

Caracterización molecular y fenotípica de recursos fitogenéticos de pimiento (*Capsicum* spp.) y búsqueda de genes asociados a caracteres fenotípicos de interés

Jorge González López

Tesis Doctoral 2021

Programa de Doctorado en Biología Celular y Molecular



UNIVERSIDADE DA CORUÑA

A mis padres

Agradecimientos

Durante estos años ha habido muchas personas que me han ayudado de una manera u otra a completar esta tesis y a quienes quiero darles las gracias.

Al Dr. Joaquín Cañizares, que me acogió en su grupo de investigación del Instituto Universitario de Conservación y Mejora de la Agrodiversidad Valenciana, donde realicé los análisis bioinformáticos. Gracias también al Dr. José Blanca y a Peio Ziarsolo por todo el tiempo que me han dedicado y lo que me enseñaron durante mi estancia.

A los miembros del Grupo de Investigación en Biología Evolutiva (GIBE), Fede, Ana, Andrés, Horacio y Manuel. Por vuestra gran ayuda, sobre todo en la última etapa de la tesis. A Marta, mi amiga, gracias por confiar en mí cuando solo era un estudiante de Grado, por lo bien que me trataste entonces y por seguir haciéndolo ahora. Gracias por tus muestras de ánimo constante y tu apoyo, esenciales durante estos años.

A la Dra. Graciela Estévez, por su ayuda con la parte estadística del tercer capítulo.

A mis compañeras de laboratorio. A Alba, gracias infinitas por todo lo que me has enseñado (incluso de mi tesis), por responder siempre con una sonrisa a las mil dudas que te planteaba y por saber dónde estaban todas las cosas del laboratorio; has sido una gran compañera. A Marta, gracias por las tardes de confidencias, por todo lo que me has escuchado, por los “rajamientos” y por las sugerencias musicales; me has hecho esta etapa mucho más llevadera. A Laura, por tu optimismo y tu alegría contagiosa, gracias por todo lo que me ayudaste cuando era un recién aterrizado en el laboratorio. A Irene, por tus consejos y por escucharme siempre que lo he necesitado; me alegro mucho de los buenos momentos que hemos vivido juntos dentro y, sobre todo, fuera del

laboratorio. A Manu y Diego, por contribuir al buen ambiente del laboratorio y compartir momentos tan buenos en el café.

A Ana M. Vázquez que diseñó la maravillosa portada de esta tesis.

A mis amigos del pádel. A Germán, por tantos años de amistad desde que éramos unos niños. A Samu y a Edu, por todos los momentos de risa que hemos compartido y alguna tarde de deporte también. A Marta, mi “vecina”, por todo lo que me ha aguantado durante estos años hablando de pimientos y de muchos otros temas, no coaccionándome en ningún momento para que la nombrase en esta sección.

A Inés, mi compañera de viaje. Sin ti esta tesis no habría sido posible. Gracias por todo lo que me has apoyado, por soportarme cuando ni yo mismo lo hacía y por animarme a seguir siempre hacia delante. Gracias por ser mi alegría. Te quiero siempre.

A mis padres, por vuestro amor y vuestro apoyo, por enseñarme a luchar y el valor del esfuerzo. Gran parte de lo que soy hoy en día os lo debo a vosotros. A mi hermana, por escucharme, comprenderme y hacerme ser mejor persona. A mi cuñado, por interesarte por mi trabajo y demostrarme siempre que estabas orgulloso de mí. A mi sobrina, por tener la capacidad de alegrarnos la vida con un simple balbuceo.

Índice

Índice	I
Índice de tablas y figuras.....	V
Lista de abreviaturas.....	X
Resúmenes.....	1
Resumen.....	3
Resumo.....	4
Abstract	5
Prólogo	7
Introducción general	11
El género <i>Capsicum</i>	13
Interés nutricional del pimiento, aplicaciones y usos e importancia económica.....	20
Procesos de domesticación, recursos fitogenéticos y bancos de germoplasma	24
Herramientas de secuenciación masiva y su aplicación en colecciones de recursos fitogenéticos de pimiento.....	28
Objetivos.....	35
Capítulo 1. Análisis de la diversidad genética de recursos fitogenéticos de pimiento (<i>Capsicum spp.</i>).....	39
Capítulo 1.1. Estudio de la variabilidad y estructura genética mediante marcadores microsatélite	41
Introducción.....	43
Material y métodos.....	45
Material vegetal	45
Extracción de ADN.....	54
Genotipado de los <i>loci</i> microsatélite.....	54
Diversidad genética.....	56
Estructura genética	58
Resultados.....	59
Diversidad genética.....	59
Estructura genética	63

Diversidad genética en los grupos encontrados en el análisis de estructura poblacional	68
Discusión.....	71
Diversidad genética.....	71
Estructura genética.....	73
Consideraciones metodológicas	78
Capítulo 1.2. Evaluación de marcadores moleculares ligados a enfermedades y pungencia	85
Introducción.....	87
Material y métodos.....	91
Material vegetal	91
Extracción de ADN.....	92
Análisis molecular	92
Resultados.....	95
Discusión.....	98
Resistencia frente a <i>Phytophthora capsici</i> (OpD04.717)	98
Resistencia frente a <i>Tobamovirus</i> (060I2END)	99
Resistencia frente a <i>Meloidogyne incognita</i> (SCAR_CD).....	99
Resistencia frente a <i>Xanthomonas campestris</i> (PR-Bs3).....	100
Resistencia frente al virus del mosaico del pepino (<i>CaTm-int1</i>).....	101
Resistencia frente al virus del bronceado del tomate (SCAC568)	101
Resistencia frente a <i>Potyvirus</i> (<i>Pvr1-S</i> , <i>Pvr1-R1</i> , <i>Pvr1-R2</i> , <i>CSO</i> y <i>T200A</i>)	102
Carácter pungente (<i>MAP1</i>)	105
Capítulo 2. Análisis de la diversidad, estructura genética y relaciones entre accesiones del género <i>Capsicum</i> mediante <i>genotyping-by-sequencing</i> (GBS).....	107
Introducción.....	109
Material y métodos.....	112
Material vegetal	112
Extracción de ADN, preparación de las librerías y secuenciación	113
Mapeo y SNV <i>calling</i>	114
Estructura poblacional	115

Diversidad genética.....	116
Resultados	116
Secuenciación, pre-procesado de las secuencias y mapeo.....	116
Identificación de SNVs.....	118
Estructura genética.....	120
Diversidad genética.....	130
Discusión.....	132
Capítulo 3. Caracterización morfológica en cinco especies del género <i>Capsicum</i> y análisis de asociación <i>genome-wide association study</i> (GWAS) entre datos fenotípicos y genotípicos en <i>C. annuum</i>	145
Introducción.....	147
Material y métodos.....	150
Datos fenotípicos	150
Datos genotípicos	157
Cálculo de la <i>Q-matrix</i> y la <i>K-matrix</i>	158
Estudio de asociación a nivel del genoma completo (<i>genome-wide association study</i> , GWAS).....	159
Desequilibrio de ligamiento (<i>linkage disequilibrium</i> , LD)	159
Mapeo contra el genoma anotado de <i>C. annuum</i>	160
Resultados	161
Variabilidad morfológica.....	161
Análisis de asociación en <i>C. annuum</i>	163
Identificación de SNPs	163
Estructura poblacional	164
GWAS.....	165
Marcadores con asociación	168
<i>S1_157644336</i>	168
<i>S8_130837307</i>	168
<i>S2_147083777</i>	169
<i>S11_4165433</i>	169
<i>S1_201025316</i>	170

<i>S1_19470319</i>	170
<i>S4_210105544</i>	170
<i>S9_213716963</i>	171
Análisis de asociación en el complejo <i>C. annuum</i>	173
Identificación de SNPs	173
Estructura poblacional	173
GWAS.....	175
Discusión.....	176
Variabilidad morfológica.....	176
Análisis de asociación en <i>C. annuum</i>	177
Marcadores con asociación	179
<i>S1_157644336</i>	179
<i>S1_201025316</i>	180
<i>S4_210105544</i>	182
<i>S9_213716963</i>	183
<i>S8_130837307, S2_147083777, S11_4165433 y S1_19470319</i>	185
Análisis de asociación en el complejo <i>C. annuum</i>	186
Epílogo	191
Conclusiones	201
Referencias	209
Anexos	241
Anexos de tablas	243
Anexos de figuras	273

Índice de tablas y figuras

Introducción general

Tablas

Tabla I.1. Datos de producción, área cosechada y rendimiento del cultivo de pimientos frescos y secos a nivel global y para los principales países productores23

Figuras

Figura I.1. Relaciones filogenéticas entre las especies del género *Capsicum* 15

Figura I.2. Flores y frutos de las variedades más representativas de las cinco especies cultivadas del género *Capsicum* 19

Figura I.3. Comparativa de los métodos de secuenciación masiva basados en la reducción del genoma RRL, RAD-seq y GBS31

Capítulo 1. Análisis de la diversidad genética de recursos fitogenéticos de pimiento (*Capsicum spp.*)

Capítulo 1.1. Estudio de la variabilidad y estructura genética mediante marcadores microsatélite

Tablas

Tabla 1.1. Acciones de *Capsicum spp.* empleadas en la presente Tesis Doctoral .48

Tabla 1.2. Marcadores microsatélite analizados en las tres colecciones de *Capsicum spp.*55

Tabla 1.3. Caracterización de los *loci* microsatélite analizados en las 230 acciones de *Capsicum spp.*61

Tabla 1.4. Diversidad genética para cada especie de *Capsicum* estudiada.....63

Tabla 1.5. Frecuencia de alelos nulos para cada marcador microsatélite y en cada uno de los grupos encontrados en el segundo nivel jerárquico de STRUCTURE..69

Tabla 1.6. Valores de los coeficientes de endogamia (F_{IS}) para cada clúster, del coeficiente de endogamia corregido para la presencia de alelos nulos ($F_{IS-INest}$) y del criterio de información de la desviación (DIC) tanto con el modelo *nfb* como con el modelo *fb*71

Figuras

Figura 1.1. Métodos de evaluación de la estructura genética más probable con el análisis STRUCTURE para las 230 accesiones de *Capsicum* y empleando los 28 loci microsatélite obtenidos con STRUCTURE HARVESTER.....63

Figura 1.2. Análisis de inferencia bayesiana realizado con STRUCTURE para las 230 accesiones y empleando 28 microsatélites64

Figura 1.3. Análisis de inferencia bayesiana realizado con STRUCTURE para las 148 accesiones de *C. annuum* y empleando 28 microsatélites65

Figura 1.4. Análisis de inferencia bayesiana realizado con STRUCTURE para las 82 accesiones de especies distintas a *C. annuum* y empleando 28 microsatélites66

Figura 1.5. Análisis factorial de correspondencia (AFC) para las 230 accesiones y empleando 28 microsatélites.....67

Capítulo 1.2. Evaluación de marcadores moleculares ligados a enfermedades y pungencia

Tablas

Tabla 1.7. Marcadores moleculares ligados a enfermedades y pungencia analizados en las dos colecciones de *Capsicum* spp.92

Tabla 1.8. Porcentaje de accesiones de cada colección con genotipo resistente para cada marcador de resistencia y con genotipo pungente para el marcador *MAP1* .96

Tabla 1.9. Porcentaje de accesiones de cada colección con cada uno de los alelos de los marcadores *Pvr1-S*, *Pvr1-R1* y *pvr1-R2* y con genotipos resistentes para los loci *Pvr2* y *Pvr4*97

Capítulo 2. Análisis de la diversidad, estructura genética y relaciones entre accesiones del género *Capsicum* mediante *genotyping-by-sequencing* (GBS)

Tablas

Tabla 2.1. Estadísticas de diversidad para cada grupo de la clasificación jerárquica establecido a partir de los PCoAs.....	131
--	-----

Figuras

Figura 2.1. Etapas en el desarrollo experimental del GBS	111
Figura 2.2. Número de secuencias mapeadas para cada valor de Mapq	117
Figura 2.3. Porcentaje de secuencias que mapean y porcentaje de secuencias que mapean con Mapq mayor de 57 para cada accesión	118
Figura 2.4. Recuento de SNVs tras el filtrado para cada uno de los doce cromosomas de <i>C. annuum</i>	119
Figura 2.5. Porcentaje de cada tipo de SNP representado según la substitución nucleotídica en transiciones y transversiones	119
Figura 2.6. Análisis de coordenadas principales global, incluyendo las 232 accesiones y empleando los 48.109 marcadores.....	120
Figura 2.7. Análisis de coordenadas principales incluyendo solo las 13 muestras agrupadas como <i>C. pubescens</i>	122
Figura 2.8. Análisis de coordenadas principales de las 23 muestras clasificadas como <i>C. baccatum</i>	123
Figura 2.9. Análisis de coordenadas principales de las 38 muestras que conforman el conjunto <i>C. chinense</i> + <i>C. frutescens</i>	124
Figura 2.10. Análisis de coordenadas principales para el segundo nivel jerárquico a partir de las agrupaciones observadas en el PCoA del conjunto <i>C. chinense</i> + <i>C. frutescens</i>	125
Figura 2.11. Análisis de coordenadas principales para las líneas que aparecen formando el grupo <i>C. annuum</i> en el PCoA global.....	126

Figura 2.12. Análisis de coordenadas principales para las 122 líneas encuadradas en el “Grupo 2”	127
Figura 2.13. Probabilidad marginal (<i>P</i>) para cada clúster (<i>K</i>) testado con el programa fastSTRUCTURE.....	128
Figura 2.14. Estructura poblacional para las 232 accesiones inferida por el programa fastSTRUCTURE.....	130

Capítulo 3. Caracterización morfológica en cinco especies del género *Capsicum* y análisis de asociación *genome-wide association study* (GWAS) entre datos fenotípicos y genotípicos en *C. annuum*

Tablas

Tabla 3.1. Descripción de los caracteres fenotípicos cuantitativos medidos para planta y fruto	151
Tabla 3.2. Descripción de los caracteres fenotípicos cualitativos estudiados para planta y fruto	152
Tabla 3.3. Descripción de los caracteres fenotípicos medidos con el programa Tomato Analyzer.....	154
Tabla 3.4. SNPs que mostraron asociaciones significativas con parámetros fenotípicos.....	172

Figuras

Figura 3.1. Diversidad morfológica encontrada en los frutos de las accesiones estudiadas	162
Figura 3.2. Estructura poblacional inferida con el programa ADMIXTURE para el conjunto de accesiones de <i>C. annuum</i> analizadas en el GWAS.....	165
Figura 3.3. Gráficos Manhattan para los parámetros con asociaciones significativas	167
Figura 3.4. Estructura poblacional inferida con el programa ADMIXTURE para el conjunto de accesiones del complejo <i>C. annuum</i> analizadas en el GWAS	174

Figura 3.5. Ejemplos de gráficos Q-Q que muestran modelos mal ajustados para el complejo <i>C. annuum</i>	175
--	-----

Anexos

Tablas

Tabla A.1. Frecuencias de alelos nulos para cada marcador microsatélite calculadas con los programas FreeNa y ML-Null Freq	243
Tabla A.2. Parámetros de diversidad genética por especie y por <i>locus</i> microsatélite	244
Tabla A.3. Genotipos de cada accesión para los diez marcadores de enfermedad y el marcador de pungencia.....	246
Tabla A.4. Estadísticas de las lecturas de GBS empleadas.....	248
Tabla A.5. Datos de fenotipado de cada accesión para los parámetros cualitativos de planta y fruto	253
Tabla A.6. Media de los datos de fenotipado de cada accesión para los parámetros cuantitativos de planta y fruto	258
Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer.....	263

Figuras

Figura A.1. Métodos de evaluación de la estructura genética inferida con el análisis STRUCTURE obtenidos con STRUCTURE HARVESTER para los grupos <i>C. annuum</i> y no- <i>C. annuum</i> y empleando 28 microsatélites	273
Figura A.2. Análisis factoriales de correspondencia (AFCs) calculados para cada uno de los subgrupos encontrados en el AFC global	274
Figura A.3. Resultados de validación cruzada (<i>cross-validation</i>) empleada para inferir el número de clústeres en el programa ADMIXTURE	275
Figura A.4. Gráficos <i>quantile-quantile</i> (Q-Q) obtenidos para los parámetros con asociaciones significativas en <i>C. annuum</i>	276

Lista de abreviaturas

- *A*: número medio de alelos por *locus*.
- ADN: ácido desoxirribonucleico.
- AFC: análisis factorial de correspondencia.
- AFLPs: *amplified fragment length polymorphisms* (polimorfismos en la longitud de los fragmentos amplificados).
- AR: riqueza alélica.
- ARMS: *amplification refractory mutation system* (sistema de amplificación de mutaciones refractarias).
- BGHZ: Banco de Germoplasma de Hortícolas de Zaragoza.
- BPGV: *Banco Português de Germoplasma Vegetal*.
- CAPS: *cleaved amplified polymorphic sequence* (secuencia polimórfica amplificada y digerida).
- CGN: *Center for Genetic Resources of Wageningen*.
- CIAM: Centro de Investigaciones Agrarias de Mabegondo.
- CMV: *cucumber mosaic virus* (virus del mosaico del pepino).
- COMAV: Instituto de Conservación y Mejora de la Agrodiversidad Valenciana.
- CTAB: *cetyl trimethyl ammonium bromide* (bromuro de cetiltrimetilamonio).
- DAMD-PCR: *direct or indirected amplification of minisatellite region DNA amplified using the polymerase chain reaction* (amplificación directa o indirecta de regiones minisatélite mediante PCR).
- DArTseq™: *Diversity arrays technology*.
- ddRAD-seq: *double digest RAD-seq* (RAD-seq con doble digestión).

- DIC: *deviance information criterion* (criterio de información de la desviación).
- dNTPs: desoxinucleótidos trifosfato.
- DP-GT: *depth genotype* (profundidad de genotipo de los SNPs).
- EDTA: ácido etilendiaminotetraacético.
- FAO: *Food and Agriculture Organization of the United Nations* (Organización de las Naciones Unidas para la Alimentación y la Agricultura).
- FAOSTAT: *Food and Agriculture Organization of the United Nations Corporate Statistical Database* (Base de datos estadísticos corporativos de la Organización de las Naciones Unidas para la Alimentación y la Agricultura).
- F_{IS} : coeficiente de endogamia.
- $F_{IS-INest}$: coeficiente de endogamia corregido para la presencia de alelos nulos, calculado con el programa INest.
- GBS: *genotyping-by-sequencing* (genotipado por secuenciación).
- GLM: *generalized linear model* (modelo lineal generalizado).
- GWAS: *genome-wide association study* (estudio de asociación a nivel del genoma completo).
- ha: hectárea.
- H_e : heterocigosidad esperada.
- hg: hectogramo.
- H_o : heterocigosidad observada.
- HTS: *high-throughput sequencing* (secuenciación de última generación).
- HWE: *Hardy-Weinberg equilibrium* (equilibrio Hardy-Weinberg).
- IIM: *bayesian individual inbreeding model* (modelo de endogamia individual bayesiana).

