MV y en los de la derecha los genotipos con mayor susceptibilidad frente a la enfermedad. En el gráfico puede observarse que si bien hay una mayor concentración de familias con mejor respuesta frente a MV hay también genotipos de respuesta intermedia y de mayor susceptibilidad frente a la enfermedad. La existencia de variabilidad en la respuesta de la población frente a MV resulta fundamental para futuros estudios de mapeo con esta población. Además, puede notarse que la elección de líneas parentales con respuesta diferencial frente a la enfermedad permitió obtener familias descendientes con respuesta variable para el carácter. Con respecto a las variables, la fecha de evaluación como era de esperarse resultó el factor principal de agrupamiento de las mismas, permitiendo diferenciarse dos grandes grupos de variables altamente correlacionadas en función de la fecha de evaluación.

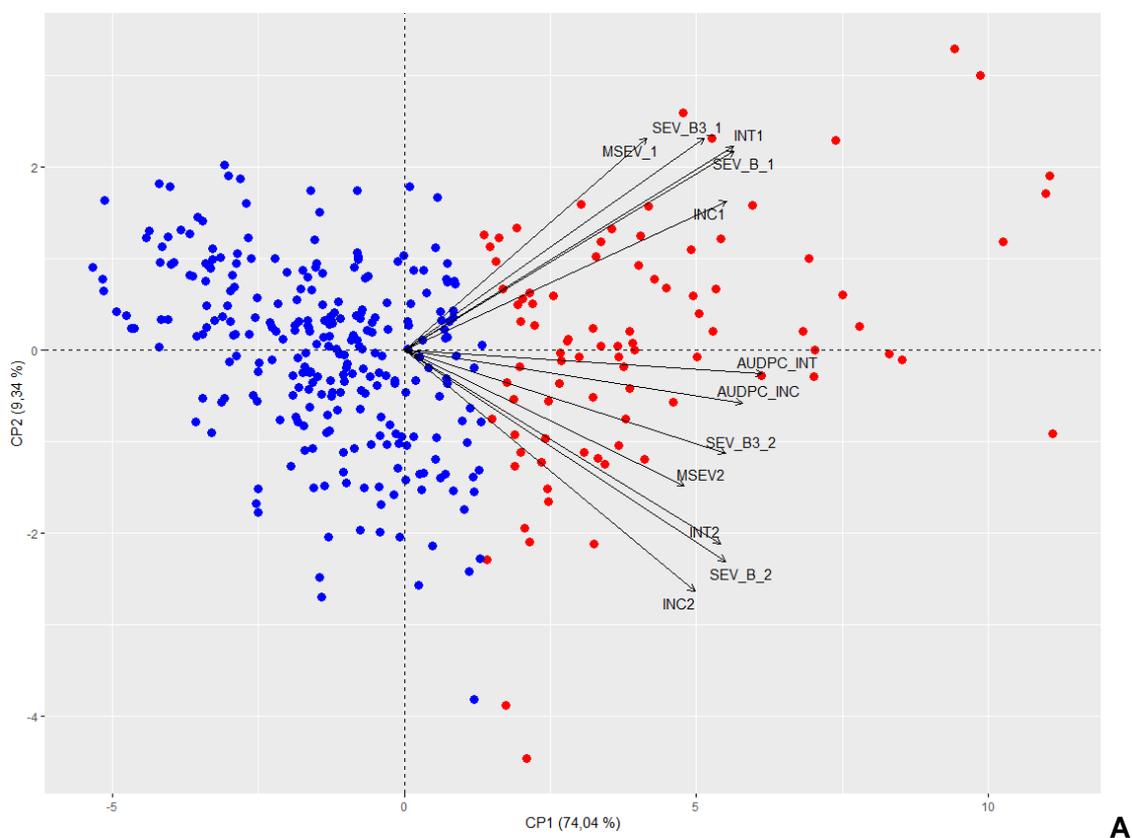
### Análisis Multivariados de Conglomerados

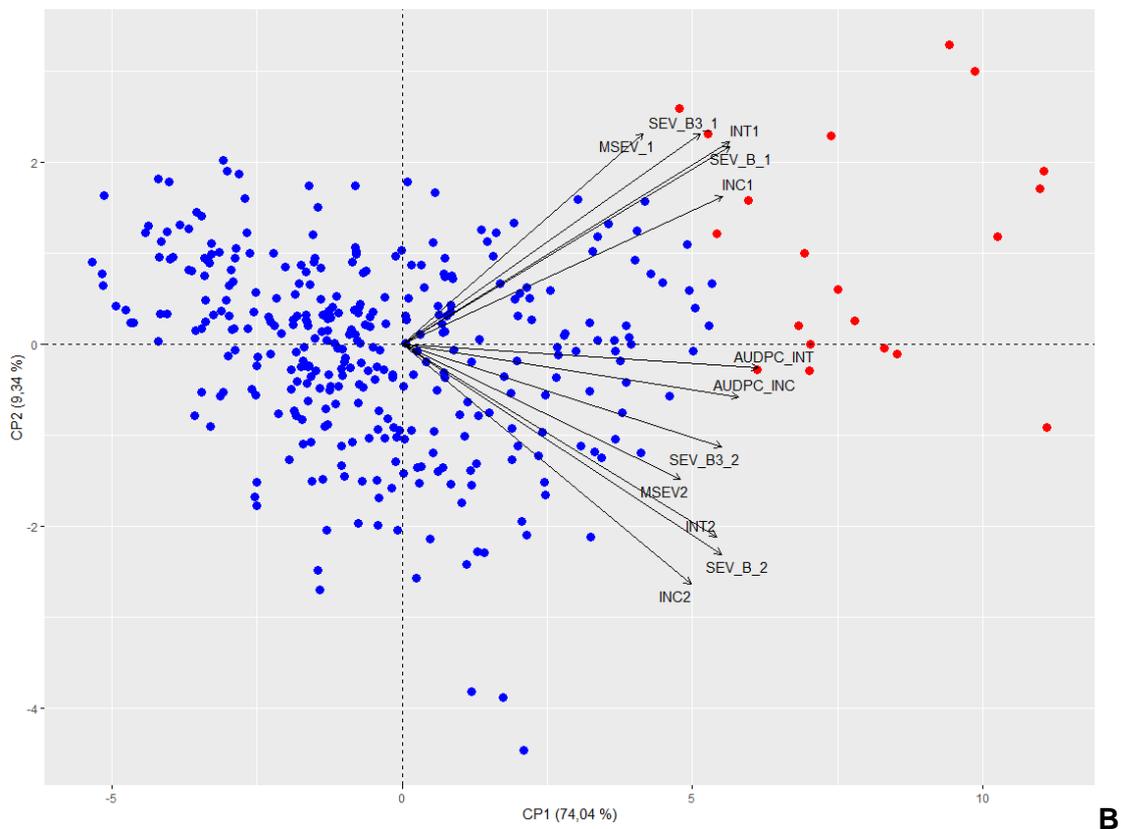
Se estimó una matriz de distancias Euclídeas entre los genotipos que constituyen la población MAGIC2 y se estimaron los números de grupos más probables con el paquete NbClust de R (Charrad et al., 2014) el cual indicó un valor de 2 grupos. En base a este resultado se

realizaron agrupamientos mediante el método no jerárquico *K-means* y mediante el método Jerárquico *Average*, siendo este último resultando el de mayor coeficiente de correlación cofenética (0.79 contra 0.71 del método *K-means*), es decir, más correlacionado con la matriz de distancias Euclídeas obtenidas en primera instancia.

En las Figura 3.9 se exponen los Biplots del ACP particionados con colores por los grupos fenotípicos multivariados definidos por métodos No-Jerárquicos y Jerárquicos respectivamente.

La pertenencia de cada genotipo a cada grupo fenotípico, definidos por ambas metodologías se detalla en el Anexo digital "MAjustadas\_VD".

**A**



**Figura 3.9.** Biplot del ACP particionados por los conglomerados definidos mediante K-Means (A) y Agrupamiento Jerárquico Average (B). El color rojo identifica al grupo de mayor susceptibilidad y el azul al grupo de mayor resistencia.

El método de agrupamiento No-Jerárquico presentó una distribución más homogénea que el agrupamiento Jerárquico del número de genotipos en ambos grupos, contando con un 25 % de los genotipos en el grupo de mayor susceptibilidad y con un 75 % en el grupo de mayor resistencia. En cambio, el agrupamiento Jerárquico dividió a los genotipos en un grupo de mayor susceptibilidad con sólo el 6 % de los genotipos y los genotipos restantes los ubicó dentro del grupo de mejor comportamiento frente a la enfermedad.

### **Selección y Ajuste de Modelos Lineales Generales Mixtos para las variables Altura de planta y DAF**

Los resultados de la evaluación y selección de los diferentes modelos ajustados para el análisis de las variables altura de planta y DAF en la población MAGIC1 y MAGIC2 se representa en las Tablas 3.6 y 3.7.

**Tabla 3.6.** Resultados de evaluación y selección de modelos para DAF y Altura de planta en la población MAGIC2. En rojo se identifican los modelos elegidos.

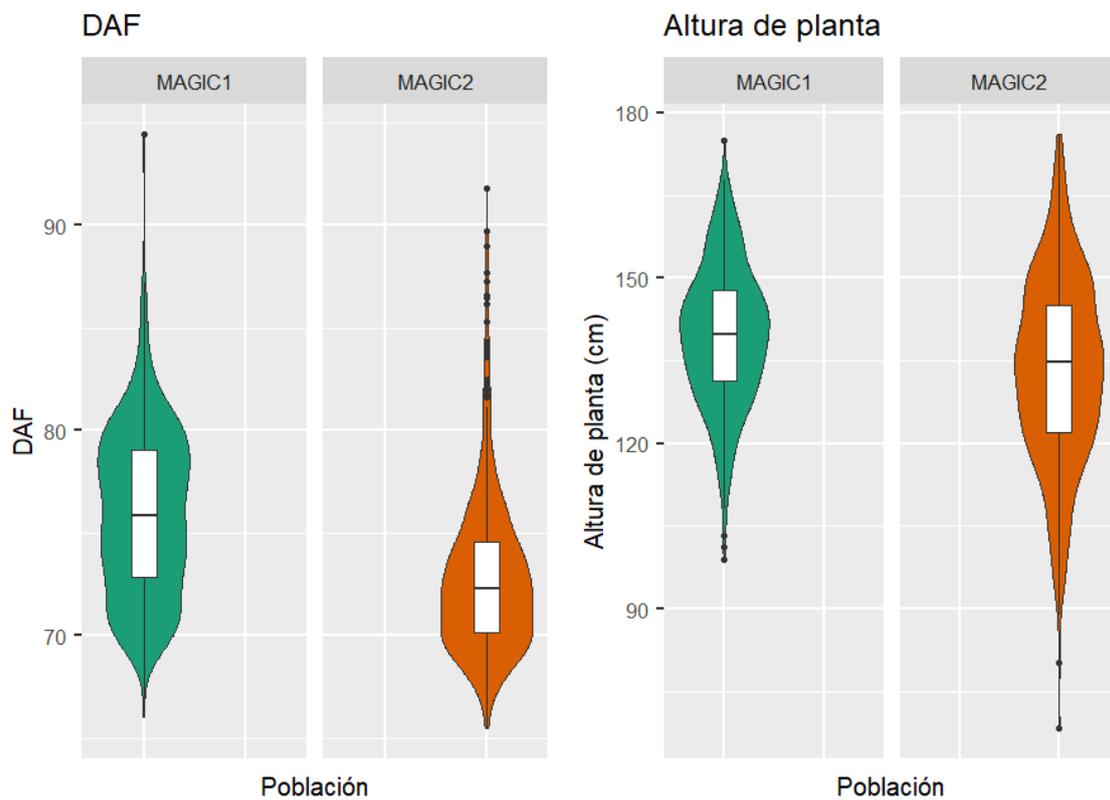
DAF	AIC	BIC	logLik	deviance	Chisq
Modelo 1.1	4457,3	6343,2	-1825,6	3651,3	
<b>Modelo 1.2</b>	4449,8	6340,4	-1820,9	3641,8	9,459
Modelo 1.1 genotipo aleatorio	4634,6	4658	-2312,3	4624,6	
<b>Modelo 1.2 genotipo aleatorio</b>	4636,3	4664,4	-2312,2	4624,3	0,2575
Altura de planta	AIC	BIC	logLik	deviance	Chisq
Modelo 1.1	6644	8522,7	-2920	5840	
<b>Modelo 1.2</b>	6601,3	8484,6	-2897,6	5795,3	44,783
Modelo 1.1 genotipo aleatorio	6603,3	6626,6	-3296,6	6593,3	
<b>Modelo 1.2 genotipo aleatorio</b>	6600,1	6628,2	-3294,1	6588,1	5,1622

**Tabla 3.7.** Resultados de evaluación y selección de modelos para DAF y Altura de planta en la población MAGIC1. En rojo se identifican los modelos elegidos.

DAF	AIC	BIC	logLik	deviance	Chisq
Modelo 1.4	2336,1	4532,3	-658,1	1316,1	
<b>Modelo 1.5</b>	-6406,7	-4206,2	3714,3	-7428,7	8744,8
Modelo 1.6	-6625,6	-4425,1	3823,8	-7647,6	218,94
Modelo 1.7	-8191,3	-5986,4	4607,6	-9215,3	1567,64
Modelo 1.4 genotipo aleatorio	3052,6	3069,8	-1522,3	3044,6	
Modelo 1.5 genotipo aleatorio	3052,7	3074,2	-1521,4	3042,7	1,882
<b>Modelo 1.6 genotipo aleatorio</b>	3015,1	3071,1	-1494,5	2989,1	31,442
Modelo 1.7 genotipo aleatorio	3032,5	3058,4	-1510,3	3020,5	22,189
Altura de planta	AIC	BIC	logLik	deviance	Chisq
Modelo 1.4	3970,4	6171,8	-1474,2	2948,4	
<b>Modelo 1.5</b>	-7964,6	-5758,8	4494,3	-8988,6	11936,92
Modelo 1.6	-7605,3	-5399,6	4314,7	-8629,3	0
Modelo 1.7	-8592,7	-6382,6	4809,3	-9618,7	989,35
Modelo 1.4 genotipo aleatorio	4640,2	4657,5	-2316,1	4632,2	
Modelo 1.5 genotipo aleatorio	4640,5	4662,1	-2315,3	4630,5	1,6942
<b>Modelo 1.6 genotipo aleatorio</b>	4630,3	4686,3	-2302,2	4604,3	20,0358
Modelo 1.7 genotipo aleatorio	4636,4	4662,2	-2312,2	4624,4	6,169

Las medias fenotípicas ajustadas a cada familia para las variables DAF y altura de planta, se detallan en el Anexo digital "Medias\_DAF\_ALT\_LTIP\_PCAP".

La comparación de la distribución de las medias ajustadas obtenidas para las variables DAF y Altura de planta para las poblaciones MAGIC1 y MAGIC2 se grafica en la Figura 3.10.

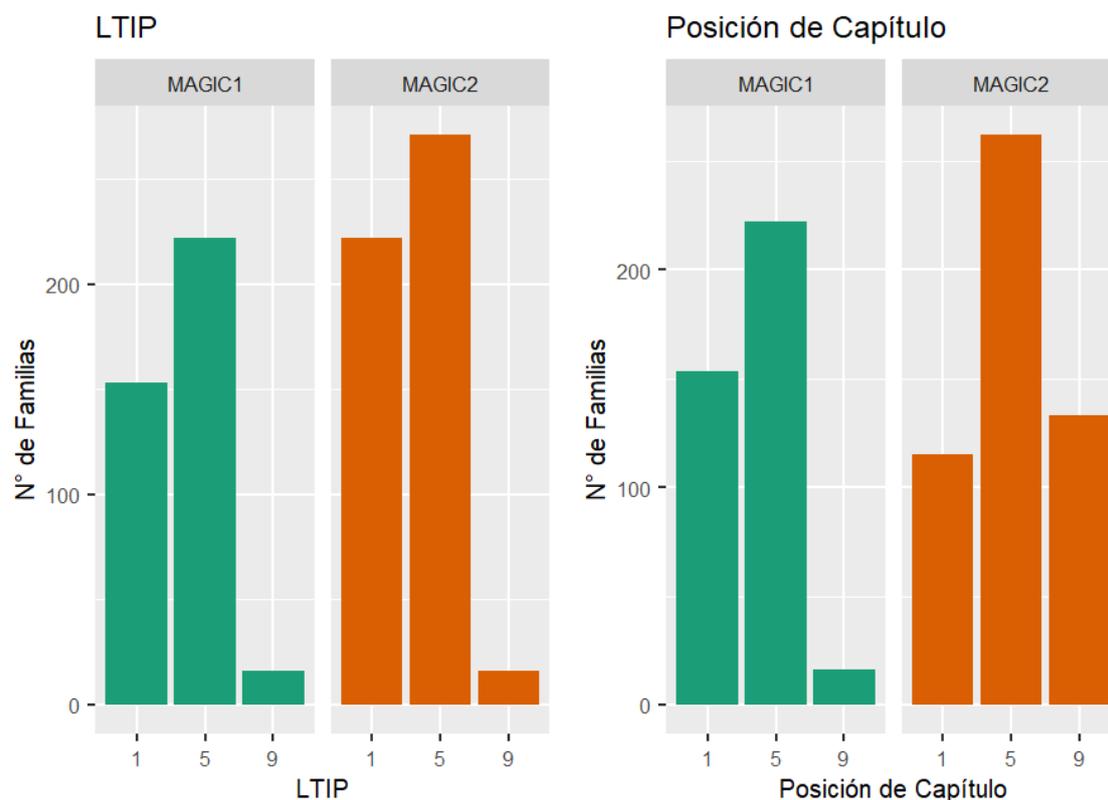


**Figura 3.10.** Valores medios y distribución de las medias ajustadas para DAF y Altura de planta para las poblaciones MAGIC1 y MAGIC2.

La Figura 3.10 muestra que la población MAGIC1 presentó valores medios más altos y significativos (a partir de una prueba T con nivel de significancia: 0,05) para DAF y para altura de planta que la población MAGIC2. Estos resultados eran esperables de acuerdo a las características de los parentales fundadores, los cuales poseen ciclos más largos y mayor altura. También, puede verse el rango de segregación que presentan ambas poblaciones para estos caracteres, siendo un poco más concentrado en la población MAGIC2.

Empleando los modelos que mejor ajustaron pudieron definirse las contribuciones de cada factor a la varianza fenotípica total y así poder estimar para ambas variables la heredabilidad en sentido amplio. En la población MAGIC1 las heredabilidades en sentido amplio fueron para la variable DAF 0,66 y para Altura de planta 0,41. En la población MAGIC2, para la variable DAF la heredabilidad en sentido amplio fue 0,26 y para la Altura de planta fue de 0,33.

Los resultados obtenidos de la evaluación de las variables categóricas LTIP y Posición de Capítulo se grafican en la Figura 3.11. En el Anexo digital “Medias\_DAF\_ALT\_LTIP\_PCAP” se detallan las categorías que obtuvieron cada uno de los genotipos caracterizados.



**Figura 3.11.** Distribución de las categorías obtenidas de las variables LTIP y Posición de Capítulo para las poblaciones MAGIC1 y MAGIC2.

La Figura 3.11 muestra un patrón de distribución similar para LTIP en ambas poblaciones, presentando un menor número de genotipos en la clase fenotípica “9”. En cambio, para la variable Posición de Capítulo si bien ambas poblaciones tienen una mayor frecuencia de familias dentro de la clase “5”, la población MAGIC2 tiene una mayor frecuencia de familias en clase “9” que la población MAGIC1, la cual presentó una frecuencia mínima para dicha clase. Esta diferencia puede explicarse por las líneas parentales de la población MAGIC2, que presentaron una posición de capítulo más decumbente que el promedio de las líneas parentales de la población MAGIC1.

#### **Análisis multivariado integrando variables fitopatométricas, fenológicas y de arquitectura de planta en la población MAGIC2**

A partir de las medias ajustadas para las variables DAF y Altura de planta y la caracterización realizada para los caracteres LTIP y posición de capítulo se abordaron diferentes metodologías multivariadas para el análisis.

En primera instancia se realizó un análisis de componentes principales integrando las variables fitopatométricas de respuesta a MV, DAF y Altura de planta. El total de la varianza

multivariada fue explicado con 14 CPs, alcanzando explicar más del 90 % de la varianza multivariada total con los primeros 5, Tabla 3.8.

**Tabla 3.8.** Componentes principales obtenidos del análisis conjunto de las variables fitopatométricas, DAF y altura de planta.

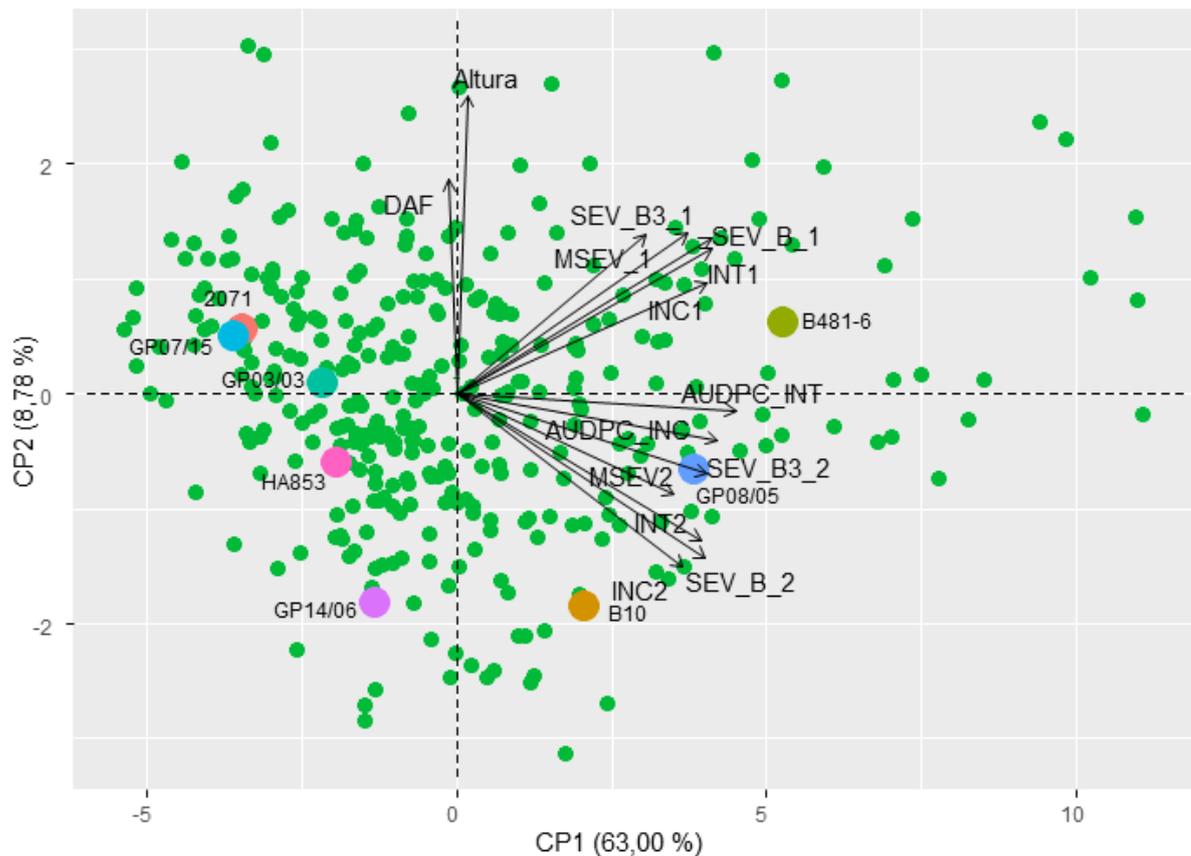
	CP1	CP2	CP3	CP4	CP5	CP6	CP7	CP8	CP9	CP10
<b>Varianza</b>	8,82	1,23	0,98	0,92	0,69	0,49	0,29	0,18	0,14	0,10
<b>Propor. de la Varianza (%)</b>	63,00	8,78	7,01	6,58	4,93	3,54	2,04	1,28	1,01	0,70
<b>Acum. De la Var (%)</b>	63,00	71,78	78,79	85,37	90,30	93,83	95,88	97,16	98,17	98,87

Se seleccionaron los primeros dos CPs, que conjuntamente explicaron el 71,78 % de la varianza multivariada total. El CP1 explica el 63,00 % y el CP2 el 8,78 % de la varianza total. Las contribuciones de las variables a los dos primeros CPs seleccionados se detallan en la Tabla 3.9.

**Tabla 3.9.** Contribuciones de las variables a los dos primeros CPs seleccionados.

Variable	CP1	CP2
INC1	8,83	3,67
MSEV_1	5,05	7,43
INT1	9,16	6,16
SEV_B_1	9,11	7,18
SEV_B3_1	7,51	7,62
INC2	7,20	8,68
MSEV2	6,65	2,93
INT2	8,43	6,28
SEV_B_2	8,72	7,90
SEV_B3_2	8,80	1,87
AUDPC_INC	9,58	0,63
AUDPC_INT	10,93	0,08
Altura de planta	0,02	26,14
DAF	0,01	13,43

En la Figura 3.12 se presenta el biplot obtenido del análisis conjunto de las variables fitopatométricas, DAF y altura de planta. En el mismo se identifican con colores diferentes a las líneas parentales de las familias F<sub>2</sub>.



**Figura 3.12.** Biplot de los componentes principales 1 y 2. Las líneas parentales se identifican con distintos colores y con etiquetas respecto de las familias  $F_2$  de la población MAGIC2.

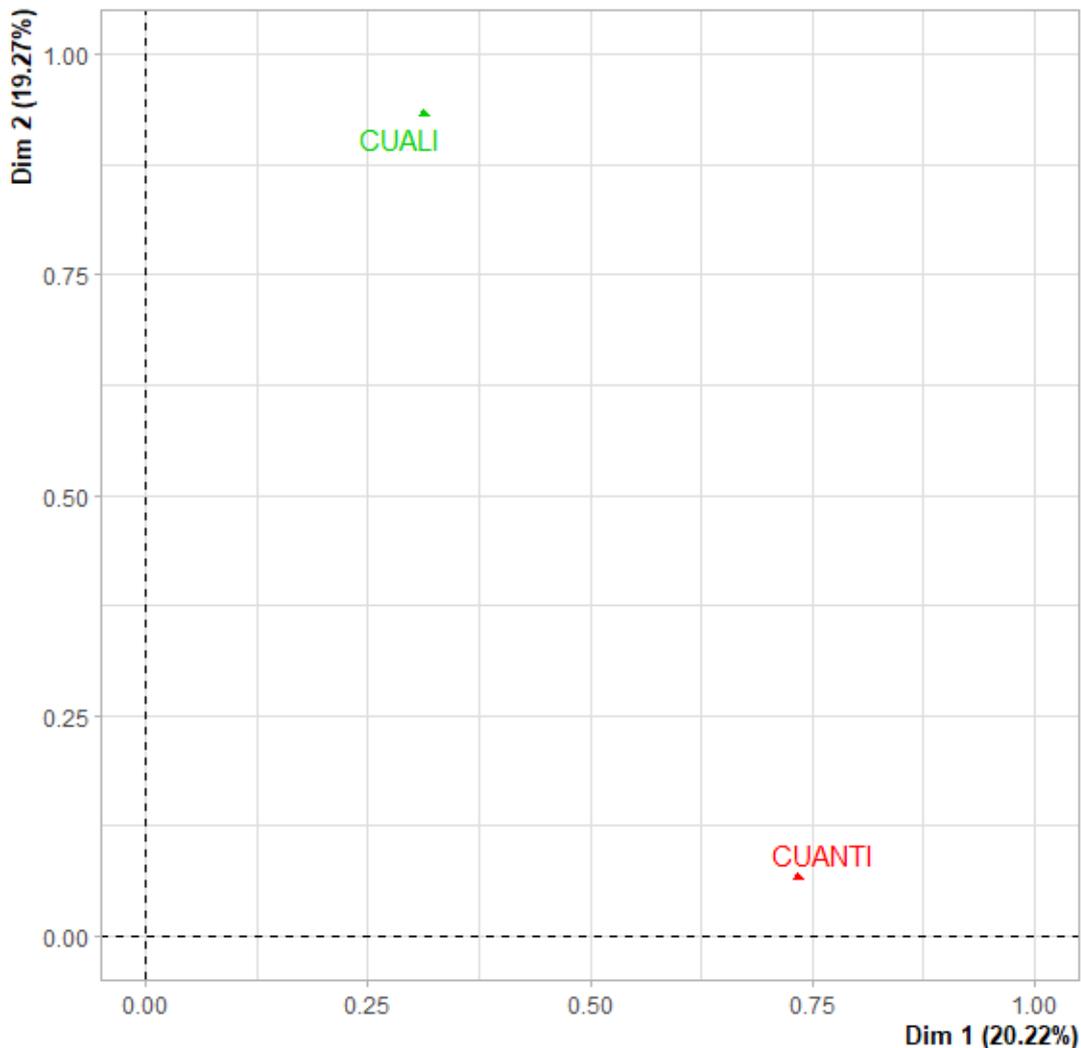
Como puede observarse en el Biplot los vectores DAF y Altura de planta están altamente correlacionados y su proyección está dada casi completamente sobre el CP2, como también puede verificarse en la Tabla 3.9. Dada la orientación y proyección de los vectores DAF y Altura de planta, en los cuadrantes inferiores se ubican los genotipos de menor longitud de ciclo y altura. Además, la figura nos permite ver una distribución equilibrada de genotipos en cuanto a la longitud del ciclo y altura de planta, como así también una tendencia de aquellos genotipos que presentan un ciclo más largo poseen un mejor comportamiento frente a la enfermedad.

Con el objetivo de profundizar la caracterización de la población MAGIC2 se incorporó al análisis multivariado las variables de arquitectura de planta, LTIP y Posición de Capítulo. Al ser estas dos variables de naturaleza categórica ordinal es que se realizó una Análisis Factorial Múltiple (AFM) que permitió el análisis conjunto de variables de tipo cuantitativas y cualitativas.

El conjunto de datos fue dividido de acuerdo a la naturaleza de las variables, definiéndose el grupo de variables de tipo cuantitativas (incluyendo las variables fitopatométricas, DAF y

Altura de Planta) denominado “CUANTI” y el otro grupo compuesto por las variables cualitativas (LTIP y Posición de capítulo) denominado “CUALI”. El total de la varianza multivariada fue explicado principalmente por las primeras cinco dimensiones, alcanzando explicar más del 85 % de la varianza multivariada total.

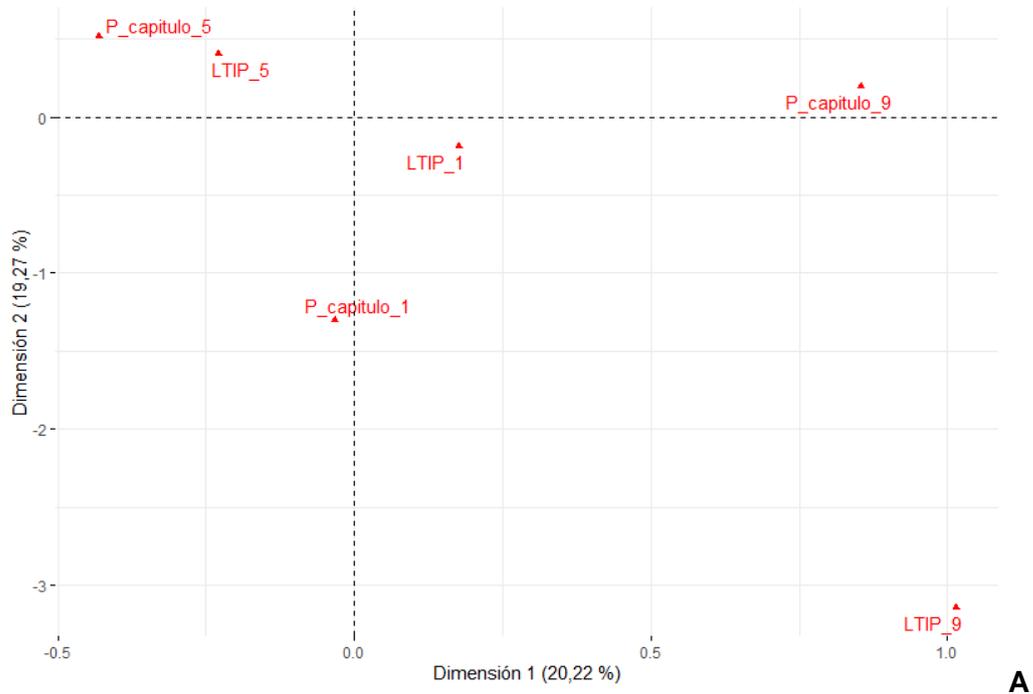
La Figura 3.13 representa la disposición de los dos grupos de variables en el plano general de las primeras dos dimensiones del AFM.

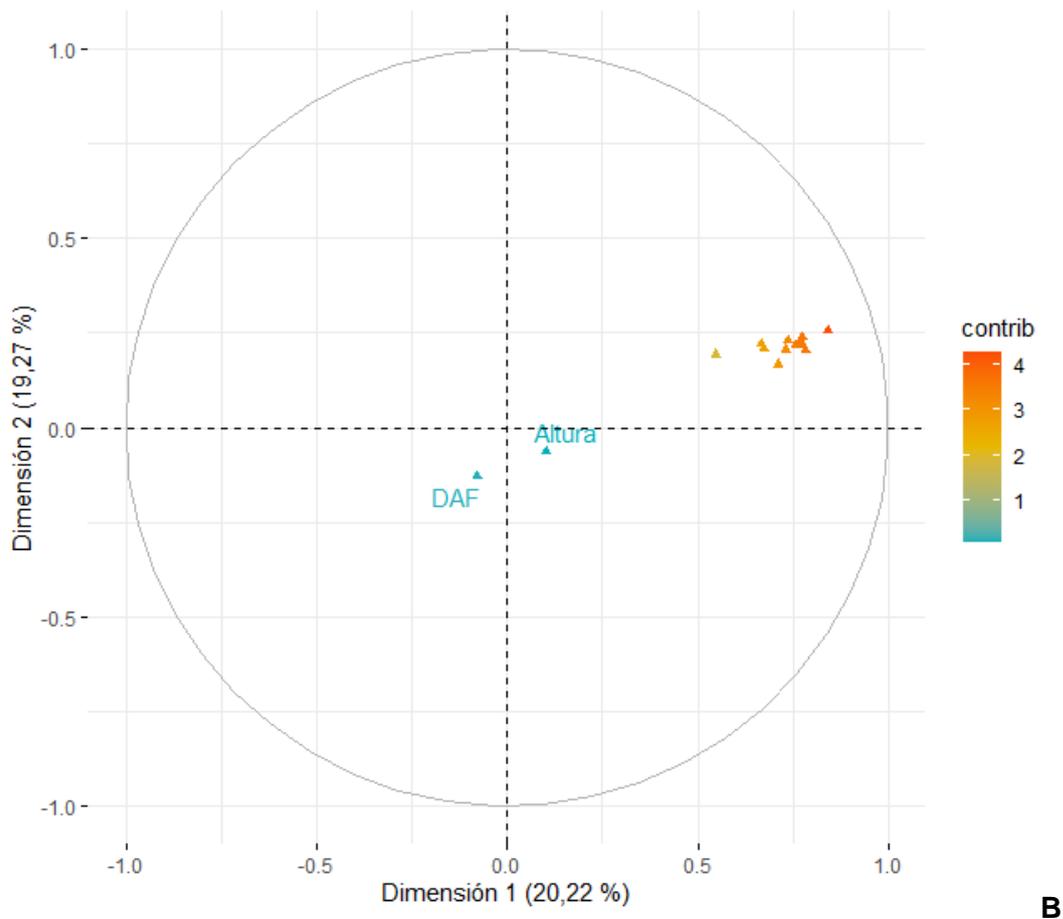


**Figura 3.13.** Representación gráfica en el plano principal de los dos grupos de variables en el AFM.

La Figura 3.13 muestra que las variables cuantitativas son explicadas por el factor 1, en cambio el grupo de variable cualitativas es explicado en mayor medida por el factor 2 pero también tiene una contribución importante sobre el factor 1. Además, debido a su ubicación distante en el plano general ambos grupos de variables no muestran interacción.

En la Figura 3.14 se presenta las relaciones entre las variables cualitativas por separado de las variables cuantitativas en el plano general del AFM.



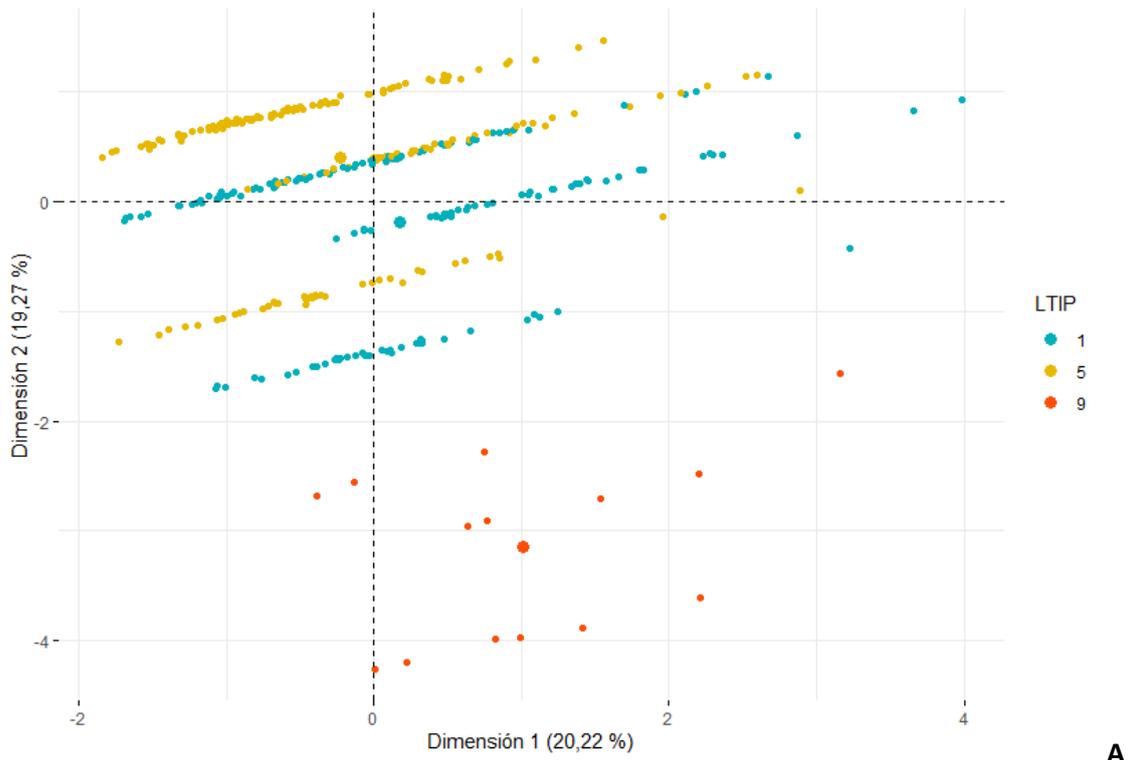


**Figura 3.14.** Representación gráfica de las variables cualitativas y cuantitativas utilizadas para el AFM. **A:** Relaciones entre las variables de tipo cualitativo. **B:** Relaciones entre las variables de tipo cuantitativo indicándose con una escala de color la contribución de cada uno sobre las dimensiones.

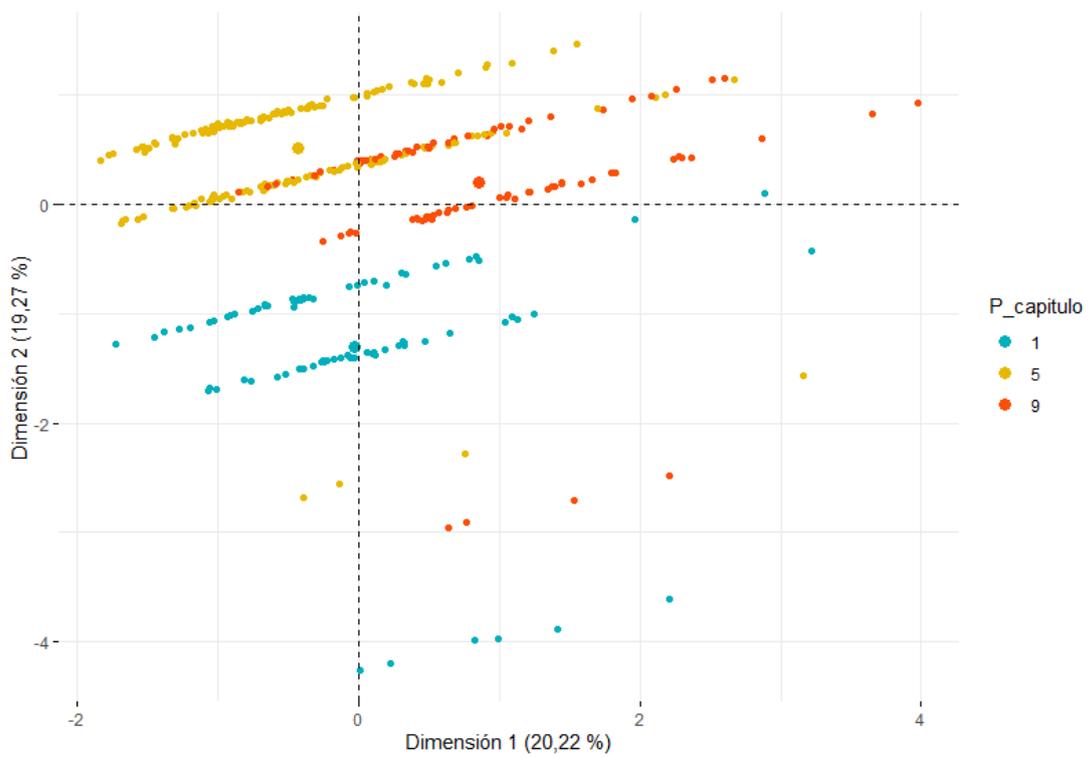
De la Figura 3.14 se desprende que los individuos que presentan los mayores valores para las variables LTIP y Posición de Capítulo se ubicarán a la derecha en los cuadrantes superior e inferior, los que presenten valores medios se ubicaran en el cuadrante superior izquierdo, denotándose aquí la estrecha relación entre las variables para la categoría media debido a su proximidad en el plano, y los individuos con valores inferiores se ubicaran en la zona media del plano en la parte inferior.

Con respecto a las variables cuantitativas los resultados son similares a los obtenidos con los ACPs desarrollados previamente, las variables que mayor contribución tuvieron sobre la Dim.1 fueron las fitopatométricas y en menor medida la altura de planta. En cambio, DAF contribuye mayoritariamente a la Dim.2.

En la Figura 3.15 se representan los individuos en el plano general obtenido por el AFM y diferenciados por las categorías que obtuvieron para las variables cualitativas utilizadas.



A



B

**Figura 3.15.** Representación de los individuos en el plano principal del AFM y su caracterización para las variables LTIP y Posición de capítulo. **A:** Se identifican las

categorías obtenidas para la variable LTIP. **B:** Se identifican las categorías obtenidas para la variable Posición de capítulo.

La Figura 3.15 muestra que el AFM permite obtener un claro ordenamiento de los individuos integrando las variables fitopatométricas, de fenología y de arquitectura de planta resultando de gran valor para la caracterización de la población MAGIC2. Las familias se ubicaron de izquierda a derecha del plano en un gradiente de mayor a menor tolerancia frente a MV. El AFM permitió evidenciar la variabilidad en las características fenológicas y de arquitectura de planta evaluadas con respecto a la respuesta a MV. Aunque, puede notarse una mayor cantidad de genotipos que reúnen baja tolerancia frente a MV y mayores valores de inclinación de capítulo.

## DISCUSIÓN

La caracterización fenotípica de las poblaciones MAGIC de girasol en desarrollo que se presentan en esta Tesis resultan de gran importancia para evaluar el potencial de estas poblaciones para estudios posteriores de mapeo de regiones genómicas para distintos caracteres de interés agronómico.

Para este trabajo de Tesis el ensayo de caracterización fenotípica de la respuesta a MV se realizó en el infectario de la EEA INTA Balcarce que tiene una larga historia de ensayos de resistencia a *Verticillium* en girasol y papa. Desde 2013 hasta la campaña 2020/21, el infectario estuvo sujeto al monocultivo de girasol con incorporación mecánica del rastrojo de los ensayos al suelo, incrementando sustantivamente la carga de inóculo año tras año. El sistema de labranza implementado en el infectario natural para la implantación de cultivos es exclusivamente convencional, lo que implica roturación, desfragmentación y movilización de suelo. Estas prácticas de preparación de suelo, tienden a incrementar la carga de inóculo patogénico (microesclerocios) al mismo tiempo que aumenta la homogeneidad de la distribución del inóculo año a año, esto permite el incremento de la incidencia y severidad de la enfermedad (Quiroz et al., 2008). El infectario de la EEA-INTA Balcarce posee una carga de inóculo muy elevada, habiéndose cuantificado en la campaña 2015/16 valores cercanos a 900 UFC/g de suelo (Erreguerena et al., 2019).

La variabilidad espacial de la carga de inóculo en el infectario fue considerada en el diseño experimental de los ensayos, optando por un diseño Alpha Lattice con genotipos susceptibles aleatoriamente distribuidos como controles, para poder ponderar eficientemente la fuente de variación espacial. Los modelos que mejor ajustaron según los criterios de AIC y de Información Bayesiana (BIC), fueron para la mayoría de las variables empleadas los que implementaron el diseño Fila–Columna. Asimismo, la heredabilidad generalizada de Cullis et al., (2006) fue estimada para la comparación de los modelos Alpha Lattice y Fila–Columna con el modelo espacial P-splines según lo recomendado por Rodríguez-Álvarez et al. (2018). Las estimaciones de las heredabilidades fueron más altas en todos los casos para los modelos Fila-Columna y Alpha Lattice, además se evaluaron las correlaciones entre las medias ajustadas por el modelo Fila–Columna y el espacial P-splines y las mismas fueron cercanas a 1, lo que demostró que la implementación del modelo Fila–Columna es suficiente para el modelado de la variabilidad espacial que presentó el experimento.

En este trabajo y en la bibliografía relevada para la MV en girasol, la variable SEV fue medida mediante una escala ordinal que indica el porcentaje de la planta exhibiendo síntomas de marchitamiento y abigarrado foliar (Shah & Madden, 2004). En la escala utilizada de 6 niveles

el "0" representa la ausencia de síntomas, el "3" alrededor del 50 % de los estratos de la planta afectados y el "5" hasta el estrato superior de la planta comprometido (Montecchia, 2019; Montecchia et al., 2021). Otras escalas utilizadas para medir severidad de MV en girasol fueron la propuesta por Bertero de Romano & Vazquez (1982) que implementó una escala de 0 a 4 dividiendo la planta en cuartos durante el estado fenológico R5. Por su parte, Galella et al (2004) implementaron una escala de 9 a 1 en la que el 1 representó la mayor susceptibilidad y el 9 la resistencia durante el estado fenológico R6. Recientemente, en un trabajo realizado en INRA (Francia) Missonnier et al (2017) utilizó una escala de 1 a 9 siendo, el 1 el nivel de ausencia de síntomas durante los estados fenológicos R5 y R9. La evaluación implementada en esta Tesis fue realizada planta por planta y promediada a nivel parcela. Luego, los valores obtenidos fueron transformados mediante la metodología propuesta por Gianola & Norton (1981) para obtener así una distribución normal. Una vez normalizados los valores de SEV, se ajustaron los modelos mixtos descriptos en los resultados para obtener las medias ajustadas de las familias.

Entre las variables analizadas en este trabajo, las que mostraron diferencias significativas entre los genotipos fueron INC1, AUDPC\_INC, INT\_2, AUDPC\_INT y SEV\_B3\_2. De estas cinco variables la que presentó una proporción más equilibrada entre las dos clases definidas fue la variable AUDPC\_INT, la clase "A" que representa a los genotipos con valores medios más altos, es decir genotipos más susceptibles, agrupó al 20,45 % de los genotipos y la clase "B" a la porción restante, por lo que esta variable *per se* resultó para este experimento la más apropiada para caracterizar e identificar diferencias entre los genotipos. La variable AUDPC\_INT como fue descrita en Materiales y Métodos es la media ponderada por parcela resultante de la multiplicación de las frecuencias obtenidas para cada nivel de SEV e integra los valores obtenidos en ambas fechas de evaluación. Los resultados obtenidos se diferencian con los obtenidos en experimentos realizados en el mismo infectario durante 5 campañas de evaluación y con una población de estudio más diversa como la PMA de girasol de INTA por Montecchia et al (2021). En este último trabajo, las variables que mejor representaron las diferencias entre los genotipos fueron la incidencia y severidad a floración.

La estimación de la heredabilidad permite conocer la contribución de la varianza debida a los componentes genéticos sobre la varianza fenotípica total de un carácter de interés en una población determinada (Falconer & Mackay, 1996). Conocer la heredabilidad es de gran importancia en estudios de mejoramiento, porque permite determinar el grado de vinculación entre los valores fenotípicos observados y los valores genéticos para el carácter estudiado en una población determinada (Bernardo, 2020; Schmidt et al., 2019). Los valores de heredabilidad por lo tanto se ven afectados por factores genéticos y ambientales, por lo que

cambios en el material genético estudiado y en el ambiente de evaluación provocarán cambios en las estimaciones del valor del carácter, estas características hacen que los valores de heredabilidad sean muy difíciles de extrapolar (Falconer & Mackay, 1996).

Entre las variables estudiadas para la caracterización de la respuesta frente a MV de la población MAGIC2 se destacan con valores más altos la AUDPC\_INT con 0,90, INT\_1 con 0,72 y AUDPC\_INC con 0,70. Estos valores altos en las variables que integran los datos relevados en ambas fechas de evaluación muestran un mayor impacto de la varianza del componente genético sobre la varianza total del carácter respecto al aporte del resto de las variables. Estos resultados son similares a los obtenidos por Montecchia et al (2021) quienes reportaron valores más altos de heredabilidad en sentido amplio para variables integradoras como AUDPC\_INC. En este trabajo también fueron obtenidos resultados similares en cuanto a la mayor contribución del componente genético a la varianza fenotípica total de la variable SEV en comparación con la variable INC. Los menores valores de heredabilidad obtenidos para INC, pueden deberse a que, por la naturaleza binomial de su distribución, se considera un valor fijo para la varianza residual tal como proponen Snijders & Bosker (2012), lo que probablemente conlleve a una subestimación de la heredabilidad. Las altas estimaciones obtenidas resultan de gran importancia e interés para la implementación de esta población en futuros estudios de mapeo fino de QTLs para discernir la arquitectura genética de la resistencia frente a MV.

Las estimaciones de heredabilidad en sentido amplio para las variables DAF y Altura de planta fueron mayores en ambos casos para la población MAGIC1 que para la población MAGIC2. Estas diferencias entre poblaciones pueden explicarse por los diferentes diseños experimentales utilizados para evaluarlas. Los diseños de tipo aumentado, en cualquiera de sus variantes, resultan ser por lo general menos precisos, por tener solamente repeticiones de algunos genotipos y estimar el error experimental en base a estos (Crossa & Federer, 2012). Sin embargo, la utilización de este tipo de diseño es frecuente en los programas de mejoramiento debido a la escasez de semilla de los genotipos a evaluar en estados tempranos de desarrollo del germoplasma y porque estos diseños brindan la posibilidad de evaluar un mayor número de genotipos, resignando precisión pero con una menor demanda de recursos (Federer & Raghavarao, 1975; Federer et al., 2001). En este trabajo fue muy útil la implementación del diseño de tipo aumentado permitiéndonos evaluar la misma población de familias F<sub>2</sub>, de la cual disponíamos cantidades limitadas de semilla, frente a MV en Balcarce y la fenología, altura de planta y los caracteres de arquitectura de planta ya mencionados en Pergamino.

Al observar las elevadas y positivas correlaciones entre las medias ajustadas de cada una de las variables fitopatométricas empleadas, se empleó el análisis de componentes principales para integrar la información que brindan estas variables. Este análisis resultó muy adecuado para obtener componentes principales que expliquen entre los dos primeros el 83,38 % de la varianza multivariada total. Estas variables que se desprenden de las combinaciones lineales de las 12 variables fitopatométricas empleadas permitieron un ordenamiento de los genotipos e identificar grupos con respuesta contrastante frente a la enfermedad. En el mismo sentido, el análisis de conglomerados permitió establecer 2 grupos diferenciados en su respuesta a la enfermedad. Los métodos de agrupamiento jerárquicos resultaron en grupos más heterogéneos que los de los formados por métodos No-Jerárquicos. Los resultados obtenidos permiten la identificación de genotipos contrastantes para futuros estudios de validación de marcadores moleculares asociados a la resistencia de la planta frente a MV, desarrollados en estudios previos del grupo (Montecchia et al., 2021).

La inclusión de las variables DAF y Altura de planta en un segundo ACP permitieron obtener componentes principales explicando los primeros, el 71,78 % de la varianza multivariada total. Este análisis permitió ver una correlación positiva y alta entre las variables DAF y Altura de planta, y una tendencia de que aquellos genotipos de ciclo a floración más largo manifestaron mejores comportamientos frente a MV. Resultados similares entre la respuesta frente a la enfermedad y el ciclo fueron reportados por Fick & Zimmer (1974) y Montecchia et al. (2021).

El estudio de la inclinación de la lámina realizada en ambas poblaciones reveló la preponderancia de genotipos con hojas de tipo planófila, registrándose una baja frecuencia de genotipos en ambas poblaciones con hojas de tipo erectófila. Si bien la frecuencia de genotipos con menor ángulo de inclinación fue baja, ambas poblaciones presentan variabilidad para este carácter, resultando de gran importancia la caracterización fenotípica previa realizada en este trabajo, para el uso de ambas poblaciones en estudios futuros que exploren los factores genéticos responsables de este carácter. La influencia de este carácter en la búsqueda de genotipos de girasol más eficientes en la intercepción lumínica y en el uso de la radiación ha sido estudiado por diferentes autores en girasol y en otros cultivos (Aguirrezábal et al., 1996; Hernández, 2010; Hernández & Orioli, 1983, 1994). Sin embargo, a pesar de la importancia que tiene este carácter para la producción del cultivo, hasta el momento el nivel de conocimiento sobre las bases genéticas responsables es relativamente bajo (Dimitrijevic & Horn, 2018) y es uno de los atributos que distintos autores coinciden en que se deben focalizar los estudios con el objetivo de obtener cultivares que se adapten a las condiciones de cambio climático que se presentan en la actualidad y que se pronostican para el futuro (Debaeke et al., 2021). El desarrollo y disponibilidad de recursos genéticos que

presenten variabilidad para este carácter y que reúnan otras características de interés para el cultivo resultan fundamentales para el entendimiento de la arquitectura genética de caracteres de interés agronómico. Asimismo, serán una fuente de recursos de gran utilidad para el aporte de germoplasma a los programas de mejoramiento con el objetivo de obtener híbridos adaptados a las condiciones de cultivo actuales y futuras.

La inclinación y curvatura del capítulo es uno de los objetivos primordiales en la selección de cultivares de girasol adaptados a las condiciones de cultivo en nuestro país y en el mundo debido a su influencia en la tolerancia de los híbridos de girasol al daño producido por aves. En la actualidad, la búsqueda de genotipos de girasol que presenten una menor exposición del capítulo a las aves, es uno de los atributos clave en los programas de mejoramiento para la obtención de híbridos adaptados a las condiciones de cultivo. Estudios previos en la búsqueda de características de tolerancia al ataque de aves en girasol no tuvieron resultados satisfactorios debido a que los genotipos con un mejor comportamiento frente a las aves tuvieron un menor rendimiento de grano y contenido de aceite (Linz et al., 2011). Khaleghizadeh (2011) por su parte, estudió el impacto de diferentes caracteres morfológicos de la planta sobre la preferencia alimentaria de las aves y cómo influyen en la eficiencia del ave en la toma del alimento. Los resultados obtenidos por este autor revelaron que los caracteres que mayor influencia tienen sobre el daño de las aves al cultivo fueron una menor altura del capítulo sobre la superficie del suelo, una mayor distancia del capítulo al tallo, la forma cóncava, el mayor diámetro y la inclinación del capítulo orientándose hacia el suelo, brácteas largas y con ángulo de inclinación abiertos y la intensidad de inserción de los aquenios al capítulo. Similares conclusiones fueron reportadas por otros autores como Bullard, (1988), Fox & Linz, (1983) y Parfitt, (1984).

Hasta el momento no se han publicado estudios referentes a la arquitectura genética responsable de los caracteres anteriormente citados para morigerar el daño de las aves. En base a lo expuesto, la caracterización preliminar realizada en ambas poblaciones MAGIC de la posición del capítulo es de gran importancia para los estudios que pueden realizarse a futuro. Los resultados obtenidos revelaron la preponderancia en ambas poblaciones de genotipos con ángulos de inclinación de capítulo entre 90 y 180°, y menores frecuencias de genotipos dentro de las otras dos categorías de clasificación empleadas en este trabajo, lo cual demuestra la presencia de variabilidad para este carácter en ambas poblaciones, recurso fundamental para futuros estudios de mapeo genético.

La técnica de Análisis Factorial Múltiple (AFM) es muy utilizada en distintas disciplinas para estudios de datos sensoriales, de ecología, económicos y sociológicos (Abdi & Valentin,

2007), aunque si bien no ha sido tan empleada en el análisis de datos agronómicos se ha reportado la implementación de esta técnica multivariada en diferentes enfoques de estudio. Los reportes de la implementación de la técnica AFM han sido en caracterizaciones de accesiones de maíz provenientes de bancos de germoplasma en diferentes ambientes (Defacio, 2016; Zuliani et al., 2012, 2018), en el estudio del comportamiento de distintas variedades de trigo pan en ensayos multiambientales (Vitelleschi & Chavasa, 2015), para el estudio de características del fruto en dos filiales ( $F_2$  y  $F_3$ ) diferentes de una población de tomate (Del Medico et al., 2020), para estudios de heredabilidad multivariada para caracteres de calidad del fruto en tomate (Del Medico et al., 2019) y para la caracterización fenotípica de clones de banana integrando en un mismo análisis variables de naturaleza cuantitativa y cualitativa (Del Medico et al., 2021).

La última implementación citada en el párrafo anterior de la técnica AFM fue la realizada también en este trabajo, permitiendo dividir a las variables consideradas en el análisis en dos categorías, las cuantitativas, altura de planta y días a floración, y las variables cualitativas, posición del capítulo y altura de la punta de la lámina en relación al nivel de inserción del pecíolo. Esta metodología permitió la caracterización fenotípica integral en un análisis de tres vías de variables cuantitativas y cualitativas, permitiendo relevar la variabilidad existente en la población MAGIC2 para los distintos caracteres evaluados y definir conjuntos de familias con características determinadas para futuros estudios de mapeo.

La caracterización fenotípica de ambas poblaciones MAGIC presentada en este capítulo demostró la variabilidad fenotípica que poseen para caracteres como resistencia a MV, posición de capítulo, inclinación de la hoja, altura de planta y DAF. La variabilidad fenotípica constituye un recurso fundamental para la utilidad de estas poblaciones en futuros estudios de mapeo fino de QTLs que se proyectan conducir una vez alcanzado un mayor grado de homocigosis en las mismas. Y asimismo el potencial que poseen dada su base genética, de ser empleadas para el estudio de diversos caracteres de interés agronómico para el mejoramiento del cultivo y para la validación de marcadores asociados a MV en estudios previos del grupo.



---

---

**CAPÍTULO 4: Implementación de imágenes de cámaras  
multiespectrales para la el fenotipado de la marchitez  
prematura por *Verticillium dahliae***

---

---

## INTRODUCCIÓN

### El rol del fenotipado en el mejoramiento genético de plantas

Los esfuerzos en las áreas de agronomía y de mejoramiento genético de los últimos 50 años han permitido triplicar el rendimiento de los cultivos. Los avances tecnológicos a disposición de los mejoradores integrados a los procesos convencionales ampliamente utilizados en programas de mejoramiento ofrecen la posibilidad de aumentar la tasa de ganancia genética. No obstante, los intentos de explotar al máximo las nuevas tecnologías está limitado por factores económicos, estructurales y tecnológicos entre otros. Focalizando en las limitantes tecnológicas, por ejemplo, la posibilidad de aprovechar los avances en genómica, en particular la capacidad de diseccionar la arquitectura genética de caracteres cuantitativos como el rendimiento y la tolerancia a estreses bióticos, están limitados por la capacidad de fenotipado (Araus & Cairns, 2014; Phillips, 2010). En este sentido, la mejora genética es esencialmente un juego de números: cuantos más cruces y ambientes sean utilizados para la selección, mayor será la probabilidad de identificar un genotipo superior. Por lo cual la capacidad de fenotipar un gran número de líneas rápidamente e identificar con precisión la mejor progenie es clave para el mejoramiento genético. Sin embargo, el fenotipado sigue siendo en gran medida una actividad artesanal, específica para cada especie, que requiere mucha mano de obra y es inevitablemente sensible al medio ambiente. Esto hace que el fenotipado sea uno de los principales cuellos de botella para el mejoramiento genético, limitando operativamente y financieramente los estudios genéticos. El genotipado, por otra parte, ha evolucionado proporcionando información genómica rápida y de bajo costo que ha allanado el camino para el desarrollo de grandes poblaciones de mapeo y paneles de diversidad con el objetivo de identificar regiones genómicas asociadas a caracteres de interés (Cabrera-Bosquet et al., 2012; Cobb et al., 2013). Dada la creciente disponibilidad de información genotípica de alta densidad, el estudio de las relaciones genotipo-fenotipo depende cada vez más de la disponibilidad de información fenotípica y ambiental de alta calidad (Cobb et al., 2013; Crossa et al., 2021). Los avances en materia de fenotipado son clave para hacer más eficiente el mejoramiento genético tradicional, molecular y biotecnológico permitiendo garantizar la mejora genética de los cultivos para la seguridad alimentaria (Araus & Cairns, 2014).

En los últimos años las plataformas de fenotipado rápido de alto rendimiento (HTPP de sus siglas en inglés *High Throughput Phenotyping Platform*) se instalaron como alternativa de programas de mejoramiento público y privado de grandes dimensiones. Se trata mayoritariamente de instalaciones totalmente automatizadas en invernaderos o cámaras de crecimiento con un control ambiental preciso. Aunque las HTPP permiten capturar información detallada y no invasiva a lo largo de todo el ciclo de la especie en estudio, los resultados de

los ambientes controlados son diferentes a las condiciones reales que experimentarán las plantas en el campo, lo que dificulta su extrapolación a datos tomados en campo (Araus & Cairns, 2014; Yang et al., 2017; Yang et al., 2013).

Por su parte, las plataformas de fenotipado de campo de alto rendimiento (HTFP de sus siglas en inglés *High Throughput Field Phenotyping*) más comunes utilizan vehículos terrestres con ruedas o aéreos que despliegan múltiples tipos de sensores para medir diferentes caracteres de las plantas en una escala de tiempo de unos pocos segundos por parcela. En el caso de estas plataformas, el proceso de toma de datos demanda mucho tiempo, lo cual limita el número de parcelas sobre las que se pueden relevar datos de manera eficiente (Zhang y Kovacs, 2012).

La alternativa de fenotipado en campo mediante teledetección es especialmente interesante, ya que permite un relevamiento eficaz y repetido de grandes poblaciones de plantas en corto tiempo y puede sustituir o complementar las mediciones sobre el terreno, que requieren elevada mano de obra. Además, las tecnologías actuales permiten alcanzar un alto nivel de resolución y precisión, y son adecuadas para detectar la variabilidad espacial. Una de las principales ventajas de estos sistemas es su versatilidad, ya que la frecuencia y la resolución de la toma de datos pueden elegirse y adaptarse a los caracteres a medir. Aunque el HTFP es muy prometedor para el mejoramiento genético de plantas, se debe considerar los costos que conlleva al momento de implementarlo en los programas de mejoramiento (Reynolds et al., 2019; Reynolds et al., 2020). El HTFP también es particularmente de gran valor para la investigación genética, ya que la disponibilidad de datos fenotípicos precisos de grandes poblaciones de plantas evaluadas a campo que ofrece, permite vincular los procesos que ocurren a nivel molecular con las características morfológicas que se expresan y utilizar el mapeo de ligamiento o GWAS para diseccionar la arquitectura genética de caracteres complejos (Bazakos et al., 2017).

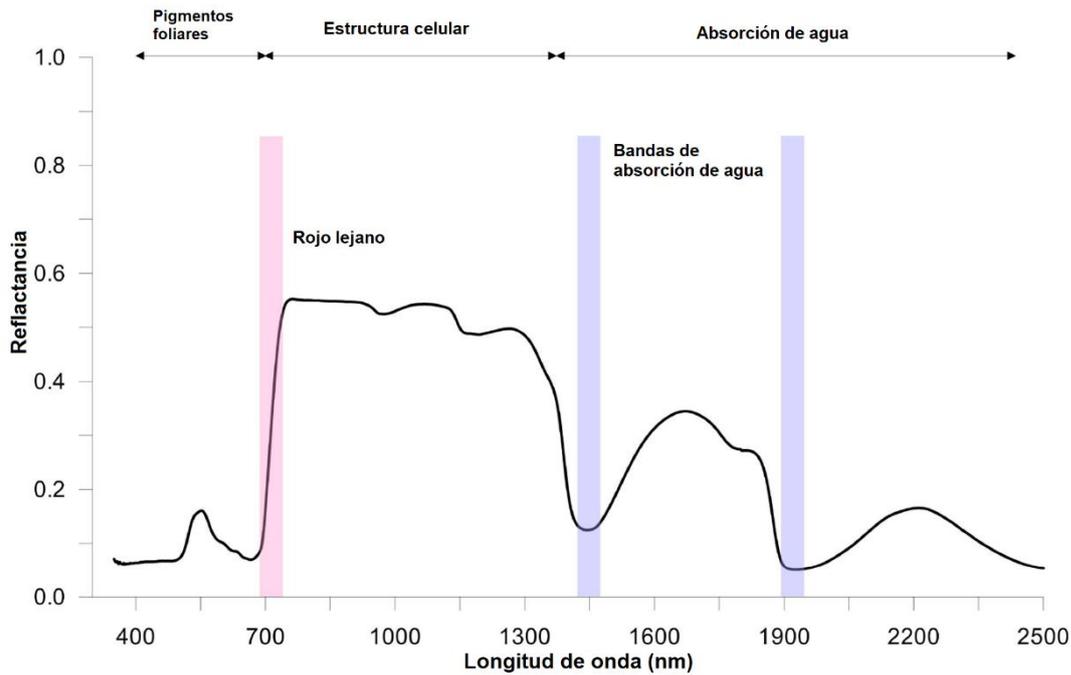
### **Teledetección**

La teledetección es el proceso de detección y seguimiento de las características físicas de un área u objeto mediante la medición de su radiación reflejada y emitida a distancia. Implica el uso de sensores para captar datos de forma rápida y no invasiva. Estos sensores cubren partes específicas del espectro electromagnético y miden la cantidad de radiación (solar entrante) que incluye la radiación de longitudes de onda cortas a muy largas que se refleja en los objetos de interés (Serrano, 2020).

En el caso de la vegetación, la reflectancia está influenciada por su fisiología, composición bioquímica y estructural, y está controlada por tres mecanismos físicos: absorción, reflexión y

transmisión. Cualquier cambio en la estructura de la hoja causado por el genotipo de la planta, la edad de la hoja, la sanidad o el estado fisiológico, puede afectar a la absorción, la transmisión y la reflectancia de las ondas electromagnéticas y contribuir a las fluctuaciones del patrón espectral. Por lo tanto, la vegetación tiene un patrón espectral específico que representa la fracción de la radiación reflejada en función de la longitud de onda y puede utilizarse para evaluar las características de crecimiento y el estado de las plantas. Las regiones del visible (VIS, 400-700 nm), del infrarrojo cercano (NIR, 700-1300 nm) y del infrarrojo de longitud de onda corta (SWIR, 1300-2500 nm) contienen información sobre el estado fisiológico de las hojas, el contenido de agua, los pigmentos y los componentes estructurales de la biomasa (Carter, 1993; Guangyu et al., 2012; Mohammed et al., 2000).

En la región visible, la cantidad de luz reflejada en cada longitud de onda depende de las concentraciones de pigmentos fotosintéticos (clorofila, carotenoides y antocianinas) en las células empalizadas de las hojas, que absorben la mayor parte de la radiación incidente (Ustin et al., 2009). La mayor absorción por parte de la clorofila en los rangos de longitud de onda correspondientes a la región azul y roja da lugar al aspecto verde del tejido foliar. Los carotenoides absorben en la región azul y verde, dando lugar a los colores naranja y rojo. En el caso de las antocianinas la absorción se produce en la región verde, dando lugar a los colores azul y púrpura. En la región del infrarrojo cercano (NIR), la reflectancia es mayor y está determinada por la estructura celular y tisular de las hojas. Además, la estructura del canopy también influye, ya que la radiación NIR que se transmite de los estratos foliares superiores a los inferiores del cultivo, pueden ser reflejada nuevamente hacia el estrato superior del canopy (Gibson et al., 2013). En la región del rojo lejano (parte ascendente de la curva del espectro a 680-750 nm), la absorción de luz por la clorofila es menor que en la parte visible del espectro y se modifica a medida que la planta madura y puede utilizarse como indicador del estrés hídrico y del estado nutricional de la planta (Filella & Peñuelas, 1993; Gibson et al., 2013). Asimismo, componentes como la lignina, la celulosa y principalmente el agua, absorben la radiación en la región de onda corta del infrarrojo cercano. Por lo tanto, la absorción en esta región se relaciona con el contenido hídrico de la hoja (Kokaly et al., 2009). Por ejemplo, en el caso de plantas sanas, reflejan más luz en la región del NIR que las plantas estresadas, mientras que las plantas estresadas reflejan más luz de longitud de onda roja que las plantas sanas. Esto se debe a la pérdida de estructura de los tejidos y a la reducción de la absorción en la banda activa de la clorofila en las plantas estresadas (Karlovska et al., 2015; Mohammed et al., 2000). En la Figura 4.1 se representa un espectro típico de luz reflejada por un cultivo.