- INRA: *Institut National de la Recherche Agronomique*.
- IPGR: *Institute for Plant Genetic Resources "K. Malkov"*.
- IPGRI: *International Plant Genetic Resources Institute*.
- IPK: *Leibniz Institute of Plant Genetics and Crop Plant Research*.
- K: número de clústeres.
- LD: *linkage disequilibrium* (desequilibrio de ligamiento o desequilibrio gamético).
- MAC: *minor allele count* (recuento del alelo menos común).
- MAF: *minor allele frequency* (frecuencia del alelo menos común).
- MCMC: *Markov chain Monte Carlo* (cadenas de Markov Monte Carlo).
- MgCl₂: cloruro de magnesio.
- MLM: *mixed linear model* (modelo lineal mixto).
- Na: número medio de alelos por *locus*.
- Nap: número medio de alelos privados por *locus*.
- NGS: *next-generation sequencing* (secuenciación de última generación).
- NFS: factores sensibles a N-etilmaleimida.
- P: porcentaje de polimorfismo.
- P95: porcentaje de marcadores en los que la frecuencia del alelo mayoritario es menor de 0,95.
- pb: pares de bases.
- PCA: *principal component analysis* (análisis de componentes principales)
- PCoA: *principal coordinates analysis* (análisis de coordenadas principales).
- PCR: *polymerase chain reaction* (reacción en cadena de la polimerasa).
- PepMov: *pepper mottle Virus* (virus del moteado del pimiento).

- PGRCU-USDA: *Plant Genetic Resources Conservation Unit by the U.S. Department of Agriculture.*
- PVY: *potato virus Y* (virus Y de la patata).
- QTL: *quantitative trait locus* (locus de rasgo cuantitativo).
- RAD-seq: *restriction-site-associated DNA sequencing* (secuenciación de ADN asociado a un sitio de restricción).
- RAPD: *random amplified polymorphic DNA* (ADN polimórfico amplificado al azar).
- RFLPs: *restriction fragment length polymorphisms* (polimorfismos en la longitud de fragmentos de restricción).
- RRLs: *reduced-representation libraries* (librerías de representación reducidas).
- S-SAPs: *sequence-specific amplification polymorphisms* (polimorfismos en la amplificación secuencia-específicos).
- SAI-UDC: Servicios de Apoio á Investigación da Universidade da Coruña.
- SCAR: *sequence characterized amplified region* (región amplificada caracterizada por secuencia).
- SSRs: *simple sequence repeats* (repeticiones de secuencia simple, microsatélites).
- SNPs: *single nucleotide polymorphisms* (polimorfismos de un solo nucleótido).
- SNVs: *single nucleotide variants* (variantes de un solo nucleótido).
- STRs: *short tandem repeats* (repeticiones cortas en tándem, microsatélites).
- t: tonelada.
- TEV: *tobacco etch virus* (virus del grabado del tabaco).
- TSWV: *tomato spotted wilt virus* (virus del bronceado del tomate).
- UDP: uridina difosfato.
- USDA-ARS: *U.S Department of Agriculture-Agriculture Research Service.*

- VCF: *variant call format* (formato “variant call”).
- WVC: *World Vegetable Center of Taiwan*.

Resúmenes

Resumen

El pimiento (*Capsicum* spp.) es uno de los alimentos de origen vegetal más importantes. Con los procesos de domesticación y selección de variedades concretas para el cultivo, la diversidad genética inicial quedó muy reducida y los recursos fitogenéticos se volvieron imprescindibles para preservarla. Esta Tesis Doctoral amplía el conocimiento de tres colecciones de recursos fitogenéticos de pimiento. En primer lugar, se realizó un análisis de la variabilidad y estructura genética con veintiocho marcadores microsatélite que reveló una marcada distinción *C. annuum*-no *C. annuum* y agrupaciones menores en base a la especie y origen geográfico y se planteó por primera vez los posibles sesgos y limitaciones en trabajos de este tipo. Además, se determinó el perfil genético de resistencia a enfermedades y pungencia para identificar accesiones candidatas a ser incluidas en programas de mejora. En segundo lugar, el análisis de 48.109 marcadores SNVs desarrollados *de novo* mediante secuenciación masiva reveló que las relaciones entre accesiones vienen dadas por la especie a la que pertenecen, su origen y caracteres morfológicos. Finalmente se analizaron fenotípicamente los recursos fitogenéticos y se llevó a cabo un análisis de asociación GWAS con los datos moleculares, identificando ocho marcadores SNP asociados a características agronómicas de interés.

Resumo

O pemento (*Capsicum* spp.) é un dos alimentos de orixe vexetal máis importantes. Cos procesos de domesticación e a selección de variedades concretas para o seu cultivo, a diversidade xenética inicial ficou moi reducida e os recursos fitoxenéticos volvéronse imprescindibles para preservala. Esta tese de doutoramento amplía o coñecemento de tres coleccións de recursos fitoxenéticos de pemento. En primeiro lugar levouse a cabo unha análise da variabilidade e estrutura xenética con vinte e oito marcadores microsatélite que desvelou unha marcada diferenciación *C. annuum*-non *C. annuum* e agrupacións menores segundo a especie e a orixe xeográfica e suxeríronse por primeira vez as posibles limitacións en traballos deste tipo. Ademais determinouse o perfil xenético de resistencia a enfermidades e punxencia co fin de identificar accesións candidatas a ser engadidas en programas de mellora. En segundo lugar, 48.109 marcadores SNVs desenrolados *de novo* mediante secuenciación masiva amosaron que as relacións entre accesións veñen dadas pola especie á cal pertencen, o seu lugar de orixe e características morfolóxicas. Finalmente analizáronse fenotipicamente os recursos fitoxenéticos e levouse a cabo unha análise de asociación GWAS cos datos moleculares, identificando oito marcadores SNP asociados a características agronómicas de interese.

Abstract

Pepper (*Capsicum* spp.) is one of the most important crops in human food. Due to domestication processes and selection of specific varieties for cultivation, genetic diversity was greatly reduced and plant genetic resources became essential to preserve it. This PhD thesis broadens the knowledge about three collections of pepper plant genetic resources. Firstly, genetic variability and structure was analyzed using twenty-eight microsatellite markers. A clear distinction *C. annuum*-non *C. annuum* was unveiled as well as minor groupings based on species and geographical origin. Possible biases and limitations in work of this type were discussed for the first time. Furthermore, the genetic profile of disease resistance and pungency was determined in order to identify the best candidate accessions for inclusion in breeding programs. Secondly, 48,109 SNV markers were *de novo* developed using high-throughput sequencing. SNV analysis revealed that relationships between pepper accessions are given by the species to which they belong, their place of origin and morphological features. Finally, plants were examined phenotypically and a *genome-wide association study* (GWAS) was carried out. Eight SNP markers were found to be associated with agronomic traits of interest.

Prólogo

El pimiento (*Capsicum* spp.) es uno de los cultivos hortícolas más importantes a nivel mundial. Tras la domesticación de las especies cultivadas y con la aparición de la agricultura moderna, el pimiento sufrió un proceso de erosión genética en el que se perdió gran parte de la variabilidad genética inicial. En la actualidad los agricultores disponen de un amplio conjunto de variedades, con elevados rendimientos y adaptadas a las exigencias de cada mercado en particular pero con poca capacidad de adaptación a nuevos estreses bióticos y abióticos. Con el fin de preservar la diversidad en especies cultivadas surgen los recursos fitogenéticos, conservados en los bancos de germoplasma. El correcto uso de los recursos fitogenéticos requiere sin embargo de una completa caracterización genética y morfológica.

En esta Tesis se caracterizó una amplia colección de recursos fitogenéticos de las cinco especies cultivadas del género *Capsicum* mediante marcadores moleculares de distintos tipos y descriptores fenotípicos y se llevó a cabo un análisis de asociación para encontrar relaciones de causalidad entre variantes genéticas y caracteres agromorfológicos.

Parte del trabajo de la presente Tesis se desarrolló en una estancia de investigación predoctoral en el Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV) de la Universitat Politècnica de València (UPV) financiada por el proyecto ED431C 2018/S7 de la Xunta de Galicia, Modalidade de Grupos de Referencia Competitiva (Grupo de Investigación en Bioloxía Evolutiva).

Introducción general

El género *Capsicum*

Las especies de pimiento se incluyen dentro del género *Capsicum*, el cual pertenece al orden Solanales, familia Solanaceae, subfamilia Solanoideae, tribu Capsiaceae. El pimiento está muy relacionado con otras plantas cultivadas de gran importancia como la patata (*Solanum tuberosum*) y el tomate (*S. lycopersicum*), ambas incluidas en la tribu Solanae, que a su vez está englobada dentro de la misma subfamilia que el pimiento (Olmstead et al., 1999; Martins & Barkman, 2005; Wu et al., 2009).

El género *Capsicum* se caracteriza por presentar plantas anuales de tipo arbustivo y de porte erecto, postrado o compacto (Andrews, 1995; Carrizo García et al., 2016). Los tallos de las plantas jóvenes son angulares, presentando secciones redondeadas o cuadradas cuando son adultas y que pueden presentar el pigmento antocianina tanto en toda su longitud como en los nodos. El grado de pubescencia es muy variable, desde su ausencia total hasta la pubescencia densa. El tipo de crecimiento es indeterminado o semi-indeterminado (Bosland & Votava, 2000). Las hojas son principalmente simples, enteras y simétricas, de color verde y ovaladas o bien elípticas o lanceoladas. Sin embargo, existe una gran variabilidad morfológica en cuanto a las hojas y, al igual que en los tallos, la pubescencia puede estar presente en distintos grados (Bosland & Votava, 2000; Cheng et al., 2018). En lo que respecta a las flores, estas son típicamente pentámeras, hermafroditas e hipóginas, con corolas rotadas de 5-7 pétalos, pudiendo algunas especies presentar corolas acampanuladas de forma excepcional. La coloración de las corolas también varía, pudiendo ser blanca, verdosa o púrpura según la especie (Bosland & Votava, 2000). Asimismo, el tiempo de floración difiere según la especie y las

condiciones de cultivo (Craufurd & Wheeler, 2009). En relación a los frutos, son de tipo baya, carnosos y globosos (Carrizo García et al., 2016), existiendo una gran variedad de formas, tamaños y colores, especialmente entre las especies cultivadas (Andrews, 1995). Dicha variabilidad también se observa en los niveles de los diferentes compuestos bioquímicos de los mismos (van Zonneveld et al., 2015). No obstante, la característica más importante del pimiento es su carácter pungente, es decir, la capacidad para producir sensación de picor, que viene determinada por la acumulación de capsaicinoides en los frutos (Govindarajan & Salzer, 1986; Kobata et al., 1999; Pino et al., 2007).

El género *Capsicum* se originó en Bolivia sur-central y posteriormente cada especie fue domesticada en diferentes zonas (McLeod et al., 1982; Eshbaugh, 1993; Perry et al., 2007). Se considera que el género está compuesto por aproximadamente 35 especies (Carrizo García et al., 2016), si bien recientemente se han añadido especies nuevas (Barboza et al., 2019, 2020), por lo que el número de integrantes aún no es definitivo. Según el registro actual del *U.S Department of Agriculture-Agriculture Research Service* (USDA-ARS, 2020) el número actual de especies podría ser superior a 50. De todas las especies del género, cinco de ellas han sido domesticadas y se cultivan en la actualidad: *C. annuum* L., *C. baccatum* L., *C. chinense* Jacq., *C. frutescens* L. y *C. pubescens* Ruiz & Pav. (Kraft et al., 2014). Las cinco especies cultivadas pertenecen a distintos complejos o *pools* genéticos, basados en el grado de compatibilidad genética y proximidad reproductiva que tienen entre sí y con las especies silvestres (Figura I.1) (Ibiza et al., 2012).

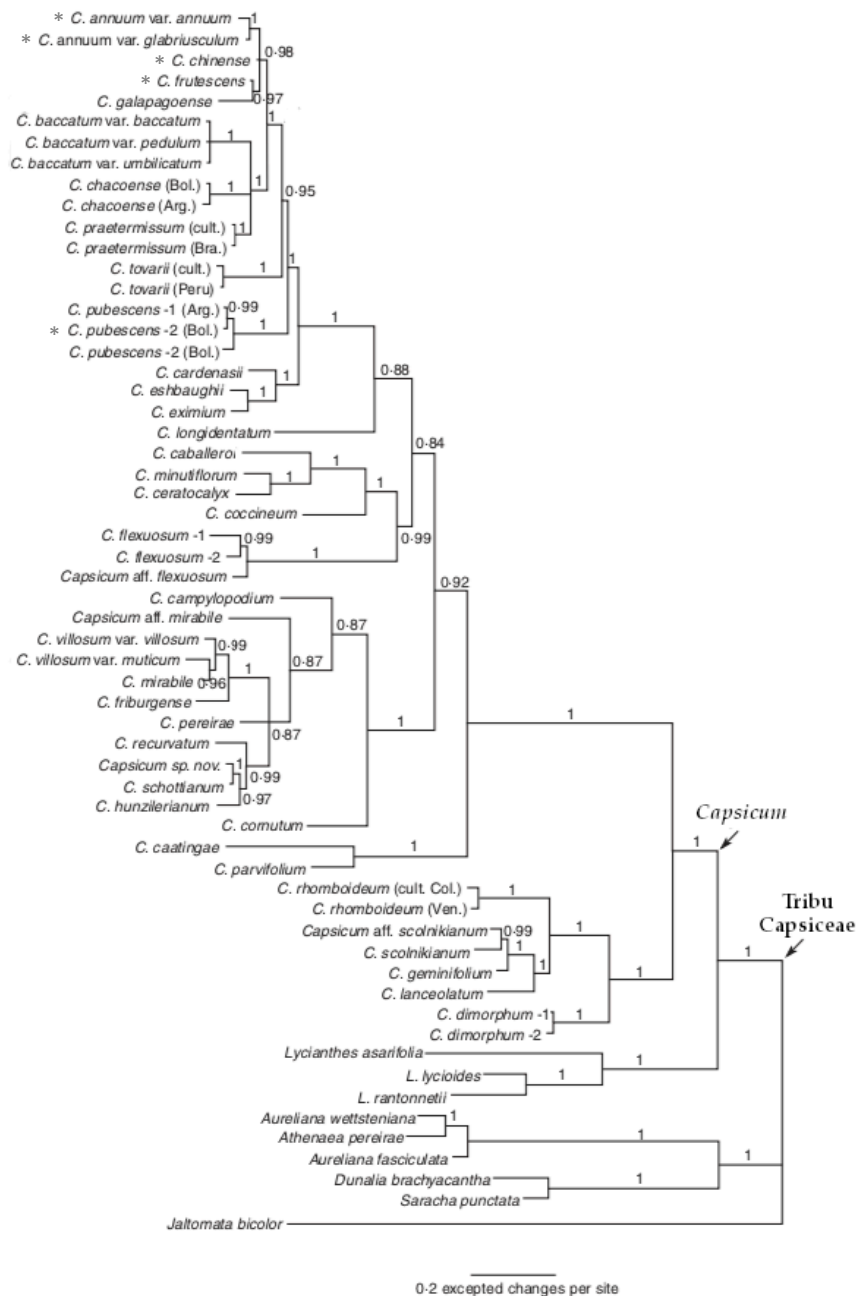


Figura I.1. Relaciones filogenéticas entre las especies del género *Capsicum* obtenidas mediante inferencia bayesiana. Para cada nodo se indican las probabilidades *a posteriori*. Los asteriscos (*) señalan las especies incluidas en la presente Tesis Doctoral. Modificada de Carrizo García et al. (2016).

Capsicum annuum (concretamente *C. annuum* var. *annuum*) es la especie cultivada de mayor éxito comercial en la actualidad, probablemente debido a que fue la que primero se introdujo en Europa y a que presenta mejores cualidades agronómicas relacionadas

con el tamaño, color, sabor y forma (Andrews, 1993). Esta especie fue domesticada en México, concretamente en la zona centro-este del país (Kraft et al., 2014) hace más de 5.000 años (Pickersgill, 2007). Tras las expediciones de Cristóbal Colón por América, *C. annuum* llegó a Europa y rápidamente se expandió por todo el mundo alcanzando un gran éxito de cultivo (Andrews, 1993). Las flores de *C. annuum* presentan la corola blanquecina (Fig. I.2A) (Andrews, 1995). A esta especie pertenecen variedades de pimiento, tanto dulces como picantes, económicamente muy relevantes tales como el “morrón” (“bell”) (Fig. I.2B), el “jalapeño”, el “ancho”, el “California Wonder”, el “dulce italiano” (Fig. I.2C), el “pimiento de Padrón” o el “choricero” (DeWitt & Bosland, 1996; Nuez et al., 1996). El pariente silvestre de *C. annuum* es *C. annuum* var. *glabriusculum* (Dunal) Heiser y Pickersgill, y junto con *C. chinense* y *C. frutescens*, los parientes silvestres de ambas y *C. galapagoensis* Hunziker, todas ellas conforman el complejo *C. annuum* (Walsh & Hoot, 2001; Ibiza et al., 2012) (Figura I.1).

Capsicum frutescens habría sido domesticada en la zona de América Central conocida como Mesoamérica, que en la actualidad comprende la mitad meridional de México, El Salvador y el oeste de Honduras, Nicaragua y Costa Rica (Pickersgill, 2007). Las flores presentan una coloración verde-blanquecina (Fig. I.2D) (Andrews, 1995). El cultivar más representativo de la especie es el “tabasco” (Fig. I.2E), base de la famosa salsa del mismo nombre (Nuez et al., 1996). Esta especie tuvo un mayor éxito de consumo en África e India (Bosland & Votava, 2000; Krishna De, 2003).