**Figura 4.1.** Espectro típico de luz reflejada por un cultivo. Figura adaptada de Galieni et al., (2021).

### Plataformas de fenotipado de alto rendimiento

En general las plataformas de fenotipado de alto rendimiento se pueden dividir en dos grandes grupos, las terrestres y las aéreas. A continuación, se describe sólo el grupo de plataformas aéreas debido a que fue el implementado en este trabajo.

El grupo de plataformas de fenotipado aéreas incluyen a los satélites, aviones y drones. Cuanto más alta esté la plataforma, mayor será el área que se pueda examinar, pero menor será la resolución espacial y el tiempo de revisita. Además, la presencia de nubes puede limitar el uso de plataformas que toman imágenes desde gran altura. Los satélites son la principal plataforma de teledetección para vigilar la vegetación natural, bosques y campos agrícolas, pero para los fines de mejoramiento genético de plantas y la investigación biológica en general presentan una baja resolución espacial (0,7 a 30 m dependiendo del satélite y de las bandas espectrales). Sin embargo, al ser capaces de supervisar grandes áreas con frecuencia, son sistemas ideales para el monitoreo de cultivos y son muy valiosos para la agricultura de precisión (Serrano, 2020).

Las plataformas de teledetección de vehículos aéreos no tripulados (UAV-RSP de sus siglas en inglés *Unmanned Aerial Vehicle Remote Sensing Platforms*) conocidos popularmente como "drones" se caracterizan por su flexibilidad, bajo costo y su baja altura de vuelo y los convierte en una herramienta de gran importancia y versatilidad para el fenotipado de cultivos (Berni et al., 2009; Liebisch et al., 2015) y la agricultura de precisión (Ballesteros et al., 2014;

Gómez-Candón et al., 2014; Zhang & Kovacs, 2012). Estas plataformas permiten combinar sensores, registrar imágenes de alta resolución, son flexibles para monitorear varios ensayos con regularidad y un gran número de parcelas simultáneamente. Se pueden distinguir dos tipos: helicópteros y de alas fijas. Los primeros pueden volar más cerca del suelo (5-20 m) pero sólo pueden utilizar sistemas de sensores fijos, lo que implica limitaciones en la resolución espacial y el rango espectral. Los de alas fijas, por su parte pueden volar a mayor altura (hasta 100-120 m), dependiendo de las restricciones impuestas por los reglamentos de vuelo locales y a mayor velocidad. En consecuencia, pueden cubrir un área mayor, pero con una resolución espacial menor. Una de las principales dificultades asociadas al uso de este tipo de plataformas, ya sean helicópteros o alas fijas, es que son muy dependientes de las condiciones meteorológicas, por lo que su uso se ve dificultado principalmente por vientos fuertes y precipitaciones (Colomina & Molina, 2014; Pádua et al., 2017; Pajares, 2015; Salamí et al., 2014; Sankaran et al., 2015; Shi et al., 2016; Yang et al., 2017).

La elección de la plataforma depende, por tanto, de los requisitos de fenotipado, y suele implicar un compromiso entre la resolución espacial, la precisión, el rendimiento y el costo. Según Mir et al (2019), las plataformas UAV-RSP son las más adecuadas para el HTFP en mejoramiento, ya que permiten la obtención de imágenes de un gran número de parcelas, con diferentes sensores, en un plazo corto de tiempo, reduciendo el riesgo de cambios en las condiciones de iluminación y con una alta resolución espacial (hasta mm). Su flexibilidad permite realizar visitas frecuentes (alta resolución temporal) o planificar vuelos en momentos críticos del desarrollo del cultivo para evaluar el rasgo de interés e investigar la respuesta biológica del cultivo a las condiciones ambientales.

### **Estrategias de análisis de imágenes espectrales y su implementación en los programas de mejoramiento**

El fenotipado es una actividad clave y fundamental para los programas de mejoramiento, con el objetivo de reducir tiempos y esfuerzos, la implementación de distintos tipos de imágenes está siendo una de las técnicas más empleadas para lograr este objetivo. Se han propuesto diferentes enfoques metodológicos basados en el espectro de reflectancia, la temperatura del canopy y la luz visible para evaluar diferentes caracteres de interés en experimentos a campo (Araus & Cairns, 2014). Los rasgos geométricos de un cultivo pueden estimarse construyendo el modelo digital de superficie (DSM de sus siglas en inglés *Digital Surface Model*) o el modelo digital de elevación (DEM de sus siglas en inglés *Digital Elevation Model*) y llevando a cabo un análisis de clasificación de imágenes, pueden estimarse caracteres de interés como altura de la planta, porcentaje de vuelco, la emergencia, entre otros (Hunt et al., 2010; Li et al., 2015). Las características de absorción y reflectancia de los cultivos pueden utilizarse para obtener

las características fisiológicas de un cultivo (Swain et al., 2010; Nigon et al., 2015; Oøvergaard et al., 2010). La temperatura del canopy está estrechamente relacionada con la transpiración del cultivo reflejando el potencial hídrico de la hoja, la conductancia estomática en condiciones de estrés abiótico y biótico (Yang et al., 2017). La combinación de datos hiperespectrales e infrarrojos térmicos también permite predecir el rendimiento de los cultivos (Berni et al., 2009; Gonzalez-Dugo et al., 2015).

Los índices espectrales de vegetación (IV) son combinaciones matemáticas, por lo general cocientes, de bandas espectrales principalmente rojas, verdes e infrarrojas (Chlingaryan et al., 2018). Han sido propuestas para encontrar relaciones funcionales entre las características de los cultivos y las observaciones obtenidas a partir de la teledetección (Wiegand et al., 1979). Desde el desarrollo del índice *Simple Ratio Index* (SR) (Birth & McVey, 1968; Jordan, 1969; Knipling, 1970) y el índice *Normalized Difference Vegetation Index* (NDVI) (Johnson, 2014; Tucker, 1979) se han desarrollado un gran número de índices de vegetación, como el *Enhanced Vegetation Index* (EVI2) (Bolton & Friedl, 2013) y el *Normalized Difference Water Index* (NDWI) (Satir & Berberoglu, 2016), por citar algunos. La disponibilidad de un gran número de índices conduce a la necesidad de elegir y combinar óptimamente los índices para su utilización en función de las características de la planta que se pretenda estudiar (Chlingaryan et al., 2018).

El análisis de las imágenes es una de las grandes limitantes del HTP, por lo que se presentan en la bibliografía una gran variedad de enfoques. Entre estos, la técnica de *Machine Learning* (ML) es una de las más adoptadas para el análisis de grandes cantidades de datos. El ML se refiere a un grupo de algoritmos que permiten reconocer patrones presentes en los datos a partir de los cuales se generan modelos que los representan. Estos modelos, permiten predecir información sobre eventos o escenarios que aún no se han observado. El uso de estos enfoques en fitomejoramiento y en investigaciones en el área vegetal en general, ofrece la oportunidad de analizar grandes conjuntos de datos para descubrir patrones de comportamiento observando simultáneamente una combinación de factores en lugar de analizar cada característica de manera individual. Inclusive poder incorporar al modelo, en el caso de análisis de estreses, los efectos integrados de los factores genéticos, agronómicos, económicos, meteorológicos y humanos (Singh et al., 2016).

El proceso de ML consiste en particionar el set de datos sobre el cual se pretende trabajar en dos matrices, una que se la denomina "*training dataset*" que es la empleada para la calibración del modelo, la matriz remanente se la denomina "*testing dataset*" y es utilizada para la prueba del modelo (Singh et al., 2016). Una vez que el modelo es validado en otros conjuntos de

datos de la misma o de una diferente población de estudio y alcanza valores de precisión adecuados, puede ser empleado como método rutinario para la identificación, clasificación, cuantificación y predicción del carácter en estudio (Singh et al., 2016). Una desventaja que presentan este tipo de algoritmos, es su susceptibilidad al sobreajuste cuando se trata de conjuntos de datos pequeños o grandes con bajos niveles de varianza. Por esta razón, los errores de validación se comparan con los errores de entrenamiento, para verificar que no haya diferencias importantes entre ambos errores (Cawley & Talbot, 2010; Herrero-Huerta et al., 2020).

Las diferentes técnicas de ML disponibles se pueden clasificar en función de cómo se aprenden las características (como es el proceso de aprendizaje) y que se aprende (el objetivo del modelado). El primer punto clasifica a las herramientas de ML en “supervisadas” y “no supervisadas”. Las técnicas supervisadas son aquellas que poseen una información previa asociada a un rasgo particular, es decir por ejemplo se sabe que una planta enferma se asocia a una imagen particular. En cambio, si la técnica no tiene ningún etiquetado previo, la técnica es de tipo no supervisada. El segundo punto clasifica a las técnicas de ML en “generativas” y “discriminativas”, las técnicas generativas captan el patrón general de los datos y permiten generar imágenes sintéticas, en cambio las técnicas discriminativas, a partir de los datos proporcionados son capaces de discriminar patrones específicos (Singh et al., 2016).

Entre las técnicas más utilizadas de ML para el análisis de imágenes en el estudio de plantas, se han reportado trabajos que emplean Análisis Discriminante, Lineal y Cuadrático (Calderón et al., 2015), Máquinas de Vector de Soporte (SVM de sus siglas en inglés *Support Vector Machine*) (Yang et al., 2016), Redes Neuronales Artificiales (ANN de sus siglas en inglés *Artificial Neural Networks*) (Mehra et al., 2016), Árboles de decisión y clasificación (Brichet et al., 2017), *Random Forest* (Herrero-Huerta et al., 2020), enfoques bayesianos (Hernández-Rabadán et al., 2014), entre otras.

#### **Antecedentes de fenotipado automático de enfermedades causadas por *V. dahliae***

El hongo *Verticillium dahliae* infecta la planta a través de las raíces y coloniza su sistema vascular, bloqueando el flujo de agua e induciendo finalmente los síntomas de marchitez (Van Alfen, 1989). Este daño da lugar a una reducción significativa de la tasa de transpiración de las hojas, lo que finalmente conduce a la clorosis y defoliación de las mismas, provocando un cambio de reflectancia espectral (Hillnhütter et al., 2010). El contenido de clorofila tiende a disminuir en las plantas infectadas, mostrando una menor reflectancia en las regiones visibles verde (550 nm) y roja (650-720 nm). Las plantas estresadas también muestran una reducción en la densidad del canopy y en el área foliar que lleva a una disminución de la reflectancia

espectral en el rango infrarrojo cercano (680-800 nm). Además, la región del infrarrojo térmico (8000-15.000 nm) es muy adecuada para la detección de la infección por *V. dahliae* debido a la disminución de la tasa de transpiración que induce el cierre de los estomas, reduciendo el enfriamiento por evaporación y aumentando la temperatura del canopeo. Teniendo en cuenta estos cambios en el espectro de las plantas infectadas, los síntomas de la enfermedad podrían detectarse a distancia en las regiones del espectro visible, del rojo lejano, el infrarrojo cercano y la región del infrarrojo térmico (Calderón et al., 2015).

Hasta el momento no se han reportado estudios que implementen índices de vegetación para el estudio y fenotipado de la Marchitez por *Verticillium dahliae* en girasol. Si se han publicados estudios sobre el uso de índices de vegetación en estudios fenotípicos de esta enfermedad en olivo (*Olea europaea* L.) (Calderón et al., 2015), algodón (*Gossypium hirsutum* L.) (Pabuayon et al., 2019) y en frutilla (Cockerton et al., 2019). En girasol el fenotipado de alto caudal ha sido empleado mediante plataformas estacionarias en invernáculo, como lo es la plataforma *Heliaphen* de INRA (Gosseau et al., 2019), el uso de plataformas mediante drones ha sido empleada sólo para la identificación de malezas en el cultivo (Peña et al., 2015). En otros cultivos se han empleado ampliamente plataformas de fenotipado con drones para el estudio de estreses bióticos, abióticos y para la predicción del rendimiento a partir de los índices de vegetación obtenidos de cámaras RGB, multiespectrales e hiperespectrales (Araus et al., 2018; Barbedo, 2019; Chawade et al., 2019). Se han reportado estudios de enfermedades en soja (*Glycine max*) (Castelao Tetila et al., 2017), papa (*Solanum tuberosum*) (Khot et al., 2015), maíz (*Zea mays*) (Loladze et al., 2019), trigo (*Triticum aestivum*) (Su et al., 2018) por citar algunos ejemplos.

En este capítulo se presenta la implementación de índices de vegetación obtenidos a partir de imágenes capturadas con una cámara multiespectral en un vuelo realizado con un dron sobre un ensayo de respuesta a MV de la población MAGIC2 descrita en el Capítulo 3 para el fenotipado de esta enfermedad en girasol. En este estudio se compararon distintos modelos predictivos basados en técnicas de *Machine Learning* para clasificar la respuesta frente a MV de 349 familias F<sub>2</sub> de la población MAGIC2 y sus 8 líneas parentales a partir de 4 índices de vegetación.

## MATERIALES Y MÉTODOS

### Material vegetal

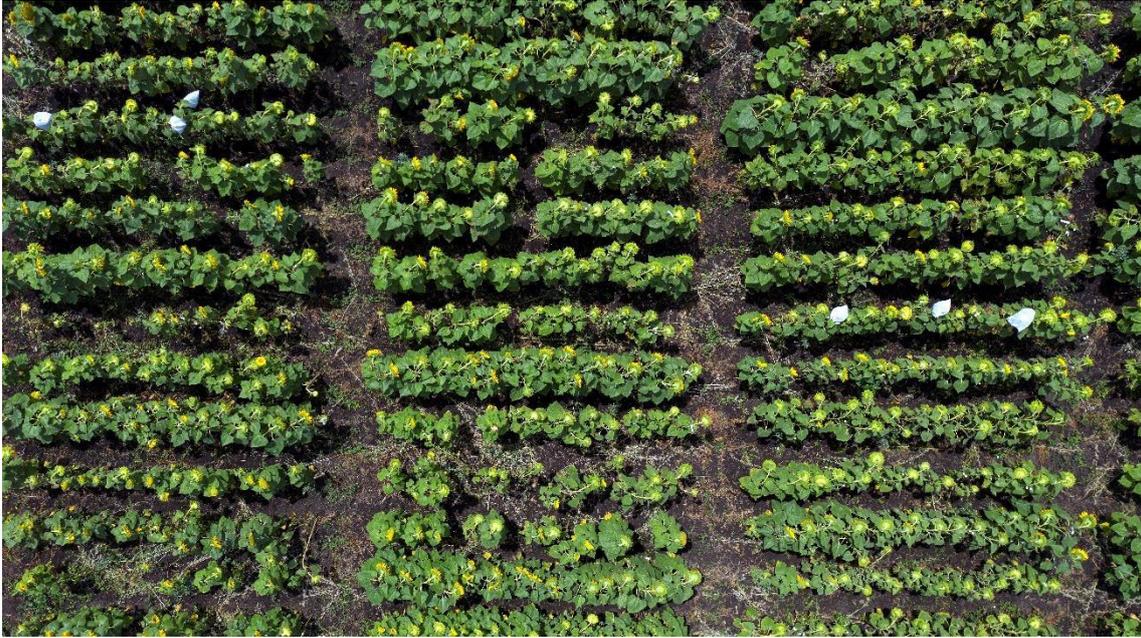
Una selección de 349 familias  $F_2$  de la población MAGIC2 y las 8 líneas parentales fueron caracterizadas fenotípicamente para su respuesta frente a la Marchitez prematura causada por el hongo de suelo *Verticillium dahliae* en el infectario natural de la EEA INTA Balcarce (37°50' 0" S, 58°15' 33" O, Buenos Aires, Argentina) (Figura 4.2) tal como se detalló en el Capítulo 3.



**Figura 4.2.** Fotografía panorámica aérea del ensayo realizado en la EEA INTA Balcarce, Buenos Aires, Argentina.

### Diseño experimental

El diseño experimental fue de tipo Alpha Lattice en dos repeticiones y se utilizó un control interno susceptible para poder relevar la carga patogénica del infectario. La fecha de siembra fue el 9 de noviembre de 2020. La siembra se realizó de forma manual sobre los surcos previamente marcados a 0.7 m con una sembradora, mediante el uso de “escopeta” y golpes de 3 semillas (aprox.) cada 20 cm para alcanzar una densidad de 5 plantas por metro lineal. Las unidades experimentales (parcelas) consistieron en un surco de 5 m de largo (Figura 4.3).



**Figura 4.3.** Unidades experimentales correspondientes a parcelas de un surco.

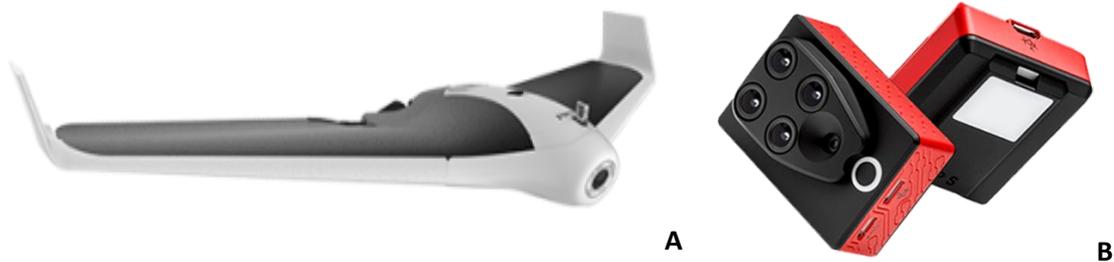
#### **Evaluaciones fenotípicas manuales**

Las evaluaciones fenotípicas manuales se describieron en el Capítulo 3 de esta Tesis.

#### **Fenotipado automático: teledetección y procesamiento de imágenes**

##### **Adquisición de imágenes multispectrales**

Para la captura de las imágenes se realizaron vuelos el día 28 de enero de 2021 (coincidente con la segunda fecha de evaluación manual) a las 12:00 hs AM para minimizar la variación en la iluminación y el ángulo cenital solar (Gu et al., 1992). Las condiciones climáticas fueron de cielo despejado con una velocidad del viento menor a 10 Km/h y una temperatura de 30,8 °C, las cuales permitieron no tener distorsión en las imágenes. Se utilizó un vehículo aéreo no tripulado Drone Parrot Disco-Pro Ag (Fig. 4.4; Tabla 4.1). La altura de vuelo fue de 50 m. Las imágenes fueron capturadas con un sensor multispectral Parrot Sequoia con 4 bandas espectrales en el Verde (550nm  $\pm$  40nm), Rojo (660nm  $\pm$  40nm), Rojo lejano (735nm  $\pm$  10nm) e Infrarrojo cercano (790nm  $\pm$  40nm). La ruta de vuelo se programó con el software Pix4d capture (Pix4D SA, 1015 Lausanne, Switzerland) y se automatizó para garantizar la cobertura de toda la superficie del ensayo.



**Figura 4.4.** A: Drone Parrot Disco-Pro Ag. B: Cámara multispectral Parrot Sequioa.

En la Tabla 4.1 se detallan las características de la cámara multispectral utilizada.

**Tabla 4.1.** Especificaciones técnicas de la cámara multispectral Parrot Sequioa.

Parámetros	Valor
Tiempo de disparo	0.1 s
Resolución espectral	1 nm
Campo de visión	25°
Tamaño del pixel	3.75 $\mu\text{m}$
Largo focal	3.98 mm
Tamaño de imagen	1280*960 pixels

Se utilizó el software Pix4D Capture (Pix4D Inc., San Francisco, CA, USA) para recolectar los datos de las imágenes utilizando la opción de cuadrícula simple y la vista nadir (90 ° vertical) [(Su & Chou, 2015). Las imágenes se adquirieron con una superposición del 25 % entre las mismas. En total se obtuvieron 662 imágenes para la superficie de estudio.

### Procesamiento de las imágenes

El conjunto de 662 imágenes crudas obtenidas de la cámara Parrot Sequioa fueron procesadas con el software Pix4D Mapper (Pix4D SA, Suiza) para generar los ortomosaicos en formato *geotiff* para cada uno de los índices de vegetación. Este programa utiliza el algoritmo de estructura a partir del movimiento (SfM de sus siglas en inglés *Structure for Motion*) el cual utiliza imágenes adquiridas en múltiples puntos de vista sin procesamiento previo para modelar la estructura tridimensional de un objeto o superficie (Fonstad et al., 2013; Westoby et al., 2012).

Para el procesamiento de las imágenes se siguieron los siguientes pasos:

(a) El procesamiento inicial implicó la identificación, extracción y correspondencia de los puntos clave; la calibración de los parámetros internos (distancia focal) y externos (orientación) de la cámara y la georreferenciación.

(b) Elaboración de nube de puntos y malla: este paso se basa en los puntos de enlace automáticos, lo que conlleva la densificación de puntos y la creación de una malla con textura 3D.

(c) Creación del Modelo Digital de Superficie (MDS) para determinar los ortomosaicos y los mapas de índices de vegetación. La creación de ortomosaicos se basó en la ortorectificación para eliminar las distorsiones de perspectiva de las imágenes y producir mapas de índices de vegetación con el valor de cada píxel con reflectancia real del área de interés.

La resolución espacial final de cada uno de los ortomosaicos obtenidos fue de 4 cm/píxel.

### Cálculo de los índices de vegetación a partir de las imágenes obtenidas

Los espectros medios reflectancia calculados para las cuatro bandas espectrales obtenidas se utilizaron para calcular los índices espectrales NDVI (*Normalized Difference Vegetation Index*), NWVI (*Normalized Water Vegetation Index*), OSAVI (*Optimized Soil-Adjusted Vegetation Index*) y LCI (*Leaf Chlorophyll Index*) con la implementación del software Pix4D Mapper como fue descrito en el punto anterior. Estos índices se obtienen a partir de las ecuaciones que se muestran en la Tabla 4.2.

**Tabla 4.2.** Ecuaciones para la obtención de los cuatro índices de vegetación utilizados.

Índices de vegetación	Ecuación
NDVI ( <i>Normalized Difference Vegetation Index</i> )	$\frac{(R800 - R670)}{(R800 + R670)}$
NWVI ( <i>Normalized Water Vegetation Index</i> )	$\frac{(R550 - R670)}{(R550 + R670)}$
OSAVI ( <i>Optimized Soil-Adjusted Vegetation Index</i> )	$\frac{(1 + 0.16) * (R800 - R670)}{(R800 + R670 + 0.16)}$
LCI ( <i>Leaf Chlorophyll Index</i> )	$\frac{(R800 - R735)}{(R800 + R670)}$

### Extracción de la información a nivel de parcela

Para obtener la información sobre cada parcela del ensayo, es necesario la extracción de los datos a nivel de parcela de los ortomosaicos. La extracción de la información fue realizada utilizando una cuadrícula superpuesta sobre la imagen del ortomosaico, con la función nueva capa de archivo *shape* del software QGIS versión 3.16 (Team QGIS, 2021). Para generar la cuadrícula se debe definir el número total de filas y columnas que debe coincidir con el número de parcelas del ensayo, y el largo y ancho de cada celda de la cuadrícula. A partir de la capa vectorial poligonal de los límites de parcela generada y utilizando la función de estadísticas

zonales del software QGIS versión 3.16 se obtuvieron valores medios por parcela de cada índice de vegetación considerando un buffer negativo (Li et al., 2021) de 10 cm. La aplicación de este buffer funciona como una corrección de la superficie de la parcela, para concentrar la extracción de los datos en el centro de cada parcela y eliminar ruidos aportados por el entresurco. Además, fueron obtenidos los coeficientes de variación para cada índice de vegetación.

### **Análisis de los datos obtenidos usando modelos predictivos de ML**

Fueron implementados modelos predictivos basados en técnicas de ML que utilizan la información proporcionada por los índices de vegetación para poder predecir las clases fenotípicas frente a MV a nivel de parcela. Para comenzar el análisis, se dividió la matriz de datos total en una denominada matriz de entrenamiento que incluyó el 80 % de la totalidad de los datos y en una matriz de prueba que incluyó al 20 % de la matriz original. La primera matriz fue empleada para el entrenamiento de los modelos y la segunda para evaluar la precisión en la predicción de las clases fenotípicas mediante la validación cruzada de las clases fenotípicas predichas y las definidas por la evaluación a campo según se describió en el capítulo 3 de esta Tesis.

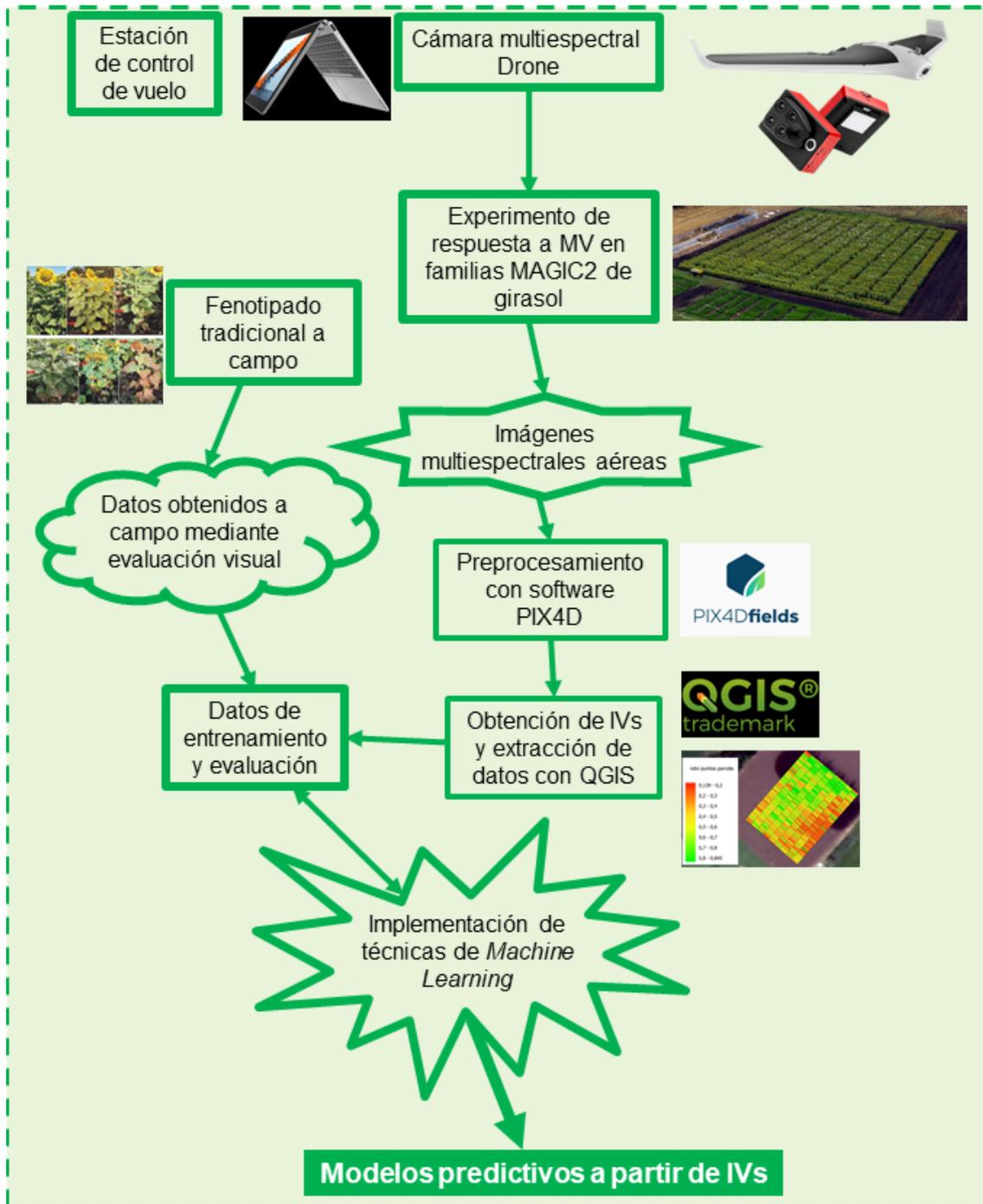
Los modelos evaluados fueron los árboles de clasificación (CART) implementados con el paquete de R Rpart (Therneau & Atkinson, 2019), Máquinas de Vector Soporte (SVM de sus siglas en inglés *Support Vector Machine*) con el paquete e1071 (Meyer et al., 2021), Análisis Discriminante Lineal (LDA de sus siglas en inglés *Linear Discriminant Analysis*) (Venables & Ripley, 2002), *Random Forest* con el paquete randomForest (Liaw & Wiener, 2002) y XGBoost (de sus siglas en inglés *eXtreme Gradient Boosting*) con el paquete CARET (Kuhn, 2020). Para la elaboración de los modelos predictivos se implementaron los cuatro índices de vegetación mencionados anteriormente.

Para entrenar y comparar los 5 modelos utilizados se empleó la técnica *K-Fold Cross Validation*. Ésta consiste en dividir al conjunto de entrenamiento en k subconjuntos y, al momento de realizar el entrenamiento, se toma cada k subconjunto como conjunto de prueba del modelo, mientras que el resto de los datos se utiliza como conjunto de entrenamiento. Este proceso se repite k veces, y en cada iteración se selecciona un conjunto de prueba diferente, mientras los datos restantes se emplean como conjunto de entrenamiento (Tong & Nikoloski, 2021). El número de subconjuntos utilizados fue de 10. Una vez finalizadas las iteraciones, se calculó la precisión media y su varianza para cada uno de los modelos entrenados. (Koh et al., 2021). La precisión fue medida a partir de los datos obtenidos en la evaluación manual y se obtuvo para cada modelo a partir de la siguiente fórmula:

$$\textit{Precisión} = \frac{\textit{Clases predichas correctas}}{\textit{Clases predichas correctas} + \textit{Clases predichas incorrectas}} \times 100$$

Para evaluar el *overfitting* o sobreajuste de los modelos, se evaluó la precisión obtenida en el conjunto de datos de entrenamiento y de prueba, en el caso que se presenten diferencias considerables en la precisión de ambas matrices, indicará la presencia de sobreajuste en los modelos (Bilbao & Bilbao, 2017; Cawley & Talbot, 2010).

La Figura 4.5 muestra a modo de resumen la rutina de trabajo empleada para obtener los modelos predictivos.



**Figura 4.5.** Esquema de trabajo para la obtención de modelos predictivos de la respuesta a la Marchitez por *V. dahliae* en girasol a partir de índices de vegetación.

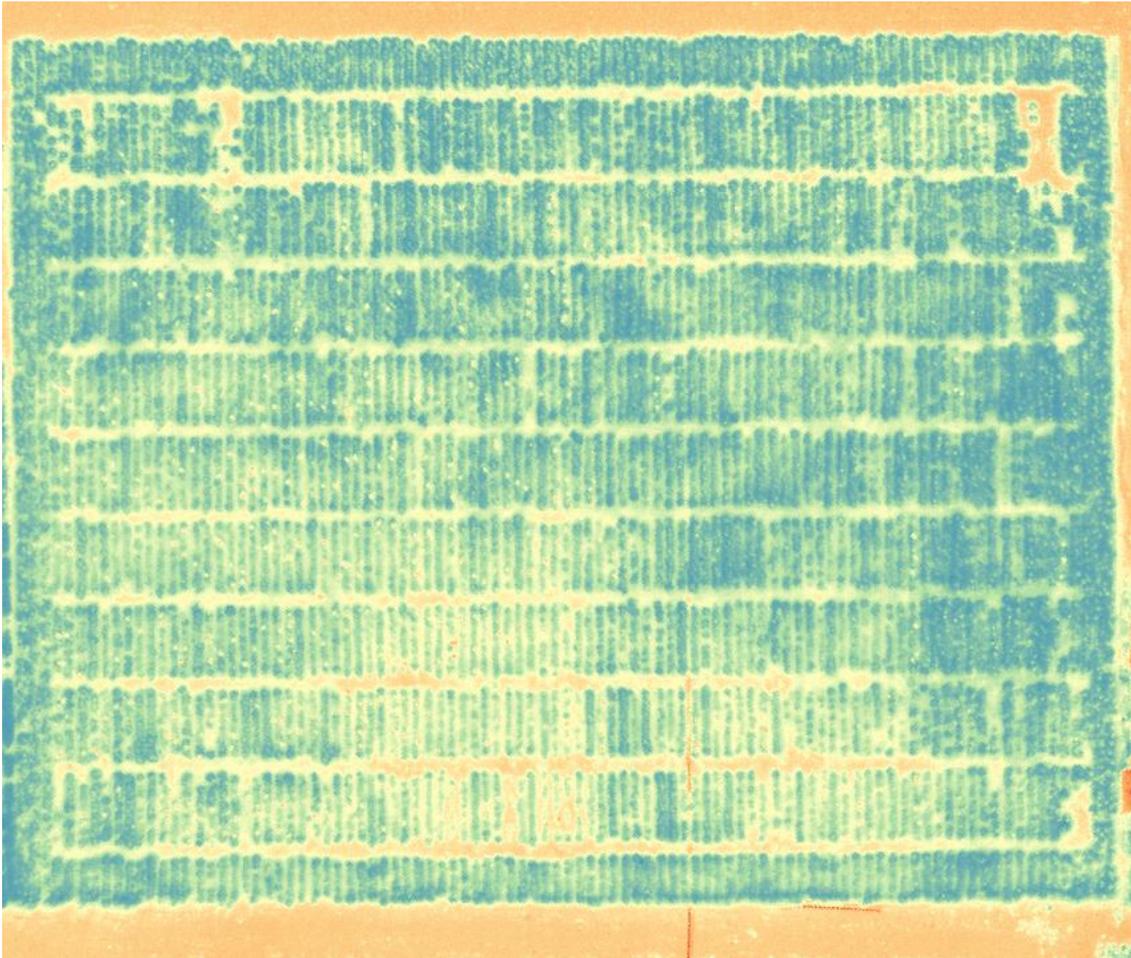
## **RESULTADOS**

### **Evaluaciones fenotípicas manuales**

En este capítulo para cuantificar la precisión de los modelos predictivos fue utilizada la variable severidad media tomada en la segunda fecha de evaluación coincidente con la toma de imágenes. El fenotipado manual realizado sobre el experimento permitió clasificar a las parcelas en dos grandes grupos de acuerdo a su respuesta frente a MV. Fueron definidas como tolerantes, aquellas parcelas que tuvieran un valor de severidad media menor a 3, y como susceptibles las que tuvieran un valor de severidad media mayor a 3. El 45,18 % de las parcelas fueron clasificadas como susceptibles y el 54,82 % fueron clasificadas como resistentes.

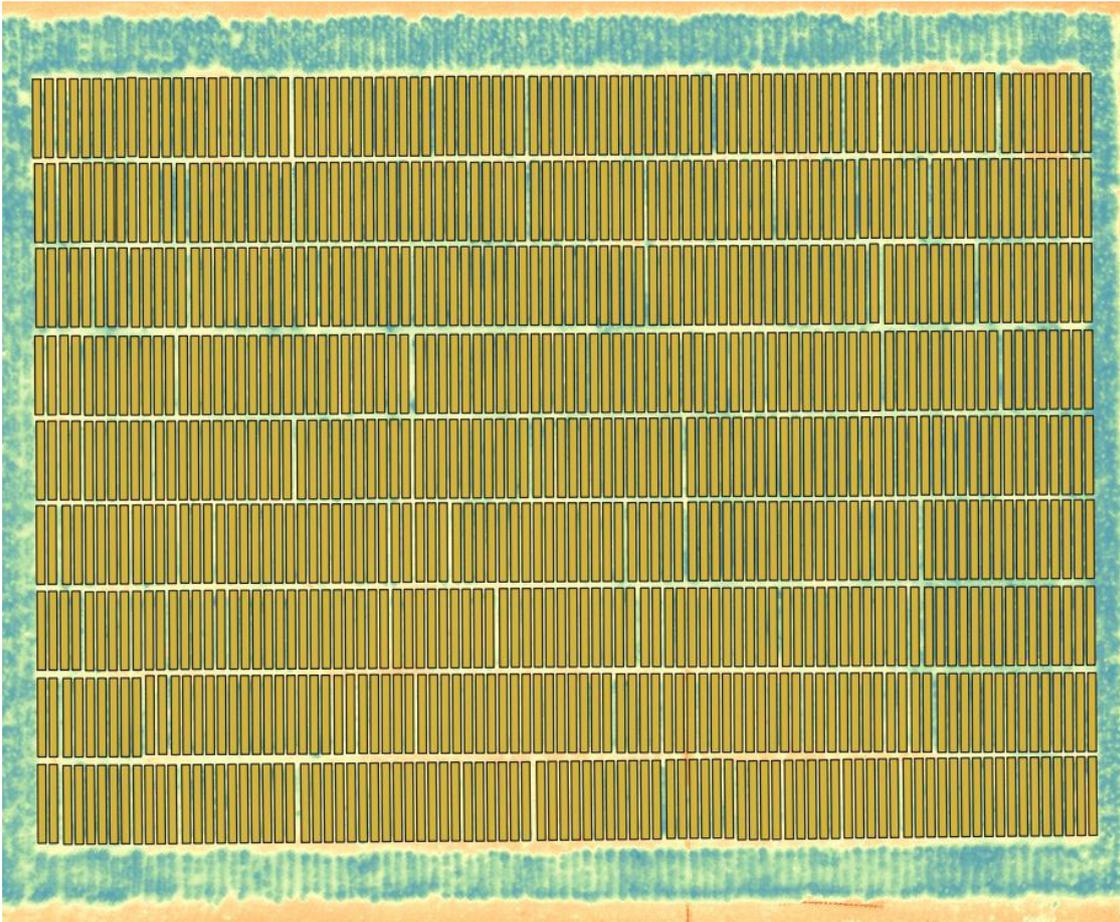
### **Adquisición y procesamiento de imágenes**

El vuelo realizado el día 28 de enero de 2021 coincidente con la segunda fecha de evaluación fenotípica manual, permitió la obtención de imágenes para la construcción de un ortomosaico a partir de la reflectancia de las cuatro bandas descritas en la sección de Materiales y Métodos del presente capítulo (Figura 4.6). A partir de la combinación de la reflectancia de las cuatro bandas espectrales mencionadas se obtuvieron los índices de vegetación NDVI, NWVI, OSAVI y LCI.



**Figura 4.6.** Imagen del sitio de evaluación luego del procesamiento con los softwares PIX4d y QGIS para la obtención del índice NDVI.

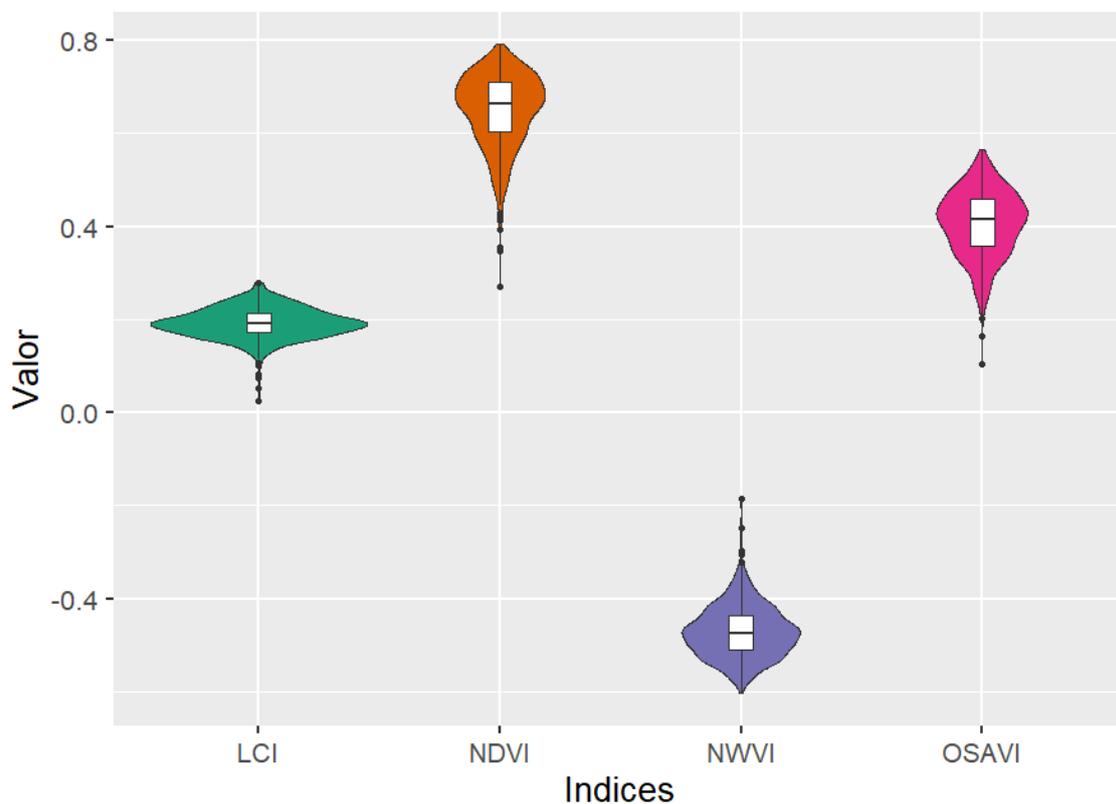
En la Figura 4.7 se muestra la grilla de polígonos construida con la función nueva capa de archivo *shape* del software QGIS y utilizada para definir cada una de las parcelas del ensayo.



**Figura 4.7.** Grilla de polígonos realizada para la definición de las parcelas del experimento.

A partir de la grilla de polígonos elaborada con QGIS para delimitar el tamaño de cada una de las parcelas y utilizando la función de estadísticas zonales del mismo software, se extrajeron los valores medios para cada uno de los IV empleados. Estos resultados fueron exportados en una planilla de formato *xls* y luego incorporados a la matriz de datos del ensayo junto con las variables fitopatométricas obtenidas del fenotipado manual.

Los resultados obtenidos indican que para NDVI la media general fue de 0.65 y el coeficiente de variación (CV) del 12.31 %, para NWVI la media fue de - 0.47 y el CV de 10.64 %, para OSAVI la media fue de 0.41 y el CV de 17.07 % y para LCI la media fue de 0.19 y el CV de 15.79 %. En la Figura 4.8 se muestra la distribución a través de las parcelas de los 4 IV obtenidos.



**Figura 4.8.** Gráfico de violín de los 4 índices de vegetación obtenidos del análisis de las imágenes multispectrales.

A partir de los IV obtenidos se analizaron las correlaciones lineales entre estos y los valores de las variables fitopatométricas de INC2, MSEV2, INT2 y AUDPC\_INC obtenidos de la evaluación fenotípica manual. Las correlaciones fueron negativas, significativas y menores a 0,45, excepto para NWVI las cuales fueron positivas debido a la propia naturaleza de obtención del índice, que hace uso de la reflectancia obtenida en la región del verde, lo cual difiere de los demás IV empleados en este trabajo. En la Tabla 4.3 se muestran los valores obtenidos para cada caso.

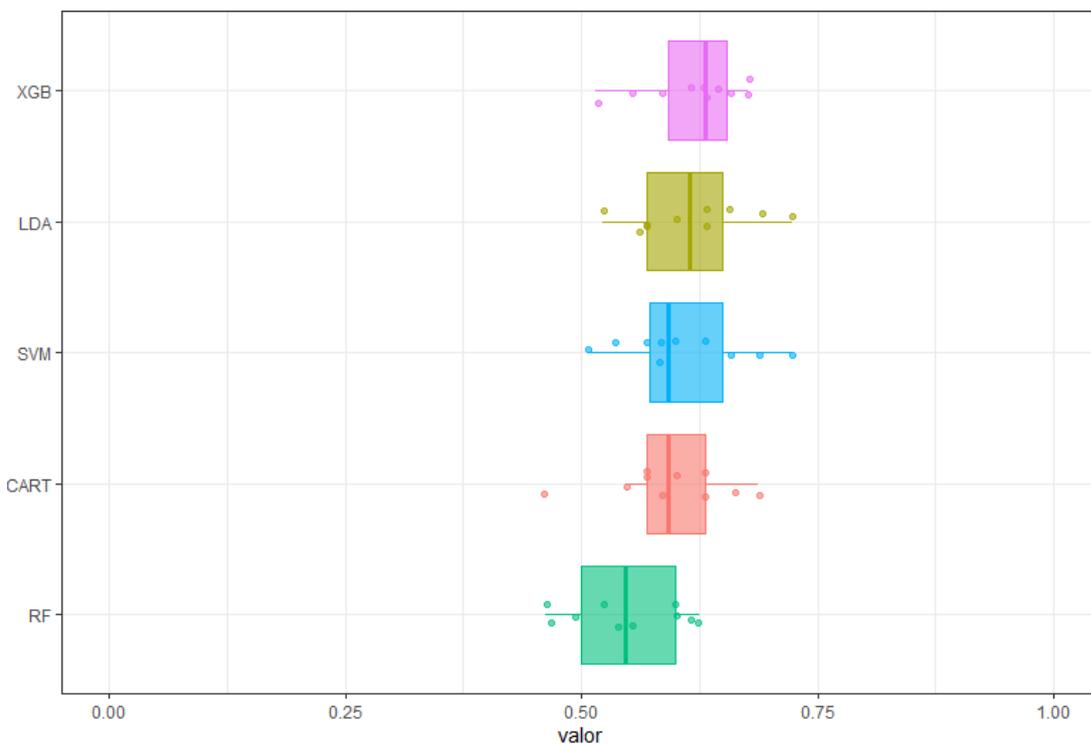
**Tabla 4.3.** Correlaciones entre los índices de vegetación y las variables INC2, MSEV2, INT2 y AUDPC\_INC.

<b>Variabes</b>	<b>NDVI</b>	<b>NWVI</b>	<b>OSAVI</b>	<b>LCI</b>
<b>INC2</b>	-0,39	0,38	-0,41	-0,26
<b>MSEV2</b>	-0,30	0,30	-0,31	-0,20
<b>INT2</b>	-0,42	0,41	-0,44	-0,27
<b>AUDPC_INC</b>	-0,37	0,38	-0,39	-0,15

### Análisis de los resultados obtenidos con los diferentes modelos predictivos de ML implementados

A partir de los datos generados en el punto anterior fueron desarrollados cinco modelos predictivos de ML (XGB, LDA, SVM, CART y RF) con el propósito de poder clasificar parcelas de girasol como tolerantes o susceptibles en función de la información espectral.

La comparación del desempeño alcanzado por los 5 modelos evaluados fue realizada implementando la metodología de validación cruzada descrita en Materiales y Métodos, y utilizando los datos obtenidos de la evaluación fenotípica manual se estimó la precisión en la predicción de la clase fenotípica obtenida de cada uno de los modelos propuestos. En la Figura 4.9 se muestra la precisión media y la dispersión obtenida para cada uno de los modelos entrenados.



**Figura 4.9.** Precisión obtenida en el entrenamiento de los distintos modelos por el método de validación cruzada K-Fold.

El modelo XGBoost fue el que obtuvo una precisión media más alta y el que menor varianza presentó durante el proceso de validación. En orden de precisión, en segundo lugar se ubicó el modelo lineal discriminante y el de menor valor fue el modelo fue el *Random Forest*.

Asimismo, la precisión fue estimada en los conjuntos de entrenamiento y de prueba para tener una medida del *overfitting* de los modelos. Los resultados se muestran en la Tabla 4.4.

**Tabla 4.4.** Precisión alcanzada por los 4 modelos predictivos evaluados.

<b>Modelo</b>	<b>Precisión en el conjunto de entrenamiento (%)</b>	<b>Precisión en el conjunto de prueba (%)</b>
Árboles de decisión	63,90	62,30
<i>Random Forest</i>	100	59,26
SVM	65,18	53,70
LDA	62,54	62,02
XGBoost	63,27	63,12

Los modelos XGBoost y arboles de decisión alcanzaron los valores más altos en la matriz de prueba, y no mostraron grandes diferencias con respecto a la precisión obtenida en la matriz de entrenamiento. En tanto que el modelo *Random Forest* mostró una amplia diferencia en la precisión obtenida para los dos conjuntos de datos, evidenciando un claro sobreajuste de este modelo.

## DISCUSIÓN

El continuo incremento de la ganancia genética de los cultivos demuestra por un lado el exitoso proceso de mejoramiento genético que se viene desarrollando tanto en los programas de mejoramiento públicos como privados y por el otro el potencial de satisfacer las demandas actuales y futuras de sustentabilidad ambiental de los sistemas productivos y seguridad alimentaria. En este sentido, es fundamental complementar la innovación impulsada fundamentalmente por la genómica y sus aplicaciones en mejoramiento como lo es la selección genómica, con nuevas tecnologías incluida la fenotipificación digital, nuevas prácticas agronómicas, nuevas tecnologías de mejoramiento, análisis de microbioma, entre otras para alcanzar las metas propuestas para la producción de alimentos y la disminución de la huella ambiental para el año 2050 (Crossa et al., 2021). Dado el contexto de cambio climático actual, la incorporación de las tecnologías de fenotipificación digital o fenómica en el área de mejoramiento genético facilita la recopilación de datos de fenotipo de alto rendimiento reduciendo los costos y el tiempo que demanda medir los distintos caracteres de interés agronómico en el campo.

La fenómica de campo tiene como objetivo estudiar el fenotipo de las plantas en una serie de condiciones ambientales. Los métodos modernos de fenómica son capaces de utilizar cámaras hiperespectrales, multiespectrales y RGB para proporcionar cientos de datos de reflectancia en bandas estrechas en numerosos ambientes y etapas del desarrollo del cultivo. Estos métodos pueden utilizarse para obtener rápidamente y con precisión datos sobre caracteres agronómicos basados en la información proporcionada por las imágenes obtenidas gracias a las nuevas tecnologías que redujeron el costo de su implementación en el área de agricultura de precisión y mejoramiento vegetal (Atkinson et al., 2018; Reynolds et al., 2019). La implementación de plataformas de fenotipado de alto caudal, como las conformadas a partir vehículos aéreos no tripulados, resultan muy atractivas para su implementación en estudios de fenotipado en diversas áreas y en particular en el mejoramiento genético vegetal (Chawade et al., 2019).

La detección de enfermedades en los cultivos se ha llevado a cabo exitosamente de forma visual por fitopatólogos experimentados en diferentes enfermedades de las plantas. Esta capacidad está limitada en los casos que requieren el fenotipado de poblaciones de mejoramiento de gran tamaño, siendo la automatización una alternativa promisoría. La automatización es posible aprovechando los cambios típicos en el comportamiento de reflectancia de las plantas que provocan las enfermedades. Los patógenos causan una reducción del contenido de clorofila de la planta impactando en la reflectancia de la región del

espectro visible y del rojo lejano (550 nm; 650-720 nm) debido a las lesiones necróticas o cloróticas. Los efectos de amarronamiento por senescencia de los tejidos vegetales infectados afectan a las regiones del espectro visible y el infrarrojo cercano (680-800 nm), así como al infrarrojo de onda corta (1400-1600 nm y 1900-2100 nm) debido a la sequedad. Las diferencias de reflectancia entre los tejidos sanos y enfermos de las plantas pueden captarse mediante imágenes espectrales y utilizarse mediante algoritmos avanzados de aprendizaje (Su et al., 2018). Por lo tanto, las plataformas de fenotipado basadas en imágenes pueden desempeñar un papel importante en la detección y el reconocimiento de las enfermedades de las plantas cuando la evaluación manual no es compatible con los tiempos óptimos para la toma de datos (Barbedo, 2019).

Hasta el momento no hay publicados estudios de fenotipado de MV en girasol implementando tecnologías de fenotipado automático. Si se han reportado estudios de fenotipado de respuesta a MV en frutilla, donde el uso de drones e imágenes permitió el desarrollo de un sistema eficaz y eficiente de fenotipado de alto rendimiento (Cockerton et al., 2019). La aplicación de esta plataforma de fenotipado redujo la cantidad de trabajo manual para la evaluación de 2500 plantas en cinco ciclos a lo largo de una temporada de evaluación. Según los autores redujeron las horas de trabajo de 37 h 30' necesarias para la toma manual a 2 h 30' usando fenotipado de alto caudal. En el caso de girasol presentado en el capítulo 3, el fenotipado manual de las 810 parcelas de girasol para MV insumió un total de 56 horas netas de trabajo, mientras que el fenotipado de alto caudal llevó un total de 8 h 30'. Debe considerarse que el tiempo insumido por el fenotipado de alto caudal en este trabajo, corresponde a la primera experiencia del grupo y a un único muestreo, por lo que el tiempo puede reducirse ajustando un protocolo para la evaluación de la MV en girasol, definiendo las condiciones óptimas para la obtención y procesamiento de las imágenes, la elección de IV apropiados y la implementación de modelos precisos para la clasificación de los genotipos.

Numerosos índices de vegetación pueden obtenerse a partir de la combinación de distintas longitudes de onda y de diferentes ecuaciones (Candiago et al., 2015). Estos pueden utilizarse para estimar y controlar parámetros de crecimiento de las plantas como el índice de área foliar, la cobertura del suelo, el estado hídrico de las hojas, las concentraciones de clorofila o de nitrógeno, la cuantificación y detección de enfermedades, entre otras variables (Cammarano et al., 2014; Haboudane et al., 2008; Pabuayon et al., 2019). En estudios fenotípicos realizados sobre MV en frutilla, el índice NDVI tuvo una fuerte asociación con la variable AUDPC utilizada para cuantificar la enfermedad en el trabajo publicado por Cockerton et al (2019). En estudios realizados en MV en olivo, se evaluaron 22 índices de vegetación diferentes, destacándose como los más informativos a NDVI, OSAVI e indicadores de la

temperatura del canopeo (Calderón et al., 2013; 2015). En girasol en cambio, las correlaciones de los índices de vegetación NDVI, OSAVI, NWWI y LCI evaluados en este capítulo con las principales variables fitopatométricas consideradas para MV fueron menores a 0,5. Sin embargo, los IV obtenidos fueron implementados para la generación de modelos de ML que lograron discriminar a nivel de parcela, genotipos de girasol tolerantes y susceptibles frente a la marchitez prematura por *V. dahliae* con una precisión por encima del 60 %, no habiéndose reportado hasta el momento estudios similares para esta enfermedad en el cultivo.

Estudios de fenotipado con drones sobre *V. dahliae* en plantaciones de olivo fueron reportados por Calderón et al (2015). Estos autores implementaron las técnicas LDA y SVM obteniendo valores de precisión entre el 60 % y 80 %, reportándose una mayor precisión para LDA en individuos con bajos niveles de severidad, y en cambio la técnica SVM tuvo una mayor precisión en individuos con mayores niveles de severidad de la enfermedad. En estudios realizados en arroz, Wang et al (2021) caracterizaron el comportamiento al vuelco de diferentes variedades y su relación con el rendimiento de grano a través del uso de imágenes hiperespectrales. Implementaron 5 modelos de aprendizaje entre ellos dos variantes de SVM, *Random Forest*, Naive Bayes y XGBoost, y también obtuvieron mayor precisión con XGBoost. Por su parte, Sandino et al (2018) emplearon drones para la obtención de imágenes multiespectrales en producciones forestales e implementaron el modelo XGBoost logrando valores de precisión mayores al 90 % en la detección de árboles afectados por enfermedades fúngicas. Asimismo, Sugiura et al (2018) reportaron valores de precisión mayores al 80 % en la clasificación de virosis en papa y Castelao Tetila et al (2017) valores cercanos al 90 % en fenotipado de enfermedades foliares de soja. Los resultados obtenidos en este capítulo presentaron niveles de precisión menores que los trabajos citados anteriormente. Sin embargo, consideramos que estos resultados son muy promisorios teniendo en cuenta que se trata de una exploración basada en un único muestreo y cuatro IV.

A partir de estos resultados, y en base a la experiencia acumulada en otros cultivos, se espera incorporar mejoras en futuros ensayos que impactarán positivamente sobre los niveles de precisión. Entre las mejoras pueden mencionarse, el uso de referencias espectrales en el lote para realizar una calibración radiométrica. Pourazar et al (2019) reportaron la importancia de esta corrección para la implementación exitosa de imágenes obtenidas por drones en proyectos de investigación de enfermedades en plantas. Por su parte, Berra et al (2017) recomiendan emplear referencias de color neutro en el terreno, por ejemplo, tarjetas grises o con porcentajes de reflectancia conocidos, para evaluar el balance de blancos, el color y los efectos de distorsión espacial. Con la implementación de estas referencias, se trata de verificar que el balance de blancos sea estable, que no se hayan producido cambios en las

condiciones de luz durante el vuelo o que éstos fueron limitados. Otro punto importante es el número y la ubicación de los puntos de control del terreno, estos se utilizan en el procesamiento de las imágenes para georreferenciar y poder ubicarlas con precisión en un sistema de coordenadas, además pueden emplearse para una definición precisa y eficiente de las parcelas. La delimitación de las parcelas suele ser una tarea laboriosa y demandante de tiempo, la cual debe ser reducida para lograr incrementar la eficiencia y precisión de los protocolos de trabajo, para esto resultará fundamental la inclusión de puntos marcados con GPS durante la siembra de los experimentos a relevar, de esta manera se lograrán parcelas precisamente equidistantes y además se dispondrá una grilla de puntos que facilitará la extracción de la información de los ortomosaicos que se generen.

La implementación de distintas metodologías de *Machine Learning* demostraron un comportamiento similar con una leve superioridad de la técnica XGBoost sobre las demás. Por su parte, la metodología de *Random Forest* mostró un claro sobreajuste del modelo entrenado, siendo la peor alternativa para los datos analizados. En general, *Random Forest* funciona mejor en la fijación del conjunto de datos de entrenamiento que en los de prueba, en comparación con XGBoost, y este último suele tener un mejor comportamiento frente al *overfitting*. Similares resultados fueron reportados por Herrero-Huerta et al (2020) en un estudio de predicción de rendimiento en soja utilizando las metodologías de RF y XGBoost. La amplia diversidad de algoritmos disponibles para el análisis y uso de la información generada por las plataformas de fenotipado de alto caudal disponibles actualmente, plantea la necesidad de realizar un estudio previo sobre cuáles de las metodologías es conveniente aplicar en cada caso de estudio para obtener modelos predictivos confiables y de alta precisión que permitan hacer más eficiente el fenotipado.

En este capítulo se exploraron cuatro índices de vegetación estructurales y diferentes metodologías de análisis. Los resultados obtenidos demuestran la viabilidad de esta metodología para detectar infecciones por *V. dahliae* y las ventajas que ofrece.

En función de esta primera experiencia, en futuros estudios se proyecta:

- Incorporar el uso de referencias espectrales en el terreno para realizar una calibración radiométrica
- Ajustar la altura óptima de vuelo y la superposición de las imágenes para lograr una resolución espacial que contribuya a incrementar la precisión de los modelos manteniendo una relación de compromiso con el tiempo necesario para el procesamiento de las imágenes.

- Realizar un mayor número de vuelos durante el ciclo del cultivo que permita un seguimiento más preciso del progreso de la enfermedad
- Implementar durante la siembra el marcado de las coordenadas de cada una de las parcelas de los ensayos con GPS para hacer más eficiente y preciso el procesamiento de las imágenes y la extracción de la información, ya que la realización de cuadrículas de manera manual como fue realizado en este capítulo resulta laborioso y demandante de tiempo.
- Incorporar a los modelos indicadores relacionados con la temperatura del cultivo, otros índices de vegetación no evaluados en este caso y que ya fueron reportados en trabajos de fenotipado de enfermedades en distintos cultivos (Albetis et al., 2017; De Castro et al., 2015; Pourazar et al., 2019), como el GNDVI, NDRE, DVI, RVI, las bandas de reflectancia del rojo, rojo lejano e infrarrojo cercano entre otros.
- Explorar otras metodologías para el análisis de este tipo de datos ampliamente reportadas en la bibliografía (Nabwire et al., 2021; Sandhu et al., 2021; Vayssade et al., 2021), como son las técnicas de *Deep Learning*.

Por otra parte, para que esta tecnología sea utilizada y adoptada por los programas de mejoramiento debe analizarse con detenimiento el costo de la misma y hasta el momento hay pocos reportes que aborden este punto. Holman (2020) en su trabajo de tesis doctoral analizó los costos que conllevan el fenotipado manual y el automático implementando una cámara multiespectral en un dron. El ensayo consistió en 10.000 parcelas de trigo que fueron evaluadas durante 5 años para caracteres de altura de planta, altura del canopeo y temperatura del canopeo sólo en el caso del fenotipado automático. Los costos obtenidos por el autor fueron de 10,31 £ por parcela para el fenotipado manual y de 6,20 £ por parcela para el fenotipado automático. Estos resultados son muy alentadores para la adopción de la tecnología por los programas de mejoramiento, ya que lograron reducir aproximadamente un 50 % los costos de fenotipado. Además, este autor analizó el tiempo demandando por ambos procesos, siendo más de 3 veces menor el tiempo insumido por el fenotipado automático. Futuros estudios deberán llevarse adelante para comparar los costos de ambas metodologías para el fenotipado de MV en girasol y también de otros caracteres para poder determinar la viabilidad de la implementación de este tipo de tecnología para su aplicación en los programas de mejoramiento del cultivo.

Los resultados obtenidos en este trabajo mostraron el potencial que tiene la implementación de índices de vegetación para el fenotipado automático de la Marchitez prematura por *V. dahliae* en girasol y la utilización de modelos que permiten en base a la información aportada por los índices espectrales, poder clasificar genotipos de girasol como tolerantes o resistentes

a la enfermedad. La implementación de una plataforma de fenotipado de alto caudal por medio de drones para la caracterización de poblaciones de gran tamaño con menor demanda de recursos y tiempo resulta promisorio e incentivador en relación a los resultados alcanzados. La mayor ventaja que presenta esta tecnología, es la rápida adquisición de datos para un elevado número de parcelas. Un análisis profundo de los costos de esta metodología y su comparación con el fenotipado manual resulta fundamental para que sea posible su adopción por los programas de mejoramiento, en proyectos de investigación y también a nivel productivo para el diagnóstico de epifitias. Se continuará con las evaluaciones en los próximos años, para poder generar un mayor volumen de datos que permita evaluar otras técnicas y otras variables, con el objetivo de obtener mayores valores de precisión por los modelos generados y poder incorporar esta tecnología como una práctica rutinaria en el programa de mejoramiento de girasol de INTA.

---

---

## **CONSIDERACIONES FINALES Y PERSPECTIVAS**

---

---

Este trabajo de Tesis ha contribuido al desarrollo de nuevos recursos genéticos para el mejoramiento del cultivo de girasol. Tomando como base la línea de trabajo en poblaciones de mapeo no estructuradas iniciada en el año 2008 en INTA que dio lugar al primer estudio de mapeo por asociación para caracteres complejos en el cultivo de girasol (Fusari et al., 2012) y estudios posteriores enfocados a la respuesta ante la PHC por *S. sclerotiorum* y la MV por *V. dahliae* (Filippi et al., 2015; Montecchia et al., 2021) se seleccionaron líneas mejoradas que reúnen QTLs mayores para estas características, sumadas a otras características de calidad y resistencia a estreses abióticos para generar poblaciones multiparentales. El atractivo de estas poblaciones radica en la posibilidad de combinar alelos favorables para un espectro amplio de características en un mismo fondo genético lo cual favorece el estudio, la validación de QTLs y el desarrollo de marcadores moleculares asociados dada la alta precisión del mapeo genético que se logra con el número de recombinaciones. Los desafíos que presenta el desarrollo de este tipo de recursos son el tiempo y el esfuerzo requeridos para obtener líneas homogéneas sin perder representatividad alélica. En este trabajo se siguió un diseño de cruzamientos básico para adecuar los recursos y los tiempos requeridos para completar el trabajo de Tesis. Las conclusiones principales de este trabajo son:

- El análisis multivariado de los datos fenotípicos relevados en los estudios para resistencia a enfermedades conducidos sobre la PMA de girasol, asociado al análisis de diversidad y distancia genética entre las mismas líneas del programa de mejoramiento de INTA, permitieron la selección de líneas genéticamente diversas para usar como parentales para el desarrollo de las poblaciones MAGIC presentadas en esta Tesis.
- El diseño aplicado para la construcción de ambas MAGICs fue el de obtención de dialelos incompletos lo cual permitió obtener los híbridos de ocho vías en el transcurso de los primeros 24 meses de desarrollo del plan de Tesis, evaluar en F<sub>2</sub> las características fenotípicas seleccionadas y alcanzar el estado de F<sub>3</sub> en la campaña 2020/21. Se planea llegar a la homocigosis práctica de las líneas recombinantes derivadas de estas poblaciones en la campaña 2023/24.
- La caracterización genotípica permitió obtener el perfil genético de las líneas homocigotas parentales seleccionadas para la formación de las poblaciones MAGIC y de los híbridos de dos, cuatro y ocho vías, evaluar la diversidad genética y chequear la calidad de los cruzamientos realizados en las distintas etapas de formación de las poblaciones multiparentales.

- La exploración y el entrenamiento en el uso de herramientas como el análisis de parentesco y la reconstrucción de bloques haplotípicos generó información útil para confirmar la calidad de los cruzamientos y proporcionó un avance en el conocimiento de las estrategias para el estudio de QTLs en este tipo de poblaciones.
- La caracterización de la respuesta de la población MAGIC2 frente a la Marchitez prematura por *Verticillium dahliae* confirmó la variabilidad fenotípica de la población y permitió identificar grupos de familias F<sub>2</sub> con respuesta contrastante a la enfermedad.
- La caracterización preliminar realizada en ambas poblaciones MAGIC de la posición del capítulo y de la inclinación de la lámina confirmó la variabilidad fenotípica de estas poblaciones siendo de gran importancia para los estudios futuros que se realicen con el objetivo de identificar las variantes genéticas que influyen sobre estos caracteres.
- Los resultados obtenidos en esta Tesis revelaron el potencial que tiene la implementación de índices de vegetación para el fenotipado de la Marchitez prematura por *Verticillium dahliae* en girasol y la utilización de modelos de aprendizaje que permiten en base a la información aportada por los índices espectrales, poder clasificar genotipos de girasol como tolerantes o resistentes a la enfermedad.
- La construcción y caracterización de las poblaciones MAGIC presentadas en esta Tesis, a partir de líneas provenientes del programa de mejoramiento de INTA y de líneas públicas, constituyen recursos genéticos únicos hasta el momento para el mejoramiento del cultivo a nivel nacional e internacional. Pudiendo ser utilizadas para estudios de diversos caracteres complejos de interés agronómico y debido a la procedencia de las líneas parentales será más factible su incorporación directa en los programas de mejoramiento.

Las poblaciones MAGIC iniciadas en el marco de esta Tesis seguirán su desarrollo dentro del programa de mejoramiento genético de girasol de INTA para alcanzar la homocigosis práctica de las progenies derivadas. Se pretende profundizar los conocimientos del grupo en las diferentes metodologías de análisis para el estudio de QTLs en estas poblaciones, obtener el perfil genómico completo de las líneas parentales y avanzar en el genotipado de las líneas recombinantes una vez alcanzada la homocigosis. Por otra parte, estas poblaciones generadas se convertirán en un recurso fundamental para estudios de validación de regiones genómicas asociadas a caracteres de interés agronómico en estudios previos del grupo. Además, de esta Tesis se desprende una nueva línea de trabajo no abordada hasta el momento por el grupo y de la que no se tienen reportes en girasol, como lo es la implementación de índices de vegetación obtenidos a partir de imágenes multiespectrales

para el fenotipado masivo. La implementación de modelos predictivos y de clasificación a partir de índices espectrales será ajustada en estudios próximos con el objetivo de validar los modelos obtenidos y realizar futuras evaluaciones en otros caracteres.