Capsicum chinense es la tercera especie domesticada del complejo *C. annuum*, proceso que ocurrió en las tierras bajas tropicales de América del Sur hace aproximadamente

4.000 años (Pickersgill, 2007). Sus flores presentan una coloración más verdosa que las otras dos especies del complejo *C. annuum* (Fig. I.2F) (Andrews, 1995). Dentro de esta especie se incluyen distintas variedades muy pungentes, entre las que destaca el “habanero” (Fig. I.2G), muy apreciado en México, y el “Scotch Bonnet” (Krishna De, 2003). Al igual que *C. frutescens* es una especie popular en África e India (Bosland & Votava, 2000). *Capsicum frutescens* y *C. chinense* están consideradas como especies hermanas, señalando algunos autores que incluso podría tratarse de una única especie (McLeod et al., 1979; Pickersgill, 1971; Walsh & Hoot, 2001).

Por su parte, el complejo *C. baccatum* está formado por la especie cultivada *C. baccatum* (estrictamente *C. baccatum* var. *pendulum*), su pariente silvestre *C. baccatum* var. *baccatum* y las especies silvestres *C. praetermissum* Heise y Smith, y *C. tovarii* Eshbaugh, Smith y Nickrent (McLeod et al., 1983; Onus & Pickersgill, 2004; Ince et al., 2010) (Figura I.1). A pesar de que *C. praetermissum* tiene flores púrpuras, algunos autores la enmarcan claramente dentro este complejo, perteneciente al grupo de “flores blancas” basándose en resultados genéticos y citogenéticos (Tong & Bosland, 1999). A su vez, aunque la posición de *C. tovarii* no está totalmente clara, varios estudios insisten en incluirla también en este complejo (Tong & Bosland, 1999; Ince et al., 2010). Sin embargo, otros apuntan a que formaría un complejo aislado y relacionado con el complejo *C. annuum*, el complejo *C. baccatum* y el complejo *C. pubescens* (Carrizo García et al., 2016). La zona con mayor diversidad de la variedad silvestre *C. baccatum* var. *baccatum* se localiza en Bolivia, y dada la alta tasa de cruzamiento con la variedad domesticada *C. baccatum* var. *pendulum*, (*C. baccatum* de aquí en adelante) se apunta a que esta región fue la zona de domesticación, proceso que habría tenido lugar hace al menos 5.000 años

(Eshbaugh, 1970; Pickersgill, 2007). *Capsicum baccatum* posee flores blanquecinas pero, a diferencia de las del complejo *C. annuum*, estas se caracterizan por exhibir manchas amarillas difusas en la base de los pétalos (Fig. I.2H) (Eshbaugh, 1968, 1970). Los pimientos conocidos como “ají” normalmente hacen referencia a variedades de *C. baccatum*, destacando el “ají amarillo” o “ají escabeche” (Fig. I.2I) (DeWitt & Bosland, 1996; Chile Pepper Institute, 1997). El cultivo de esta especie tiene mayor éxito en Sudamérica y la zona de los Andes (Bosland & Votava, 2000).

La última especie domesticada es *C. pubescens*. Esta especie forma un *pool* junto con las especies silvestres *C. cardenasii* Heiser y Smith y *C. eximium* Hunziker (Eshbaugh, 1979, 1982; Krishna De, 2003) aunque análisis filogenéticos las sitúan en clados diferenciados (Carrizo García et al., 2016) (Figura I.1). La zona de domesticación se situaría en Sudamérica (Pickersgill, 2007), posiblemente en Bolivia teniendo en cuenta que la mayor correlación isoenzimática entre *C. pubescens* y sus parientes silvestres se da con accesiones de esta región (Jensen et al., 1979; Eshbaugh, 1982; McLeod et al., 1983) y que los frutos de menor tamaño (fenotipo asociado a caracteres primitivos) se localizan en Bolivia (Eshbaugh, 1979). *Capsicum pubescens* se distingue morfológicamente de las demás especies por presentar flores de color púrpura (Fig. I.2J) y frutos con semillas negras o pardas (Eshbaugh, 1993). Las variedades cultivadas más conocidas de *C. pubescens* son el “rocoto”, el “peron” y el “manzano” (Fig. I.2K) (Nuez et al., 1996; Bosland & Votava, 2000). Al igual que *C. baccatum*, el cultivo de *C. pubescens* en la actualidad está restringido a Sudamérica y, más concretamente, a la región andina (Bosland & Votava, 2000), si bien se pueden encontrar pequeños cultivos en Guatemala y México, en la zona de Chiapas (Krishna De, 2003). Esto es debido a que el cultivo de

C. baccatum presenta ciertas particularidades que hacen que sea difícil implementarlo en otras regiones, ya que requiere ambientes frescos y libres de heladas y una temporada de crecimiento prolongada, así como la propia naturaleza carnosa del fruto que lo lleva a un rápido deterioro (Krishna De, 2003).

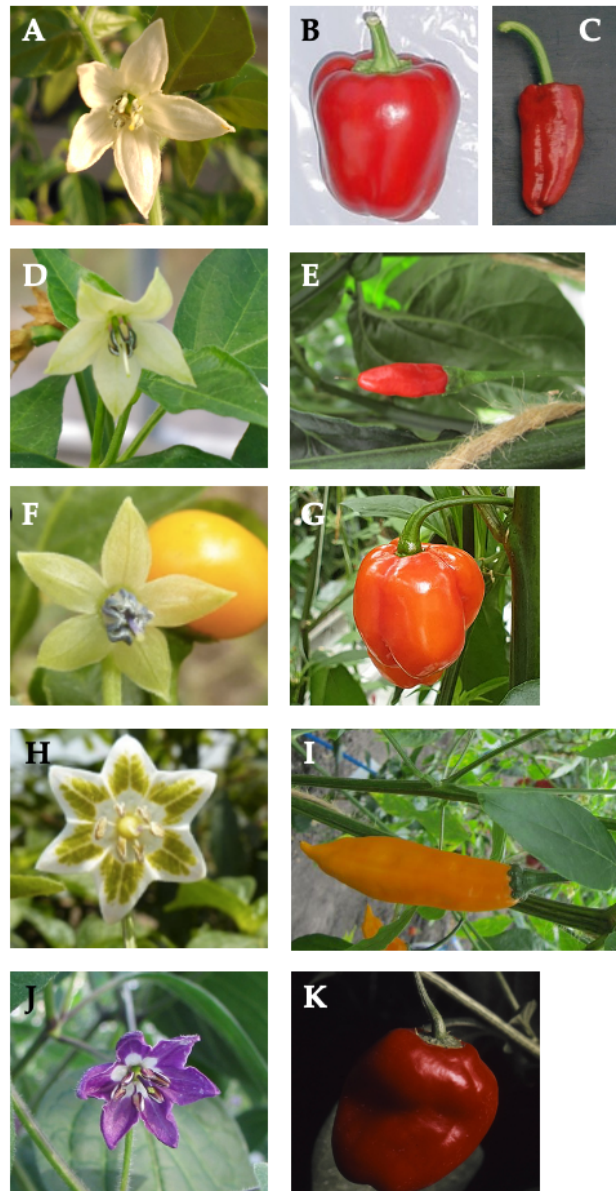


Figura I.2. Flores y frutos de las variedades más representativas de las cinco especies cultivadas del género *Capsicum*. A. Flor de *C. annuum*. B. “Morrón” (*C. annuum*). C. “Dulce italiano” (*C. annuum*). D. Flor de *C. frutescens*. E. “Tabasco” (*C. frutescens*). F. Flor de *C. chinense*. G. “Habanero” (*C. chinense*). H. Flor de *C. baccatum*. I. “Ají amarillo” (*C. baccatum*). J. Flor de *C. pubescens*. K. “Manzano” (*C. pubescens*). Fuentes de las imágenes: A, D, imágenes propias. B, C, E, G, I, J, K, (Wageningen University and Research, 2020). F, H (van Zonneveld et al., 2015).

Interés nutricional del pimiento, aplicaciones y usos e importancia económica

El pimiento presenta en su composición una serie de compuestos que le confieren un importante valor nutricional. La pungencia o picante es una de las características más destacadas del pimiento y viene dada por un tipo de metabolitos secundarios denominados capsaicinoides (Scientific Committee on Food on Capsaicin, 2002). Los capsaicinoides engloban a una familia de 25 compuestos relacionados entre sí que son análogos de alcaloides (Liu et al., 2013). En torno al 80-90% de los capsaicinoides presentes en el pimiento son capsaicina y dihidrocapsaicina (Govindarajan & Salzer, 1986; Chapa-Oliver & Mejía-Teniente, 2016). La capsaicina siempre ha tenido implicaciones en salud humana y se ha comprobado que presenta función analgésica (Simone et al., 1989; Brederson et al., 2013), antioxidante (Galano & Martínez, 2012), antiinflamatoria (Kim et al., 2003), anti-obesidad (Kang et al., 2007), cardioprotectora (Srinivasan, 2016) y anticancerígena (Bley et al., 2012; Clark & Lee, 2016).

El característico color rojo de algunos pimientos viene dado por los pigmentos capsantina y capsorubina, que pertenecen al grupo de los carotenoides (Ittah et al., 1993; Minguez-Mosquera & Hornero-Mendez, 1994; Kim et al., 2004). La fracción de color amarilla está determinada por zeaxantina, violaxantina, anteraxantina, β -criptoxantina, β -caroteno y capsotuteína (Minguez-Mosquera & Hornero-Mendez, 1994; Hornero-Méndez et al., 2000). El β -caroteno y la β -criptoxantina son formas de provitamina A (Topuz & Ozdemir, 2007). Estos compuestos tienen un importante papel en salud humana, ya que presentan capacidad antioxidante y de eliminar radicales libres, actuando en la prevención de ciertos tipos de cáncer, de enfermedades

cardiovasculares y oculares, de la degeneración de la piel y del envejecimiento (Zhang et al., 1991; Topuz & Ozdemir, 2007). Otras vitaminas presentes en alta concentración en el pimiento son el ácido ascórbico (vitamina C) y la vitamina E (tocoferol) (Baenas et al., 2019). La incorporación de ambas vitaminas a través de la dieta es muy importante, ya que presentan una actividad antioxidante particularmente alta, reduciendo los niveles de radicales libres y disminuyendo las reacciones de peroxidación, lo que se traduce también en una reducción del riesgo de enfermedades cardiovasculares y de algunos tipos de cáncer (Navarro et al., 2006). Además, en el pimiento están presentes distintos compuestos fenólicos como los flavonoides y las antocianinas, de nuevo con actividad antioxidante y anticancerígena (Sun et al., 2007; Jeong et al., 2011; Di Sotto et al., 2018;).

Todos estos compuestos le confieren al pimiento valiosas propiedades nutricionales que han llevado a posicionarlo como uno de los principales cultivos en el mundo. En alimentación, el pimiento se puede encontrar directamente fresco, seco, en conserva, en especias (como pimentón o formando parte de la mezcla del curry), en encurtidos, en aceites o en salsas picantes como el tabasco (Andrews, 1995; Thompson, 1995; Nuez et al., 1996). Pese a que los frutos son las partes más consumidas, en algunas regiones también se consumen las hojas de pimiento como verduras de hoja (Liu et al., 2013). Por especies, *C. annuum* es la especie predominante en todos los mercados del mundo, aunque otras como *C. chinense* y *C. frutescens* también son muy comercializadas en algunas regiones. *Capsicum baccatum* y *C. pubescens* tienen un cultivo minoritario restringido a América Latina (Andrews, 1995; Liu et al., 2013).

En el año 2018 se produjeron en el mundo 4.164.594 toneladas de pimientos secos y 36.771.482 toneladas de producto fresco en un área total de cultivo de 3.766.757 hectáreas (Tabla I.1) (FAOSTAT, 2018). El primer productor mundial de pimientos frescos es China, con más de 18 millones de toneladas en el año 2018 y claramente muy por encima del segundo país, México. Tras Turquía e Indonesia, España ocupa la quinta posición con más de un millón de toneladas recogidas en 20.580 hectáreas de cultivo. De hecho, de los cinco principales productores, España es el país que mejor rendimiento presenta con 619.756 hg/ha, muy por encima de los demás países, evidenciando un cultivo muy eficiente. El pimiento es un importante cultivo comercial para pequeños agricultores de muchos países asiáticos (China, India, Paquistán, Bangladesh, Tailandia o Indonesia) y de África (Egipto, Etiopía, Nigeria o Ghana) (Rai et al., 2013), por lo que estos países tienen el potencial de acumular grandes producciones, si bien precisan mucha más extensión de cultivo para obtenerlas.

El país con mayor producción de pimientos secos o deshidratados y mayor extensión dedicada a su obtención es India, con 1.808.011 de toneladas recolectadas en 781.737 hectáreas, seguido de China, Etiopía, Tailandia y Paquistán, por lo que esta forma de cultivo se concentra en Asia y África, generando riqueza en ambos continentes.

Tabla I.1. Datos de producción, área cosechada y rendimiento del cultivo de pimientos frescos y secos a nivel global y para los principales países productores, según los datos de FAOSTAT (2018). Los datos globales de producción y área cosechada son la suma de todos los países de los que se cuenta con datos y el rendimiento está calculado como el promedio de todos los países.

	Producción (t)	Área cosechada (ha)	Rendimiento (hg/ha)
Pimientos frescos			
Global	36.771.482	1.990.423	184.742
China	18.184.711	769.078	236.448
México	3.379.289	156.799	215.517
Turquía	2.554.974	91.973	277.795
Indonesia	2.542.358	308.547	82.398
España	1.275.457	20.580	619.756
Pimientos secos			
Global	4.164.594	1.776.334	23.445
India	1.808.011	781.737	23.128
China	321.290	47.753	67.281
Etiopía	294.299	152.642	19.280
Tailandia	247.010	91.675	26.944
Paquistán	148.114	65.275	22.691

En algunos de los principales países productores como México, India o Tailandia el consumo de pimiento está estrechamente ligado a su cultura, siendo su cocina globalmente conocida por el uso del pimiento como condimento (Liu et al., 2013). El pimiento se convierte por tanto en un importante activo económico pero también social y etnográfico en algunas zonas del mundo.

A pesar de que el principal uso del pimiento es en alimentación humana, algunos de sus metabolitos pueden ser empleados para otras aplicaciones. Los capsaicinoides son la base de muchos productos como champús para prevenir la caída del cabello, cremas analgésicas, aceites corporales, esprays de defensa o repelentes para ratones (Bosland, 1996; Surh, 2002; Srinivasan, 2016). Además, estos compuestos tienen capacidad antimicrobiana (Xing et al., 2006), inhibiendo por ejemplo el crecimiento del hongo *Fusarium* (Tewksbury et al., 2008), por lo que pueden ser empleados como biopesticida.

Procesos de domesticación, recursos fitogenéticos y bancos de germoplasma

El primer registro de selección artificial en plantas para su domesticación y consumo data de hace aproximadamente 10.000 años en el sudeste asiático, donde el ser humano comienza a hacer cruzamientos selectivos de plantas con características de interés (por ejemplo tallos más resistentes) y propagándolas posteriormente (Raman, 2017). La domesticación más drástica empleando estos métodos fue la del maíz, que habría tenido lugar hace 6.500 años en América a través de la selección artificial de una herbácea denominada Teosinte, que posee pequeñas espigas con granos diminutos, hasta obtener el maíz actual con espigas y granos grandes (Doebley et al., 2006). De forma similar al maíz, todas las especies vegetales cultivadas en la actualidad han sufrido grandes modificaciones a partir de sus parentales silvestres. Esto ha sucedido con el arroz, el trigo, el tomate o el girasol, que han sido seleccionados para conseguir plantas más grandes, más robustas o con frutos más grandes (Doebley et al., 2006). Estos procesos de domesticación se llevan a cabo empleando tan solo unos pocos individuos parentales, por lo que dejan una huella importante en los patrones de diversidad genética de las

plantas, perdiendo la mayoría de la variabilidad presente en las poblaciones parentales. En otras palabras, en cada generación solo se propagan las semillas de las mejores plantas según el criterio humano, creando cuellos de botella que implican que las nuevas especies o variedades tengan una diversidad muy limitada (Doebley, 1989).

Cada una de las especies cultivadas de pimiento se domesticó de forma independiente y posteriormente fue exportada a otros lugares donde se diversificó. De todas las especies de pimiento, la que mayor éxito obtuvo en el cultivo fue *C. annuum*, que tras su domesticación fue introducida en Europa a finales del siglo XV y posteriormente dispersada por los países mediterráneos hasta llegar a África, India y China (Andrews, 1995; Bosland & Votava, 2000). Las otras especies se difundieron por distintas regiones, *C. chinense* y *C. frutescens* por África y Asia; y *C. baccatum* y *C. pubescens* por Sudamérica (Bosland & Votava, 2000). En todos estos lugares el pimiento fue nuevamente seleccionado en base a características de interés propias de cada región (Ortiz et al., 2010) y esta selección incrementó de nuevo la pérdida de diversidad ya generada en sus procesos iniciales de domesticación (Hammer et al., 2003). Como resultado, actualmente es posible encontrar una gran variedad morfológica en las diferentes variedades comerciales de pimiento. Sin embargo, estas variedades son extremadamente susceptibles a nuevas presiones ambientales (Votava et al., 2005) como consecuencia de su baja diversidad genética, producto de su selección artificial. Por ello, las variedades tradicionales cobran gran importancia ya que están adaptadas a un ambiente local pero mantienen un mayor grado de variabilidad genética (Esquinas-Alcázar, 1993). Con el fin de preservar las variedades tradicionales sin necesidad de mantenerlas en su ecosistema surgen las colecciones de recursos fitogenéticos, que la Organización de las Naciones

Unidas para la Alimentación y la Agricultura (FAO, 2001) define como “cualquier material de origen vegetal con valor presente o futuro, incluyendo tanto variedades tradicionales como cultivares modernos y plantas silvestres”.

Las colecciones de recursos fitogenéticos se conservan en bancos de germoplasma, donde se recopila el material vegetal, se conserva, se caracteriza y evalúa y se intercambia con investigadores y otros bancos (Martín, 2015). En el mundo hay unos 1.750 bancos de germoplasma que recogen aproximadamente 7.400.000 accesiones (también conocidas como entradas o líneas) de recursos fitogenéticos (FAO, 2001). Los bancos que más accesiones de pimiento conservan son el *World Vegetable Center* (WVC, Shanhua, Taiwán) con 7.853 entradas, el *Plant Genetic Resources Conservation Unit by the U.S. Department of Agriculture* (PGRCU-USDA, Georgia, Estados Unidos) con 5.012, el *Seed Savers Exchange* (Iowa, Estados Unidos) con 2.321 líneas y el *Institute for Plant Genetic Resources “K. Malkov”* (IPGR, Sadovo, Bulgaria) que mantiene 1.918 líneas de pimiento (Genesys, 2020).