---

---

## **BIBLIOGRAFÍA**

---

---

- Aballay, M. M., Aguirre, N. C., Filippi, C. V., Valentini, G. H., & Sánchez, G. (2021). Fine-tuning the performance of ddRAD-seq in the peach genome. *Scientific Reports* 2021 11:1, 11(1), 1–13. <https://doi.org/10.1038/s41598-021-85815-0>
- Abdel-Haleem, H., Ji, P., Boerma, H. R., & Li, Z. (2013). An R package for SNP marker-based parent-offspring tests. *Plant Methods*, 9(1), 44–44. <https://doi.org/10.1186/1746-4811-9-44>
- Abdi, H., & Valentin, D. (2007). Multiple factor analysis. In N. Salkind (Ed.), *Encyclopedia of Measurement and Statistics* (pp. 657–663). Thousand Oaks (CA): Sage.
- Abecasis, G. R., Cherny, S. S., Cookson, W. O., & Cardon, L. R. (2002). Merlin — Rapid analysis of dense genetic maps using sparse gene flow trees. *Nature Genetics*, 30(1), 97–101. <https://doi.org/10.1038/ng786>
- Acquaah, G. (2012). Principles of Plant Genetics and Breeding: Second Edition. In *Principles of Plant Genetics and Breeding: Second Edition*. <https://doi.org/10.1002/9781118313718>
- Aguirre, N. C., Filippi, C. V., Zaina, G., Rivas, J. G., Acuña, C. V., Villalba, P. V., García, M. N., González, S., Rivarola, M., Martínez, M. C., Puebla, A. F., Morgante, M., Hopp, H. E., Paniego, N. B., & Poltri, S. N. M. (2019). Optimizing DDRADseq in non-model species: A Case Study in *Eucalyptus dunnii* Maiden. *Agronomy*, 9(9). <https://doi.org/10.3390/agronomy9090484>
- Aguirrezábal, L., Orioli, G., Hernández, L., Pereyra, V., & Miravé, J. (1996). *Girasol: aspectos fisiológicos que determinan el rendimiento*. Editorial Unidad Integrada Balcarce.
- Albetis, J., Duthoit, S., Guttler, F., Jacquin, A., Goulard, M., Poilvé, H., Féret, J. B., & Dedieu, G. (2017). Detection of Flavescence dorée grapevine disease using Unmanned Aerial Vehicle (UAV) multispectral imagery. *Remote Sensing*, 9(4), 1–20. <https://doi.org/10.3390/rs9040308>
- Alvarez, D., Mazzalay, A., Heinz, N., González, J., Dominguez, M., Filippi, C., Quiroz, F., Troglia, C., Montecchia, J., Fass, M., Gonzalez, S., Rivarola, M., Fernandez, P., Corro-molas, A., Heinz, R., Lia, V., & Paniego, N. (2021). Sunflower: Breeding and genetic resources at INTA, Argentina. *1st Plant Breeding Symposium*, 112–115.
- Amadeu, R. R., Munoz, P. R., Zheng, C., & Endelman, J. B. (2020). QTL Mapping in Outbred Tetraploid (and Diploid) Diallel Populations. In *bioRxiv* (p. 2020.12.18.423479). <https://doi.org/10.1101/2020.12.18.423479>
- Amorim, A., & Pereira, L. (2005). Pros and cons in the use of SNPs in forensic kinship

- investigation: A comparative analysis with STRs. *Forensic Science International*, 150(1), 17–21. <https://doi.org/10.1016/j.forsciint.2004.06.018>
- Anderson, E. C., & Garza, J. C. (2006). The power of single-nucleotide polymorphisms for large-scale parentage inference. *Genetics*, 172(4), 2567–2582. <https://doi.org/10.1534/genetics.105.048074>
- Anderson, S. L., Mahan, A. L., Murray, S. C., & Klein, P. E. (2018). Four Parent Maize (FPM) Population: Effects of Mating Designs on Linkage Disequilibrium and Mapping Quantitative Traits. *The Plant Genome*, 11(2), 170102. <https://doi.org/10.3835/plantgenome2017.11.0102>
- Andrade, F., Aguirrezábal, L. A., & Rizalli, R. H. (2002). Crecimiento y rendimiento comparado. In F. Andrade & V. Sadras (Eds.), *Bases para el manejo del maíz, el girasol y la soja* (pp. 57–96).
- Andrews, K. R., Good, J. M., Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. In *Nature Reviews Genetics* (Vol. 17, Issue 2, pp. 81–92). Nature Publishing Group. <https://doi.org/10.1038/nrg.2015.28>
- Araus, J.L., Kefauver, S. C., Zaman-Allah, M., Olsen, M. S., & Cairns, J. E. (2018). Translating high-throughput phenotyping into genetic gain. *Trends Plant Sci.*, 23(5), 451–466. <https://doi.org/10.1016/j.tplants.2018.02.001>
- Araus, José Luis, & Cairns, J. E. (2014). Field high-throughput phenotyping: The new crop breeding frontier. In *Trends in Plant Science* (Vol. 19, Issue 1, pp. 52–61). Elsevier. <https://doi.org/10.1016/j.tplants.2013.09.008>
- Arrones, A. (2019). *Desarrollo Y Fenotipado De Una Población Multiparental Magic De Berenjena*. Universitat Politècnica de València.
- Arrones, Andrea, Vilanova, S., Plazas, M., Mangino, G., Pascual, L., Díez, M. J., Prohens, J., & Gramazio, P. (2020). The Dawn of the Age of Multi-Parent MAGIC Populations in Plant Breeding: Novel Powerful Next-Generation Resources for Genetic Analysis and Selection of Recombinant Elite Material. *Biology*, 9(8), 229. <https://doi.org/10.3390/biology9080229>
- Ashley, M. V. (2010). Plant parentage, pollination, and dispersal: How DNA microsatellites have altered the landscape. *Critical Reviews in Plant Sciences*, 29(3), 148–161. <https://doi.org/10.1080/07352689.2010.481167>
- Astle, W., & Balding, D. J. (2009). Population Structure and Cryptic Relatedness in Genetic

- Association Studies. <https://doi.org/10.1214/09-STS307>, 24(4), 451–471.  
<https://doi.org/10.1214/09-STS307>
- Atkinson, J. A., Jackson, R. J., Bentley, A. R., Ober, E., & Wells, D. M. (2018). Field Phenotyping for the Future. *Annual Plant Reviews Online*, 1, 719–736.  
<https://doi.org/10.1002/9781119312994.apr0651>
- Aylor, D. L., Valdar, W., Foulds-Mathes, W., Buus, R. J., Verdugo, R. A., Baric, R. S., Ferris, M. T., Frelinger, J. A., Heise, M., Frieman, M. B., Gralinski, L. E., Bell, T. A., Didion, J. D., Hua, K., Nehrenberg, D. L., Powell, C. L., Steigerwalt, J., Xie, Y., Kelada, S. N. P., ... Churchill, G. A. (2011). Genetic analysis of complex traits in the emerging Collaborative Cross. *Genome Research*, 21(8), 1213–1222. <https://doi.org/10.1101/gr.111310.110>
- Baack, E. J., Sapir, Y., Chapman, M. A., Burke, J. M., & Rieseberg, L. H. (2008). Selection on domestication traits and quantitative trait loci in crop-wild sunflower hybrids. *Molecular Ecology*, 17(2), 666–677. <https://doi.org/10.1111/j.1365-294X.2007.03596.x>
- Baack, E. J., Whitney, K. D., & Rieseberg, L. H. (2005). Hybridization and genome size evolution: Timing and magnitude of nuclear DNA content increases in *Helianthus* homoploid hybrid species. *New Phytologist*, 167(2), 623–630. <https://doi.org/10.1111/j.1469-8137.2005.01433.x>
- Bachlava, E., Taylor, C. A., Tang, S., Bowers, J. E., Mandel, J. R., Burke, J. M., & Knapp, S. J. (2012). Snp discovery and development of a high-density genotyping array for sunflower. *PLoS ONE*, 7(1), 29814. <https://doi.org/10.1371/journal.pone.0029814>
- Badouin, H., Gouzy, J., Grassa, C. J., Murat, F., Staton, S. E., Cottret, L., Lelandais-Brière, C., Owens, G. L., Carrère, S., Mayjonade, B., Legrand, L., Gill, N., Kane, N. C., Bowers, J. E., Hubner, S., Bellec, A., Bérard, A., Bergès, H., Blanchet, N., ... Langlade, N. B. (2017). The sunflower genome provides insights into oil metabolism, flowering and Asterid evolution. *Nature*, 546(7656), 148–152. <https://doi.org/10.1038/nature22380>
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E. U., Cresko, W. A., & Johnson, E. A. (2008). Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, 3(10), 3376. <https://doi.org/10.1371/journal.pone.0003376>
- Bajgain, P., Rouse, M. N., Tsilo, T. J., Macharia, G. K., Bhavani, S., & Jin, Y. (2016). Nested Association Mapping of Stem Rust Resistance in Wheat Using Genotyping by Sequencing. *PLoS ONE*, 11(5), 155760. <https://doi.org/10.1371/journal.pone.0155760>

- Ballesteros, R., Ortega, J. F., Hernández, D., & Moreno, M. A. (2014). Applications of georeferenced high-resolution images obtained with unmanned aerial vehicles. Part I: Description of image acquisition and processing. *Precision Agriculture*, *15*(6), 579–592. <https://doi.org/10.1007/s11119-014-9355-8>
- Bandillo, N., Raghavan, C., Muyco, P. A., Anna, M., Sevilla, L., Lobina, I. T., Dilla-ermita, C. J., Tung, C., Mccouch, S., Thomson, M., Mauleon, R., Singh, R. K., Gregorio, G., Redoña, E., & Leung, H. (2013). *Multi-parent advanced generation inter-cross ( MAGIC ) populations in rice : progress and potential for genetics research and breeding*. 1–15.
- Barbedo, J. G. A. (2019). A Review on the Use of Unmanned Aerial Vehicles and Imaging Sensors for Monitoring and Assessing Plant Stresses. *Drones 2019, Vol. 3, Page 40*, *3*(2), 40. <https://doi.org/10.3390/DRONES3020040>
- Barreto, H. J., Edmeades, G., Chapman, S., & Crossa, J. (1997). The alpha lattice design in plant breeding and agronomy: Generation and analysis. In *Training Material CIMMYT*.
- Baruca Arbeiter, A., Jakše, J., & Bandelj, D. (2014). Paternity analysis of the olive variety “istrska Belica” and identification of pollen donors by microsatellite markers. *Scientific World Journal*, *2014*. <https://doi.org/10.1155/2014/208590>
- Bates, D., Mächler, M., Bolker, B. M., & Walker, S. C. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. <https://doi.org/10.18637/jss.v067.i01>
- Bazakos, C., Hanemian, M., Trontin, C., Jiménez-Gómez, J. M., & Loudet, O. (2017). New Strategies and Tools in Quantitative Genetics: How to Go from the Phenotype to the Genotype. *Annual Review of Plant Biology*, *68*(January), 435–455. <https://doi.org/10.1146/annurev-arplant-042916-040820>
- Bernardo, R. (2008). Molecular markers and selection for complex traits in plants: Learning from the last 20 years. In *Crop Science* (Vol. 48, Issue 5, pp. 1649–1664). John Wiley & Sons, Ltd. <https://doi.org/10.2135/cropsci2008.03.0131>
- Bernardo, R. (2016). Bandwagons I, too, have known. *Theoretical and Applied Genetics* *2016* *129*:12, *129*(12), 2323–2332. <https://doi.org/10.1007/S00122-016-2772-5>
- Bernardo, R. (2020). Reinventing quantitative genetics for plant breeding: something old, something new, something borrowed, something BLUE. *Heredity*, *125*, 375–385. <https://doi.org/10.1038/s41437-020-0312-1>
- Bernardos, J., & Farrel, M. (2012). Evaluación del daño por la paloma torcaza (Zenaida

- auriculata) en girasol y pérdida de cosecha en la provincia de La Pampa, campaña 2011–2012. In *Estación Experimental Anguil, INTA*.
- Berni, J. A. J., Zarco-Tejada, P. J., Sepulcre-Cantó, G., Fereres, E., & Villalobos, F. (2009). Mapping canopy conductance and CWSI in olive orchards using high resolution thermal remote sensing imagery. *Remote Sensing of Environment*, 113(11), 2380–2388. <https://doi.org/10.1016/j.rse.2009.06.018>
- Berra, E. F., Gaulton, R., & Barr, S. (2017). Commercial Off-the-Shelf Digital Cameras on Unmanned Aerial Vehicles for Multitemporal Monitoring of Vegetation Reflectance and NDVI. *IEEE Transactions on Geoscience and Remote Sensing*, 55(9), 4878–4886. <https://doi.org/10.1109/TGRS.2017.2655365>
- Berry, S. T. et al. (1996). Construction of a high density, composite RFLP linkage map for cultivated sunflower *Helianthus annuus*. *14th International Sunflower Conference*, 1150–1160.
- Berry, S. T. et al. (1997). Presentation of the Advanta sunflower RFLP linkage map for public research. *19th Sunflower Res. Workshop*, 113–118.
- Berry, S. T., Leon, A. J., Hanfrey, C. C., Challis, P., Burkholz, A., Barnes, S. R., Rufener, G. K., Lee, M., & Caligari, P. D. S. (1995). Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. *Theoretical and Applied Genetics*, 91(2), 195–199. <https://doi.org/10.1007/BF00220877>
- Bertero de Romano, A., & Vazquez, A. (1982). A new race of *Verticillium dahliae*. *10th International Sunflower Conference*, 177–178.
- Bertero de Romano, A., & Vázquez, A. (2003). Origin of the argentine sunflower varieties. *Helia*, 26(38), 127–136. <https://doi.org/10.2298/hel0338127d>
- Bhat, J. A., Yu, D., Bohra, A., Ganie, S. A., & Varshney, R. K. (2021). Features and applications of haplotypes in crop breeding. *Communications Biology*, 4(1), 1–12. <https://doi.org/10.1038/s42003-021-02782-y>
- Bilbao, I., & Bilbao, J. (2017). Overfitting problem and the over-training in the era of data: Particularly for Artificial Neural Networks. *Eighth International Conference on Intelligent Computing and Information Systems (ICICIS)*, 173–177. <https://doi.org/doi:10.1109/INTELCIS.2017.8260032>.
- Birth, G. S., & McVey, G. R. (1968). Measuring the Color of Growing Turf with a Reflectance Spectrophotometer1. *Agronomy Journal*, 60(6), 640–643.

- <https://doi.org/10.2134/AGRONJ1968.00021962006000060016X>
- Bishop, C. D., & Cooper, R. M. (1983). An ultrastructural study of root invasion in three vascular wilt diseases. *Physiological Plant Pathology*, 22(1), 15-19. [https://doi.org/10.1016/s0048-4059\(83\)81034-0](https://doi.org/10.1016/s0048-4059(83)81034-0)
- Blanc, G., Charcosset, A., Mangin, B., Gallais, A., & Moreau, L. (2006). Connected populations for detecting quantitative trait loci and testing for epistasis: an application in maize. *Theoretical and Applied Genetics* 2006 113:2, 113(2), 206–224. <https://doi.org/10.1007/S00122-006-0287-1>
- Bolker, B. (2008). *Ecology Ecological Models and Data in R*. Princeton University Press. <https://doi.org/10.1086/644667>
- Bolton, D. K., & Friedl, M. A. (2013). Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agricultural and Forest Meteorology*, 173, 74–84. <https://doi.org/10.1016/J.AGRFORMET.2013.01.007>
- Bowers, J. H., Nameth, S. T., Riedel, R. M., & Rowe, R. C. (1996). Infection and colonization of potato roots by *Verticillium dahliae* as affected by *Pratylenchus penetrans* and *P. crenatus*. In *Phytopathology* (Vol. 86, Issue 6, pp. 614–621). <https://doi.org/10.1094/Phyto-86-614>
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. S. (2007). TASSEL: Software for association mapping of complex traits in diverse samples. *Bioinformatics*, 23(19), 2633–2635. <https://doi.org/10.1093/bioinformatics/btm308>
- Bramardi, S. (2000). *Estrategias para el análisis de datos en la caracterización de recursos genéticos*. Universidad Politécnica de Valencia, España.
- Brichet, N., Fournier, C., Turc, O., Strauss, O., Artzet, S., Pradal, C., Welcker, C., Tardieu, F., & Cabrera-Bosquet, L. (2017). A robot-assisted imaging pipeline for tracking the growths of maize ear and silks in a high-throughput phenotyping platform. *Plant Methods* 2017 13:1, 13(1), 1–12. <https://doi.org/10.1186/S13007-017-0246-7>
- Brim, C. A. (1966). A Modified Pedigree Method of Selection in Soybeans <sup>1</sup>. *Crop Science*, 6(2), 220–220. <https://doi.org/10.2135/cropsci1966.0011183X000600020041x>
- Broman, K. W. (2006). *Use of Hidden Markov Models for QTL Mapping*. <http://biostats.bepress.com/jhubiostat/paper125>
- Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, S., Yandell, B. S., & Churchill, G. A. (2019a). R/qtl2: Software for mapping quantitative trait loci with high-

- dimensional data and multiparent populations. *Genetics*, 211(2), 495–502. <https://doi.org/10.1534/genetics.118.301595>
- Broman, K. W., Gatti, D. M., Simecek, P., Furlotte, N. A., Prins, P., Sen, S., Yandell, B. S., & Churchill, G. A. (2019b). R/qt12: Software for Mapping Quantitative Trait Loci with High-Dimensional Data and Multiparent Populations. *Genetics*, 211(2), 495–502. <https://doi.org/10.1534/genetics.118.301595>
- Broman, K. W., Gatti, D. M., Svenson, K. L., Sen, S., & Churchill, G. A. (2019). Cleaning genotype data from diversity outbred mice. In *G3: Genes, Genomes, Genetics* (Vol. 9, Issue 5, pp. 1571–1579). <https://doi.org/10.1534/g3.119.400165>
- Brown, A. (1989). Core collections: a practical approach to genetic resources management. *Genome*, 31(2), 818–824. <https://doi.org/https://doi.org/10.1139/g89-144>
- Bruggers, R. L., Rodriguez, E., & Zaccagnini, M. E. (1998). Planning for bird pest problem resolution: a case study. *International Biodeterioration & Biodegradation*, 42(2–3), 173–184. [https://doi.org/10.1016/S0964-8305\(98\)00046-8](https://doi.org/10.1016/S0964-8305(98)00046-8)
- Brunel, D. (1994). A microsatellite marker in *Helianthus annuus* L. *Plant Molecular Biology*, 24(2), 397–400. <https://doi.org/10.1007/BF00020177>
- Bullard, R. W. (1988). Characteristics of bird-resistance in agricultural crops. *Proc. Vertebr. Pest Conf., March*.
- Burke, J. M., Tang, S., Knapp, S. J., & Rieseberg, L. H. (2002). Genetic analysis of sunflower domestication. *Genetics*, 161(3), 1257–1267. <https://doi.org/10.1093/genetics/161.3.1257>
- Cadic, E., Coque, M., Vear, F., Grezes-Besset, B., Pauquet, J., Piquemal, J., Lippi, Y., Blanchard, P., Romestant, M., Pouilly, N., Rengel, D., Gouzy, J., Langlade, N., Mangin, B., & Vincourt, P. (2013). Combined linkage and association mapping of flowering time in Sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics*, 126(5), 1337–1356. <https://doi.org/10.1007/s00122-013-2056-2>
- Calderón, R., Navas-Cortés, J. A., Lucena, C., & Zarco-Tejada, P. J. (2013). High-resolution airborne hyperspectral and thermal imagery for early detection of *Verticillium* wilt of olive using fluorescence, temperature and narrow-band spectral indices. *Remote Sensing of Environment*, 139, 231–245. <https://doi.org/10.1016/j.rse.2013.07.031>
- Calderón, Rocío, Navas-Cortés, J. A., & Zarco-Tejada, P. J. (2015). Early detection and quantification of *verticillium* wilt in olive using hyperspectral and thermal imagery over

- large areas. *Remote Sensing*, 7(5), 5584–5610. <https://doi.org/10.3390/rs70505584>
- Cammarano, D., Fitzgerald, G. J., Casa, R., & Basso, B. (2014). Assessing the Robustness of Vegetation Indices to Estimate Wheat N in Mediterranean Environments. *Remote Sensing* 2014, Vol. 6, Pages 2827-2844, 6(4), 2827–2844. <https://doi.org/10.3390/RS6042827>
- Campanelli, G., Sestili, S., Acciarri, N., Montemurro, F., Palma, D., Leteo, F., & Beretta, M. (2019). Multi-Parental Advances Generation Inter-Cross Population, to Develop Organic Tomato Genotypes by Participatory Plant Breeding. *Agronomy*, 9(3), 119. <https://doi.org/10.3390/agronomy9030119>
- Candiago, S., Remondino, F., De Giglio, M., Dubbini, M., & Gattelli, M. (2015). Evaluating multispectral images and vegetation indices for precision farming applications from UAV images. *Remote Sensing*, 7(4), 4026–4047. <https://doi.org/10.3390/rs70404026>
- Carter, G. A. (1993). Responses of Leaf Spectral Reflectance To Plant Stress. *American Journal of Botany*, 80(3), 239–243. <https://doi.org/10.1002/j.1537-2197.1993.tb13796.x>
- Castaño, F. (2018). *The sunflower crop in Argentina: past, present and potential future*. 25(1), 1–10.
- Castaño, Fernando, Vear, F., & de Labrouhe, D. T. (1993). Resistance of sunflower inbred lines to various forms of attack by *Sclerotinia sclerotiorum* and relations with some morphological characters. *Euphytica*, 68(1–2), 85–98. <https://doi.org/10.1007/BF00024158>
- Castelao Tetila, E., Brandoli Machado, B., Belete, N. A. D. S., Guimaraes, D. A., & Pistori, H. (2017). Identification of Soybean Foliar Diseases Using Unmanned Aerial Vehicle Images. *IEEE Geoscience and Remote Sensing Letters*, 14(12), 2190–2194. <https://doi.org/10.1109/LGRS.2017.2743715>
- Catchen, J., Hohenlohe, P. A., Bassham, S., Amores, A., & Cresko, W. A. (2013). Stacks: An analysis tool set for population genomics. *Molecular Ecology*, 22(11), 3124–3140. <https://doi.org/10.1111/mec.12354>
- Cavanagh, C., Morell, M., Mackay, I., & Powell, W. (2008). *From mutations to MAGIC: resources for gene discovery, validation and delivery in crop plants*. 215–221. <https://doi.org/10.1016/j.pbi.2008.01.002>
- Cawley, G. C., & Talbot, N. L. C. (2010). *On Over-fitting in Model Selection and Subsequent Selection Bias in Performance Evaluation*. 11, 2079–2107.

- Celik, I., Bodur, S., Frary, A., & Doganlar, S. (2016). Genome-wide SNP discovery and genetic linkage map construction in sunflower (*Helianthus annuus* L.) using a genotyping by sequencing (GBS) approach. *Molecular Breeding*, 36(9), 1–9. <https://doi.org/10.1007/s11032-016-0558-8>
- Chapman, M. A., Leebens-Mack, J. H., & Burke, J. M. (2008). Positive Selection and Expression Divergence Following Gene Duplication in the Sunflower CYCLOIDEA Gene Family. *Molecular Biology and Evolution*, 25(7), 1260–1273. <https://doi.org/10.1093/molbev/msn001>
- Charrad, M., Ghazzali, N., Boiteau, V., & Niknafs, A. (2014). Nbclust: An R package for determining the relevant number of clusters in a data set. *Journal of Statistical Software*, 61(6), 1–36. <https://doi.org/10.18637/jss.v061.i06>
- Chawade, A., Van Ham, J., Blomquist, H., Bagge, O., Alexandersson, E., & Ortiz, R. (2019). High-throughput field-phenotyping tools for plant breeding and precision agriculture. *Agronomy*, 9(5), 1–18. <https://doi.org/10.3390/agronomy9050258>
- Chlingaryan, A., Sukkariéh, S., & Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: A review. *Computers and Electronics in Agriculture*, 151(November 2017), 61–69. <https://doi.org/10.1016/j.compag.2018.05.012>
- Christie, M. R., Tennessen, J. A., & Blouin, M. S. (2013). Bayesian parentage analysis with systematic accountability of genotyping error, missing data and false matching. *Bioinformatics*, 29(6), 725–732. <https://doi.org/10.1093/bioinformatics/btt039>
- Churchill, G. A., Airey, D. C., Allayee, H., Angel, J. M., Attie, A. D., Beatty, J., Beavis, W. D., Belknap, J. K., Bennett, B., Berrettini, W., Bleich, A., Bogue, M., Broman, K. W., Buck, K. J., Buckler, E., Burmeister, M., Chesler, E. J., Cheverud, J. M., Clapcote, S., ... Zou, F. (2004). The Collaborative Cross, a community resource for the genetic analysis of complex traits. In *Nature Genetics* (Vol. 36, Issue 11, pp. 1133–1137). Nature Publishing Group. <https://doi.org/10.1038/ng1104-1133>
- Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S. J., Lu, X., & Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly*, 6(2), 80–92. <https://doi.org/10.4161/fly.19695>
- Cobb, J. N., DeClerck, G., Greenberg, A., Clark, R., & McCouch, S. (2013). Next-generation phenotyping: Requirements and strategies for enhancing our understanding of genotype-

- phenotype relationships and its relevance to crop improvement. In *Theoretical and Applied Genetics* (Vol. 126, Issue 4, pp. 867–887). Springer Verlag. <https://doi.org/10.1007/s00122-013-2066-0>
- Cockerham, C. (1963). Estimation of genetic variances. *Statistical Genetics and Plant Breeding*, 982, 53–94.
- Cockerton, H. M., Li, B., Vickerstaff, R. J., Eyre, C. A., Sargent, D. J., Armitage, A. D., Marina-Montes, C., Garcia-Cruz, A., Passey, A. J., Simpson, D. W., & Harrison, R. J. (2019). Identifying *Verticillium dahliae* resistance in strawberry through disease screening of multiple populations and image based phenotyping. *Frontiers in Plant Science*, 10, 924. <https://doi.org/10.3389/fpls.2019.00924>
- Cockram, J., & Mackay, I. (2018). Genetic Mapping Populations for Conducting High-Resolution Trait Mapping in Plants. In R. Varshney, M. Pandey, & A. Chitkineni (Eds.), *Plant Genetics and Molecular Biology* (pp. 109–138). Springer. <https://doi.org/10.1007/978-3-319-91313-1>
- Collard, B. C. Y., Jahufer, M. Z. Z., Brouwer, J. B., & Pang, E. C. K. (2005). An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. In *Euphytica* (Vol. 142, Issues 1–2, pp. 169–196). Springer. <https://doi.org/10.1007/s10681-005-1681-5>
- Collard, B., & Mackill, D. J. (2008). Marker-assisted selection: An approach for precision plant breeding in the twenty-first century. In *Philosophical Transactions of the Royal Society B: Biological Sciences* (Vol. 363, Issue 1491, pp. 557–572). Royal Society. <https://doi.org/10.1098/rstb.2007.2170>
- Colomina, I., & Molina, P. (2014). Unmanned aerial systems for photogrammetry and remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92, 79–97. <https://doi.org/10.1016/J.ISPRSJPRS.2014.02.013>
- Cramer, E. R. A., Hall, M. L., De Kor, S. R., Lovette, I. J., & Vehrencamp, S. L. (2011). Infrequent extra-pair paternity in the banded wren, a synchronously breeding tropical passerine. *Condor*, 113(3), 637–645. <https://doi.org/10.1525/cond.2011.100233>
- Crites, G. D. (1993). Domesticated Sunflower in Fifth Millennium B.P. Temporal Context: New Evidence from Middle Tennessee. *American Antiquity*, 58(1), 146–148. <https://doi.org/10.2307/281459>
- Crossa, Jose, & Federer, W. (2012). I.4 Screening Experimental Designs for Quantitative Trait

- Loci, Association Mapping, Genotype-by Environment Interaction, and Other Investigations. *Frontiers in Physiology*, 0, 156. <https://doi.org/10.3389/FPHYS.2012.00156>
- Crossa, Jose, Fritsche-Neto, R., Montesinos-Lopez, O. A., Costa-Neto, G., Dreisigacker, S., Montesinos-Lopez, A., & Bentley, A. R. (2021). The Modern Plant Breeding Triangle: Optimizing the Use of Genomics, Phenomics, and Enviromics Data. *Frontiers in Plant Science*, 0, 332. <https://doi.org/10.3389/FPLS.2021.651480>
- Crossa, José, Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de los Campos, G., Burguño, J., González-Camacho, J. M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., & Varshney, R. K. (2017). Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. *Trends in Plant Science*, 22(11), 961–975. <https://doi.org/10.1016/j.tplants.2017.08.011>
- Cuadras, C. M. (2019). *Nuevos Métodos de Análisis Multivariante*. CMC Editions.
- Cullis, B. R., Smith, A., & Coombes, N. E. (2006). On the design of early generation variety trials with correlated data. *Journal of Agricultural, Biological, and Environmental Statistics*, 11(4), 381–393.
- Cupertino, F. B., Leal, J. B., Vidal, P. O., & Gaiotto, F. A. (2009). Parentage testing of hybrid full-sib families of Eucalyptus with microsatellites. <Http://Dx.Doi.Org/10.1080/02827580802613073>, 24(1), 2–7. <https://doi.org/10.1080/02827580802613073>
- Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. In *Nature Reviews Genetics* (Vol. 12, Issue 7, pp. 499–510). Nature Publishing Group. <https://doi.org/10.1038/nrg3012>
- Davies, R. W., Flint, J., Myers, S., & Mott, R. (2016). Rapid genotype imputation from sequence without reference panels. *Nature Genetics*, 48(8), 965–969. <https://doi.org/10.1038/ng.3594>
- De Castro, A. I., Ehsani, R., Ploetz, R., Crane, J. H., & Abdulridha, J. (2015). Optimum spectral and geometric parameters for early detection of laurel wilt disease in avocado. *Remote Sensing of Environment*, 171, 33–44. <https://doi.org/10.1016/j.rse.2015.09.011>
- Debaeke, P., Casadebaig, P., & Langlade, N. B. (2021). New challenges for sunflower

- ideotyping in changing environments and more ecological cropping systems. *OCL*, 28, 29. <https://doi.org/10.1051/OCL/2021016>
- Defacio, R. (2016). *Evaluación comparativa de distintas estrategias de análisis de datos para la caracterización y ordenamiento de la variabilidad genética de poblaciones locales de maíz (Zea mays L.)* [Universidad Nacional de Rosario]. [http://rehip.unr.edu.ar/xmlui/bitstream/handle/2133/13925/Tesis - versión final a imprimir.pdf?sequence=3](http://rehip.unr.edu.ar/xmlui/bitstream/handle/2133/13925/Tesis%20-%20versión%20final%20a%20imprimir.pdf?sequence=3)
- Dehmer, K. J., & Friedt, W. (1998). Evaluation of different microsatellite motifs for analysing genetic relationships in cultivated sunflower (*Helianthus annuus* L.). *Plant Breeding*, 117(1), 45–48. <https://doi.org/10.1111/j.1439-0523.1998.tb01446.x>
- Del Medico, A. P., Cabodevila, V. G., Vitelleschi, M. S., & Pratta, G. R. (2019). Multivariate estimate of heritability for quality traits in tomatoes by the multiple factor analysis. *Pesquisa Agropecuaria Brasileira*, 54. <https://doi.org/10.1590/S1678-3921.pab2019.v54.00064>
- Del Medico, A., Tenaglia, G., Vitelleschi, M. S., Lavalle, A., & Pratta, G. (2021). Análisis Factorial Múltiple Mixto para la Identificación de clones de banana (*Musa* spp). *XXV Jornadas "Investigaciones En La Facultad" de Ciencias Económicas y Estadística*, 4.
- Dell'Acqua, M., Gatti, D. M., Pea, G., Cattonaro, F., Coppens, F., Magris, G., Hlaing, A. L., Aung, H. H., Nelissen, H., Baute, J., Frascaroli, E., Churchill, G. A., Inzé, D., Morgante, M., & Pè, M. E. (2015). Genetic properties of the MAGIC maize population: A new platform for high definition QTL mapping in *Zea mays*. *Genome Biology*, 16(1), 1–23. <https://doi.org/10.1186/s13059-015-0716-z>
- Descalsota, G. I. L., Swamy, B. P. M., Zaw, H., Inabangan-Asilo, M. A., Amparado, A., Mauleon, R., Chadha-Mohanty, P., Arocena, E. C., Raghavan, C., Leung, H., Hernandez, J. E., Lalusin, A. B., Mendioro, M. S., Diaz, M. G. Q., & Reinke, R. (2018). Genome-Wide Association Mapping in a Rice MAGIC Plus Population Detects QTLs and Genes Useful for Biofortification. *Frontiers in Plant Science*, 9, 1347. <https://doi.org/10.3389/fpls.2018.01347>
- Di Rienzo, J., Guzmán, W., & Casanoves, F. (2001). *D.G.C. Test de Comparación de Medias*. (InfoStat Versión 1.1/Profesional). Grupo InfoStat, Facultad de Ciencias Agropecuarias, Universidad Nacional de Córdoba, Córdoba, Argentina.
- Diaz, S., Ariza-Suarez, D., Izquierdo, P., Lobaton, J. D., de la Hoz, J. F., Acevedo, F., Duitama, J., Guerrero, A. F., Cajiao, C., Mayor, V., Beebe, S. E., & Raatz, B. (2020). Genetic

- mapping for agronomic traits in a MAGIC population of common bean (*Phaseolus vulgaris* L.) under drought conditions. *BMC Genomics*, 21(1), 799. <https://doi.org/10.1186/s12864-020-07213-6>
- Dimitrijevic, A., & Horn, R. (2018). Sunflower hybrid breeding: From markers to genomic selection. In *Frontiers in Plant Science* (Vol. 8, p. 2238). Frontiers Media S.A. <https://doi.org/10.3389/fpls.2017.02238>
- Diouf, I., & Pascual, L. (2021). Multiparental Population in Crops: Methods of Development and Dissection of Genetic Traits. In Tripodi P (Ed.), *Crop Breeding* (pp. 13–32). Humana, New York, NY. [https://doi.org/https://doi.org/10.1007/978-1-0716-1201-9\\_2](https://doi.org/https://doi.org/10.1007/978-1-0716-1201-9_2)
- Diouf, Isidore, Derivot, L., Koussevitzky, S., Carretero, Y., Bitton, F., Moreau, L., & Causse, M. (2020). Genetic basis of phenotypic plasticity and genotype × environment interactions in a multi-parental tomato population. *Journal of Experimental Botany*, 71(18), 5365–5376. <https://doi.org/10.1093/jxb/eraa265>
- Dixon, W. J. (1950). Analysis of Extreme Values. *The Annals of Mathematical Statistics*, 21(4), 488–506. <http://www.jstor.org/stable/2236602>
- Dixon, W. J. (1951). Ratios Involving Extreme Values. *The Annals of Mathematical Statistics*, 22(1), 68–78. <https://doi.org/10.1214/aoms/1177729693>
- Donald, C., & Hamblin, J. (1983). The Convergent Evolution of Annual Seed Crops in Agriculture. *Advances in Agronomy*, 36, 97–143.
- 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(8), 599–615. <https://doi.org/10.1139/G08-038>
- Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E. (2011). A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS ONE*, 6(5), e19379. <https://doi.org/10.1371/journal.pone.0019379>
- Endelman, J. B., & Plomion, C. (2014). LPmerge: an R package for merging genetic maps by linear programming. *Bioinformatics*, 30(11), 1623–1624. <https://doi.org/10.1093/BIOINFORMATICS/BTU091>
- Erreguerena, I. A., Rojo, R., Quiroz, F. J., Clemente, G. E., & Escande, A. R. (2019). Relationship between *Verticillium dahliae* inoculum and sunflower wilt in Argentina.

- Canadian Journal of Plant Pathology*, 41(4), 576–584.  
<https://doi.org/10.1080/07060661.2019.1620860>
- Escofier, B., & Pagès, J. (1992). *Análisis Factoriales Simples y Múltiples*. Ed. Universidad del País Vasco.
- Fahlgren, N., Gehan, M. A., & Baxter, I. (2015). Lights, camera, action: High-throughput plant phenotyping is ready for a close-up. *Current Opinion in Plant Biology*, 24, 93–99.  
<https://doi.org/10.1016/j.pbi.2015.02.006>
- Falconer, D., & Mackay, T. (1996). *Introducción a La Genética Cuantitativa*. Acribia S.A Zaragoza España.
- Fass, M. (2018). *Identificación y caracterización de mutantes de genes candidato involucrados en la resistencia al patógeno Sclerotinia sclerotiorum en girasol*. Universidad Nacional de Buenos Aires.
- Fass, M. I., Rivarola, M., Ehrenbolger, G. F., Maringolo, C. A., Montecchia, J. F., Quiroz, F., García-García, F., Blázquez, J. D., Hopp, H. E., Heinz, R. A., Paniego, N. B., & Lia, V. V. (2020). Exploring sunflower responses to Sclerotinia head rot at early stages of infection using RNA-seq analysis. *Scientific Reports*, 10(1), 1–14. <https://doi.org/10.1038/s41598-020-70315-4>
- Federer, W., & Raghavarao, D. (1975). On Augmented Designs. *Biometrics*, 31(1), 29–35.  
<http://journal.frontiersin.org/Article/10.3389/fpls.2015.00945/abstract>
- Federer, W. T., Reynolds, M., & Crossa, J. (2001). Combining results from augmented designs over sites. *Agronomy Journal*, 93(2), 389–395.  
<https://doi.org/10.2134/agronj2001.932389x>
- Fehr, W. R. (1991). *Principles of cultivar development: theory and technique*. New York, NY: Macmillan.
- Ferdosi, M. H., Kinghorn, B. P., Werf, J. H. J. Van Der, Lee, S. H., & Gondro, C. (2014). *hsphase : an R package for pedigree reconstruction , detection of recombination events , phasing and imputation of half-sib family groups hsphase : an R package for pedigree reconstruction , detection of recombination events , phasing and imputation of ha*. 15(1), 1–10. <https://doi.org/10.1186/1471-2105-15-172>
- Fernandez, P., Di Rienzo, J., Fernandez, L., Hopp, H. E., Paniego, N., & Heinz, R. A. (2008). Transcriptomic identification of candidate genes involved in sunflower responses to chilling and salt stresses based on cDNA microarray analysis. *BMC Plant Biology*, 8(1),

- 1–18. <https://doi.org/10.1186/1471-2229-8-11>
- Fernández, P., Paniego, 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*(1), 1–9. <https://doi.org/10.1186/1471-2164-4-40>
- Fernandez, P., Rienzo, J. A. D., Moschen, S., Dosio, G. A. A., Aguirrezábal, L. A. N., Hopp, H. E., Paniego, N., & Heinz, R. A. (2011). Comparison of predictive methods and biological validation for qPCR reference genes in sunflower leaf senescence transcript analysis. *Plant Cell Reports*, *30*(1), 63–74. <https://doi.org/10.1007/s00299-010-0944-3>
- Fernandez, P., Soria, M., Blesa, D., DiRienzo, J., Moschen, S., Rivarola, M., Clavijo, B. J., Gonzalez, S., Peluffo, L., Príncipi, D., Dosio, G., Aguirrezabal, L., García-García, F., Conesa, A., Hopp, E., Dopazo, J., Heinz, R. A., & Paniego, N. (2012). Development, Characterization and Experimental Validation of a Cultivated Sunflower (*Helianthus annuus* L.) Gene Expression Oligonucleotide Microarray. *PLoS ONE*, *7*(10), e45899. <https://doi.org/10.1371/journal.pone.0045899>
- Fick, G. N., & Zimmer, D. E. (1974). Monogenic Resistance to Verticillium Wilt in Sunflowers. *Crop Science*, *14*(6), 895–896. <https://doi.org/10.2135/cropsci1974.0011183x001400060037x>
- Filella, I., & Peñuelas, J. (1993). The red edge position and shape as indicators of plant chlorophyll content, biomass and hydric status. *Int. J. Remote Sens.*, *15*(7), 1459–1470. <https://doi.org/10.1080/01431169408954177>
- Filippi, Carla, Zubrzycki, J., Lía, V., Heinz, R. A., Paniego, N. B., & Hopp, H. E. (2014). Genetics and Genomics Applied to Sunflower Breeding. In *Sunflowers: Growth and Development, Environmental Influences and Pests/Diseases* (pp. 61–94). Nova Science.
- Filippi, C. V., Merino, G. A., Montecchia, J. F., Aguirre, N. C., Rivarola, M., Naamati, G., Fass, M. I., Álvarez, D., Di Rienzo, J., Heinz, R. A., Moreira, B. C., Lia, V. V., & Paniego, N. B. (2020). Genetic diversity, population structure and linkage disequilibrium assessment among international sunflower breeding collections. *Genes*, *11*(3). <https://doi.org/10.3390/genes11030283>
- Filippi, C. V., Zubrzycki, J. E., Di Rienzo, J. A., Quiroz, F., Fusari, C. M., Alvarez, D., Maringolo, C. A., Cordes, D., Escande, A., Hopp, H. E., Heinz, R. A., Lia, V. V., & Paniego, N. B. (2017). Phenotyping sunflower genetic resources for sclerotinia head rot response: Assessing variability for disease resistance breeding. *Plant Disease*, *101*(11), 1941–

1948. <https://doi.org/10.1094/PDIS-12-16-1784-RE>
- Filippi, C. V., Zubrzycki, J. E., Di Rienzo, J. A., Quiroz, F. J., Puebla, A. F., Alvarez, D., Maringolo, C. A., Escande, A. R., Hopp, H. E., Heinz, R. A., Paniego, N. B., & Lia, V. V. (2020). Unveiling the genetic basis of Sclerotinia head rot resistance in sunflower. *BMC Plant Biology*, *20*(1), 322. <https://doi.org/10.1186/s12870-020-02529-7>
- Filippi, C V, Aguirre, N., Rivas, J. G., Zubrzycki, J., Puebla, A., Cordes, D., Moreno, M. V., Fusari, C. M., Alvarez, D., Heinz, R. A., Hopp, H. E., Paniego, N. B., & Lia, V. V. (2015). Population structure and genetic diversity characterization of a sunflower association mapping population using SSR and SNP markers. *BMC Plant Biology*, *15*(1). <https://doi.org/10.1186/s12870-014-0360-x>
- Fiorani, F., & Schurr, U. (2013). Future scenarios for plant phenotyping. *Annual Review of Plant Biology*, *64*, 267–291. <https://doi.org/10.1146/annurev-arplant-050312-120137>
- Fischer, R. A., & Edmeades, G. O. (2010). Breeding and cereal yield progress. *Crop Science*, *50*(April), S-85-S-98. <https://doi.org/10.2135/cropsci2009.10.0564>
- Fisher, P. J., Malthus, B., Walker, M. C., Corbett, G., & Spelman, R. J. (2009). The number of single nucleotide polymorphisms and on-farm data required for whole-herd parentage testing in dairy cattle herds. *Journal of Dairy Science*, *92*(1), 369–374. <https://doi.org/10.3168/jds.2008-1086>
- Flanagan, S. P., & Jones, A. G. (2018). The future of parentage analysis: From microsatellites to SNPs and beyond. *Molecular Ecology*, *28*(3), 544–567. <https://doi.org/10.1111/mec.14988>
- Flint-Garcia, S. A., Thornsberry, J. M., & Edwards, S. B. (2003). Structure of Linkage Disequilibrium in Plants. *Annual Review of Plant Biology*, *54*, 357–374. <https://doi.org/10.1146/annurev.arplant.54.031902.134907>
- Flores Berrios, E., Gentzbittel, L., Kayyal, H., Alibert, G., & Sarrafi, A. (2000). AFLP mapping of QTLs for in vitro organogenesis traits, using recombinant inbred lines in sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics*, *101*(8), 1299–1306. <https://doi.org/10.1007/s001220051610>
- Fonstad, M. A., Dietrich, J. T., Courville, B. C., Jensen, J. L., & Carbonneau, P. E. (2013). Topographic structure from motion: A new development in photogrammetric measurement. *Earth Surface Processes and Landforms*, *38*(4), 421–430. <https://doi.org/10.1002/esp.3366>

- Forster, B. P., Heberle-Bors, E., Kasha, K. J., & Touraev, A. (2007). The resurgence of haploids in higher plants. In *Trends in Plant Science* (Vol. 12, Issue 8, pp. 368–375). Elsevier. <https://doi.org/10.1016/j.tplants.2007.06.007>
- Fox, G., & Linz, G. (1983). Evaluation of red-winged blackbird resistant sunflower germplasm. *Bird Control Semin.* 9, 181–189.
- Fradin, E. F., & Thomma, B. P. H. J. (2006). Physiology and molecular aspects of *Verticillium* wilt diseases caused by *V. dahliae* and *V. albo-atrum*. In *Molecular Plant Pathology* (Vol. 7, Issue 2, pp. 71–86). John Wiley & Sons, Ltd. <https://doi.org/10.1111/j.1364-3703.2006.00323.x>
- Franco, T. L., & Hidalgo, R. (2003). Análisis estadístico de datos de caracterización morfológica de recursos fitogenéticos. *IBPGR*, 89. [https://www.biodiversityinternational.org/fileadmin/\\_migrated/uploads/tx\\_news/Análisis\\_estadístico\\_de\\_datos\\_de\\_caracterización\\_morfológica\\_de\\_recursos\\_fitogenéticos\\_894.pdf](https://www.biodiversityinternational.org/fileadmin/_migrated/uploads/tx_news/Análisis_estadístico_de_datos_de_caracterización_morfológica_de_recursos_fitogenéticos_894.pdf)
- Funk, V. A., Bayer, R. J., Keeley, S., Chan, R., Watson, L., Gemeinholzer, B., Schilling, E., Panero, J. L., Baldwin, B. G., & Garcia-Jacas, N. (2005). Everywhere but Antarctica: using a supertree to understand the diversity and distribution of the Compositae. *Biol. Skr.* 55: 343-373.
- Fusari, C. M., Di Rienzo, J. A., Troglia, C., Nishinakamasu, V., Moreno, M. V., Maringolo, C., Quiroz, F., Álvarez, D., Escande, A., Hopp, E., Heinz, R., Lia, V. V., & Paniego, N. B. (2012). Association mapping in sunflower for sclerotinia head rot resistance. *BMC Plant Biology*, 12. <https://doi.org/10.1186/1471-2229-12-93>
- 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. <https://doi.org/10.1186/1471-2229-8-7>
- Fusari, C. M., Lia, V. V., Nishinakamasu, V., Zubrzycki, J. E., Puebla, A. F., Maligne, A. E., Hopp, H. E., Heinz, R. A., & Paniego, N. B. (2011). Single nucleotide polymorphism genotyping by heteroduplex analysis in sunflower (*Helianthus annuus* L.). *Molecular Breeding*, 28(1), 73–89. <https://doi.org/10.1007/s11032-010-9462-9>
- Gabriel, S. B., Schaffner, S. F., Nguyen, H., Moore, J. M., Roy, J., Blumenstiel, B., Higgins, J., DeFelice, M., Lochner, A., Faggart, M., Liu-Cordero, S. N., Rotimi, C., Adeyemo, A., Cooper, R., Ward, R., Lander, E. S., Daly, M. J., & Altshuler, D. (2002). The structure of

- haplotype blocks in the human genome. *Science*, 296(5576), 2225–2229. [https://doi.org/10.1126/SCIENCE.1069424/SUPPL\\_FILE/PAP.PDF](https://doi.org/10.1126/SCIENCE.1069424/SUPPL_FILE/PAP.PDF)
- Galella, M., Bazzalo, M., & León, A. (2004). Comparison of the pathogenicity of *Verticillium dahliae* isolates from Argentina and the USA. *16th International Sun- Flower Conference*, 177–178.
- Galièni, A., D'Ascenzo, N., Stagnari, F., Pagnani, G., Xie, Q., & Pisante, M. (2021). Past and Future of Plant Stress Detection: An Overview From Remote Sensing to Positron Emission Tomography. *Frontiers in Plant Science*, 11, 1975. <https://doi.org/10.3389/FPLS.2020.609155/BIBTEX>
- Gardner, F., Pearce, R., & Mitchel, R. (1985). *Physiology of crop plants*. Iowa State University Press. USA.
- Garin, V. (2019). *A statistical framework for the detection of quantitative trait loci in plant multi-parent populations composed of crosses* [Wageningen University]. <https://doi.org/https://doi.org/10.18174/494464>
- Gedil, M. A., Slabaugh, M. B., Berry, S., Johnson, R., Michelmore, R., Miller, J., Gulya, T., & Knapp, S. J. (2001). Candidate disease resistance genes in sunflower cloned using conserved nucleotide-binding site motifs: Genetic mapping and linkage to the downy mildew resistance gene PI1. *Genome*, 44(2), 205–212. <https://doi.org/10.1139/gen-44-2-205>
- Gentzbittel, L., Vear, F., Zhang, Y. X., Bervillé, A., & Nicolas, P. (1995). Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics*, 90(7–8), 1079–1086. <https://doi.org/10.1007/BF00222925>
- Gianola, D., & Norton, H. W. (1981). Scaling threshold characters. *Genetics*, 99(2), 357–364. <https://doi.org/10.1093/genetics/99.2.357>
- Gibson, P. J., Power, C. H., & Keating, J. (2013). Introductory Remote Sensing: Principles and Concepts. *Introductory Remote Sensing: Principles and Concepts*. <https://doi.org/10.4324/9780203714522>
- Gómez-Candón, D., De Castro, A. I., & López-Granados, F. (2014). Assessing the accuracy of mosaics from unmanned aerial vehicle (UAV) imagery for precision agriculture purposes in wheat. *Precision Agriculture*, 15(1), 44–56. <https://doi.org/10.1007/s11119-013-9335-4>
- Gonzalez-Dugo, Victoria, Hernandez, P., Solis, I., & Zarco-Tejada, P. J. (2015). Using high-

- resolution hyperspectral and thermal airborne imagery to assess physiological condition in the context of wheat phenotyping. *Remote Sensing*, 7(10), 13586–13605. <https://doi.org/10.3390/rs71013586>
- Gonzalez, J. (2016). Mejoramiento genético de girasol (*Helianthus annuus*) en INTA - EEA Pergamino, a partir de distintas fuentes de germoplasma. *RTA*, 10(31), 49–52.
- González, J., Mancuso, N., Alvarez, D., Cordes, D., & Vázquez, A. (2015). Contribution of the Argentine germplasm to the improvement of sunflower. In *Helia* (Vol. 38, Issue 62, pp. 121–140). Institute of Field and Vegetable Crops. <https://doi.org/10.1515/helia-2014-0026>
- González, Julio, Mancuso, N., Ludueña, P., & Ivancovich, A. (2008). Verticilosis en germoplasma de girasol. *17° International Sunflower Conference*.
- Gosseau, F., Blanchet, N., Varès, D., Burger, P., Campergue, D., Colombet, C., Gody, L., Liévin, J. F., Mangin, B., Tison, G., Vincourt, P., Casadebaig, P., & Langlade, N. (2019). Heliaphen, an outdoor high-throughput phenotyping platform for genetic studies and crop modeling. *Frontiers in Plant Science*, 9(January), 1–19. <https://doi.org/10.3389/fpls.2018.01908>
- Goudet, J. (2005). hierfstat, a package for r to compute and test hierarchical F-statistics. *Molecular Ecology Notes*, 5(1), 184–186. <https://doi.org/10.1111/j.1471-8286.2004.00828.x>
- Gower, J. C. (1971). A General Coefficient of Similarity and Some of Its Properties. In *Biometrics* (Vol. 27, Issue 4).
- Graham, N., Ismael, A., & Stovold, T. (2019). DNA fingerprinting to reconstruct parentage from trees in the forest – a second proof of concept. *SCION*, 1–5.
- Gross, P. L., & Hanzel, J. J. (1991). Stability of Morphological Traits Conferring Bird Resistance to Sunflower across Different Environments. *Crop Science*, 31(4), 997–1000. <https://doi.org/10.2135/CROPSCI1991.0011183X003100040031X>
- Grover, A., & Sharma, P. C. (2016). Development and use of molecular markers: Past and present. In *Critical Reviews in Biotechnology* (Vol. 36, Issue 2, pp. 290–302). Taylor and Francis Ltd. <https://doi.org/10.3109/07388551.2014.959891>
- Gu, X. F., Guyot, G., & Verbrugge, M. (1992). Evaluation of measurement errors in ground surface reflectance for satellite calibration. *Int. J. Remote Sens.*, 13(14), 2531–2546. <https://doi.org/10.1080/01431169208904062>
- Guangyu, C., Taihui, Z., Jian, M., Yi, S., & Xin, C. (2012). Changes in Spectral Reflectance of

- Vegetation in Response to Specific Nutrient Supply. *Lecture Notes in Electrical Engineering*, 155 LNEE, 671–675. [https://doi.org/10.1007/978-3-642-28744-2\\_87](https://doi.org/10.1007/978-3-642-28744-2_87)
- Gulya, T. (1985). Registration of five disease resistant sunflower germplasm. *Crop Science*, 25, 719–720.
- Haboudane, D., Tremblay, N., Miller, J. R., & Vigneault, P. (2008). Remote estimation of crop chlorophyll content using spectral indices derived from hyperspectral data. *IEEE Transactions on Geoscience and Remote Sensing*, 46(2), 423–436. <https://doi.org/10.1109/TGRS.2007.904836>
- Haddadi, P., Ebrahimi, A., Langlade, N. B., Yazdi-samadi, B., Berger, M., Calmon, A., Naghavi, M. R., Vincourt, P., & Sarrafi, A. (2012). Genetic dissection of tocopherol and phytosterol in recombinant inbred lines of sunflower through quantitative trait locus analysis and the candidate gene approach. *Molecular Breeding*, 29(3), 717–729. <https://doi.org/10.1007/s11032-011-9585-7>
- 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. <https://doi.org/10.4314/ajb.v9i40>.
- Hadfield, J. D., Richardson, D. S., & Burke, T. (2006). Towards unbiased parentage assignment: Combining genetic, behavioural and spatial data in a Bayesian framework. *Molecular Ecology*, 15(12), 3715–3730. <https://doi.org/10.1111/j.1365-294X.2006.03050.x>
- Haley, C. S., & Knott, S. A. (1992). A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity*, 69(4), 315–324. <https://doi.org/10.1038/hdy.1992.131>
- Hall, A., Feoli, C., Ingaramo, J., & Balzarini, M. (2010). Proyecto brechas ASAGIR: variación interzonal e interanual de las diferencias entre rendimientos de girasol alcanzables y logrados para el período 1999-2007. *5to. Congreso Argentino de Girasol: Con La Mirada En La Competitividad*, 364.
- Harter, A. V., Gardner, K. A., Falush, D., Lentz, D. L., Bye, R. A., & Rieseberg, L. H. (2004). Origin of extant domesticated sunflowers in eastern North America. *Nature*, 430(6996), 201–205. <https://doi.org/10.1038/nature02710>

- He, J., Zhao, X., Laroche, A., Lu, Z. X., Liu, H. K., & Li, Z. (2014). Genotyping-by-sequencing (GBS), An ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. In *Frontiers in Plant Science* (Vol. 5, Issue SEP, p. 484). Frontiers Media S.A. <https://doi.org/10.3389/fpls.2014.00484>
- Heesacker, A., Kishore, V. K., Gao, W., Tang, S., Kolkman, J. M., Gingle, A., Matvienko, M., Kozik, A., Michelmore, R. M., Lai, Z., Rieseberg, L. H., & Knapp, S. J. (2008). SSRs and INDELs mined from the sunflower EST database: Abundance, polymorphisms, and cross-taxa utility. *Theoretical and Applied Genetics*, 117(7), 1021–1029. <https://doi.org/10.1007/s00122-008-0841-0>
- Heiser, C. B., Smith, D. M., Clevenger, S. B., & Martin, W. C. (1969). The north american sunflowers (*Helianthus*). *Memoirs of the Torrey Botanical Club*, 22(3), 1–218.
- Hemshrot, A., Poets, A. M., Tyagi, P., Lei, L., Carter, C. K., Hirsch, C. N., Li, L., Brown-Guedira, G., Morrell, P. L., Muehlbauer, G. J., & Smith, K. P. (2019). Development of a multiparent population for genetic mapping and allele discovery in six-row barley. *Genetics*, 213(2), 595–613. <https://doi.org/10.1534/genetics.119.302046>
- Hernández-Rabadán, D. L., Ramos-Quintana, F., & Guerrero Juk, J. (2014). Integrating SOMs and a Bayesian Classifier for Segmenting Diseased Plants in Uncontrolled Environments. *Scientific World Journal*, 2014. <https://doi.org/10.1155/2014/214674>
- Hernández, L. (2010). Leaf angle and light interception in sunflower (*Helianthus annuus* L.). Role of the petiole's mechanical and anatomical properties. *Phyton*, 79, 109–115. [www.revistaphyton.fund-romuloraggio.org.ar](http://www.revistaphyton.fund-romuloraggio.org.ar)
- Hernández, L., & Orioli, G. (1983). Estudio comparativo de la estructura foliar, intercepción y rendimiento en girasol. *Anales Edaf. y Agrobiol.*, 42, 2137–2148.
- Hernández, L., & Orioli, G. (1994). El ideotipo del girasol (*Helianthus annuus* L.). *Agriscientia*, 11, 87–98. <https://doi.org/10.31047/1668.298x.v11.n0.2445>
- Herrero-Huerta, M., Rodriguez-Gonzalvez, P., & Rainey, K. M. (2020). Yield prediction by machine learning from UAS-based mulit-sensor data fusion in soybean. *Plant Methods*, 16(1). <https://doi.org/10.1186/s13007-020-00620-6>
- Hillnhütter, C., Schweizer, A., Kühnhold, V., & Sikora, R. (2010). Remote sensing for the detection of soil-borne plant parasitic nematodes and fungal pathogens. In E. Oerke, C. Gerhards, & R. Menz (Eds.), *Precision Crop Protection - The Challenge and Use of Heterogeneity* (pp. 151–165). Springer.

- Holman, F. (2020). *Development and Evaluation of Unmanned Aerial Vehicles for High Throughput Phenotyping of Field-based Wheat Trials*. King's College London.
- Huang, B., Verbyla, K. L., & Verbyla, A. P. (2015). MAGIC populations in crops : current status and future prospects. *Theoretical and Applied Genetics*, 999–1017. <https://doi.org/10.1007/s00122-015-2506-0>
- Huang, B Emma, Verbyla, K. L., & Verbyla, A. P. (2015). MAGIC populations in crops : current status and future prospects. *Theoretical and Applied Genetics*, 999–1017. <https://doi.org/10.1007/s00122-015-2506-0>
- Huang, Bevan E., George, A. W., Forrest, K. L., Kilian, A., Hayden, M. J., Morell, M. K., & Cavanagh, C. R. (2012). A multiparent advanced generation inter-cross population for genetic analysis in wheat. *Plant Biotechnology Journal*, 10(7), 826–839. <https://doi.org/10.1111/j.1467-7652.2012.00702.x>
- Huang, X., Paulo, M. J., Boer, M., Effgen, S., Keizer, P., Koornneef, M., & Van Eeuwijk, F. A. (2011). Analysis of natural allelic variation in Arabidopsis using a multiparent recombinant inbred line population. *Proceedings of the National Academy of Sciences of the United States of America*, 108(11), 4488–4493. <https://doi.org/10.1073/pnas.1100465108>
- Huisman, O. C. (1982). Interrelations of Root Growth Dynamics to Epidemiology of Root-Invading Fungi. *Annual Review of Phytopathology*, 20(1), 303–327. <https://doi.org/10.1146/annurev.py.20.090182.001511>
- Hunt, E. R., Dean Hively, W., Fujikawa, S. J., Linden, D. S., Daughtry, C. S. T., & McCarty, G. W. (2010). Acquisition of NIR-green-blue digital photographs from unmanned aircraft for crop monitoring. *Remote Sensing*, 2(1), 290–305. <https://doi.org/10.3390/rs2010290>
- Huynh, B., Agyare, R. Y., Guo, Y., Herniter, I., Lo, S., Wanamaker, S. I., Xu, S., & Timothy, J. (2018). *A multi-parent advanced generation inter-cross ( MAGIC ) population for genetic analysis and improvement of cowpea ( Vigna unguiculata L . Walp .)*. 1129–1142. <https://doi.org/10.1111/tpj.13827>
- Iraqi, F. A., Mahajne, M., Salaymah, Y., Sandovski, H., Tayem, H., Vered, K., Balmer, L., Hall, M., Manship, G., Morahan, G., Pettit, K., Scholten, J., Tweedie, K., Wallace, A., Weerasekera, L., Cleak, J., Durrant, C., Goodstadt, L., Mott, R., ... Schwartz, D. (2012). The genome architecture of the collaborative cross mouse genetic reference population. *Genetics*, 190(2), 389–401. <https://doi.org/10.1534/genetics.111.132639>
- Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C., Deng, D.

- D., Hinchliffe, D. J., Jones, D. C., & Fang, D. D. (2016a). A MAGIC population-based genome-wide association study reveals functional association of GhRBB1\_A07 gene with superior fiber quality in cotton. *BMC Genomics*, *17*(1), 903. <https://doi.org/10.1186/s12864-016-3249-2>
- Islam, M. S., Thyssen, G. N., Jenkins, J. N., Zeng, L., Delhom, C. D., McCarty, J. C., Deng, D. D., Hinchliffe, D. J., Jones, D. C., & Fang, D. D. (2016b). A MAGIC population-based genome-wide association study reveals functional association of GhRBB1\_A07 gene with superior fiber quality in cotton. *BMC Genomics*, *17*(1). <https://doi.org/10.1186/s12864-016-3249-2>
- Jaganathan, D., Bohra, A., Thudi, M., & Varshney, R. K. (2020). Fine mapping and gene cloning in the post-NGS era: advances and prospects. In *Theoretical and Applied Genetics* (Vol. 133, Issue 5, pp. 1791–1810). Springer. <https://doi.org/10.1007/s00122-020-03560-w>
- Jan, C. C., Vick, B. A., Miller, J. F., Kahler, A. L., & Butler, E. T. (1998). Construction of an RFLP linkage map for cultivated sunflower. *Theoretical and Applied Genetics*, *96*(1), 15–22. <https://doi.org/10.1007/s001220050703>
- Jansen, R. C. (1994). Controlling the type I and type II errors in mapping quantitative trait loci. *Genetics*, *138*(3).
- Jansen, Ritsert C., Jannink, J. L., & Beavis, W. D. (2003). Mapping quantitative trait loci in plant breeding populations: Use of parental haplotype sharing. *Crop Science*, *43*(3), 829–834. <https://doi.org/10.2135/cropsci2003.8290>
- Jeong, I. S., Yoon, U. H., Lee, G. S., Ji, H. S., Lee, H. J., Han, C. D., Hahn, J. H., An, G., & Kim, T. H. (2013). SNP-based analysis of genetic diversity in anther-derived rice by whole genome sequencing. *Rice*, *6*(1), 1–12. <https://doi.org/10.1186/1939-8433-6-1>
- Johnson, D. (2000). *Métodos multivariados aplicados al análisis de datos*. International Thomson.
- Johnson, D. M. (2014). An assessment of pre- and within-season remotely sensed variables for forecasting corn and soybean yields in the United States. *Remote Sensing of Environment*, *141*, 116–128. <https://doi.org/10.1016/J.RSE.2013.10.027>
- Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, *24*(11), 1403–1405. <https://doi.org/10.1093/bioinformatics/btn129>
- Jones, A. G., & Ardren, W. R. (2003). Methods of parentage analysis in natural populations. In

- Molecular Ecology* (Vol. 12, Issue 10, pp. 2511–2523). John Wiley & Sons, Ltd. <https://doi.org/10.1046/j.1365-294X.2003.01928.x>
- Jones, A. G., Small, C. M., Paczolt, K. A., & Ratterman, N. L. (2010). A practical guide to methods of parentage analysis. In *Molecular Ecology Resources* (Vol. 10, Issue 1, pp. 6–30). John Wiley & Sons, Ltd. <https://doi.org/10.1111/j.1755-0998.2009.02778.x>
- Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, *50*(4), 663–666. <https://doi.org/10.2307/1936256>
- K. C. Swain, S. J. Thomson, & H. P. W. Jayasuriya. (2010). Adoption of an Unmanned Helicopter for Low-Altitude Remote Sensing to Estimate Yield and Total Biomass of a Rice Crop. *Transactions of the ASABE*, *53*(1), 21–27. <https://doi.org/10.13031/2013.29493>
- Kalinowski, S. T., Taper, M. L., & Marshall, T. C. (2007). Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular Ecology*, *16*(5), 1099–1106. <https://doi.org/10.1111/j.1365-294X.2007.03089.x>
- Kallunki, J. A., & Heywood, V. H. (1994). *Flowering Plants of the World* (46(4)). Brittonia.
- Kamvar, Z. N., Tabima, J. F., & Grünwald, N. J. (2014). Poppr: An R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ*, *2014*(1), 1–14. <https://doi.org/10.7717/peerj.281>
- Kao, C. H., Zeng, Z. B., & Teasdale, R. D. (1999). Multiple interval mapping for quantitative trait loci. *Genetics*, *152*(3), 1203–1216. <https://doi.org/10.1093/genetics/152.3.1203>
- Karlovska, A., Grīnfelde, I., Alsina, I., Priedītis, G., & Roze, D. (2015). Plant reflected spectra depending on biological characteristics and growth conditions. *7th International Scientific Conference Rural Development*, *0*(0). <https://doi.org/10.15544/RD.2015.045>
- Kassambara, A., & Mundt, F. (2020). *Factoextra: Extract and Visualize the Results of Multivariate Data Analyses*. <https://cran.r-project.org/package=factoextra>.
- Katinas, L., Gutiérrez, D. G., Grossi, M. A., & Crisci, J. V. (2007). *Panorama de la familia Asteraceae (= Compositae) en la República Argentina*.
- Kearsey, M. J., & Farquhar, A. G. L. (1998). QTL analysis in plants; where are we now? *Heredity*, *80*(2), 137–142. <https://doi.org/10.1046/j.1365-2540.1998.00500.x>
- Kessner, D., & Novembre, J. (2015). Power Analysis of Artificial Selection Experiments Using Efficient Whole Genome Simulation of Quantitative Traits. *Genetics*, *199*(4), 991–1005.