Los estudios de diversidad en recursos fitogenéticos son vitales para obtener una correcta identificación de los mismos que permita llevar a cabo una conservación adecuada. Mediante estos estudios se pueden realizar clasificaciones taxonómicas precisas, investigar las relaciones entre especies e inferir los procesos de diversificación de los cultivares en sus correspondientes centros (González-Pérez et al., 2014). La caracterización de los recursos fitogenéticos le añade además valor a los mismos, permitiendo su inclusión en programas de mejora (Rai et al., 2013). Estudios de diversidad genética en recursos fitogenéticos han permitido mejorar características de

interés muy variadas de las especies analizadas. Por ejemplo, gracias a un análisis con 19 marcadores microsatélite en recursos fitogenéticos del árbol *Dalbergia odorifera* (Fabaceae) se obtuvo una *core collection* que retiene la mayor variabilidad genética con el mínimo número de accesiones, lo que permitirá optimizar las labores de conservación y propagación, centrándose solo en aquellas más variables y con mayor calidad de madera (Liu et al., 2019). En tomate este mismo tipo de marcadores se puede emplear para identificar y diferenciar variedades de interés, lo que ha permitido un gran avance en el mercado de una de las especies vegetales más importantes del mundo (Iquebal et al., 2013) al evitar fraudes en las variedades cultivadas. Para analizar la diversidad de los recursos fitogenéticos se han utilizado distintas estrategias, como descriptores fenotípicos, pero en la actualidad los análisis basados en marcadores moleculares de ADN son los más empleados ya que no están sujetos a la influencia de condiciones ambientales o factores relacionados con el desarrollo de las plantas (Varshney et al., 2005). Pese a que en muchos estudios se plantea que los resultados obtenidos serán provechosos en la conservación y utilidad de los recursos fitogenéticos, no hemos encontrado en ninguno las vías por las que se pondrá en conocimiento esa valiosa información a los bancos de germoplasma. Los responsables de algunos bancos indicaron que no es habitual que se les informe de los resultados obtenidos y que solo tienen acceso a esta información cuando el propio banco participa como institución investigadora en el proyecto o si los datos se incluyen en bases de datos (IPK, CGN, comunicación personal). Por ese motivo planteamos que esa comunicación podría ser directa, informando a los bancos de germoplasma de los resultados obtenidos empleando los recursos que ellos conservan; o indirecta, a través de la publicación de

los trabajos y que los responsables de los bancos recapitularan aquellos en los que se les cita.

Herramientas de secuenciación masiva y su aplicación en colecciones de recursos fitogenéticos de pimiento

Las herramientas de secuenciación masiva (*next-generation sequencing*, NGS; *high-throughput sequencing*, HTS) han sufrido un rápido desarrollo desde su aparición en 2005 que ha revolucionado la investigación en diversas áreas de la biología, tales como la agricultura y la biomedicina, permitiendo importantes innovaciones como la selección genómica para aumentar el rendimiento de los cultivos y aumentar la precisión en diagnósticos y tratamientos de enfermedades (Wickland et al., 2017). Además de facilitar la caracterización de nuevos marcadores microsatélite, las técnicas de NGS permiten generar gran cantidad de marcadores de tipo SNPs (*single nucleotide polymorphisms*, polimorfismos de un único nucleótido). Además de que el número de SNPs es mayor que el que se utiliza cuando se trabaja con microsatélites, AFLPs (*amplified fragment length polymorphisms*, polimorfismos en la longitud de los fragmentos amplificados) o RFLPs (*restriction fragment length polymorphisms*, polimorfismos en la longitud de fragmentos de restricción), el genotipado con SNPs es un proceso más rápido y más automatizado que el llevado a cabo con los marcadores tradicionales (Davey et al., 2011).

La secuenciación de un genoma completo puede generar millones de SNPs pero para muchas aplicaciones, como aquellas que evalúan desequilibrio de ligamiento o

desequilibrio gamético (*linkage disequilibrium*, LD), tal cantidad de marcadores puede resultar innecesaria (Wickland et al., 2017). Por otra parte, la secuenciación de grandes regiones puede introducir “ruido” que limite la fiabilidad del estudio y haga más difícil obtener resultados robustos (Quince et al., 2011; Davey et al., 2013;). Por estas razones se han hecho populares las metodologías de secuenciación de representación reducida del genoma, que implican la secuenciación de un subconjunto de zonas del genoma para reducir la complejidad (Davey et al., 2011).

Existen diferentes metodologías de secuenciación basadas en la reducción del genoma mediante enzimas de restricción, muy similares entre sí pero que difieren en algunas etapas concretas del proceso (Figura I.3). En las librerías de representación reducidas (*reduced-representation libraries*, RRLs) el ADN de muchos individuos se fragmenta con una enzima de restricción de corte frecuente, se seleccionan los fragmentos por su tamaño y posteriormente se secuencian, produciendo una cobertura parcial del genoma pero a lo largo de toda su longitud. Pese a que las RRLs se empezaron a utilizar con secuenciación Sanger (Altshuler et al., 2000), su adaptación al NGS ha tenido gran éxito (Van Tassell et al., 2008; Osorio-Guarín et al., 2018). Otra aproximación que puede ser considerada una variante de las RRLs es la secuenciación de ADN asociado a un sitio de restricción (*restriction-site-associated DNA sequencing*, RAD-seq). En el RAD-seq el ADN se digiere con enzimas de restricción y sobre los trozos generados se ligan adaptadores *barcode* que permitan identificar a los individuos, lo que posibilita que todas las muestras se concentren en una única librería, ahorrando costes y esfuerzos. Posteriormente los fragmentos generados se cortan al azar para obtener longitudes óptimas para la plataforma de secuenciación y se realizan PCRs selectivas para amplificar tan solo las

secuencias que contenían un sitio de corte para la enzima de restricción (Baird et al., 2008; Davey & Blaxter, 2010). Como una variante del RAD-seq se desarrolló el RAD-seq con digestión doble (*double digest RAD-seq*, ddRAD-seq) (Peterson et al., 2012), que elimina la etapa de corte al azar de los fragmentos de ADN e incorpora una doble digestión con dos enzimas de restricción para generar mayor número de fragmentos que serán seleccionados posteriormente en base a su tamaño. Otra metodología es la de genotipado por secuenciación (*genotyping-by-sequencing*, GBS) (Elshire et al., 2011). En este caso el ADN se digiere con enzimas de corte frecuente y a los fragmentos se les ligan adaptadores *barcode* y adaptadores generales. Finalmente aquellos fragmentos que contengan ambos tipos de adaptadores y con un tamaño adecuado serán secuenciados. Una variante muy interesante de GBS es la tecnología *Diversity Arrays Technology* (DArTseq™), que se basa en la especificidad de las enzimas de restricción empleadas para generar los fragmentos que van a ser secuenciados (Killian et al., 2012). DArTseq™ utiliza la enzima *PstI*, sensible a la metilación y que, por consiguiente, no corta las regiones altamente metiladas (y por tanto asociadas a genes) y predominantemente repetitivas de los genomas de organismos vegetales (Rabinowicz et al., 1999). Dado que en la mayoría de plantas los fragmentos generados por *PstI* son demasiados para llevar a cabo un experimento eficiente, DArTseq™ aplica una combinación de *PstI* con más enzimas, siendo la mezcla enzimática específica para cada organismo (Killian et al., 2012). La selección de las enzimas empleadas por DArTseq™ para separar las secuencias de bajo número de copias de la fracción repetitiva del genoma permite seleccionar las zonas del genoma que se corresponden predominantemente con regiones transcripcionalmente activas (Diversity Arrays Technology Pty Ltd, 2020) y que por lo tanto están sometidas a presiones selectivas.

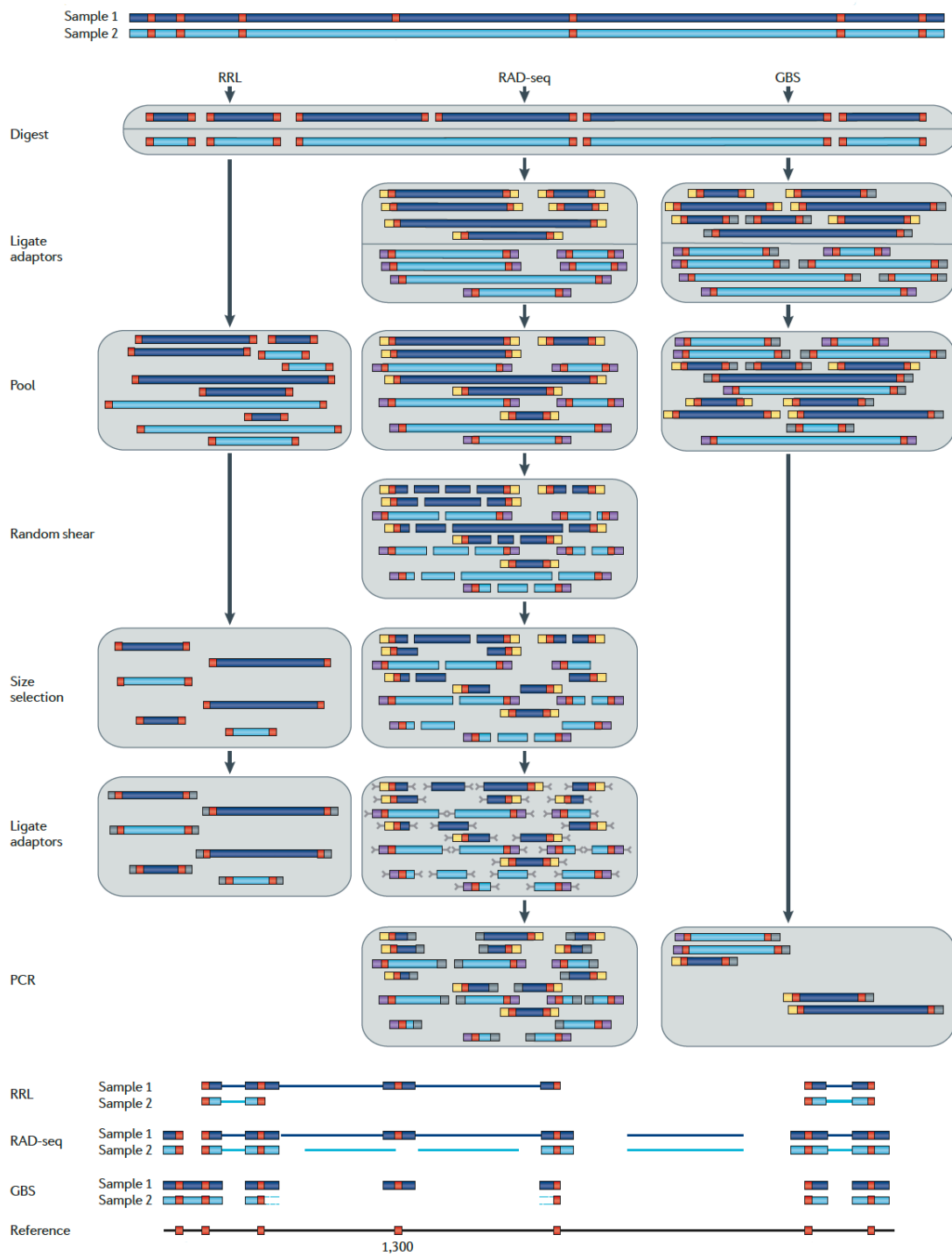


Figura I.3. Comparativa de los métodos de secuenciación masiva basados en la reducción del genoma RRL, RAD-seq y GBS. Modificada de Davey et al. (2011).

Estas herramientas han sido empleadas en pimiento para realizar análisis de estructura y diversidad genética y mapeo de QTLs (*quantitative trait loci*, loci de rasgo cuantitativo) (Eun et al., 2016; Taranto et al., 2016; Pereira-Dias et al., 2019; Tamisier et al., 2020). Por otra parte, la combinación de marcadores SNPs con información fenotípica permitió llevar a cabo estudios de asociación a nivel del genoma (*genome-wide association studies*, GWAS) con el objetivo de identificar regiones concretas del genoma que estén implicadas en variaciones morfológicas o funcionales (Han et al., 2018; Solomon et al., 2019; Tamisier et al., 2020). Para los análisis GWAS se revela muy interesante la utilización de datos provenientes de DArTseq™, ya que se estarían seleccionando solamente las secciones del genoma asociadas a genes expresados y desechando las zonas repetitivas, que no van a proporcionar información de interés en este tipo de estudios y que en el caso del pimiento comprenden aproximadamente el 81% del total (Qin et al., 2014).

Para aprovechar todo el potencial de estas herramientas, en muchos casos es necesario contar con genomas de referencia contra los que mapear las secuencias. En pimiento hay disponibles varios genomas ensamblados (Kim et al., 2014; Qin et al., 2014; Hulse-Kemp et al., 2018). El genoma empleado en esta Tesis Doctoral (Hulse-Kemp et al., 2018) es el más moderno de los disponibles. Es el más ordenado y continuo de todos y emplea un ensamblaje por fases, es decir, representa de manera separada los cromosomas homólogos al menos a escala de genes y esto permite comprender correctamente la expresión alelo-específico y la heterocigosidad (Weisenfeld et al., 2018). Esto supone una ventaja cuando se llevan a cabo análisis de asociación.

Para el trabajo bioinformático con grandes conjuntos de datos generados con NGS, se hicieron necesarias otras adaptaciones y mejoras. Es el caso de los programas que implementan análisis clúster, que se han venido usando durante años en estudios poblacionales. Entre ellos destaca STRUCTURE (Pritchard et al., 2000) como el *software* más utilizado. Sin embargo, con el empleo de un número cada vez mayor de marcadores, se hizo necesario crear nuevos programas de análisis clúster que pudieran procesar mayor número de datos en un tiempo aceptable. Pese a que STRUCTURE se puede utilizar con grandes conjuntos de datos, el tiempo necesario es mucho mayor que el de los nuevos programas como fastSTRUCTURE (Raj et al., 2014). Stift et al. (2019) realizaron una comparativa entre todos los programas disponibles para realizar este tipo de análisis y concluyeron que, pese a que STRUCTURE era el que aportaba mejores resultados, el tiempo de procesamiento era mucho mayor (varias horas por réplica con STRUCTURE frente a un par de segundos con fastSTRUCTURE). Otro programa empleado a menudo con grandes ficheros de datos es ADMIXTURE (Alexander et al., 2009). Sin embargo, este presenta un sesgo importante cuando se trabaja con datos perdidos y en estos casos tiende a dar como resultado una estructura genética no robusta. Por su parte, con fastSTRUCTURE el riesgo de sesgo es menor aún con datos perdidos (Stift et al., 2019).

Objetivos

Los objetivos planteados en esta Tesis Doctoral han sido los siguientes:

- I. Analizar la diversidad genética de una colección de pimientos (*Capsicum* spp.) procedentes de la región andina, Península Ibérica y la región mediterránea con un panel de marcadores microsatélite, marcadores moleculares asociados a genes o QTLs de resistencia a patógenos y marcadores genéticos relacionados con el carácter pungente de pimiento.
- II. Evaluar la colección mediante la tecnología de Genotyping-by-Sequencing (GBS) con el fin de establecer el grado de variabilidad, la estructura genética del conjunto y las relaciones ancestrales entre las distintas regiones geográficas evaluadas.
- III. Caracterizar fenotípicamente parte de la colección, a nivel de planta y fruto, mediante descriptores convencionales y herramientas fenómicas de alto rendimiento.
- IV. Realizar un análisis de asociación a nivel de genoma (GWAS) empleando el panel de SNPs generados a través de GBS y el conjunto de caracteres fenotípicos, con el fin de identificar regiones genómicas de interés agro-morfológico.

Capítulo 1

Análisis de la diversidad genética de recursos fitogenéticos de pimiento

(Capsicum spp.)

Capítulo 1.1

Estudio de la variabilidad y estructura

genética mediante marcadores

microsatélite

Introducción

La correcta caracterización de las accesiones de las colecciones de germoplasma y el estudio de las relaciones entre ellas es el punto de partida para una apropiada gestión de los recursos fitogenéticos. Los estudios de diversidad permiten evitar la erosión genética de los recursos y preservarlos, por lo que son fundamentales en especies cultivadas como el pimiento.

Las herramientas moleculares permiten evaluar la variabilidad genética de las especies, a fin de identificar las mejores variedades para su selección e inclusión en programas de mejora. Además, la identificación molecular es más fiable que la morfológica, que es la llevada a cabo habitualmente en los bancos de germoplasma, dada la plasticidad que pueden presentar los caracteres morfológicos, a menudo sometidos a la influencia de factores ambientales (Ince et al., 2010). La discriminación taxonómica basada en caracteres morfológicos se hace aún más difícil dentro de las especies del complejo *C. annuum*, integradas en un continuo morfológico (Walsh & Hoot, 2001; van Zonneveld et al., 2015) que a menudo puede llevar a identificaciones erróneas.

Las distintas especies de *Capsicum* se han caracterizado tradicionalmente empleando isoenzimas (Loaiza-Figueroa et al., 1989), análisis citogenéticos (Pickersgill, 1971), polimorfismos en la longitud de fragmentos de restricción (*restriction fragment length polymorphisms*, RFLPs; Prince et al., 1992), polimorfismos en la longitud de los fragmentos amplificados (*amplified fragment length polymorphisms*, AFLPs; Guzman et al., 2005), ADN polimórfico amplificado al azar (*random amplified polymorphic DNA*, RAPDs; Adetula, 2006), amplificación directa o indirecta de regiones minisatélite mediante PCR

(direct or indirected amplification of minisatellite region DNA amplified using the polymerase chain reaction, DAMD-PCR; Ince et al., 2009), polimorfismos en la amplificación secuencia-específicos (*sequence-specific amplification polymorphisms*, S-SAPs; Tam et al., 2009), secuencias polimórficas amplificadas y digeridas (*cleaved amplified polymorphic sequences*, CAPSs; Di Dato et al., 2015) y marcadores microsatélite (*simple sequence repeats*, SSRs o STRs; Ibiza et al., 2012).