- <https://doi.org/10.1534/GENETICS.115.175075>
- Khaleghizadeh, A. (2011). Effect of morphological traits of plant, head and seed of sunflower hybrids on house sparrow damage rate. *Crop Protection*, 30(3), 360–367. <https://doi.org/10.1016/J.CROPRO.2010.12.023>
- Khot, L. R., Sankaran, S., Carter, A. H., Johnson, D. A., & Cummings, T. F. (2015). UAS imaging-based decision tools for arid winter wheat and irrigated potato production management. <https://doi.org/10.1080/01431161.2015.1117685>, 37(1), 125–137. <https://doi.org/10.1080/01431161.2015.1117685>
- Kiani, S. P., Grieu, A. P., Maury, A. P., Hewezi, A. T., Gentzbittel, A. L., & Sarrafi, A. 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*. <https://doi.org/10.1007/s00122-006-0419-7>
- Kiani, S. P., Talia, P., Maury, P., Grieu, P., Heinz, R., Perrault, A., Nishinakamasu, V., Hopp, E., Gentzbittel, L., Paniago, 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(4), 773–787. <https://doi.org/10.1016/j.plantsci.2006.12.007>
- Kim, C., Guo, H., Kong, W., Chandnani, R., Shuang, L. S., & Paterson, A. H. (2016). Application of genotyping by sequencing technology to a variety of crop breeding programs. In *Plant Science* (Vol. 242, pp. 14–22). Elsevier Ireland Ltd. <https://doi.org/10.1016/j.plantsci.2015.04.016>
- King, E. G., Macdonald, S. J., & Long, A. D. (2012). Properties and power of the *Drosophila* synthetic population resource for the routine dissection of complex traits. *Genetics*, 191(3), 935–949. <https://doi.org/10.1534/genetics.112.138537>
- Kinman, M. L. (1970). New development in USDA and state experimental station sunflower breeding programmes. *4th International Sunflower Conference*, 181–183.
- Knaus, B. J., & Grünwald, N. J. (2017). `vcfr`: a package to manipulate and visualize variant call format data in R. *Molecular Ecology Resources*, 17(1), 44–53. <https://doi.org/10.1111/1755-0998.12549>
- Knipling, E. B. (1970). Physical and physiological basis for the reflectance of visible and near-infrared radiation from vegetation. *Remote Sensing of Environment*, 1(3), 155–159. [https://doi.org/10.1016/S0034-4257\(70\)80021-9](https://doi.org/10.1016/S0034-4257(70)80021-9)

- Koh, J. C. O., Spangenberg, G., & Kant, S. (2021). Automated Machine Learning for High-Throughput Image-Based Plant Phenotyping. *Remote Sensing 2021, Vol. 13, Page 858, 13(5)*, 858. <https://doi.org/10.3390/RS13050858>
- Kokaly, R. F., Asner, G. P., Ollinger, S. V., Martin, M. E., & Wessman, C. A. (2009). Characterizing canopy biochemistry from imaging spectroscopy and its application to ecosystem studies. *Remote Sensing of Environment, 113(SUPPL. 1)*, S78–S91. <https://doi.org/10.1016/J.RSE.2008.10.018>
- Kolkman, J. M., Berry, S. T., Leon, A. J., 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. <https://doi.org/10.1534/genetics.107.074054>
- Korontzis, G., Malosetti, M., Zheng, C., Maliepaard, C., Mulder, H. A., Lindhout, P., Veerkamp, R. F., & van Eeuwijk, F. A. (2020). QTL detection in a pedigreed breeding population of diploid potato. *Euphytica, 216(9)*. <https://doi.org/10.1007/s10681-020-02674-y>
- Kover, P. X., Valdar, W., Trakalo, J., Scarcelli, N., Ehrenreich, I. M., Purugganan, M. D., Durrant, C., & Mott, R. (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genetics, 5(7)*, 1000551. <https://doi.org/10.1371/journal.pgen.1000551>
- Kuhn, M. (2020). *caret: Classification and Regression Training* (R package version 6.0-86). <https://cran.r-project.org/package=caret>
- Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution, 35(6)*, 1547–1549. <https://doi.org/10.1093/molbev/msy096>
- Ladejobi, O., Elderfield, J., Gardner, K. A., Gaynor, R. C., Hickey, J., Hibberd, J. M., Mackay, I. J., & Bentley, A. R. (2016). Maximizing the potential of multi-parental crop populations. In *Applied and Translational Genomics* (Vol. 11, pp. 9–17). Elsevier B.V. <https://doi.org/10.1016/j.atg.2016.10.002>
- Lambeth, C., Lee, B. C., O'Malley, D., & Wheeler, N. (2001). Polymix breeding with parental analysis of progeny: An alternative to full-sib breeding and testing. *Theoretical and Applied Genetics, 103(6–7)*, 930–943. <https://doi.org/10.1007/s001220100627>
- Langmead, B., & Salzberg, S. L. (2012). *Fast gapped-read alignment with Bowtie 2*. *9(4)*, 357–360. <https://doi.org/10.1038/nmeth.1923>

- Lê, S., Josse, J. & Husson, F. (2008). FactoMineR: An R Package for Multivariate Analysis. *Journal of Statistical Software*. 25(1), 25(1), 18.
- Leclercq, P. (1966). Une stérilité mâle utilisable pour la production d'hybrides simples de tournesol. *Annales de l'Amélioration Des Plantes*, 16, 135–144.
- Leclercq, P. (1969). Une sterilité male cytoplasmique chez le tournesol. *Ann.Amel.Plantes*, 19, 99–106. <https://ci.nii.ac.jp/naid/10005384804>
- Letunic, I., & Bork, P. (2019). Interactive Tree of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Research*, 47(W1), 256–259. <https://doi.org/10.1093/nar/gkz239>
- Leung, H., Raghavan, C., Zhou, B., Oliva, R., Choi, I. R., Lacorte, V., Jubay, M. L., Cruz, C. V., Gregorio, G., Singh, R. K., Ulat, V. J., Borja, F. N., Mauleon, R., Alexandrov, N. N., McNally, K. L., & Sackville Hamilton, R. (2015). Allele mining and enhanced genetic recombination for rice breeding. In *Rice* (Vol. 8, Issue 1, pp. 1–11). Springer New York LLC. <https://doi.org/10.1186/s12284-015-0069-y>
- Li, J., Zhang, F., Qian, X., Zhu, Y., & Shen, G. (2015). Quantification of rice canopy nitrogen balance index with digital imagery from unmanned aerial vehicle. *Remote Sensing Letters*, 6(3), 183–189. <https://doi.org/10.1080/2150704X.2015.1021934>
- Li, K.-Y., Burnside, N. G., Lima, R. S. de, Peciña, M. V., Sepp, K., Pinheiro, V. H. C., Lima, B. R. C. A. de, Yang, M.-D., Vain, A., & Sepp, K. (2021). An Automated Machine Learning Framework in Unmanned Aircraft Systems: New Insights into Agricultural Management Practices Recognition Approaches. *Remote Sensing 2021, Vol. 13, Page 3190*, 13(16), 3190. <https://doi.org/10.3390/RS13163190>
- Li, R., Lyons, M. A., Wittenburg, H., Paigen, B., & Churchill, G. A. (2005). Combining Data From Multiple Inbred Line Crosses Improves the Power and Resolution of Quantitative Trait Loci Mapping. *Genetics*, 169(3), 1699–1709. <https://doi.org/10.1534/GENETICS.104.033993>
- Li, W., Boer, M. P., Zheng, C., Joosen, R. V. L., & van Eeuwijk, F. A. (2021). An IBD-based mixed model approach for QTL mapping in multiparental populations. *Theoretical and Applied Genetics*, 134(11), 3643–3660. <https://doi.org/10.1007/s00122-021-03919-7>
- Liaw, A., & Wiener, M. (2002). Classification and Regression by randomForest. *R News*, 2(3), 18–22. <https://cran.r-project.org/doc/Rnews/>
- Liebisch, F., Kirchgessner, N., Schneider, D., Walter, A., & Hund, A. (2015). Remote, aerial

- phenotyping of maize traits with a mobile multi-sensor approach. *Plant Methods*, 11(1), 1–20. <https://doi.org/10.1186/s13007-015-0048-8>
- Liller, C. B., Walla, A., Boer, M. P., Hedley, P., Macaulay, M., Effgen, S., von Korff, M., van Esse, G. W., & Koornneef, M. (2017). Fine mapping of a major QTL for awn length in barley using a multiparent mapping population. *Theoretical and Applied Genetics*, 130(2), 269–281. <https://doi.org/10.1007/s00122-016-2807-y>
- Linder, C. R., Taha, I., Seiler, G. J., Snow, A. A., & Rieseberg, L. H. (1998). Long-term introgression of crop genes into wild sunflower populations. *Theoretical and Applied Genetics*, 96(3–4), 339–347. <https://doi.org/10.1007/s001220050746>
- Linz, G. M., Homan, H. J., Werner, S. J., Hagy, H. M., & Bleier, W. J. (2011). Assessment of Bird-management Strategies to Protect Sunflowers. *BioScience*, 61(12), 960–970. <https://doi.org/10.1525/BIO.2011.61.12.6>
- Liu, E. Y., Zhang, Q., McMillan, L., de Villena, F. P. M., & Wang, W. (2010). Efficient genome ancestry inference in complex pedigrees with inbreeding. *Bioinformatics*, 26(12), i199. <https://doi.org/10.1093/bioinformatics/btq187>
- Loladze, A., Rodrigues, F. A., Toledo, F., Vicente, F. S., Gérard, B., & Boddupalli, M. P. (2019). Application of remote sensing for phenotyping tar spot complex resistance in maize. *Frontiers in Plant Science*, 10, 552. <https://doi.org/10.3389/fpls.2019.00552>
- López Gialdi, A. I., Moschen, S., Villán, C. S., López Fernández, M. P., Maldonado, S., Paniago, N., Heinz, R. A., & Fernandez, P. (2016). Identification and characterization of contrasting sunflower genotypes to early leaf senescence process combining molecular and physiological studies (*Helianthus annuus* L.). *Plant Science*, 250, 40–50. <https://doi.org/10.1016/j.plantsci.2016.05.017>
- Mackay, I. J., Bansept-Basler, P., Bentley, A. R., Cockram, J., Gosman, N., Greenland, A. J., Horsnell, R., Howells, R., O'Sullivan, D. M., Rose, G. A., & Howell, P. J. (2014). An eight-parent multiparent advanced generation inter-cross population for winter-sown wheat: Creation, properties, and validation. *G3: Genes, Genomes, Genetics*, 4(9), 1603–1610. <https://doi.org/10.1534/g3.114.012963>
- Mackay, I., & Powell, W. (2007). Methods for linkage disequilibrium mapping in crops. In *Trends in Plant Science* (Vol. 12, Issue 2, pp. 57–63). Elsevier. <https://doi.org/10.1016/j.tplants.2006.12.001>
- Mackay, T. F. C. (2001). The genetic architecture of quantitative traits. In *Annual Review of*

- Genetics* (Vol. 35, pp. 303–339). Annual Reviews 4139 El Camino Way, P.O. Box 10139, Palo Alto, CA 94303-0139, USA  
<https://doi.org/10.1146/annurev.genet.35.102401.090633>
- Mah, J., Linz, G. M., & Hanzel, J. J. (1990). Relative effectiveness of individual sunflower traits for reducing red-winged blackbird depredation. *Crop Protection*, 9(5), 359–362.  
[https://doi.org/10.1016/0261-2194\(90\)90007-T](https://doi.org/10.1016/0261-2194(90)90007-T)
- Malmberg, M. M., Shi, F., Spangenberg, G. C., Daetwyler, H. D., & Cogan, N. O. I. (2018). Diversity and genome analysis of Australian and global oilseed brassica napus L. Germplasm using transcriptomics and whole genome re-sequencing. *Frontiers in Plant Science*, 9(April), 1–14. <https://doi.org/10.3389/fpls.2018.00508>
- Malosetti, M., van Eeuwijk, F. A., Boer, M. P., Casas, A. M., Elía, M., Moralejo, M., Bhat, P. R., Ramsay, L., & Molina-Cano, J. L. (2011). Gene and QTL detection in a three-way barley cross under selection by a mixed model with kinship information using SNPs. *Theoretical and Applied Genetics*, 122(8), 1605–1616. <https://doi.org/10.1007/s00122-011-1558-z>
- Manching, H., & Wisser, R. J. (2021). SPEARS: Standard Performance Evaluation of Ancestral haplotype Reconstruction through Simulation. *Bioinformatics*, 37(6), 868–870. <https://doi.org/10.1093/bioinformatics/btaa749>
- Mandel, J. R., Dechaine, J. M., Marek, L. F., & Burke, J. M. (2011). Genetic diversity and population structure in cultivated sunflower and a comparison to its wild progenitor, *Helianthus annuus* L. *Theoretical and Applied Genetics*, 123(5), 693–704. <https://doi.org/10.1007/s00122-011-1619-3>
- Mandel, Jennifer R., Nambesasan, S., Bowers, J. E., Marek, L. F., Ebert, D., Rieseberg, L. H., Knapp, S. J., & Burke, J. M. (2013). Association Mapping and the Genomic Consequences of Selection in Sunflower. *PLoS Genetics*, 9(3), e1003378. <https://doi.org/10.1371/journal.pgen.1003378>
- Maringolo, C. (2007). *Regiones Cromosómicas Asociadas a Resistencia a Podredumbre Húmeda del Capítulo de Girasol (Sclerotinia sclerotiorum (Lib.) de Bary)*. Universidad Nacional de Mar del Plata, Unidad Integrada Balcarce, Argentina.
- Masuka, B., Magorokosho, C., Olsen, M., Atlin, G. N., Bänziger, M., Pixley, K. V., Vivek, B. S., Labuschagne, M., Matemba-Mutasa, R., Burgenõ, J., Macrobert, J., Prasanna, B. M., Das, B., Makumbi, D., Tarekegne, A., Crossa, J., Zaman-Allah, M., Van Biljon, A., & Cairns, J. E. (2017). Gains in maize genetic improvement in eastern and southern Africa: II. CIMMYT open-pollinated variety breeding pipeline. *Crop Science*, 57(1), 180–191.

- <https://doi.org/10.2135/cropsci2016.05.0408>
- Maurer, A., Draba, V., Jiang, Y., Schnaithmann, F., Sharma, R., Schumann, E., Kilian, B., Reif, J. C., & Pillen, K. (2015). Modelling the genetic architecture of flowering time control in barley through nested association mapping. *BMC Genomics*, *16*(1), 290. <https://doi.org/10.1186/s12864-015-1459-7>
- McMullen, M. D., Kresovich, S., Villeda, H. S., Bradbury, P., Li, H., Sun, Q., Flint-Garcia, S., Thornsberry, J., Acharya, C., Bottoms, C., Brown, P., Browne, C., Eller, M., Guill, K., Harjes, C., Kroon, D., Lepak, N., Mitchell, S. E., Peterson, B., ... Buckler, E. S. (2009). Genetic properties of the maize nested association mapping population. *Science*, *325*(5941), 737–740. <https://doi.org/10.1126/science.1174320>
- Medico, A. P. Del, Cabodevila, V. G., Vitelleschi, M. S., & Pratta, G. R. (2020). Characterization of tomato generations according to a three-way data analysis. *Bragantia*, *79*(1), 8–18. <https://doi.org/10.1590/1678-4499.20190047>
- Mehra, L. K., Cowger, C., Gross, K., & Ojiambo, P. S. (2016). Predicting Pre-planting Risk of Stagonospora nodorum blotch in Winter Wheat Using Machine Learning Models. *Frontiers in Plant Science*, *0*(MAR2016), 390. <https://doi.org/10.3389/FPLS.2016.00390>
- Melo, A. T. O., & Hale, I. (2019). Apparent': A simple and flexible R package for accurate SNP-based parentage analysis in the absence of guiding information. *BMC Bioinformatics*, *20*(1), 1–10. <https://doi.org/10.1186/s12859-019-2662-3>
- Merino, G. A. (2018). *Imputación de genotipos faltantes en datos de secuenciación masiva*. Universidad Nacional de Córdoba.
- Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., & Leisch, F. (2021). *e1071: Misc Functions of the Department of Statistics*. Probability Theory Group (Formerly: E1071). <https://cran.r-project.org/package=e1071>
- Mihaljevic, M. Ž., Maletic, E., Preiner, D., Zduni, G., Bubola, M., Zyprian, E., & Pejic, I. (2020). Genetic Diversity, Populatand Parentage Analysis of Croatian Grapevine Germplasm. *Genes*, *11*(737), 30.
- Miller, J. F., & Fick, G. N. (1978). Adaptation of Reciprocal Full-Sib Selection in Sunflower Breeding Using Gibberellic Acid Induced Male Sterility 1 . *Crop Science*, *18*(1), 161–162. <https://doi.org/10.2135/cropsci1978.0011183x001800010045x>
- Mir, R. R., Reynolds, M., Pinto, F., Khan, M. A., & Bhat, M. A. (2019). High-throughput phenotyping for crop improvement in the genomics era. *Plant Science*, *282*(January), 60–

72. <https://doi.org/10.1016/j.plantsci.2019.01.007>
- Missonnier, H., Jacques, A., Bang, J. S., Daydé, J., & Mirleau-Thebaud, V. (2017). Accounting for biotic spatial variability in fields: Case of resistance screening against sunflower *Verticillium* wilt. *PLoS ONE*, *12*(8), 1–14. <https://doi.org/10.1371/journal.pone.0181050>
- Mohammed, G. H., Noland, T. L., Irving, P. H., Sampson, P. H., Zarco-Tejada, P. J., & Miller, J. R. (2000). Natural and stress-induced effects on leaf spectral reflectance in Ontario species. In *Forest Research Report* (Vol. 156).
- Mol, L., Scholte, K., & Vos, J. (1995). Effects of crop rotation and removal of crop debris on the soil population of two isolates of *Verticillium dahliae*. *Plant Pathology*, *44*(6), 1070–1074. <https://doi.org/10.1111/j.1365-3059.1995.tb02666.x>
- Moncada Herrera, J. (2007). El Análisis Factorial Múltiple: Un Paso más en la Superación de la Dicotomía Cualitativo-Cuantitativo. *6° Congreso Chileno de Antropología*, 734–747. <https://www.aacademica.org>.
- Montecchia, J. (2019). *Identificación y caracterización de fuentes de resistencia genética a la Marchitez Anticipada causada por Verticillium dahliae en Girasol*. Universidad Nacional de Buenos Aires.
- Montecchia, J. F., Fass, M. I., Cerrudo, I., Quiroz, F. J., Nicosia, S., Maringolo, C. A., Rienzo, J. Di, Troglia, C., Hopp, H. E., Escande, A., González, J., Álvarez, D., Heinz, R. A., Lia, V. V., & Paniego, N. B. (2021). On-field phenotypic evaluation of sunflower germplasm: Breeding for broad-spectrum resistance to *Verticillium* leaf mottle and wilt. *Scientific Reports*, 1–14. <https://doi.org/10.1038/s41598-021-91034-4>
- Moreno, M. V. (2010). *DIVERSIDAD GENETICA EN GIRASOL CULTIVADO: análisis de una colección de germoplasma local para su aplicación en programas de mejoramiento*. Universidad de Buenos Aires.
- Moreno, M. V., Nishinakamasu, V., Loray, M. A., Alvarez, D., Gioco, J., Vicario, A., Hopp, H. E., Heinz, R. A., Paniego, N., & Lia, V. V. (2013). Genetic characterization of sunflower breeding resources from Argentina: Assessing diversity in key open-pollinated and composite populations. *Plant Genetic Resources: Characterisation and Utilisation*, *11*(3), 238–249. <https://doi.org/10.1017/S1479262113000075>
- Morgan, A. P., Gatti, D. M., Najarian, M. L., Keane, T. M., Galante, R. J., Pack, A. I., Mott, R., Churchill, G. A., & De Villena, F. P. M. (2017). Structural variation shapes the landscape of recombination in mouse. *Genetics*, *206*(2), 603–619.

<https://doi.org/10.1534/GENETICS.116.197988/-/DC1>

- Moschen, S., Di Rienzo, J. A., Higgins, J., Tohge, T., Watanabe, M., González, S., Rivarola, M., García-García, F., Dopazo, J., Hopp, H. E., Hoefgen, R., Fernie, A. R., Paniego, N., Fernández, P., & Heinz, R. A. (2017). Integration of transcriptomic and metabolic data reveals hub transcription factors involved in drought stress response in sunflower (*Helianthus annuus* L.). *Plant Molecular Biology*, *94*(4–5), 549–564. <https://doi.org/10.1007/s11103-017-0625-5>
- Moschen, S., Higgins, J., Di Rienzo, J. A., Heinz, R. A., Paniego, N., & Fernandez, P. (2016). Network and biosignature analysis for the integration of transcriptomic and metabolomic data to characterize leaf senescence process in sunflower. *BMC Bioinformatics*, *17*(5), 10–12. <https://doi.org/10.1186/s12859-016-1045-2>
- Moschen, S., Marino, J., Nicosia, S., Higgins, J., Alseekh, S., Astigueta, F., Bengoa Luoni, S., Rivarola, M., Fernie, A. R., Blanchet, N., Langlade, N. B., Paniego, N., Fernández, P., & Heinz, R. A. (2019). Exploring gene networks in two sunflower lines with contrasting leaf senescence phenotype using a system biology approach. *BMC Plant Biology*, *19*(1). <https://doi.org/10.1186/s12870-019-2021-6>
- Mott, R., Talbot, C. J., Turri, M. G., Collins, A. C., & Flint, J. (2000). A method for fine mapping quantitative trait loci in outbred animal stocks. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(23), 12649–12654. <https://doi.org/10.1073/pnas.230304397>
- Muranty, H. (1996). Power of tests for quantitative trait loci detection using full-sib families in different schemes. *Heredity* *1996* *76*:2, *76*(2), 156–165. <https://doi.org/10.1038/hdy.1996.23>
- Murovec, J., & Bohanec, B. (2012). Haploids and Doubled Haploids in Plant Breeding. In *Plant Breeding*. InTech. <https://doi.org/10.5772/29982>
- Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., & Buckler, E. S. (2009). Association Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *The Plant Cell*, *21*(8), 2194–2202. <https://doi.org/10.1105/TPC.109.068437>
- Nabwire, S., Suh, H.-K., Kim, M. S., Baek, I., & Cho, B.-K. (2021). Review: Application of Artificial Intelligence in Phenomics. *Sensors 2021*, Vol. 21, Page 4363, *21*(13), 4363. <https://doi.org/10.3390/S21134363>

- Nei, M. (1972). Genetic Distance between Populations. *The American Naturalist*, 106, 283–292.
- Nei, M. (1978). Estimation of average heterozygosity and genetic distance from a small number of individuals. *Genetics*, 89(3), 583–590.
- Nelson, R. (1973). *Breeding plants for disease resistance concepts and applications*. University Park: Pennsylvania State University.
- Nigon, T. J., Mulla, D. J., Rosen, C. J., Cohen, Y., Alchanatis, V., Knight, J., & Rud, R. (2015). Hyperspectral aerial imagery for detecting nitrogen stress in two potato cultivars. *Computers and Electronics in Agriculture*, 112, 36–46. <https://doi.org/10.1016/j.compag.2014.12.018>
- Norman, P., Asfaw, A., Tongoona, P., Danquah, A., Danquah, E., De Koeyer, D., & Asiedu, R. (2018). Can Parentage Analysis Facilitate Breeding Activities in Root and Tuber Crops? *Agriculture MDPI*, 8(95), 24. <https://doi.org/10.3390/agriculture8070095>
- Norman, P. E., Paterne, A. A., Danquah, A., Tongoona, P. B., Danquah, E. Y., Koeyer, D. De, Ikeogu, U. N., Asiedu, R., & Asfaw, A. (2020). Paternity assignment in white guinea yam (*Dioscorea rotundata*) half-sib progenies from polycross mating design using SNP markers. *Plants*, 9(4). <https://doi.org/10.3390/plants9040527>
- Ogawa, D., Yamamoto, E., Ohtani, T., Kanno, N., Tsunematsu, H., Nonoue, Y., Yano, M., Yamamoto, T., & Yonemaru, J.-I. (2018). Haplotype-based allele mining in the Japan-MAGIC rice population. *Scientific Reports*, 8(4379), 1–11. <https://doi.org/10.1038/s41598-018-22657-3>
- Olsson, S., & Nordbring-Hertz, B. (1985). Microsclerotial germination of *Verticillium dahliae* as affected by rape rhizosphere. In *FEMS Microbiology Ecology* (Vol. 31). <https://academic.oup.com/femsec/article/1/5/293/544410>
- Ongom, P. O., & Ejeta, G. (2018). mating design and genetic structure of a multi-parent advanced generation intercross (MAGIC) population of sorghum (*Sorghum bicolor* (L.) moench). *G3: Genes, Genomes, Genetics*, 8(1), 331–341. <https://doi.org/10.1534/g3.117.300248>
- Oøvergaard, S. I., Isaksson, T., Kvaalc, K., & Korsæth, A. (2010). Comparisons of two hand-held, multispectral field radiometers and a hyperspectral airborne imager in terms of predicting spring wheat grain yield and quality by means of powered partial least squares regression. *Journal of Near Infrared Spectroscopy*, 18(4), 247–261.

<https://doi.org/10.1255/jnirs.892>

- Oraguzie, N. C., & Wilcox, P. L. (2007). An Overview of Association Mapping. In *Association Mapping in Plants* (pp. 1–9). Springer New York. [https://doi.org/10.1007/978-0-387-36011-9\\_1](https://doi.org/10.1007/978-0-387-36011-9_1)
- Pabuayon, I. L. B., Sun, Y., Guo, W., & Ritchie, G. L. (2019). High-throughput phenotyping in cotton: a review. *Journal of Cotton Research*, 2(1), 1–9. <https://doi.org/10.1186/s42397-019-0035-0>
- Pádua, L., Vanko, J., Hruška, J., Adão, T., Sousa, J. J., Peres, E., & Morais, R. (2017). UAS, sensors, and data processing in agroforestry: a review towards practical applications. *https://doi.org/10.1080/01431161.2017.1297548*, 38(8–10), 2349–2391. <https://doi.org/10.1080/01431161.2017.1297548>
- Pagès, J. (2014). *Multiple Factor Analysis by Example Using R M*. Chapman & Hall/CRC Press.
- Pajares, G. (2015). Overview and current status of remote sensing applications based on unmanned aerial vehicles (UAVs). *Photogrammetric Engineering and Remote Sensing*, 81(4), 281–329. <https://doi.org/10.14358/PERS.81.4.281>
- Palacio, F., Apodaca, M., & Crisci, J. (2020). *ANÁLISIS MULTIVARIADO PARA DATOS BIOLÓGICOS Teoría y su aplicación utilizando el lenguaje R*. Vazquez Mazzini Editores.
- Paniego, N., Echaide, M., Muñoz, M., Fernández, L., Torales, S., Faccio, P., Fuxan, I., Carrera, M., Zandomeni, R., Suárez, E. Y., & Hopp, H. E. (2002). Microsatellite isolation and characterization in sunflower (*Helianthus annuus* L.). *Genome*, 45(1), 34–43. <https://doi.org/10.1139/g01-120>
- Paradis, E., & Schliep, K. (2019). ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics*, 35, 526–528.
- Parfitt, D. (1984). Relationship of Morphological Plant Characteristics of Sunflower To Bird Feeding. *Canadian Journal of Plant Science*, 64(1), 37–42. <https://doi.org/10.4141/cjps84-006>
- Park, C., Marx, G., Moon, Y., Wiesenborn, D., Chang, K., & Hofman, V. (1997). Alternatives Uses of Sunflower. In A. A. Schneiter (Ed.), *Sunflower Technology and Production* (pp. 765–807). ASA, CSSA, SSSA.
- Pascual, L., Desplat, N., Huang, B. E., Desgroux, A., Bruguier, L., Bouchet, J., Le, Q. H., Chauchard, B., Verschave, P., & Causse, M. (2015). *Potential of a tomato MAGIC*

- population to decipher the genetic control of quantitative traits and detect causal variants in the resequencing era.* 565–577. <https://doi.org/10.1111/pbi.12282>
- Patterson, N., Price, A. L., & Reich, D. (2006). Population Structure and Eigenanalysis. *PLoS Genetics*, 2(12), e190. <https://doi.org/10.1371/JOURNAL.PGEN.0020190>
- Paux, E., Sourdille, P., Mackay, I., & Feuillet, C. (2012). Sequence-based marker development in wheat: Advances and applications to breeding. In *Biotechnology Advances* (Vol. 30, Issue 5, pp. 1071–1088). Elsevier. <https://doi.org/10.1016/j.biotechadv.2011.09.015>
- Peakall, R., & Smouse, P. E. (2012). GenAIEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research--an update. *Bioinformatics*, 28(19), 2537–2539. <https://doi.org/10.1093/bioinformatics/bts460>
- Pedraza, M. V., Pereyra, V. R., & Escande, A. R. (2004). Infection Courts and Length of Susceptible Period Related to Sunflower Head Rot (*Sclerotinia Sclerotiorum*) Resistance. *Helia*, 27(40), 171–182. <https://doi.org/10.2298/hel0440171p>
- Peerbolte, R., & Paleman, J. (1998). The CARTISOL sunflower RFLP map (146 loci) extended with 291 RFLP markers. *18th Sunflower Research Forum*, 11–12.
- Pegg, G., & Brady, B. (2002). *Verticillium Wilts*. CABI.
- Peluffo, L., Lia, V., Troglia, C., Maringolo, C., Norma, P., Escande, A., Esteban Hopp, H., Lytovchenko, A., Fernie, A. R., Heinz, R., & Carrari, F. (2010). Metabolic profiles of sunflower genotypes with contrasting response to *Sclerotinia sclerotiorum* infection. *Phytochemistry*, 71(1), 70–80. <https://doi.org/10.1016/j.phytochem.2009.09.018>
- Pembleton, L. W., Cogan, N. O. I., & Forster, J. W. (2013). StAMPP: an R package for calculation of genetic differentiation and structure of mixed-ploidy level populations. *Molecular Ecology Resources*, 13(5), 946–952. <https://doi.org/10.1111/1755-0998.12129>
- Peña, J. M., Torres-Sánchez, J., Serrano-Pérez, A., de Castro, A. I., & López-Granados, F. (2015). Quantifying efficacy and limits of unmanned aerial vehicle (UAV) technology for weed seedling detection as affected by sensor resolution. *Sensors (Switzerland)*, 15(3), 5609–5626. <https://doi.org/10.3390/s150305609>
- Pereyra, V., & Escande, A. (1994). *Manual de reconocimiento de enfermedades del Girasol en la Argentina*. INTA.
- Pereyra, V., Quiroz, F., Agüero, M., & Escande, A. (1999). Relación del rendimiento de girasol con la intensidad de síntomas provocados por *V. dahliae*. *X Jornadas Fitosanitarias Argentinas*, 35.

- Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012). Double Digest RADseq: An Inexpensive Method for De Novo SNP Discovery and Genotyping in Model and Non-Model Species. *PLoS ONE*, 7(5), e37135. <https://doi.org/10.1371/journal.pone.0037135>
- Phillips, R. L. (2010). Mobilizing science to break yield barriers. *Crop Science*, 50, S-99-S-108. <https://doi.org/10.2135/cropsci2009.09.0525>
- Pilorgé, E. (2020). Sunflower in the global vegetable oil system: situation, specificities and perspectives. *Oilseeds and Fats, Crops and Lipids*, 27(1), 34. <https://doi.org/10.1051/ocl/2020028>
- Poland, J. A., & Rife, T. W. (2012a). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*, 5(3). <https://doi.org/10.3835/plantgenome2012.05.0005>
- Poland, J. A., & Rife, T. W. (2012b). Genotyping-by-Sequencing for Plant Breeding and Genetics. *The Plant Genome*, 5(3). <https://doi.org/10.3835/plantgenome2012.05.0005>
- Pootakham, W., Jomchai, N., Ruang-areerate, P., Shearman, J. R., Sonthirod, C., Sangsrakru, D., Tragoonrung, S., & Tangphatsornruang, S. (2015). Genome-wide SNP discovery and identification of QTL associated with agronomic traits in oil palm using genotyping-by-sequencing (GBS). *Genomics*, 105(5–6), 288–295. <https://doi.org/10.1016/j.ygeno.2015.02.002>
- Postman, J., Volk, G., & Aldwinckle, H. (2010). Standardized plant disease evaluations will enhance resistance gene discovery. *HortScience*, 45(9), 1317–1320. <https://doi.org/10.21273/hortsci.45.9.1317>
- Pourazar, H., Samadzadegan, F., & Dadrass Javan, F. (2019). Aerial multispectral imagery for plant disease detection: radiometric calibration necessity assessment. *European Journal of Remote Sensing*, 52(sup3), 17–31. <https://doi.org/10.1080/22797254.2019.1642143>
- Price, A. H. (2006). Believe it or not, QTLs are accurate! *Trends in Plant Science*, 11(5), 213–216. <https://doi.org/10.1016/j.tplants.2006.03.006>
- Prioletta, S., Pereyra, V., Pedraza, M., & Escande, A. (1995). Efecto de la exposición del esclerocio a baja temperatura y de las condiciones ambientales durante la carpogénesis sobre la producción de apotecios de *Sclerotinia sclerotiorum*. *Reunión Nacional de Oleaginosas. Pergamino. Buenos Aires. Argentina*.
- Pustovoit, G. (1978). *Main results of sunflower selection for group resistance in VNIIMK*.
- Putt, E. (1978). *Sunflower Science and Technology* (J. Carter (ed.)). American Society of

- Agronomy, Crop Science Society of America, Soil Science Society of America.
- Putt, E. (1997). Early History of Sunflower. In A. A. Schneiter (Ed.), *Sunflower Technology and Production* (pp. 1–20).
- Qi, L., Long, Y., Talukder, Z. I., Seiler, G. J., Block, C. C., & Gulya, T. J. (2016). Genotyping-by-sequencing uncovers the introgression alien segments associated with sclerotinia basal stalk rot resistance from wild species-i. *Helianthus argophyllus* and *H. petiolaris*. *Frontiers in Genetics*, 7(DEC), 219. <https://doi.org/10.3389/fgene.2016.00219>
- Quiroz, F., Pereyra, V., & Escande, A. (1998). Comportamiento de cultivares de girasol frente a infecciones de *Sclerotinia Sclerotiorum* inducidas por pulverización de ascosporas o por inóculo natural. *Reunión Nacional de Oleaginosos*, 35.
- Quiroz, F., Molas, A. C., Rojo, R., Pérez Fernández, J., & Escande, A. (2008). Effects of no tillage and genetic resistance on sunflower wilt by *Verticillium dahliae*. *Soil & Tillage Research*, 99, 66–75. <https://doi.org/10.1016/j.still.2007.12.007>
- R Core Team. (2020). *R: A Language and Environment for Statistical Computing*. <https://www.r-project.org/>
- Rabiner, L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *IEEE* 77, 257–286.
- Rachid Al-Chaarani, G., Gentzbittel, L., Huang, X. Q., & Sarrafi, A. (2004). Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). *Theoretical and Applied Genetics*, 109(7), 1353–1360. <https://doi.org/10.1007/s00122-004-1770-1>
- Raghavan, C., Mauleon, R., Lacorte, V., Jubay, M., Zaw, H., Bonifacio, J., Singh, R. K., Emma Huang, B., & Leung, H. (2017). Approaches in characterizing genetic structure and mapping in a rice multiparental population. *G3: Genes, Genomes, Genetics*, 7(6), 1721–1730. <https://doi.org/10.1534/g3.117.042101>
- Rakshit, S., Rakshit, A., & Patil, J. (2012). Multiparent intercross populations in analysis of quantitative traits. *J Genet*, 91, 111–117. <https://doi.org/https://doi.org/10.1007/s12041-012-0144-8>
- Ram, R., Mehta, M., Balmer, L., Gatti, D. M., & Morahan, G. (2014). Rapid identification of major-effect genes using the collaborative cross. *Genetics*, 198(1), 75–86. <https://doi.org/10.1534/genetics.114.163014>
- Rasheed, A., Hao, Y., Xia, X., Khan, A., Xu, Y., Varshney, R. K., & He, Z. (2017). Crop

- Breeding Chips and Genotyping Platforms: Progress, Challenges, and Perspectives. In *Molecular Plant* (Vol. 10, Issue 8, pp. 1047–1064). Cell Press. <https://doi.org/10.1016/j.molp.2017.06.008>
- Rawson, H. M., Dunstone, R. L., Long, M. J., & Begg, J. E. (1984). Canopy development, light interception and seed production in sunflower as influenced by temperature and radiation. *Australian Journal of Plant Physiology*, 11(4), 255–265. <https://doi.org/10.1071/PP9840255>
- Rebai, A., & Goffinet, B. (1993). Power of tests for QTL detection using replicated progenies derived from a diallel cross. *Theoretical and Applied Genetics*, 86(8), 1014–1022. <https://doi.org/10.1007/BF00211055>
- Reynolds, D., Baret, F., Welcker, C., Bostrom, A., Ball, J., Cellini, F., Lorence, A., Chawade, A., Khafif, M., Noshita, K., Mueller-Linow, M., Zhou, J., & Tardieu, F. (2019). What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Science*, 282(December 2017), 14–22. <https://doi.org/10.1016/j.plantsci.2018.06.015>
- Reynolds, M., Chapman, S., Crespo-Herrera, L., Molero, G., Mondal, S., Pequeno, D. N. L., Pinto, F., Pinera-Chavez, F. J., Poland, J., Rivera-Amado, C., Saint Pierre, C., & Sukumaran, S. (2020). Breeder friendly phenotyping. *Plant Science*, 295(December 2019), 110396. <https://doi.org/10.1016/j.plantsci.2019.110396>
- Riday, H. (2011). Paternity Testing A Non-Linkage Based Marker-Assisted Selection Scheme for Outbred Forage. *Crop Science*, 51, 631–641.
- Riday, H., Johnson, D. W., Heyduk, K., Raasch, J. A., Darling, M. E., & Sandman, J. M. (2013). Paternity testing in an autotetraploid alfalfa breeding polycross. *Euphytica*, 194(3), 335–349. <https://doi.org/10.1007/s10681-013-0938-7>
- Rieseberg, L. (1998). Genetic Mapping as a Tool for Studying Speciation. In D. E. Soltis, P. S. Soltis, & J. J. Doyle (Eds.), *Molecular Systematics of Plants II* (pp. 459–487). Springer, Boston, MA. [https://doi.org/https://doi.org/10.1007/978-1-4615-5419-6\\_16](https://doi.org/https://doi.org/10.1007/978-1-4615-5419-6_16)
- Rieseberg, L. H. (1991). Homoploid Reticulate Evolution in Helianthus (Asteraceae): Evidence from Ribosomal Genes. *American Journal of Botany*, 78(9), 1218. <https://doi.org/10.2307/2444926>
- Rieseberg, L. H., Choi, H., Chan, R., & Spore, C. (1993). Genomic map of a diploid hybrid species. *Heredity*, 70(3), 285–293. <https://doi.org/10.1038/hdy.1993.41>
- Risch, N., & Merikangas, K. (1996). The future of genetic studies of complex human diseases.