Respecto al análisis de la diversidad y estructura genética de los recursos fitogenéticos de pimiento, los microsatélites han sido muy utilizados al ser marcadores codominantes, altamente polimórficos, con modo de herencia biparental y por lo tanto con alto poder discriminatorio, especialmente entre especies (Varshney et al., 2005). En pimiento, los *loci* microsatélite se han empleado en los últimos años en estudios que buscan determinar las relaciones inter e intraespecíficas en diversas colecciones de recursos fitogenéticos de la India (Yumnam et al., 2012; Rai et al., 2013), de España (González-Pérez et al., 2014) o de Brasil (Carvalho et al., 2017), entre otros. Los estudios con microsatélites permitieron identificar accesiones duplicadas en los bancos de germoplasma y analizar el grado de diversidad en *C. annuum* (Hanacek et al., 2009) y en las otras cuatro especies cultivadas (Ibiza et al., 2012). Las relaciones encontradas muestran una agrupación principalmente en base a la especie pero también según características morfológicas y el origen geográfico de las accesiones (González-Pérez et al., 2014; Nicolai et al., 2013; Rai et al., 2013). En el presente trabajo se evaluará mediante marcadores microsatélite la diversidad de tres colecciones: una de la zona Andina, en la cual se originó el género *Capsicum*; otra de la Península Ibérica, zona de entrada del pimiento en Europa y que constituye un importante centro de diversificación

secundario; y una última colección de países de la cuenca mediterránea, a través de la que diversificó el pimiento posteriormente. El análisis de los recursos fitogenéticos de pimiento en sus centros de diversificación permitirá identificar las accesiones más valiosas para programas de mejora. Esta información posibilitaría además desarrollar estrategias de conservación adecuadas en los bancos de germoplasma.

Material y métodos

Material vegetal

Para el análisis con marcadores microsatélite se utilizaron 230 líneas de pimiento de las cinco especies cultivadas (señaladas con una "X" en la columna "Microsatélites" de la Tabla 1.1). De ellas, 151 son *C. annuum* (cuatro de ellas *C. annuum* var. *glabriusculum*), 28 *C. chinense*, 24 *C. baccatum*, 16 *C. pubescens* y 11 *C. frutescens*.

Las accesiones empleadas provienen de los principales bancos de germoplasma mundiales con colecciones fitogenéticas de pimiento. Dichos bancos, ordenados de mayor a menor en función del número de líneas que aportan a nuestro estudio, son el Banco de Germoplasma de Hortícolas de Zaragoza (BGHZ) (Zaragoza, España), el USDA-ARS *Plant Genetic Resources Conservation Unit* (USDA) (Georgia, Estados Unidos), el *Center for Genetic Resources* (CGN) (Wageningen, Países Bajos), el *Leibniz Institute of Plant Genetics and Crop Plant Research* (IPK) (Gatersleben, Alemania), el Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV) (Valencia, España), el *Banco Português de Germoplasma Vegetal* (BPGV) (Braga, Portugal), el Centro de

Investigaciones Agrarias de Mabegondo (CIAM) (A Coruña, España) y el *Institut National de la Recherche Agronomique* (INRA) (Francia).

Las 230 líneas se agruparon en tres colecciones de recursos fitogenéticos según su zona de origen. La primera colección está formada por variedades de la región andina (Perú, Ecuador y Bolivia), en la cual tiene su origen el pimiento. La segunda colección está formada por líneas de la Península Ibérica (España y Portugal y sus archipiélagos), zona de entrada del pimiento en Europa y que por lo tanto constituye un importante centro de diversificación secundaria. La tercera y última colección está compuesta por líneas de pimiento de la región mediterránea, zona de expansión de *C. annuum* tras su entrada por España. Los datos de SSR de 96 líneas de *C. annuum* ibéricas han sido cedidos por la Dra. Cristina Silvar para la elaboración de esta Tesis Doctoral. Para la colección andina ya se han publicado resultados parciales (Silvar & García-González, 2016), con 16 de los 28 marcadores empleados en esta Tesis. En el presente capítulo se completan los datos analizando los restantes 12 marcadores (i. e. *Hpms1-281*, *GPMS178*, *Hpms1-143*, *EPMS472*, *Hpms2-45*, *HpmsE120*, *Hpms1-6*, *EPMS426*, *HpmsE013*, *GPMS159*, *Hpms2-2* y *HpmsE075*). Para las líneas andinas CGN22783, CGN21527, PM663, BGV008233, BGV013703, CGN23768, CAP357, CGN17045, CGN19211, BGV013699 y PI 593920 se analizaron *de novo* todos los marcadores y estos datos se emplearon en lugar de los publicados en Silvar & García-González (2016). El conjunto de datos completo de la colección andina (16 marcadores publicados previamente y 12 marcadores analizados *de novo*) se empleó en los análisis del presente capítulo junto con los datos de la Península Ibérica cedidos, los 28 marcadores obtenidos *de novo* para las accesiones de la región mediterránea y para las once accesiones andinas citadas anteriormente.

Para cada accesión se presentan datos de la especie a la que pertenece, el nombre común, los datos de pasaporte (localidad, provincia, país, latitud y longitud de la zona de recolección) y el banco de germoplasma del que proceden. Toda esta información ha sido proporcionada por los propios bancos de germoplasma. Además se indica para cada línea con una "X" si se generaron datos de microsatélites, marcadores de enfermedad y pungencia (capítulo 1.2), GBS (capítulo 2), fenotipado de planta y fruto y GWAS (capítulo 3) (Tabla 1.1).

Las semillas procedentes de los bancos de germoplasma se sembraron en semilleros individuales, en una mezcla de substrato orgánico y perlita en relación 2:1. Las plantas se cultivaron en cámara de cultivo con fotoperiodo de 16 h luz/8 h de oscuridad, temperaturas de 25°C /18°C (día/noche) y riego periódico hasta obtener plantas lo suficientemente grandes como para poseer el tejido necesario para la extracción de ADN.

Tabla 1.1. Accesiones de *Capsicum* spp. empleadas en la presente Tesis Doctoral.

Accesión	Especie	Nombre común	Localidad	Provincia	País	Latitud	Longitud	Banco	Microsatélites	Marcadores enfermedad	GBS	Fenotipado	GWAS
BGHZ0082	<i>C. annuum</i>	Guindilla de jardín	El Paso	Santa Cruz de Tenerife	España	28°39'01" N	17°52'42" O	BGHZ	X	-	X	X	X
BGHZ0097	<i>C. annuum</i>	Guindilla blanca	Chinchón	Madrid	España	40°08'33" N	03°25'16" O	BGHZ	X	-	X	X	X
BGHZ0099	<i>C. annuum</i>	Guindilla murciana	Alhama de Murcia	Murcia	España	37°51'13" N	01°25'24" O	BGHZ	X	-	X	-	-
BGHZ0102	<i>C. annuum</i>	Pimiento pobrot	Tremp	Lleida	España	42°10'02" N	0°53'44" E	BGHZ	X	X	X	X	X
BGHZ0104	<i>C. annuum</i>	Pimiento del piquillo	Lodosa	Navarra	España	42°25'27" N	0°20'43" O	BGHZ	X	-	X	-	-
BGHZ0106	<i>C. annuum</i>	Pimiento ocal	Malpartida de Plasencia	Cáceres	España	39°58'33" N	06°02'32" O	BGHZ	X	-	X	X	X
BGHZ0113	<i>C. annuum</i>	Pimiento bocarrón	Azaña	Badajoz	España	38°15'28" N	05°40'30" O	BGHZ	X	X	X	X	X
BGHZ0179	<i>C. annuum</i>	Pimiento picante	Toro	Zamora	España	41°31'34" N	05°23'41" O	BGHZ	X	X	X	X	X
BGHZ0181	<i>C. annuum</i>	Pimiento choricero	Gernika-Lumo	Vizcaya	España	43°18'52" N	02°40'53" O	BGHZ	X	-	X	X	X
BGHZ0206	<i>C. annuum</i>	Pimiento del país	Valverde del Camino	Huelva	España	37°34'25" N	06°45'08" O	BGHZ	X	X	X	X	X
BGHZ0208	<i>C. annuum</i>	Pimiento rotoño	Rota	Cádiz	España	36°37'17" N	06°21'42" O	BGHZ	X	-	X	-	-
BGHZ0213	<i>C. annuum</i>	Pebrre Blanc de Muro	Manacor	Islas Baleares	España	39°54'12" N	03°12'38" E	BGHZ	X	X	X	X	X
BGHZ0215	<i>C. annuum</i>	Guindilla	Fresno de la Vega	Leon	España	42°20'43" N	05°32'05" O	BGHZ	X	-	X	-	-
BGHZ0224	<i>C. annuum</i>	Pimiento morrón	Calahorra	La Rioja	España	42°18'10" N	01°57'35" O	BGHZ	X	X	X	X	X
BGHZ0280	<i>C. annuum</i>	Guindilla	Chozas de Canales	Toledo	España	40°05'59" N	04°02'32" O	BGHZ	X	-	X	X	X
BGHZ0282	<i>C. annuum</i>	Pimiento dulce pequeño	Villacarrillo	Jaén	España	38°06'37" N	03°05'06" O	BGHZ	X	X	X	X	X
BGHZ0283	<i>C. annuum</i>	Cometa	Rute	Córdoba	España	37°19'35" N	04°22'08" O	BGHZ	X	X	X	-	-
BGHZ0284	<i>C. annuum</i>	Guindilla amarilla	Ibarra	Guipúzcoa	España	43°07'59" N	02°03'38" O	BGHZ	X	-	X	X	X
BGHZ0287	<i>C. annuum</i>	Pimiento casarejo	Nájera	La Rioja	España	42°25'02" N	02°44'04" O	BGHZ	X	X	X	-	-
BGHZ0331	<i>C. annuum</i>	Pimiento de Torroella	Torroella de Montgrí	Girona	España	42°02'40" N	03°07'38" E	BGHZ	X	X	X	-	-
BGHZ0337	<i>C. annuum</i>	Pimiento Valenciano	Cullera	Valencia	España	39°09'59" N	00°15'12" O	BGHZ	X	X	X	X	X
BGHZ0338	<i>C. annuum</i>	Pimiento Miguelino	Málaga	Málaga	España	36°43'18" N	04°25'07" O	BGHZ	X	X	X	X	X
BGHZ0413	<i>C. annuum</i>	Pimiento de Alboloduy	Alboloduy	Almería	España	37°02'05" N	02°37'14" O	BGHZ	X	X	X	X	X
BGHZ0418	<i>C. annuum</i>	Pimiento largo de Reus	Vinyols i els Arcs	Tarragona	España	41°06'56" N	01°02'25" E	BGHZ	X	X	X	X	X
BGHZ0447	<i>C. annuum</i>	Guindilla	Gea de Albarracín	Teruel	España	40°24'43" N	01°20'49" O	BGHZ	X	-	X	X	X
BGHZ0463	<i>C. annuum</i>	Pimiento negro	Torre-Pacheco	Murcia	España	37°46'51" N	00°52'03" O	BGHZ	X	X	X	X	X
BGHZ0775	<i>C. annuum</i>	Pimiento largo morro de vaca	La Roda	Albacete	España	39°12'30" N	02°09'29" O	BGHZ	X	X	X	X	X
BGHZ0782	<i>C. annuum</i>	Guindilla bodega amarilla	Los Cortijos	Ciudad Real	España	39°19'00" N	04°04'00" O	BGHZ	X	X	X	X	X
BGHZ0958	<i>C. annuum</i>	Pimiento loras	Herrera de Pisuerga	Palencia	España	42°35'45" N	04°19'44" O	BGHZ	X	X	X	X	X
BGHZ0966	<i>C. annuum</i>	Picante	Daimiel	Ciudad Real	España	39°04'07" N	03°36'37" O	BGHZ	X	-	X	-	-
BGHZ0969	<i>C. annuum</i>	Guindilla/Bodilla	Pastrana	Guadalajara	España	40°25'03" N	02°55'24" O	BGHZ	X	-	X	X	X
BGHZ0971	<i>C. annuum</i>	Pimiento rizado	Colomera	Granada	España	37°22'32" N	03°42'47" O	BGHZ	X	X	X	X	X
BGHZ1109	<i>C. annuum</i>	Pimiento del terreno	Perlora	Asturias	España	43°34'53" N	05°44'55" O	BGHZ	X	X	X	X	X
BGHZ1114	<i>C. annuum</i>	Pebreteras	Betanuria	Las Palmas	España	28°25'25" N	14°03'21" O	BGHZ	X	-	X	X	X
BGHZ1174	<i>C. annuum</i>	Pimiento grande de plaza	Remolinos	Zaragoza	España	41°50'21" N	01°10'26" O	BGHZ	X	-	X	X	X
BGHZ1374	<i>C. annuum</i>	Pimiento colorado picante	Casas Ibáñez	Albacete	España	39°17'18" N	01°28'12" O	BGHZ	X	-	X	-	-
BGHZ1375	<i>C. annuum</i>	Pimiento bellotero	Candeleda	Ávila	España	40°09'13" N	05°14'16" O	BGHZ	X	X	X	X	X
BGHZ1519	<i>C. annuum</i>	Pimiento choricero de feir	Ribamontán al Mar	Cantabria	España	43°27'42" N	03°43'06" O	BGHZ	X	X	X	-	-
BGHZ1521	<i>C. annuum</i>	Pimiento comicabra	Jaraz de la Vera	Cáceres	España	40°03'40" N	05°45'21" O	BGHZ	X	X	X	X	X
BGHZ1753	<i>C. annuum</i>	Guindilla larguilla	Muñoveros	Segovia	España	41°10'26" N	03°57'02" O	BGHZ	X	-	X	X	X

X*, datos obtenidos de Silvar & García-González (2016).

Tabla 1.1. Accesiones de *Capsicum* spp. empleadas en la presente Tesis Doctoral. (continuación).

Accesión	Especie	Nombre común	Localidad	Provincia	País	Latitud	Longitud	Banco	Microsatélites	Marcadores enfermedad	GBS	Fenotipado	GWAS
BGHZ1756	<i>C. annuum</i>	Pimiento gordo de tirabeque	Herrera del Duque	Badajoz	España	39°10'00" N	0°52'54" O	BGHZ	X	X	X	-	-
BGHZ2154	<i>C. annuum</i>	Guindilla de casa	Ampuero	Cantabria	España	43°20'40" N	0°32'15" O	BGHZ	X	X	X	X	X
BGHZ2169	<i>C. annuum</i>	Pimiento pimiento blanco	Cadizar	Granada	España	36°56'46" N	0°31'04" O	BGHZ	X	-	X	-	-
BGHZ2546	<i>C. annuum</i>	Pimiento de bola	Puente la Reina	Navarra	España	42°40'24" N	0°48'46" O	BGHZ	X	X	X	X	X
BGHZ2561	<i>C. annuum</i>	Pimiento morrón corto	Asorga	León	España	42°27'32" N	0°49'32" O	BGHZ	X	X	X	X	X
BGHZ3269	<i>C. annuum</i>	Bollita picante	Carrascalejo	Ciudad Real	España	39°38'48" N	0°51'30" O	BGHZ	X	-	X	-	-
BGHZ3337	<i>C. annuum</i>	Prebe de bayuna d'Eivissa	Ibiza	Islas Baleares	España	38°54'28" N	0°12'6" E	BGHZ	X	X	X	X	X
BGHZ3388	<i>C. annuum</i>	Pimiento	Cangas del Narcea	Asturias	España	43°01'36" N	0°43'55" O	BGHZ	X	X	X	-	-
BGHZ5912	<i>C. annuum</i>	Pimiento morrón de conserva	Elche	Alicante	España	38°16'07" N	0°41'48" O	BGHZ	X	X	X	X	X
BR113-05	<i>C. annuum</i>	Blanco rosál	O Rosal	Pontevedra	España	41°56'29" N	0°45'07" O	CIAM	X	X	X	X	X
C-40	<i>C. annuum</i>	Pimiento del Bierzo	Bembibre	León	España	42°56'38" N	0°42'55" O	BGHZ	X	-	X	-	-
C-41	<i>C. annuum</i>	Bola de Murcia	San Javier	Murcia	España	37°48'26" N	0°45'00" O	BGHZ	X	-	X	-	-
C-43	<i>C. annuum</i>	Pimiento de Luzea	Baracaldo	Vizcaya	España	43°15'00" N	2° 59'00" O	BGHZ	X	-	X	-	-
CAP1661	<i>C. annuum</i>	De Bayna	Manresa	Barcelona	España	41°43'35" N	0°49'45" E	IPK	X	X	X	X	X
CGN17254	<i>C. annuum</i>	De tres cantos/Odecuerno	Alzira	Valencia	España	39°09'36" N	0°28'12" O	CGN	X	X	X	X	X
CGN17235	<i>C. annuum</i>	Guindillas	Onda	Castellón	España	39°57'36" N	0°16'58,8" O	CGN	X	X	X	X	X
CGN20507	<i>C. annuum</i>	Pimiento de 4 cascos	Castro del Río	Córdoba	España	37°45'00" N	4°30'00" O	CGN	X	X	X	X	X
CGN23293	<i>C. annuum</i>	Pimiento blanquillo	Véjar de la frontera	Cádiz	España	36°15'00" N	5°55'58,8" O	CGN	X	X	X	X	X
COI2B	<i>C. annuum</i>	Couto	Narón	A Coruña	España	43°31'29" N	0°48'10" O	CIAM	X	X	X	-	-
OI27-05	<i>C. annuum</i>	Ombra	Ourense	Ourense	España	41°53'18" N	0°28'17" O	CIAM	X	X	X	X	X
PA-145	<i>C. annuum</i>	Padrón	Padrón	A Coruña	España	42°44'20" N	8° 39' 37" O	CIAM	X	X	X	X	X
P309-05	<i>C. annuum</i>	Peñeira	Ribadeo	Lugo	España	43°33'06" N	0°7'04'00" O	CIAM	X	X	X	X	X
0308-BPGV	<i>C. annuum</i>	Pimiento	Guimarães	Beira	Portugal	41° 30'00" N	0°8' 19'00" O	BPGV	X	X	X	-	-
04793-BPGV	<i>C. annuum</i>	Pimiento	Guarda	Guarda	Portugal	40° 47'00" N	0°7' 20'00" O	BPGV	X	X	X	-	-
05572-BPGV	<i>C. annuum</i>	Pimiento picante	S. João da Pesqueira	Viscu	Portugal	41° 09'00" N	0°7' 28'00" O	BPGV	X	X	X	-	-
05740-BPGV	<i>C. annuum</i>	Pimiento	Castro Marim	Faro	Portugal	37° 22'00" N	0°7' 27'00" O	BPGV	X	X	X	-	-
06235-BPGV	<i>C. annuum</i>	Pimiento verde	Satão	Viscu	Portugal	40° 47'00" N	0°7' 41'00" O	BPGV	X	X	X	-	-
07149-BPGV	<i>C. annuum</i>	Pimiento	Vera Cruz	Évora	Portugal	38° 13'00" N	0°7' 40'00" O	BPGV	X	X	X	-	-
07822-BPGV	<i>C. annuum</i>	Pimiento	Alandroal	Évora	Portugal	38° 44' 26" N	0°7' 14' 23" O	BPGV	X	X	X	-	-
11355-BPGV	<i>C. annuum</i>	Pimiento	Azambuja	Lisboa	Portugal	39° 13' 15,775" N	0°8' 53' 30,945" O	BPGV	X	X	X	-	-
11688-BPGV	<i>C. annuum</i>	Pimentão	Abrantes	Santarém	Portugal	39° 20' 34,954" N	0°8' 04' 09,827" O	BPGV	X	X	X	-	-
11805-BPGV	<i>C. annuum</i>	Pimiento	Marvão	Portalegre	Portugal	39° 21' 57,30" N	0°7' 20' 09,66" O	BPGV	X	X	X	-	-
11849-BPGV	<i>C. annuum</i>	Pimentões	Sousel	Portalegre	Portugal	38° 57' 34,44" N	0°7' 45' 24,42" O	BPGV	X	X	X	-	-
12348-BPGV	<i>C. annuum</i>	Noras	Miranda do Douro	Bragança	Portugal	41° 28' 15,341" N	0°6' 21' 20,255" O	BPGV	X	X	X	-	-
13027-BPGV	<i>C. annuum</i>	Pimiento	Seia	Guarda	Portugal	40° 17' 33,3" N	0°7' 46' 46,5" O	BPGV	X	X	X	-	-
13113-BPGV	<i>C. annuum</i>	Pimiento	Gouveia	Guarda	Portugal	40° 32' 25,4" N	0°7' 37' 41,5" O	BPGV	X	X	X	-	-
12950-BPGV	<i>C. annuum</i>	Pimiento	Oliveira de Azeméis	Aveiro	Portugal	40° 47' 45,798" N	0°8' 29' 20,604" O	BPGV	X	X	X	-	-
CAP1371	<i>C. annuum</i>	Pimiento de Alvalade	-	-	Portugal	-	-	IPK	X	X	X	X	X
CAP1539	<i>C. annuum</i>	Piri-Piri	Loulé (mercado)	Algarve	Portugal	37° 07' 59" N	8° 58' 00" O	IPK	X	X	X	X	X
CGN16805	<i>C. annuum</i>	Coeur de Boeuf	-	-	Portugal	-	-	CGN	X	X	X	X	X

X*, datos obtenidos de Silvar & García-González (2016).

Tabla 1.1. Acciones de *Capsicum* spp. empleadas en la presente Tesis Doctoral. (continuación).

Accesión	Especie	Nombre común	Localidad	Provincia	País	Latitud	Longitud	Banco	Microsatélites	Marcadores enfermedad	GBS	Fenotipado	GWAS
CGN16897	<i>C. annuum</i>	Cruxo de Plaza	-	-	Portugal	-	-	CGN	X	X	X	X	X
CGN16899	<i>C. annuum</i>	Sequeira Mendes	-	-	Portugal	-	-	CGN	X	X	X	X	X
CGN21475	<i>C. annuum</i>	Bola de Elvas	-	-	Portugal	-	-	CGN	X	X	X	X	-
CGN22161	<i>C. annuum</i>	Quintal d'Alger	-	-	Portugal	-	-	CGN	X	X	X	X	X
CGN16894	<i>C. annuum</i>	Bastidon	-	-	Francia	-	-	CGN	X	X	X	-	-
CGN17172	<i>C. annuum</i>	Corno di Toro Rouge	-	-	Francia	-	-	CGN	X	X	X	X	X
CGN21486	<i>C. annuum</i>	Doux d'Espagne	-	-	Francia	-	-	CGN	X	X	X	X	X
CGN23239	<i>C. annuum</i>	Rouge Long Ordinaire/Lange Rode	-	-	Francia	-	-	CGN	X	X	X	X	X
CGN23244	<i>C. annuum</i>	Sucette de Provence	-	-	Francia	-	-	CGN	X	X	X	-	-
CAP68	<i>C. annuum</i>	Peperoni	Marina di Camerota	Salerno	Italia	40°01'59" N	15°16'59" E	IPK	X	X	X	X	X
CAP242	<i>C. annuum</i>	Peperoni	Asprigliano	Cosenza	Italia	39°13'59" N	16°19'59" E	IPK	X	X	X	X	X
CAP913	<i>C. annuum</i>	-	Castellina Marittima	Pisa	Italia	43°24'59" N	10°34'59" E	IPK	X	X	X	X	X
CGN16890	<i>C. annuum</i>	Bado di Salana	-	-	Italia	-	-	CGN	X	X	X	X	X
CGN16881	<i>C. annuum</i>	Cuneo	-	-	Italia	-	-	CGN	X	X	X	X	X
CGN22138	<i>C. annuum</i>	Lungo di Chivasso o Corno di Toro Giallo	-	-	Italia	-	-	CGN	X	X	X	-	-
CGN21581	<i>C. annuum</i>	Slovenskaja 5	-	-	Eslovenia	-	-	CGN	X	X	X	X	X
CGN16807	<i>C. annuum</i>	Alnapapprika	-	-	Hungría	-	-	CGN	X	X	X	X	X
CGN16817	<i>C. annuum</i>	Paradicsomalakur Zold Szentesi	-	-	Hungría	-	-	CGN	-	-	X	X	X
CGN17147	<i>C. annuum</i>	Kuppos	-	-	Hungría	-	-	CGN	-	-	X	X	X
CGN17227	<i>C. annuum</i>	Bodrogalmi	-	-	Hungría	-	-	CGN	-	-	X	X	X
CGN22127	<i>C. annuum</i>	Hatvani	-	-	Hungría	-	-	CGN	-	-	X	X	X
CGN16835	<i>C. annuum</i>	-	-	-	Hungría	-	-	CGN	-	-	X	X	X
CGN17192	<i>C. annuum</i>	Babura Somborka	-	-	Serbia/Montenegro	-	-	CGN	X	X	X	X	X
CGN17193	<i>C. annuum</i>	Tursijara IA	-	-	Serbia/Montenegro	-	-	CGN	X	X	X	-	-
CGN22183	<i>C. annuum</i>	Zinka	-	-	Serbia/Montenegro	-	-	CGN	X	X	X	X	X
CGN23235	<i>C. annuum</i>	Karadjordjeva 5-F	-	-	Serbia/Montenegro	-	-	CGN	X	X	X	-	-
CAP1049	<i>C. annuum</i>	Spec kapi	Shkallnur	Durrës	Albania	41°17'24" N	19°31'00" E	IPK	X	X	X	X	X
CAP1051	<i>C. annuum</i>	Spec d'egges	Shkallnur	Durrës	Albania	41°17'24" N	19°31'00" E	IPK	X	X	X	X	X
CAP491	<i>C. annuum</i>	Corbadzijski	-	-	Bulgaria	-	-	IPK	-	-	X	X	X
CAP1254	<i>C. annuum</i>	Sipka Zitronen	-	-	Bulgaria	-	-	IPK	-	-	X	X	X
CGN16808	<i>C. annuum</i>	Belaja Kapija	-	-	Bulgaria	-	-	CGN	-	-	X	X	X
CGN17214	<i>C. annuum</i>	Zelen	-	-	Bulgaria	-	-	CGN	-	-	X	X	X
CGN16833	<i>C. annuum</i>	-	-	-	Turquia	-	-	CGN	X	X	X	X	X
CGN17010	<i>C. annuum</i>	Dolma Biber	-	-	Turquia	-	-	CGN	X	X	X	X	X
CGN17190	<i>C. annuum</i>	Biber Dolmalik Dortlu	-	-	Turquia	-	-	CGN	X	X	X	X	X
CGN17191	<i>C. annuum</i>	Biber Charliston	-	-	Turquia	-	-	CGN	X	X	X	X	X
CGN22154	<i>C. annuum</i>	-	-	-	Turquia	-	-	CGN	X	X	X	X	X
CGN24358	<i>C. annuum</i>	Long Semi-Hot Myra	-	-	Turquia	-	-	CGN	X	X	X	X	X
CGN16916	<i>C. annuum</i>	Bikora	-	-	Israel	-	-	CGN	X	X	X	X	X
CGN16917	<i>C. annuum</i>	Zohar	-	-	Israel	-	-	CGN	X	X	X	X	X

X*, datos obtenidos de Silvar & García-González (2016).

Tabla 1.1. Accesiones de *Capsicum* spp. emplantadas en la presente Tesis Doctoral. (continuación).

Accesión	Especie	Nombre común	Localidad	Provincia	País	Latitud	Longitud	Banco	Microsatélites	Marcadores enfermedad	GBS	Fenotipado	GWAS
CGN21480	<i>C. annuum</i>	Dalia	-	-	Israel	-	-	CCN	X	X	X	X	X
CGN22825	<i>C. annuum</i>	Otli	-	-	Israel	-	-	CCN	X	X	X	X	X
CAP977	<i>C. annuum</i>	Fifili achmar	Gebba	Degache	Túnez	35°14'5.957" N	11°07'4.465" E	IPK	X	X	X	X	X
CAP979	<i>C. annuum</i>	Fifili achmar	Guetaya	Keblili	Túnez	33°25'36.204" N	8°48'39.149" E	IPK	X	X	X	X	X
CAP1538	<i>C. annuum</i>	-	Kairouan	Kairouan	Túnez	35°40'27.466" N	10°05'58.38" E	IPK	X	X	X	X	X
CAP606	<i>C. annuum</i>	-	Tajarhi	Fezzan	Libia	24°21'44.946" N	14°28'20.784" E	IPK	X	X	X	-	-
CAP607	<i>C. annuum</i>	-	Az Zighan	Fezzan	Libia	27°24'11.603" N	14°55'12.075" E	IPK	-	-	X	X	X
CAP614	<i>C. annuum</i>	-	Imassah	Fezzan	Libia	26°23'43.202" N	19°47'02.616" E	IPK	X	X	X	X	X
BGV005852	<i>C. annuum</i>	-	Al Barkat	Fezzan	Libia	27°00'3.582" N	14°25'47.981" E	IPK	X	X	X	X	X
BGV005857	<i>C. annuum</i>	Aji	Malacatos	Loja	Ecuador	04°10'00" S	79°12'00" O	COMAV	X*	-	X	X	-
BGV005981	<i>C. annuum</i>	Aji	Malacatos	Loja	Ecuador	04°13'00" S	79°16'00" O	COMAV	X*	-	X	X	X
BGV006008	<i>C. annuum</i>	Pimiento	-	-	Ecuador	03°02'00" S	78°47'00" O	COMAV	X*	-	X	X	X
BGV006065	<i>C. annuum</i>	Pimiento	-	-	Ecuador	-	-	COMAV	X*	-	X	X	-
BGV011957	<i>C. annuum</i>	-	-	-	Ecuador	-	-	COMAV	X*	-	X	X	X
P241670	<i>C. annuum</i>	-	Quito	Pichincha	Ecuador	00°13'00" S	78°50'00" O	COMAV	X*	-	X	X	-
P585246	<i>C. annuum</i>	-	Gonzanama	Loja	Ecuador	04°13'50.237" S	79°26'33.952" O	USDA	X*	-	X	X	X
P593983	<i>C. annuum</i>	-	Arenillas	El Oro	Ecuador	3° 45' 36.975" S	80°13'33.86 O	USDA	X*	-	X	X	X
BGV013687	<i>C. annuum</i>	Aji de botón	Portoviejo	Manabí	Ecuador	1°03'06" S	80°27'3.999 O	USDA	X*	-	X	X	X
BGV013691	<i>C. annuum</i>	Pimiento crotlo	-	-	Perú	-	-	COMAV	X*	-	X	X	X
CAP1364	<i>C. annuum</i>	Aji	Iquitos	Maynas	Perú	3°44'58.765" S	73°15'22.896" O	IPK	X*	-	X	X	X
CAP1366	<i>C. annuum</i>	-	Lima	Lima	Perú	12°02'15" S	77°02'33" O	IPK	X*	-	X	X	-
P257144	<i>C. annuum</i>	Aji dulce	-	-	Perú	-	-	USDA	X*	-	-	-	-
CGN20808	<i>C. annuum</i>	-	-	-	Bolivia	-	-	CCN	X*	-	X	X	X
P213915	<i>C. annuum</i>	-	-	-	Bolivia	-	-	USDA	X*	-	X	X	X
P260436	<i>C. annuum</i>	-	-	-	Bolivia	-	-	USDA	X*	-	-	-	-
P439212	<i>C. annuum</i>	-	-	-	Bolivia	-	-	USDA	X*	-	X	X	X
P596061	<i>C. annuum</i>	-	El Puente	Chuquisaca	Bolivia	21°00'00" S	65°10'59" O	USDA	X*	-	X	X	-
C-87	<i>C. annuum</i>	Chili de árbol	-	-	México	-	-	BGHZ	X*	-	X	X	X
C-94	<i>C. annuum</i>	Tipo Jalapeño	-	-	México	-	-	BGHZ	X*	-	X	X	X
C-141	<i>C. annuum</i>	Chili serrano	-	-	México	-	-	BGHZ	X*	-	X	X	-
C-204	<i>C. annuum</i>	Chili de árbol	-	-	México	-	-	BGHZ	X*	-	X	X	X
CGN21460	<i>C. annuum</i>	-	-	-	México	-	-	CCN	X*	-	X	X	X
CGN21526	<i>C. annuum</i>	-	Xochimilco	Ciudad de México	México	19°15'00" N	99°07'47.999" O	CCN	X*	-	X	X	X
CGN21527	<i>C. annuum</i>	-	Puebla de Zaragoza	Puebla	México	19°03'00" N	98°09'36 O	CCN	X	-	X	X	X
CGN22102	<i>C. annuum</i>	-	-	-	México	-	-	CCN	X*	-	X	X	X
CGN22783	<i>C. annuum</i>	-	-	-	México	-	-	CCN	X	-	X	X	X
CGN23254	<i>C. annuum</i>	-	-	-	México	-	-	CCN	X*	-	X	X	X
CGN23765	<i>C. annuum</i>	Criollos de Morelos	-	-	México	-	-	CCN	X*	-	X	X	-

X*, datos obtenidos de Silvar & García-González (2016).

Tabla 1.1. Acciones de *Capsicum* spp. empleadas en la presente Tesis Doctoral. (continuación).

Accesión	Especie	Nombre común	Localidad	Provincia	País	Latitud	Longitud	Banco	Microsatélites	Marcadores enfermedad	GBS	Fenotipado	GWAS
CGN24355	<i>C. annuum</i>	Criollos de Morelos	-	Morelos	México	-	-	CCN	X*	-	X	X	X
PM647	<i>C. annuum</i> var. <i>glabriscaulum</i>	Oaxaca	-	Oaxaca	México	-	-	INRA	X*	-	X	X	X
PM663	<i>C. annuum</i> var. <i>glabriscaulum</i>	Chihuahua	Pueblo de Zaragoza	-	México	19°03'00" N	98°09'36" O	INRA	X	-	X	-	-
PM669	<i>C. annuum</i> var. <i>glabriscaulum</i>	Panama city	Ciudad de Panamá	-	Panamá	8°59'10.309" N	79°51'18.593" O	INRA	X*	-	X	-	-
PM670	<i>C. annuum</i> var. <i>glabriscaulum</i>	Pitalito	Pitalito	-	Colombia	1°51'11.956" N	76°02'52.987" O	INRA	X*	-	X	X	X
BGV005890	<i>C. chinense</i>	Aji	-	Sucumbios	Ecuador	0°05'05" N	76°52'58" O	COMAV	X*	-	X	X	X
CCN17040	<i>C. chinense</i>	Aji picante	Lago agrario	-	Ecuador	0°05'05" N	76°52'58" O	CCN	X*	-	X	X	X
CGN23259	<i>C. chinense</i>	-	-	-	Ecuador	-	-	CCN	X*	-	X	X	X
PM60725	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM585252	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM585253	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM585278	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM593922	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM593925	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM593929	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM595908	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
BGV013705	<i>C. chinense</i>	Aji niscocho	-	Manabí	Ecuador	0°58'00" S	78°55'22.998" O	USDA	X*	-	X	X	X
CGN17220	<i>C. chinense</i>	Misquito colorado	-	Manabí	Ecuador	0°58'00" S	78°55'22.998" O	USDA	X*	-	X	X	X
CGN19211	<i>C. chinense</i>	Lemon drop	-	Manabí	Ecuador	0°58'00" S	78°55'22.998" O	USDA	X*	-	X	X	X
CGN23365	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM26078	<i>C. chinense</i>	Aji Chundo	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM315015	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
BGV14749	<i>C. chinense</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
CGN17045	<i>C. chinense</i>	-	-	-	Ecuador	-	-	COMAV	X*	-	X	X	X
CGN22155	<i>C. chinense</i>	-	-	-	Ecuador	-	-	COMAV	X*	-	X	X	X
PM543184	<i>C. chinense</i>	Aji/Bido/Udu	Yapacaní	-	Bolivia	16°48'00" S	64°24'00" O	CCN	X*	-	X	X	X
PM543188	<i>C. chinense</i>	Aji dulce	San Jose de Uchupiamonas	-	Bolivia	14°15'00" S	68°10'00" O	USDA	X*	-	X	X	X
PM57337	<i>C. chinense</i>	Aribibi gusano	Cobija (mercado)	-	Bolivia	11°02'08.8" S	68°46'52.7" O	USDA	X*	-	X	X	X
PM22427	<i>C. frutescens</i>	-	Carmen Florida	-	Bolivia	14°28'00" S	67°34'00" O	USDA	X*	-	X	X	X
PM355808	<i>C. frutescens</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM585254	<i>C. frutescens</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM585256	<i>C. frutescens</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM585257	<i>C. frutescens</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM593920	<i>C. frutescens</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	X
PM593924	<i>C. frutescens</i>	Ucho	Dorado de Cascales	-	Ecuador	0°02'35" N	77°12'03" O	USDA	X*	-	X	X	X
PM595907	<i>C. frutescens</i>	Aji bravofmili ucho	Pasumo	-	Ecuador	0°02'35" N	77°12'03" O	USDA	X*	-	X	X	X
BGV013699	<i>C. frutescens</i>	Aji de ratón	Mocache	-	Ecuador	0°06'36" S	79°26'48" O	USDA	X*	-	X	X	X
CAP472	<i>C. frutescens</i>	-	-	-	Perú	-	-	COMAV	X	-	X	X	X
CAP524	<i>C. frutescens</i>	-	-	-	Perú	-	-	IPK	X*	-	X	X	X
CAP868	<i>C. frutescens</i>	Aji misucocho/Aji amaicho	Iquitos	-	Perú	3°44'58.765" S	73°15'22.896" O	IPK	X*	-	X	X	X
			Lima	-	Perú	12°02'15" S	77°02'33" O	IPK	X*	-	X	X	X
			Trujillo	-	Perú	8°06'20" S	79°01'59" O	IPK	X*	-	X	X	X

X*, datos obtenidos de Silvar & García-González (2016).