- In *Science* (Vol. 273, Issue 5281, pp. 1516–1517). American Association for the Advancement of Science. <https://doi.org/10.1126/science.273.5281.1516>
- Rochette, N. C., Rivera-Colón, A. G., & Catchen, J. M. (2019). Stacks 2: Analytical methods for paired-end sequencing improve RADseq-based population genomics. *Molecular Ecology*, *28*(21), 4737–4754. <https://doi.org/10.1111/mec.15253>
- Rodríguez-Álvarez, M. X., Boer, M. P., van Eeuwijk, F. A., & Eilers, P. H. C. (2018). Correcting for spatial heterogeneity in plant breeding experiments with P-splines. *Spatial Statistics*, *23*, 52–71. <https://doi.org/10.1016/j.spasta.2017.10.003>
- Rudolf-Pilih, K., Petkovšek, M., Jakše, J., Štajner, N., Murovec, J., & Bohanec, B. (2019). Proposal of a New Hybrid Breeding Method Based on Genotyping, Inter-Pollination, Phenotyping and Paternity Testing of Selected Elite F1 Hybrids. *Frontiers in Plant Science*, *10*(September), 1–12. <https://doi.org/10.3389/fpls.2019.01111>
- Sala, C. A., Bulos, M., Altieri, E., & Ramos, M. L. (2012). Sunflower: Improving Crop Productivity and Abiotic Stress Tolerance. In *Improving Crop Resistance to Abiotic Stress* (Vol. 2, pp. 1203–1249). Wiley-VCH. <https://doi.org/10.1002/9783527632930.ch47>
- Salamí, E., Barrado, C., & Pastor, E. (2014). UAV flight experiments applied to the remote sensing of vegetated areas. *Remote Sensing*, *6*(11), 11051–11081. <https://doi.org/10.3390/rs61111051>
- Sallam, A., & Martsch, R. (2015). Association mapping for frost tolerance using multi-parent advanced generation inter-cross (MAGIC) population in faba bean (*Vicia faba* L.). *Genetica*, *143*(4), 501–514. <https://doi.org/10.1007/s10709-015-9848-z>
- Sambrook, J., Fritsch, E., & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor laboratory press. Cold Spring Harbor.
- Sandhu, K. S., Lozada, D. N., Zhang, Z., Pumphrey, M. O., & Carter, A. H. (2021). Deep Learning for Predicting Complex Traits in Spring Wheat Breeding Program. *Frontiers in Plant Science*, *0*, 2084. <https://doi.org/10.3389/FPLS.2020.613325>
- Sandino, J., Pegg, G., Gonzalez, F., & Smith, G. (2018). Aerial mapping of forests affected by pathogens using UAVs, hyperspectral sensors, and artificial intelligence. *Sensors (Switzerland)*, *18*(4), 1–17. <https://doi.org/10.3390/s18040944>
- Sankaran, S., Khot, L. R., Espinoza, C. Z., Jarolmasjed, S., Sathuvalli, V. R., Vandemark, G. J., Miklas, P. N., Carter, A. H., Pumphrey, M. O., Knowles, R. R. N., & Pavek, M. J. (2015). Low-altitude, high-resolution aerial imaging systems for row and field crop phenotyping:

- A review. In *European Journal of Agronomy* (Vol. 70, pp. 112–123). Elsevier. <https://doi.org/10.1016/j.eja.2015.07.004>
- Sannemann, W., Huang, B. E., Mathew, B., & Léon, J. (2015). Multi-parent advanced generation inter-cross in barley: high-resolution quantitative trait locus mapping for flowering time as a proof of concept. *Molecular Breeding*, 35(3). <https://doi.org/10.1007/s11032-015-0284-7>
- Satir, O., & Berberoglu, S. (2016). Crop yield prediction under soil salinity using satellite derived vegetation indices. *Field Crops Research*, 192, 134–143. <https://doi.org/10.1016/J.FCR.2016.04.028>
- Sausse, C., & Lévy, M. (2021). Bird damage to sunflower: International situation and prospects. *OCL - Oilseeds and Fats, Crops and Lipids*, 28, 34. <https://doi.org/10.1051/ocl/2021020>
- Scheben, A., Batley, J., & Edwards, D. (2017). Genotyping-by-sequencing approaches to characterize crop genomes: choosing the right tool for the right application. *Plant Biotechnology Journal*, 15(2), 149–161. <https://doi.org/10.1111/pbi.12645>
- Schmidt, P., Hartung, J., Bennewitz, J., & Hans-Peter, P. (2019). Heritability in plant breeding on a genotype-difference basis. *Genetics*, 212(4), 991–1008. <https://doi.org/10.1534/genetics.119.302134>
- Schnathorst, W. C. (1981). Life cycle and epidemiology of *Verticillium*. In M. E. Mace, A. A. Bell, & C. H. Beckman (Eds.), *Fungal Wilt Diseases of Plants* (pp. 81–111). New York: Academic Press.
- Schneiter, A. A., & Miller, J. F. (1981). Description of Sunflower Growth Stages 1. *Crop Science*, 21(6), 901–903. <https://doi.org/10.2135/cropsci1981.0011183x002100060024x>
- Schreiber, L., & Green, R. (1963). Effect of root exudates on germination of conidia and microsclerotia of *Verticillium albo-atrum* inhibited by the soil fungistatic principle. *Phytopathology*, 53, 260–264.
- Scott, M. F., Ladejobi, O., Amer, S., Bentley, A. R., Biernaskie, J., Boden, S. A., Clark, M., Dell, M., Laura, A., Carla, E. D., Nick, V. F., Gardner, K. A., Mackay, I. J., Sullivan, D. O., Manish, L. P., & Scott, M. F. (2020). Multi-parent populations in crops: a toolbox integrating genomics and genetic mapping with breeding. *Heredity*. <https://doi.org/10.1038/s41437-020-0336-6>
- Seiler, G. (1997). Anatomy and morphology of sunflower. In A. Schneiter (Ed.), *Sunflower Technology and production* (pp. 67–111). American Society of Agronomy-Crop Science

- Society of America-Soil Science Society of America Publishers. Madison.
- Seiler, G. & Rogers, C. (1987). Influence of sunflower morphological characteristics on achene depredation by birds. *Agriculture, Ecosystems and Environment*, 20(1), 59–70. [https://doi.org/10.1016/0167-8809\(87\)90028-4](https://doi.org/10.1016/0167-8809(87)90028-4)
- Seiler, Gerald, & Rieseberg, L. (1997). Systematics, Origin and Germplasm Resources of the Wild and Domesticated Sunflower. In A. A. Schneiter (Ed.), *Sunflower Technology and Production*.
- Serrano, I. B. (2020). *High-throughput field phenotyping using a drone with RGB imagery Exploiting the spectral , spatial* (Issue October). Universidad de KU Leuven.
- Shah, D. a, & Madden, L. V. (2004). Nonparametric analysis of ordinal data. *Phytopathology*, 94(1), 33–43.
- Shaner, G., & Finney, R. E. (1977). The Effect of Nitrogen Fertilization on the Expression of Slow-Mildewing Resistance in Knox Wheat. *Phytopathology*, 67, 1051–1056.
- Shi, Y., Alex Thomasson, J., Murray, S. C., Ace Pugh, N., Rooney, W. L., Shafian, S., Rajan, N., Rouze, G., Morgan, C. L. S., Neely, H. L., Rana, A., Bagavathiannan, M. V., Henrickson, J., Bowden, E., Valasek, J., Olsenholler, J., Bishop, M. P., Sheridan, R., Putman, E. B., ... Yang, C. (2016). Unmanned Aerial Vehicles for High-Throughput Phenotyping and Agronomic Research. *PLOS ONE*, 11(7), e0159781. <https://doi.org/10.1371/JOURNAL.PONE.0159781>
- Silva, P. S. L. e, Tomaz, F. L. de S., Siqueira, P. L. de O. F., Silva, P. I. B. e, & Lima, L. A. C. de. (2019). Yield loss in sunflower cultivars due to bird attack 1. *Revista Ciência Agronômica*, 50(1), 114–122. <https://doi.org/10.5935/1806-6690.20190014>
- Singh, A., Ganapathysubramanian, B., Singh, A. K., & Sarkar, S. (2016). Feature Review Machine Learning for High-Throughput Stress Phenotyping in Plants. *Trends in Plant Science*, 21(2). <https://doi.org/10.1016/j.tplants.2015.10.015>
- Sinoquet, H., & Andrieu, B. (1993). The geometrical structure of plant canopies: characterisation and direct measurement methods. In C. Varlet-Grancher, R. Bonhomme, & H. Sinoquet (Eds.), *Crop structure and light microclimate* (pp. 131–158). INRA Editions.
- Skoric, D., Mihaljcevic, M., & Marinkovic, R. (1978). New sources of fertility restoration (Rf genes) and Downy mildew (PI genes) in sunflowers. *VIII International Sunflower Conference*, 423–426.
- Smith, B. D., & Yarnell, R. A. (2009). Initial formation of an indigenous crop complex in Eastern

- North America at 3800 B.P. *Proceedings of the National Academy of Sciences of the United States of America*, 106(16), 6561–6566. <https://doi.org/10.1073/pnas.0901846106>
- Snijders, T., & Bosker, R. (2012). Discrete Dependent Variables. In *Multilevel Analysis an Introduction to Basic and Advanced Multilevel Modeling* (2<sup>o</sup>). Sage Publishers.
- Stadlmeier, M., Hartl, L., & Mohler, V. (2018). Usefulness of a Multiparent Advanced Generation Intercross Population With a Greatly Reduced Mating Design for Genetic Studies in Winter Wheat. *Frontiers in Plant Science*, 9, 1825. <https://doi.org/10.3389/fpls.2018.01825>
- Staton, S. E., Bakken, B. H., Blackman, B. K., Chapman, M. A., Kane, N. C., Tang, S., Ungerer, M. C., Knapp, S. J., Rieseberg, L. H., & Burke, J. M. (2012). The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. In *Plant Journal* (Vol. 72, Issue 1, pp. 142–153). John Wiley & Sons, Ltd. <https://doi.org/10.1111/j.1365-313X.2012.05072.x>
- Stevens, P. (2010). *Angiosperm phylogeny Website*. <http://www.mobot.org/mobot/research/apweb/>.
- Su, J., Liu, C., Coombes, M., Hu, X., Wang, C., Xu, X., Li, Q., Guo, L., & Chen, W. H. (2018). Wheat yellow rust monitoring by learning from multispectral UAV aerial imagery. *Computers and Electronics in Agriculture*, 155, 157–166. <https://doi.org/10.1016/J.COMPAG.2018.10.017>
- Su, T. C., & Chou, H. T. (2015). Application of multispectral sensors carried on unmanned aerial vehicle (UAV) to trophic state mapping of small reservoirs: A case study of Tain-Pu reservoir in Kinmen, Taiwan. *Remote Sensing*, 7(8), 10078–10097. <https://doi.org/10.3390/rs70810078>
- Subbaiyan, G. K., Waters, D. L. E., Katiyar, S. K., Sadananda, A. R., Vaddadi, S., & Henry, R. J. (2012). Genome-wide DNA polymorphisms in elite indica rice inbreds discovered by whole-genome sequencing. *Plant Biotechnology Journal*, 10, 623–634. <https://doi.org/doi:10.1111/j.1467-7652.2011.00676.x>
- Sugiura, R., Tsuda, S., Tsuji, H., & Murakami, N. (2018). Virus-Infected Plant Detection in Potato Seed Production Field by UAV Imagery. *ASABE Annual International Meeting*, 2–6.
- Svenson, K. L., Gatti, D. M., Valdar, W., Welsh, C. E., Cheng, R., Chesler, E. J., Palmer, A. A., McMillan, L., & Churchill, G. A. (2012). High-resolution genetic mapping using the

- mouse Diversity Outbred population. *Genetics*, 190(2), 437–447. <https://doi.org/10.1534/genetics.111.132597>
- Talia, P., Greizerstein, E. J., Hopp, H. E., Paniego, N., Poggio, L., & Heinz, R. A. (2011). Detection of single copy sequences using BAC-FISH and C-PRINS techniques in sunflower chromosomes. *Biocell*, 35(1), 19–28. <https://doi.org/10.32604/biocell.2011.35.019>
- Talia, P., Nishinakamasu, V., Hopp, H. E., Heinz, R. A., & Paniego, N. (2010). Genetic mapping of EST-SSRs, SSR and InDels to improve saturation of genomic regions in a previously developed sunflower map. *Electronic Journal of Biotechnology*, 13(6). <https://doi.org/10.2225/vol13-issue6-fulltext-14>
- Talukder, Z. I., Gong, L., Hulke, B. S., Pegadaraju, V., Song, Q., Schultz, Q., & Qi, L. (2014). A high-density SNP map of sunflower derived from RAD-sequencing facilitating fine-mapping of the rust resistance gene R12. *PLoS ONE*, 9(7), 98628. <https://doi.org/10.1371/journal.pone.0098628>
- Talukder, Z. I., Underwood, W., Ma, G., Seiler, G. J., Misar, C. G., Cai, X., & Qi, L. (2020). Genetic dissection of phomopsis stem canker resistance in cultivated sunflower using high density SNP linkage map. *International Journal of Molecular Sciences*, 21(4), 1497. <https://doi.org/10.3390/ijms21041497>
- Talukder, Z. I., Underwood, W., Misar, C. G., Seiler, G. J., Liu, Y., Li, X., Cai, X., & Qi, L. (2021). Unraveling the Sclerotinia Basal Stalk Rot Resistance Derived From Wild Helianthus argophyllus Using a High-Density Single Nucleotide Polymorphism Linkage Map. *Frontiers in Plant Science*, 11, 2240. <https://doi.org/10.3389/fpls.2020.617920>
- 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(8), 1124–1136. <https://doi.org/10.1007/s00122-002-0989-y>
- Tang, Shunxue, Leon, A., Bridges, W. C., & Knapp, S. J. (2006). Quantitative trait loci for genetically correlated seed traits are tightly linked to branching and pericarp pigment loci in sunflower. *Crop Science*, 46(2), 721–734. <https://doi.org/10.2135/cropsci2005.0006-7>
- Taylor, J. (2018). *Statistics for the Australian grains industry technical report series efficient linkage map construction using R/ASMap*.
- Team QGIS. (2021). *QGIS geographic information system. Open Source Geospatial Found Proj*. <https://www.qgis.org/en/site/index.html>

- Terzic, S., Boniface, M., Marek, L., Alvarez, D., Baumann, K., Gavrilova, V., Joita-pacureanu, M., Sujatha, M., Valkova, D., Velasco, L., Langlade, N., Muños, S., Rieseberg, L., Hulke, B. S., Seiler, G., & Vear, F. (2020). Gene banks for wild and cultivated sunflower genetic resources. *OCL - Oilseeds and Fats, Crops and Lipids*.
- Therneau, T., & Atkinson, B. (2019). *rpart: Recursive Partitioning and Regression Trees*. <https://cran.r-project.org/package=rpart>
- Thrall, P. H., & Burton, J. J. (1997). Host-Pathogen Dynamics in a Metapopulation Context: The Ecological and Evolutionary Consequences of Being Spatial. *Journal of Ecology*, 85(6), 743–753.
- Timpson, N. J., Greenwood, C. M. T., Soranzo, N., Lawson, D. J., & Richards, J. B. (2018). Genetic architecture: The shape of the genetic contribution to human traits and disease. In *Nature Reviews Genetics* (Vol. 19, Issue 2, pp. 110–124). Nature Publishing Group. <https://doi.org/10.1038/nrg.2017.101>
- Tong, H., & Nikoloski, Z. (2021). Machine learning approaches for crop improvement: Leveraging phenotypic and genotypic big data. *Journal of Plant Physiology*, 257, 153354. <https://doi.org/10.1016/J.JPLPH.2020.153354>
- Torkamaneh, D., Boyle, B., & Belzile, F. (2018). Efficient genome-wide genotyping strategies and data integration in crop plants. In *Theoretical and Applied Genetics* (Vol. 131, Issue 3, pp. 499–511). Springer Verlag. <https://doi.org/10.1007/s00122-018-3056-z>
- Trapani, N., Hall, A. J., Sadras, V. O., & Vilella, F. (1992). Ontogenetic changes in radiation use efficiency of sunflower (*Helianthus annuus* L.) crops. *Field Crops Research*, 29(4), 301–316. [https://doi.org/10.1016/0378-4290\(92\)90032-5](https://doi.org/10.1016/0378-4290(92)90032-5)
- Tripodi, P. (2021). Methods of Development of Biparental Mapping Populations in Horticultural Crops. In P. Tripodi (Ed.), *Crop Breeding*. Humana, New York, NY.
- Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, 8(2), 127–150. [https://doi.org/10.1016/0034-4257\(79\)90013-0](https://doi.org/10.1016/0034-4257(79)90013-0)
- Turner, T. R., James, E. K., & Poole, P. S. (2013). The plant microbiome. In *Genome Biology* (Vol. 14, Issue 6, pp. 1–10). BioMed Central Ltd. <https://doi.org/10.1186/gb-2013-14-6-209>
- Ustin, S. L., Gitelson, A. A., Jacquemoud, S., Schaepman, M., Asner, G. P., Gamon, J. A., & Zarco-Tejada, P. (2009). Retrieval of foliar information about plant pigment systems from

- high resolution spectroscopy. *Remote Sensing of Environment*, 113(SUPPL. 1), S67–S77. <https://doi.org/10.1016/J.RSE.2008.10.019>
- Valdar, W., Flint, J., & Mott, R. (2006a). Simulating the collaborative cross: Power of quantitative trait loci detection and mapping resolution in large sets of recombinant inbred strains of mice. *Genetics*, 172(3), 1783–1797. <https://doi.org/10.1534/genetics.104.039313>
- Valdar, W., Flint, J., & Mott, R. (2006b). Simulating the Collaborative Cross: Power of Quantitative Trait Loci Detection and Mapping Resolution in Large Sets of Recombinant Inbred Strains of Mice. *Genetics*, 172(3), 1783–1797. <https://doi.org/10.1534/genetics.104.039313>
- Valdar, W., Solberg, L. C., Gauguier, D., Burnett, S., Klenerman, P., Cookson, W. O., Taylor, M. S., Rawlins, J. N. P., Mott, R., & Flint, J. (2006). Genome-wide genetic association of complex traits in heterogeneous stock mice. *Nature Genetics*, 38(8), 879–887. <https://doi.org/10.1038/ng1840>
- Van Alfen, N. K. (1989). Reassessment of plant wilt toxins. *Ann. Rev. Phytopathol.*, 27, 533–550.
- Vayssade, J.-A., Paoli, J.-N., Gée, C., & Jones, G. (2021). DeepIndices: Remote Sensing Indices Based on Approximation of Functions through Deep-Learning, Application to Uncalibrated Vegetation Images. *Remote Sensing 2021, Vol. 13, Page 2261, 13(12)*, 2261. <https://doi.org/10.3390/RS13122261>
- Vear, F., Serieys, H., Petit, A., Serre, F., Boudon, J. P., Roche, S., Walser, P., & Tourvieille de Labrouhe, D. (2008). Origins of major genes for downy mildew resistance in sunflower. *17th International Sunflower Conference, June*.
- Velazco, J. G., María, ·, Rodríguez-Álvarez, X., Boer, M. P., David, ·, Jordan, R., Paul, ·, Eilers, H. C., Malosetti, M., & Van Eeuwijk, F. A. (2017). Modelling spatial trends in sorghum breeding field trials using a two-dimensional P-spline mixed model. *Theoretical and Applied Genetics*, 130, 1375–1392. <https://doi.org/10.1007/s00122-017-2894-4>
- Venables, W. N., & Ripley, B. D. (2002). *Modern Applied Statistics with S* (Fourth). Springer. <https://www.stats.ox.ac.uk/pub/MASS4/>
- Verbyla, Arunas P., & Cullis, B. R. (2012). Multivariate whole genome average interval mapping: QTL analysis for multiple traits and/or environments. *Theoretical and Applied Genetics*, 125(5), 933–953. <https://doi.org/10.1007/s00122-012-1884-9>

- Verbyla, Arunas P., Cullis, B. R., & Thompson, R. (2007). The analysis of QTL by simultaneous use of the full linkage map. *Theoretical and Applied Genetics*, *116*(1), 95–111. <https://doi.org/10.1007/s00122-007-0650-x>
- Verbyla, Arūnas P., George, A. W., Cavanagh, C. R., & Verbyla, K. L. (2014). Whole-genome QTL analysis for MAGIC. *Theoretical and Applied Genetics*, *127*(8), 1753–1770. <https://doi.org/10.1007/s00122-014-2337-4>
- Verbyla, Arunas P., Taylor, J. D., & Verbyla, K. L. (2012). RWGAIM: An efficient high-dimensional random whole genome average (QTL) interval mapping approach. *Genetics Research*, *94*(6), 291–306. <https://doi.org/10.1017/S0016672312000493>
- Verhoeven, K. J. F., Jannink, J.-L., & McIntyre, L. M. (2006). Using mating designs to uncover QTL and the genetic architecture of complex traits. *Heredity*, *96*(2), 139–149. <https://doi.org/10.1038/sj.hdy.6800763>
- Vidal, M., Plomion, C., Harvengt, L., Raffin, A., Boury, C., & Bouffier, L. (2015). Paternity recovery in two maritime pine polycross mating designs and consequences for breeding. *Tree Genetics and Genomes*, *11*(5). <https://doi.org/10.1007/s11295-015-0932-4>
- Visscher, P. M., Brown, M. A., McCarthy, M. I., & Yang, J. (2012). Five years of GWAS discovery. In *American Journal of Human Genetics* (Vol. 90, Issue 1, pp. 7–24). Elsevier. <https://doi.org/10.1016/j.ajhg.2011.11.029>
- Vitelleschi, M., & Chavasa, V. (2015). Divulgación Análisis factorial múltiple para la caracterización de variedades de trigo pan en diferentes ambientes. *Revista FABICIB*, *19*, 113–120.
- Vitti, D., & Zuil, S. (2012). Evaluaciones del daño generado por aves en girasol. *Voces y Ecos*, *29*, 11–13.
- Wada, T., Oku, K., Nagano, S., Isobe, S., Suzuki, H., Mori, M., Takata, K., Hirata, C., Shimomura, K., Tsubone, M., Katayama, T., Hirashima, K., Uchimura, Y., Ikegami, H., Sueyoshi, T., Obu, K. I., Hayashida, T., & Shibato, Y. (2017). Development and characterization of a strawberry MAGIC population derived from crosses with six strawberry cultivars. *Breeding Science*, *67*(4), 370–381. <https://doi.org/10.1270/jsbbs.17009>
- Wang, Jian, Wu, B., Kohnen, M. V., Lin, D., Yang, C., Wang, X., Qiang, A., Liu, W., Kang, J., Li, H., Shen, J., Yao, T., Su, J., Li, B., & Gu, L. (2021). Classification of rice yield using UAV-based hyperspectral imagery and lodging feature. *Plant Phenomics*, *2021*.

- <https://doi.org/10.34133/2021/9765952>
- Wang, Junping, Badenhorst, P., Phelan, A., Pembleton, L., Shi, F., Cogan, N., Spangenberg, G., & Smith, K. (2019). Using Sensors and Unmanned Aircraft Systems for High-Throughput Phenotyping of Biomass in Perennial Ryegrass Breeding Trials. *Frontiers in Plant Science*, 10(October), 1–9. <https://doi.org/10.3389/fpls.2019.01381>
- Wang, X. R., Torimaru, T., Lindgren, D., & Fries, A. (2010). Marker-based parentage analysis facilitates low input “breeding without breeding” strategies for forest trees. *Tree Genetics and Genomes*, 6(2), 227–235. <https://doi.org/10.1007/s11295-009-0243-8>
- Warburton, M. L., Rauf, S., Marek, L., Hussain, M., Ogunola, O., & Sanchez Gonzalez, J. D. J. (2017). The use of crop wild relatives in maize and sunflower breeding. In *Crop Science* (Vol. 57, Issue 3, pp. 1227–1240). Crop Science Society of America. <https://doi.org/10.2135/cropsci2016.10.0855>
- Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., & Reynolds, J. M. (2012). “Structure-from-Motion” photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179, 300–314. <https://doi.org/10.1016/j.geomorph.2012.08.021>
- Wiegand, C. L., Richardson, A. J., & Kanemasu, E. T. (1979). Leaf Area Index Estimates for Wheat from LANDSAT and Their Implications for Evapotranspiration and Crop Modeling1. *Agronomy Journal*, 71(2), 336–342. <https://doi.org/10.2134/AGRONJ1979.00021962007100020027X>
- Wolfram Research, I. (2012). *Mathematica, version 9.0*. Wolfram Research, Champaign, IL.
- Wu, X.-L., & Jannink, J.-L. (2004). Optimal sampling of a population to determine QTL location, variance, and allelic number. *Theoretical and Applied Genetics* 2004 108:7, 108(7), 1434–1442. <https://doi.org/10.1007/S00122-003-1569-5>
- Wu, Y., Bhat, P. R., Close, T. J., & Lonardi, S. (2008). Efficient and Accurate Construction of Genetic Linkage Maps from the Minimum Spanning Tree of a Graph. *PLOS Genetics*, 4(10), e1000212. <https://doi.org/10.1371/JOURNAL.PGEN.1000212>
- Xie, C., Gessler, D. D. G., & Xu, S. (1998). Combining different line crosses for mapping quantitative trait loci using the identical by descent-based variance component method. *Genetics*, 149(2), 1139–1146. <https://doi.org/10.1093/genetics/149.2.1139>
- Xu, S. (1996). Mapping quantitative trait loci using four-way crosses. *Genet. Res. Camb*, 68, 175–181. <https://doi.org/10.1017/S0016672300034066>

- Xu, S. (1998). Mapping quantitative trait loci using multiple families of line crosses. *Genetics*, *148*(1), 517–524. <https://doi.org/10.1093/genetics/148.1.517>
- Xu, Y., Li, P., Zou, C., Lu, Y., Xie, C., Zhang, X., Prasanna, B. M., & Olsen, M. S. (2017). Enhancing genetic gain in the era of molecular breeding. *Journal of Experimental Botany*, *68*(11), 2641–2666. <https://doi.org/10.1093/jxb/erx135>
- Yan, W., Zhao, H., Yu, K., Wang, T., Khattak, A. N., & Tian, E. (2020). Development of a multiparent advanced generation intercross (MAGIC) population for genetic exploitation of complex traits in *Brassica juncea*: Glucosinolate content as an example. *Plant Breeding*, *139*(4), 779–789. <https://doi.org/10.1111/pbr.12820>
- Yang, C. J., Edmondson, R. N., Piepho, H.-P., Powell, W., & Mackay, I. (2021). Crafting for a better MAGIC: systematic design and test for Multiparental Advanced Generation Intercross population. *G3 Genes/Genomes/Genetics*, August. <https://doi.org/10.1093/g3journal/jkab295>
- Yang, G., Liu, J., Zhao, C., Li, Z., Huang, Y., Yu, H., Xu, B., Yang, X., Zhu, D., Zhang, X., Zhang, R., Feng, H., Zhao, X., Li, Z., Li, H., & Yang, H. (2017). Unmanned aerial vehicle remote sensing for field-based crop phenotyping: Current status and perspectives. In *Frontiers in Plant Science* (Vol. 8, p. 1111). Frontiers Media S.A. <https://doi.org/10.3389/fpls.2017.01111>
- Yang, J., Gong, W., Shi, S., Du, L., Sun, J., & Song, S. L. (2016). Laser-induced fluorescence characteristics of vegetation by a new excitation wavelength. *Spectroscopy Letters*, *49*(4), 263–267. <https://doi.org/10.1080/00387010.2016.1138311>
- Yang, W., Duan, L., Chen, G., Xiong, L., & Liu, Q. (2013). Plant phenomics and high-throughput phenotyping: Accelerating rice functional genomics using multidisciplinary technologies. In *Current Opinion in Plant Biology* (Vol. 16, Issue 2, pp. 180–187). Elsevier Current Trends. <https://doi.org/10.1016/j.pbi.2013.03.005>
- 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*(1), 367. <https://doi.org/10.2135/cropsci2003.3670>
- Yu, J., & Buckler, E. S. (2006). Genetic association mapping and genome organization of maize. In *Current Opinion in Biotechnology* (Vol. 17, Issue 2, pp. 155–160). Elsevier Current Trends. <https://doi.org/10.1016/j.copbio.2006.02.003>

- Yu, J., Holland, J. B., McMullen, M. D., & Buckler, E. S. (2008). Genetic Design and Statistical Power of Nested Association Mapping in Maize. *Genetics*, 178(1), 539–551. <https://doi.org/10.1534/GENETICS.107.074245>
- Zaffaroni, E., & Schneiter, A. A. (1989). Water-Use Efficiency and Light Interception of Semidwarf and Standard-Height Sunflower Hybrids Grown In Different Row Arrangements. *Agronomy Journal*, 81(5), 831–836. <https://doi.org/10.2134/AGRONJ1989.00021962008100050026X>
- Zaw, H., Raghavan, C., Pocsedio, A., Swamy, B. P. M., Jubay, M. L., Singh, R. K., Bonifacio, J., Mauleon, R., Hernandez, J. E., Mendioro, M. S., Gregorio, G. B., & Leung, H. (2019). Exploring genetic architecture of grain yield and quality traits in a 16-way indica by japonica rice MAGIC global population. *Scientific Reports*, 9(1), 19605–19605. <https://doi.org/10.1038/s41598-019-55357-7>
- Zeng, Z. B. (1994). Precision mapping of quantitative trait loci. *Genetics*, 136(4), 1457–1468. <https://doi.org/10.1093/genetics/136.4.1457>
- Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. In *Precision Agriculture* (Vol. 13, Issue 6, pp. 693–712). Springer. <https://doi.org/10.1007/s11119-012-9274-5>
- Zhang, Z., Wang, W., & Valdar, W. (2014). Bayesian modeling of haplotype effects in multiparent populations. *Genetics*, 198(1), 139–156. <https://doi.org/10.1534/genetics.114.166249>
- Zheng, C. (2015a). Reconstruction of Genome Ancestry Blocks in Multiparental Populations. *G3: Genes, Genomes, Genetics*, 200(August), 1073–1087. <https://doi.org/10.1534/genetics.115.177873>
- Zheng, C. (2015b). Modeling X-linked ancestral origins in multiparental populations. *G3: Genes, Genomes, Genetics*, 5(5), 777–801. <https://doi.org/10.1534/g3.114.016154>
- Zheng, C., Boer, M. P., & van Eeuwijk, F. A. (2015). Reconstruction of genome ancestry blocks in multiparental populations. *Genetics*, 200(4), 1073–1087. <https://doi.org/10.1534/genetics.115.177873>
- Zheng, C., Boer, M. P., & Van Eeuwijk, F. A. (2014). A general modeling framework for genome ancestral origins in multiparental populations. *Genetics*, 198(1), 87–101. <https://doi.org/10.1534/genetics.114.163006>
- Zheng, L. Y., Guo, X., Sen, He, B., Sun, L. J., Peng, Y., Dong, S. S., Liu, T. F., Jiang, S.,

- Ramachandran, S., Liu, C. M., & Jing, H. C. (2011). Genome-wide patterns of genetic variation in sweet and grain sorghum (*Sorghum bicolor*). *Genome Biology*, 12(11). <https://doi.org/10.1186/gb-2011-12-11-r114>
- Zhou, J. (2021). *Development of UAV-based High-Throughput Phenotyping Technologies towards improving soybean breeding efficiency*. University of Missouri-Columbia.
- Zubrzycki, J. E., Maringolo, C. A., Filippi, C. V., Quiróz, F. J., Nishinakamasu, V., Puebla, A. F., Julio, A. D. R., Escande, A., Lia, V. V., Heinz, R. A., Hopp, H. E., Cervigni, G. D. L., & Paniego, N. B. (2017). Main and epistatic QTL analyses for *Sclerotinia* Head Rot resistance in sunflower. *PLoS ONE*, 12(12). <https://doi.org/10.1371/journal.pone.0189859>
- Zubrzycki, N., Filippi, C., Fusari, C., Puebla, A., Fernández, P., Hopp, E., Heinz, R., Lia, V., & Paniego, N. (2012). Desarrollo e implementación de un ensayo de genotipificación masiva de SNPs en girasol. *ALAG XV Congreso Latinoamericano de Genética*.
- Zuliani, P., Defacio, R., Lavalle, A., & Bramardi, S. (2018). Comparación de técnicas de análisis multivariado mediante simulación para caracterización de recursos fitogenéticos en función de caracteres susceptibles a interacción genotipo-ambiente. *Revista FAVE - Ciencias Agrarias*, 17(1), 75–86.
- Zuliani, P., Lavalle, A., Bramardi, S. J., & Defacio, R. (2012). Caracterización de poblaciones nativas de maíz mediante Análisis de Procrustes Generalizado y Análisis Factorial Múltiple. *Rev. FCA UNCUYO*, 49–64.

Anexo

---

---

## **ANEXO**

---

---

El modelo SpATS adopta la formulación P-spline ANOVA (PSANOVA) para representar las P-splines 2D en el marco del modelo mixto. El principal resultado de la parametrización PS-ANOVA es la descomposición de la superficie lisa en una suma de componentes lineales y funciones suaves univariantes y bivariantes que se demuestra a continuación:

$$f(f, c) = \beta_{s1}f + \beta_{s2}c + \beta_{s3}fc + f_1(f) + f_2(c) + h_3(f)c + fh_4(c) + f_5(f, c)$$

La descomposición anterior equivale a lo indicado en el capítulo 4:

$$Xs\beta_s = \beta_{s1}f + \beta_{s2}c + \beta_{s3}fc$$

$$ZsS = f_1(f) + f_2(c) + h_3(f)c + fh_4(c) + f_5(f, c)$$

Donde la superficie espacial está representada por: tendencias lineales a través de las posiciones de fila  $\beta_{s1}$  y columna  $\beta_{s2}$  y una tendencia de lineal de interacción  $\beta_{s3}$ ; dos tendencias principales suavizadas en las filas  $f_1(f)$  y columnas  $f_2(c)$ ; dos términos de interacción lineal suavizados, donde la pendiente de una tendencia lineal a lo largo de una covariable ( $f$  o  $c$ ) puede variar suavemente en función de la otra covariable  $h_3(f)$  o  $h_4(c)$  respectivamente; y  $f_5(f, c)$  es la interacción pura suavizada entre las tendencias de la columna y las tendencias de las filas.

**Anexo digital link:**

<http://hdl.handle.net/20.500.12123/11627>