Tabla 1.1. Accesiones de *Capsicum* spp. empleadas en la presente Tesis Doctoral. (continuación).

Accesión	Especie	Nombre común	Localidad	Provincia	País	Latitud	Longitud	Banco	Microsatélites	Marcadores enfermedad	GBS	Fenotipado	GWAS
CGN22792	<i>C. frutescens</i>	Aji chuncho	-	-	Perú	-	-	CGN	X*	-	X	X	X
P439488	<i>C. frutescens</i>	-	-	-	Bolivia	-	-	USDA	X*	-	X	X	X
P543203	<i>C. frutescens</i>	Aji gusanito	Riberaleta (mercado)	Vaca Diez	Bolivia	1°01'00" S	66°07'59,998" O	USDA	X*	-	X	X	X
P560941	<i>C. frutescens</i>	Aji	Rurrenabaque	Beni	Bolivia	14°28'00" S	67°34'00" O	USDA	X*	-	X	X	X
BGY006064	<i>C. baccatum</i>	Aji	-	-	Ecuador	-	-	COMAV	X*	-	X	X	-
P257133	<i>C. baccatum</i>	-	Gonzalo Diaz de Pineda	Napo	Ecuador	0°49'12" S	77°43'48" O	USDA	X*	-	X	X	-
P257135	<i>C. baccatum</i>	-	Latacunga	Cotopaxi	Ecuador	0°55'48" S	78°37'12" O	USDA	X*	-	X	X	-
P355813	<i>C. baccatum</i>	-	Ambato	Tungurahua	Ecuador	1°14'24" S	78°37'12" O	USDA	X*	-	X	X	-
P585239	<i>C. baccatum</i>	-	Salcedo	Cotopaxi	Ecuador	0°55'00" S	79°56'00" O	USDA	X*	-	X	X	-
P585241	<i>C. baccatum</i>	-	Guamote	Chimborazo	Ecuador	1°55'00" S	7842'00" O	USDA	X*	-	X	X	-
P585244	<i>C. baccatum</i>	-	Arenillas	El Oro	Ecuador	3°46'00" S	80°01'00" O	USDA	X*	-	X	X	-
P585249	<i>C. baccatum</i>	-	Saraguro	Loja	Ecuador	3°30'00" S	79°22'00" O	USDA	X*	-	X	X	-
P593932	<i>C. baccatum</i>	Aji picante	Pira (mercado)	El Oro	Ecuador	3°41'48" S	79°41'12" O	USDA	X*	-	X	X	-
P595905	<i>C. baccatum</i>	-	Santa Rosa	Imbabura	Ecuador	0°21'15" N	78°30'46" O	USDA	X*	-	X	X	-
BGY013703	<i>C. baccatum</i>	Aji montana	-	-	Perú	-	-	COMAV	X	-	X	X	-
CGN17174	<i>C. baccatum</i>	Aji	-	-	Perú	-	-	CGN	X*	-	X	X	-
CGN21479	<i>C. baccatum</i>	-	Chachapoyas	Chachapoyas	Perú	6°13'57,299" S	77°52'15,008" O	CGN	X*	-	X	X	-
P215729	<i>C. baccatum</i>	Omnicolor	-	-	Perú	-	-	USDA	X*	-	X	X	-
P290982	<i>C. baccatum</i>	Aji Mirasol	-	-	Perú	-	-	USDA	X*	-	X	X	-
P315020	<i>C. baccatum</i>	-	-	-	Perú	-	-	USDA	X*	-	X	X	-
BGY008233	<i>C. baccatum</i>	Aji	-	-	Bolivia	-	-	COMAV	X	-	X	X	-
BGY013986	<i>C. baccatum</i>	-	-	-	Bolivia	-	-	COMAV	X*	-	X	X	-
CGN17042	<i>C. baccatum</i>	-	Cochabamba	Cercado	Bolivia	16°48'00" S	64°24'00" O	CGN	X*	-	X	X	-
CGN23205	<i>C. baccatum</i>	-	Monteagudo	Hernando Siles	Bolivia	19°48'00" S	63°58'47,998" O	CGN	X*	-	X	X	-
P590305	<i>C. baccatum</i>	Aji verde	Sud Yungas (mercado)	La Paz	Bolivia	16°24'00" S	67°31'00" O	USDA	X*	-	X	X	-
BGY013300	<i>C. pubescens</i>	Aji rocoto	-	-	Ecuador	-	-	COMAV	X*	-	X	X	-
P335394	<i>C. pubescens</i>	-	-	-	Ecuador	-	-	USDA	X*	-	X	X	-
P585262	<i>C. pubescens</i>	-	Gonzanama	Loja	Ecuador	4°13'00" S	79°25'00" O	USDA	X*	-	X	X	-
P585264	<i>C. pubescens</i>	-	Sigsig	Azuay	Ecuador	0°30'20" S	78°47'00" O	USDA	X*	-	X	X	-
P585265	<i>C. pubescens</i>	-	Otavalo	Imbabura	Ecuador	0°10'00" N	78°12'00" O	USDA	X*	-	-	-	-
P585267	<i>C. pubescens</i>	-	San Felipe de Oña	Azuay	Ecuador	3°27'00" S	79°09'00" O	USDA	X*	-	-	-	-
P585269	<i>C. pubescens</i>	-	Saraguro	Loja	Ecuador	4°24'00" S	79°28'00" O	USDA	X*	-	X	X	-
P585271	<i>C. pubescens</i>	-	San Pablo de Tenta	Loja	Ecuador	0°34'00" S	79°18'00" O	USDA	X*	-	X	X	-
P585273	<i>C. pubescens</i>	-	Montufar	Carchi	Ecuador	0°34'59" N	79°48'00" O	USDA	X*	-	X	X	-
P585275	<i>C. pubescens</i>	-	Quijes	Napo	Ecuador	0°29'10,838" S	77°58'19,97" O	USDA	X*	-	X	X	-
BGY013702	<i>C. pubescens</i>	Aji tariche	-	-	Perú	-	-	COMAV	X*	-	X	X	-
CAP217	<i>C. pubescens</i>	-	-	-	Perú	-	-	IPK	X*	-	X	X	-
CAP357	<i>C. pubescens</i>	-	-	-	Perú	-	-	IPK	X	-	X	X	-
CAP1486	<i>C. pubescens</i>	-	-	-	Perú	-	-	IPK	X*	-	X	X	-
CGN22108	<i>C. pubescens</i>	-	Huánuco	Huánuco	Perú	9°54'36" S	76°13'48" O	CGN	X*	-	X	X	-
CGN23768	<i>C. pubescens</i>	-	Tarma	Junín	Perú	11°24'00" S	75°42'00" O	CGN	X	-	X	X	-

X*, datos obtenidos de Silvar & Garcia-Gonzalez (2016).

Extracción de ADN

La extracción de ADN genómico se realizó a partir de aproximadamente 1 g de tejido foliar joven de una planta de cada accesión con el método del bromuro de cetiltrimetilamonio (CTAB)/cloroformo/isoamil alcohol (Doyle & Doyle, 1987). La calidad y concentración del ADN extraído se determinó con un espectrofotómetro NANODROP ND-1000 (NanoDrop Technologies, Inc.). Se prepararon diluciones de ADN a una concentración estándar de 50 ng/ μ l.

Genotipado de los *loci* microsatélite

Para las 230 líneas se obtuvieron los genotipos de 28 *loci* microsatélite (Tabla 1.2) desarrollados previamente para *C. annuum* (Lee et al., 2004; Minamiyama et al., 2006; Yi et al., 2006; Nagy et al., 2007). Los marcadores fueron seleccionados de tal modo que estuvieran distribuidos uniformemente a lo largo de los doce cromosomas del pimiento y que mostrasen amplificación positiva en diferentes especies del género *Capsicum*, siendo a la vez marcadores polimórficos (González-Pérez et al., 2014; Silvar & García-González, 2016).

Tabla 1.2. Marcadores microsatélite analizados en las tres colecciones de *Capsicum* spp.

Marcador	Cromosoma	Motivo de repetición	Fuente
<i>Hpms1-139</i>	1	(CT) ₂ (AG) ₁₅	Lee et al. (2004)
<i>Hpms1-281</i>	1	(TTG) ₆	Lee et al. (2004)
<i>Hpms1-148</i>	1	(GA) ₁₄	Lee et al. (2004)
<i>GPMS178</i>	2	(ATT) ₁₉	Nagy et al. (2007)
<i>Hpms1-143</i>	2	(AG) ₁₂	Lee et al. (2004)
<i>Hpms1-173</i>	3	(GA) ₁₆ (TG) ₂	Lee et al. (2004)
<i>HpmsE016</i>	3	(TACA) ₃	Yi et al. (2006)
<i>EPMS472</i>	3	T ₁₆	Nagy et al. (2007)
<i>HpmsAT2-14</i>	4	(AAT) ₁₆ IMP	Lee et al. (2004)
<i>CAMS351</i>	4	(TG) ₃ ...(AG) ₂₆	Minamiyama et al. (2006)
<i>Hpms2-45</i>	5	(TTG) ₉	Lee et al. (2004)
<i>HpmsE015</i>	5	(GCA) ₈	Yi et al. (2006)
<i>Hpms1-5</i>	6	(AT) ₁₁ (GT) ₁₇	Lee et al. (2004)
<i>HpmsE120</i>	6	(CCA) ₆	Yi et al. (2006)
<i>HpmsCaSIG19</i>	7	(CT) ₆ (AT) ₈	Lee et al. (2004)
<i>Hpms1-6</i>	7	(AT) ₂ (GT) ₄ (AT) ₈	Lee et al. (2004)
<i>EPMS426</i>	7	(AT) ₁₅	Nagy et al. (2007)
<i>Hpms1-155</i>	8	AAA(TA) ₃ CA(GA) ₂₁	Lee et al. (2004)
<i>EPMS310</i>	8	(CAT) ₁₃	Nagy et al. (2007)
<i>Hpms2-24</i>	9	(CT) ₁₇ (CA) ₅ A ₂₁	Lee et al. (2004)
<i>Hpms1-3</i>	9	(AT) ₁₀ (GT) ₁₂	Lee et al. (2004)
<i>CAMS336</i>	10	(TC) ₁₆	Minamiyama et al. (2006)
<i>HpmsE013</i>	10	(T) ₂₀	Yi et al. (2006)
<i>GPMS159</i>	10	(TAA) ₂₀	Nagy et al. (2007)
<i>Hpms2-2</i>	11	(GT) ₉	Lee et al. (2004)
<i>CAMS405</i>	11	(TC) ₁₈	Minamiyama et al. (2006)
<i>GPMS197</i>	12	(GA) ₃ (TAT) ₁₆	Nagy et al. (2007)
<i>HpmsE075</i>	12	(ACC) ₆	Yi et al. (2006)

La amplificación de los microsatélites se realizó mediante la técnica de la reacción en cadena de la polimerasa (PCR). Concretamente, se empleó el protocolo de marcaje utilizando la cola oligonucleotídica universal M13 (Schuelke, 2000) marcada con diferentes fluoróforos (6-FAM, HEX o NED). Así, para la amplificación de cada *locus* microsatélite se utilizaron tres *primers*: un *primer* forward específico con la cola M13 (5'-GGAAACAGCTATGACCAT-3') en su extremo 5', un *primer* reverse específico y el oligonucleótido M13 marcado.

La PCR se llevó a cabo en un volumen final de 15 μ l que contenía 100 ng de ADN genómico, 1X buffer PCR (NZYTech), 2,5 mM $MgCl_2$ (NZYTech), dNTPs (NZYTech) a una concentración de 0,2 mM cada uno, 0,02 μ M de *primer* forward, 0,2 μ M de *primer* reverse, 0,18 μ M *primer* universal M13 y 0,5 U *Taq* polimerasa (NZYTaQ DNA Polymerase, NZYTech). Todos los fragmentos fueron amplificados usando un programa de PCR *touchdown* que consistió en una desnaturalización inicial a 94°C durante 5 min; 12 ciclos de 94°C durante 30 s, 62-56°C (descendiendo a un ritmo de 0,5°C por ciclo) durante 30 s, 72°C durante 30 s; 35 ciclos 94°C durante 30 s, 56°C durante 30 s, 72°C durante 30 s; y una extensión final a 72°C durante 10 min. La detección de los productos de PCR se realizó mediante electroforesis en gel de agarosa al 2% en 1X TAE (Tris-acetato 40 mM pH 8, EDTA 1 mM) teñido con *GreenSafe Premium* (NZYTech) y revelado bajo luz ultravioleta en un fotodocumentador GelDoc XR+ (Bio-Rad). El análisis de fragmentos se realizó en los Servicios de Apoyo á Investigación da Universidade da Coruña (SAI-UDC) en un equipo ABI PRISM 3130xl (Applied Biosystems) y empleando un estándar de tamaño *GeneScan 500 (-250) ROX internal size standar* (AppliedBiosystems). El análisis de los perfiles de fluorescencia y la asignación de los alelos se realizó con el programa Geneious 5.6.7 (Kearse et al., 2012).

Diversidad genética

La variabilidad genética fue estimada con descriptores estándar. El número de alelos (N_a), número de alelos privados (N_{ap}), rango alélico, heterocigosidad observada (H_o) y heterocigosidad esperada (H_e) por *locus* se calcularon empleando el *software* GenAlEx 6.5.1 (Peakall & Smouse, 2012). Asimismo, con el objetivo de caracterizar las especies se

determinó para cada una el número total de alelos, número de alelos privados, número medio de alelos por *locus* (A), porcentaje de polimorfismo (P), heterocigosidad observada (H_o) y heterocigosidad esperada (H_e). La riqueza alélica media por especie (AR) se calculó con el programa FSTAT 2.9.4 (Goudet, 2003). La frecuencia de alelos nulos se calculó con FreeNA (Chapuis & Estoup, 2007) y el algoritmo EM (Dempster et al., 1977) con 10.000 réplicas de *bootstrap* y con el programa ML-Null Freq (Kalinowski & Taper, 2006). El coeficiente de endogamia (F_{IS}) se estimó con GENEPOP 4.2 (Rousset, 2008).

Dentro de cada uno de los subgrupos encontrados en el análisis “multistep” de STRUCTURE se estimaron ciertos parámetros genéticos con el objetivo de caracterizar los subgrupos. Así, para cada clúster del segundo nivel jerárquico de STRUCTURE se estimaron las desviaciones del equilibrio Hardy-Weinberg (*Hardy-Weinberg equilibrium*, HWE) para cada combinación *locus/grupo* usando GenAlEx 6.5.1 (Peakall & Smouse, 2012). El desequilibrio gamético o desequilibrio de ligamiento (*linkage disequilibrium*, LD) entre pares de *loci* se testó mediante el test exacto de Fisher implementado en GENEPOP 4.2 (Rousset, 2008). El nivel de significación se calculó mediante el método de las cadenas de Markov Monte Carlo (*Markov chain Monte Carlo*, MCMC) empleando 10.000 “dememorizaciones”, 5.000 “batches” y 5.000 iteraciones por *batch*. Para las pruebas múltiples se aplicó la corrección secuencial de Bonferroni (Holm, 1979). El cálculo de la frecuencia de alelos nulos se realizó con el programa ML-Null Freq (Kalinowski & Taper, 2006). El coeficiente de endogamia (F_{IS}) se calculó con la herramienta GENEPOP 4.2 (Rousset, 2008). Para evaluar el posible efecto de los alelos nulos en los niveles de endogamia se empleó el programa INest 2.2 (Chybicki & Burczyk, 2009) utilizando el

modelo de endogamia individual bayesiana (*bayesian individual inbreeding model*, IIM). Este programa calcula el coeficiente de endogamia corregido para la presencia de alelos ($F_{IS-INest}$). Además el programa se empleó para calcular el criterio de información de la desviación (*deviance information criterion*, DIC) tanto en el modelo *nfb* que tiene en cuenta el posible efecto de los alelos nulos en los valores de homocigosis como en el modelo *fb*, que los excluye. Los análisis se llevaron a cabo empleando 500.000 cadenas de Markov Monte Carlo (MCMC) y un *burn-in* de 50.000.

Estructura genética

Para identificar la estructura genética del conjunto de accesiones utilizadas se llevaron a cabo dos aproximaciones complementarias. Por una parte se realizó un análisis factorial de correspondencia (AFC) con el programa GENETIX 4.05 (Belkhir et al., 2004) usando el módulo “AFC sur populations”.

Por otra parte, se llevó a cabo un análisis bayesiano de agrupamiento utilizando el *software* STRUCTURE 2.3.4 (Pritchard et al., 2000) siguiendo la metodología empleada con anterioridad en estudios similares (González-Pérez et al., 2014; Silvar & García-González, 2016). Las simulaciones se llevaron a cabo utilizando el modelo “admixture”, con frecuencias alélicas correlacionadas y fijando el número de clústeres (K) entre 1 y 10, con 10 réplicas independientes para cada valor de K . Para las cadenas de Markov Monte Carlo (MCMC) se estableció un *burn-in* de 500.000, seguido de 1.000.000 de iteraciones. El valor de K más probable se determinó mediante el logaritmo de la probabilidad de los datos y el estadístico ΔK (Evanno et al., 2005) implementados en STRUCTURE

HARVESTER 0.6.94 (Earl & vonHoldt, 2012). Para obtener la matriz promedio de los coeficientes de pertenencia de cada muestra para el valor estimado de K se utilizó el programa CLUMPP 1.1.2 (Jakobsson & Rosenberg, 2007). Los coeficientes de pertenencia se representaron gráficamente con STRUCTURE PLOT 2.0 (Ramasamy et al., 2014). Con el objetivo de observar subestructura de manera precisa se siguió un procedimiento en múltiples pasos (“multistep”), que consistió en correr el programa STRUCTURE en primer lugar con el conjunto de datos completo para identificar la estructura general y posteriormente volver a correr el programa de manera separada para los subconjuntos de muestras de cada partición encontrada en el primer paso.

Resultados

Diversidad genética

Veintiséis de los 28 *loci* microsatélite fueron amplificados con éxito en todas las especies. El marcador *Hpms1-143* no pudo ser amplificado en las accesiones de *C. pubescens* y *C. baccatum*, así como el *Hpms1-5* en *C. baccatum*. Pese a ello, ambos *loci* se incluyeron en los análisis posteriores ya que supusieron menos de un 4% de datos perdidos (*missing data*) en el global de datos de estas dos especies.

Se identificaron un total de 378 alelos para los 28 *loci* analizados. El número total de alelos por *locus* varió entre 5 (*Hpms2-45* y *HpmsE075*) y 23 (*HpmsE016* y *Hpms1-5*), con una media de 13,5 alelos por *locus*. Sesenta y seis de los 378 alelos (17,46%) se correspondieron con alelos privados (solo aparecieron en una especie). Todos los

marcadores mostraron alelos privados excepto *Hpms1-281*, *EPMS472*, *Hpms2-45*, *HpmsCaSIG19*, *HpmsE013* y *HpmsE075*. La heterocigosidad observada (H_o) varió entre 0,005 (*Hpms1-3*) y 0,581 (*GPMS178*). Por su parte, la heterocigosidad esperada (H_e) osciló entre 0,496 (*HpmsE015*) y 0,886 (*GPMS197*). Todos los marcadores presentaron evidencia de alelos nulos en una frecuencia $\geq 0,063$ y 26 de los 28 empleados en frecuencia $\geq 0,2$ (Tabla 1.3) tanto con el programa FreeNA como con el programa ML-Null Freq, siendo los resultados proporcionados por ambos programas muy similares (Tabla A.1).

Tabla 1.3. Caracterización de los loci microsatélite analizados en las 230 accesiones de *Capsicum spp.* *N*, número de accesiones con genotipo; *Na*, número de alelos; *Nap*; número de alelos privados; *H_o*, heterocigosidad observada; *H_e*, heterocigosidad esperada. *Alelos nulos*, frecuencia de alelos nulos calculada con el programa FreeNA.

<i>Locus</i>	Rango alélico (pb)	<i>N</i>	<i>Na</i> (<i>Nap</i>)	<i>H_o</i>	<i>H_e</i>	Alelos nulos
<i>Hpms1-139</i>	302-326	225	11 (1)	0,082	0,835	0,411
<i>Hpms1-281</i>	140-179	230	12 (0)	0,013	0,618	0,383
<i>Hpms1-143</i>	206-248	193	11 (1)	0,02	0,797	0,433
<i>Hpms1-173</i>	176-200	220	12 (1)	0,066	0,825	0,417
<i>CAMS405</i>	244-266	216	12 (3)	0,058	0,753	0,398
<i>GPMS197</i>	294-360	216	22 (4)	0,064	0,886	0,436
<i>CAMS351</i>	193-237	223	22 (5)	0,036	0,856	0,442
<i>Hpms1-5</i>	296-360	167	23 (5)	0,017	0,822	0,443
<i>Hpms2-45</i>	151-166	229	5 (0)	0,051	0,542	0,331
<i>CAMS336</i>	162-206	223	17 (4)	0,074	0,854	0,421
<i>HpmsE013</i>	252-280	224	8 (0)	0,022	0,684	0,399
<i>Hpms2-2</i>	151-179	225	9 (3)	0,039	0,685	0,386
<i>EPMS310</i>	152-194	229	12 (2)	0,03	0,732	0,409
<i>Hpms2-24</i>	152-222	226	15 (3)	0,043	0,837	0,433
<i>Hpms1-3</i>	221-245	210	12 (2)	0,005	0,548	0,365
<i>HpmsCaSIG19</i>	215-245	227	10 (0)	0,065	0,719	0,385
<i>Hpms1-6</i>	165-221	216	15 (2)	0,25	0,684	0,259
<i>Hpms1-155</i>	202-244	227	14 (4)	0,043	0,799	0,421
<i>HpmsE015</i>	152-167	228	6 (1)	0,453	0,496	0,063
<i>HpmsE016</i>	166-256	223	23 (5)	0,074	0,855	0,422
<i>HpmsE120</i>	194-237	222	7 (3)	0,022	0,614	0,370
<i>HpmsAT2-14</i>	184-226	217	13 (2)	0,049	0,701	0,387
<i>EPMS426</i>	116-190	223	22 (9)	0,089	0,762	0,383
<i>EPMS472</i>	308-340	213	6 (0)	0,036	0,741	0,406
<i>Hpms1-148</i>	197-237	229	16 (2)	0,078	0,78	0,395
<i>GPMS178</i>	254-308	219	19 (1)	0,581	0,868	0,162
<i>HpmsE075</i>	218-230	224	5 (0)	0,038	0,613	0,365
<i>GPMS159</i>	271-346	223	19 (3)	0,065	0,878	0,433

En lo que respecta a la variabilidad de los *loci* en cada especie, los 28 microsatélites resultaron ser altamente polimórficos, con un polimorfismo medio de 95,84% (Tabla 1.4). El grado de polimorfismo entre las especies estudiadas varió entre 89,3% (*C. pubescens*) y 100% (*C. annuum* y *C. chinense*). La mayoría de los *loci* presentaron elevada frecuencia de alelos nulos (>0.2) tanto en la especie para la que fueron desarrollados específicamente (*C. annuum*) como para las demás analizadas (Tabla A.2).

La especie que presentó un mayor número de alelos fue *C. annuum* (213 alelos) y la que menos *C. frutescens* con 84 alelos. Ambas especies también fueron las que presentaron mayor y menor número de alelos privados, 68 y 13, respectivamente. Sin embargo, hay que tener en cuenta que estos resultados pueden verse influidos por la gran variación en el número de muestras de cada especie, 147 y 11 accesiones respectivamente. Esto se demuestra con los valores de riqueza alélica obtenidos, que fueron similares para todas las especies y oscilaron entre 2,077 (*C. frutescens*) y 3,205 (*C. annuum* var. *glabriusculum*).

Todas las especies mostraron unos valores medios de H_o similares, variando entre 0,070 (*C. baccatum*) y 0,114 (*C. chinense*) y de H_e entre 0,367 (*C. frutescens*) y 0,621 (*C. annuum* var. *glabriusculum*), que parece sugerir un patrón de desviación de la heterocigosidad observada media respecto de la esperada.

Tabla 1.4. Diversidad genética para cada especie de *Capsicum* estudiada. *N*, número de accesiones estudiadas; *Na*, número de alelos; *Nap*; número de alelos privados; *A*, número medio de alelos por *locus*; *AR*, riqueza alélica; *P*, porcentaje de polimorfismo; *H_o*, heterocigosidad observada media; *H_e*, heterocigosidad esperada media. *F_{IS}*, coeficiente de endogamia.

Especie	<i>N</i>	<i>Na</i> (<i>Nap</i>)	<i>A</i>	<i>AR</i>	<i>P</i>	<i>H_o</i>	<i>H_e</i>	<i>F_{IS}</i>
<i>C. annuum</i>	147	213 (68)	7,607	2,557	100,00%	0,085	0,521	0,838
<i>C. chinense</i>	28	157 (31)	5,607	2,636	100,00%	0,114	0,533	0,794
<i>C. baccatum</i>	24	109 (21)	3,893	2,418	92,86%	0,070	0,452	0,837
<i>C. pubescens</i>	16	89 (19)	3,179	2,092	89,29%	0,088	0,416	0,801
<i>C. frutescens</i>	11	84 (13)	3,000	2,077	96,43%	0,113	0,367	0,725
<i>C. annuum</i> var. <i>glabriusculum</i>	4	92 (15)	3,286	3,205	96,43%	0,080	0,621	0,897

Estructura genética

En relación al número de particiones para el conjunto de datos según STRUCTURE el estadístico ΔK de Evanno apunta a una fuerte división en $K=2$ grupos (Figura 1.1A).

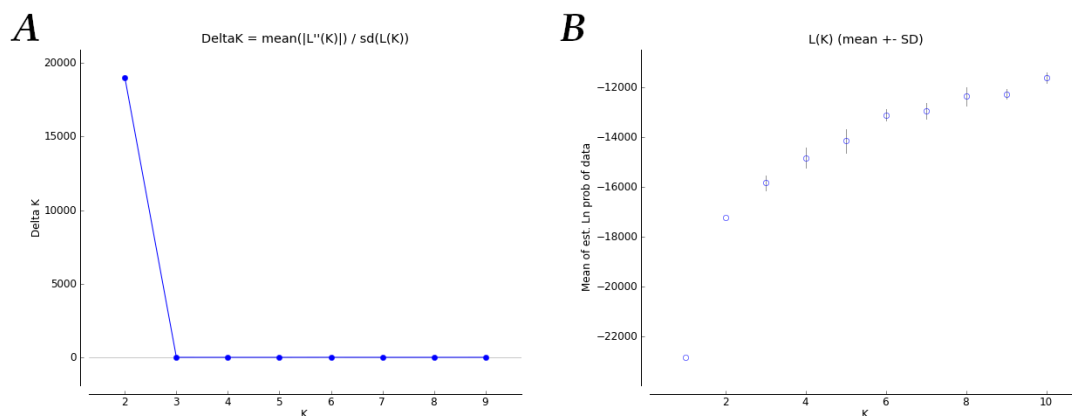


Figura 1.1. Métodos de evaluación de la estructura genética más probable con el análisis STRUCTURE para las 230 accesiones de *Capsicum* y empleando los 28 loci microsatélite obtenidos con STRUCTURE HARVESTER. A. Estadístico ΔK de Evanno en función de *K*, tras diez réplicas. B. Media del logaritmo de la probabilidad de los datos $\text{Pr}(X/K)$, las barras representan la desviación estándar.

El primer clúster, mayoritario, está formado principalmente por todas las accesiones de *C. annuum*, mientras que el segundo está compuesto por las restantes especies (Figura 1.2). De las cuatro accesiones de *C. annuum* var. *glabriusculum* incluidas, tres de ellas muestran pertenencia al clúster no-*C. annuum* (PM669, PM663 y PM670) y una de ellas al de *C. annuum* (PM647).

En el primer clúster aparecen algunas accesiones, catalogadas en los bancos de germoplasma como otras especies pero que en realidad pertenecerían claramente a *C. annuum*: CGN22155, PI 543188, PI 573337 y CAP524 clasificadas inicialmente las tres primeras como *C. chinense* y como *C. frutescens* la última. Del mismo modo, en el clúster formado por las otras cinco especies aparecen accesiones clasificadas como *C. annuum* en los bancos de germoplasma pero que no lo serían en realidad.

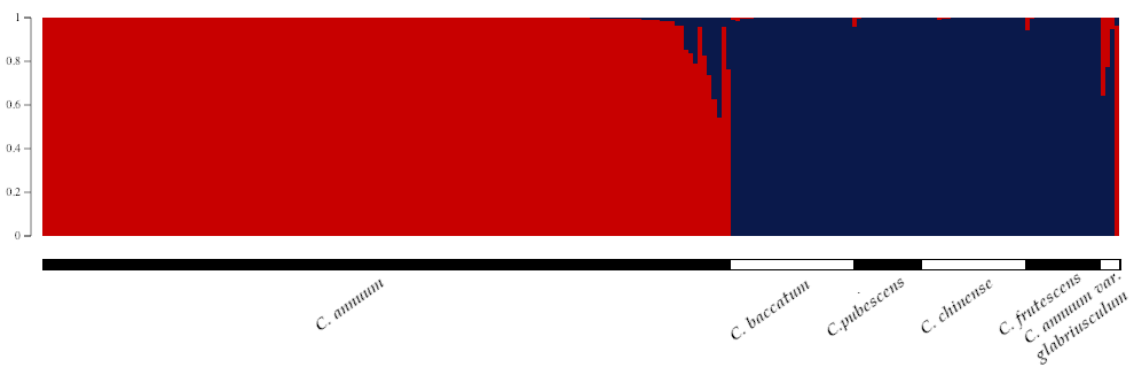


Figura 1.2. Análisis de inferencia bayesiana realizado con STRUCTURE para las 230 accesiones y empleando 28 microsatélites. Cada color representa un clúster y cada barra vertical representa un individuo con el correspondiente coeficiente de pertenencia a cada clúster.

La gráfica de la media del logaritmo de la probabilidad de los datos (Figura 1.1B) sugiere una posible subestructura de todo el conjunto estudiado para valores más altos de K , por lo que se realizaron sucesivos análisis de STRUCTURE para cada uno de los dos grupos encontrados por separado con el fin de observar patrones de subestructura dentro de cada uno de los dos clústeres siguiendo una metodología jerárquica. De este modo, mediante una metodología “multistep” se puede obtener información precisa sobre subestructura genética que no se estaría observando el estudio solo se ciñera al primer

nivel. En estos análisis se incluyeron las líneas inicialmente mal catalogadas para corroborar su pertenencia real a otras especies.

Para el análisis realizado incluyendo solo las accesiones de *C. annuum*, se observó una subestructura en tres clústeres (Figura A.1A y Figura A.1B). Los tres clústeres se corresponden principalmente con el origen geográfico de las tres colecciones estudiadas (Figura 1.3). El primer clúster está formado por la mayoría de accesiones originarias de la Península Ibérica y aparecen además cuatro accesiones de México, una de Perú y una de Ecuador. El segundo clúster estaría formado por las accesiones de la colección mediterránea, que engloba a los restantes países europeos estudiados (Francia, Italia, Serbia/Montenegro, Albania, Turquía, Eslovenia, Libia, Israel y Túnez) y una española. El último clúster está formado por las accesiones de la región andina (Perú, Ecuador y Bolivia) y las siete restantes de México, cuatro de España y una portuguesa.



Figura 1.3. Análisis de inferencia bayesiana realizado con STRUCTURE para las 148 accesiones de *C. annuum* y empleando 28 microsatélites. Cada color representa un clúster y cada barra vertical representa un individuo con el correspondiente coeficiente de pertenencia a cada clúster.

El análisis bayesiano realizado para las 82 accesiones de no-*C. annuum* reveló una estructura en 5 clústeres (Figura A.1C y Figura A.1D) que se corresponden con la clasificación taxonómica de las accesiones (Figura 1.4). En este análisis de STRUCTURE se incluyeron siete accesiones catalogadas como *C. annuum* en los bancos de germoplasma pero que en el análisis global (Figura 1.2) aparecían en el clúster de no-*C. annuum*. De este modo se comprobó para todas ellas que se trata de líneas que no

pertenecen a *C. annuum* porque presentan coeficientes de pertenencia próximas al 100% para otras especies. Estas líneas son CAP1364, CGN20808 y PI 593933 reclasificadas como *C. chinense* y BGV005852, BGV011957, CAP1366 y PI 596061 como *C. baccatum*. Además, otras accesiones de especies distintas a *C. annuum* mostraron su pertenencia real a otra especie. Estas son CAP868, PI 543203, CAP472 y PI 224427 catalogadas en los bancos como *C. frutescens* y según nuestros resultados como *C. chinense* y PI 315020 inicialmente *C. baccatum* en el banco de germoplasma pero que en realidad es *C. chinense*.

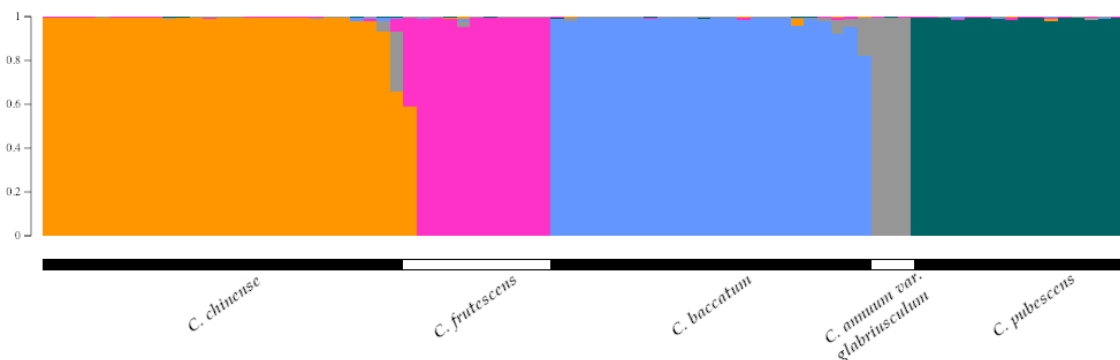


Figura 1.4. Análisis de inferencia bayesiana realizado con STRUCTURE para las 82 accesiones de especies distintas a *C. annuum* y empleando 28 microsatélites. Cada color representa un clúster y cada barra vertical representa un individuo con el correspondiente coeficiente de pertenencia a cada clúster.

En el AFC para todo el conjunto de accesiones se observaron cuatro agrupaciones (Figura 1.5): la formada por las accesiones de *C. annuum*, la correspondiente a las de *C. baccatum*, por otro lado las de *C. pubescens*, y en último lugar las accesiones de *C. chinense* y *C. frutescens* próximas entre sí, haciendo imposible separarlas a este nivel. Con respecto a las accesiones de *C. annuum* var. *glabriusculum*, una de ellas apareció agrupada con las de *C. annuum*, otras con las de *C. chinense* + *C. frutescens* y las dos restantes en posiciones intermedias entre estos grupos. El AFC permitió además

comprobar que todas las reclasificaciones taxonómicas realizadas tras el resultado del análisis de agrupación bayesiana son correctas, ya que todas las accesiones reclasificadas aparecieron con pertenencia clara a la nueva especie indicada con anterioridad.

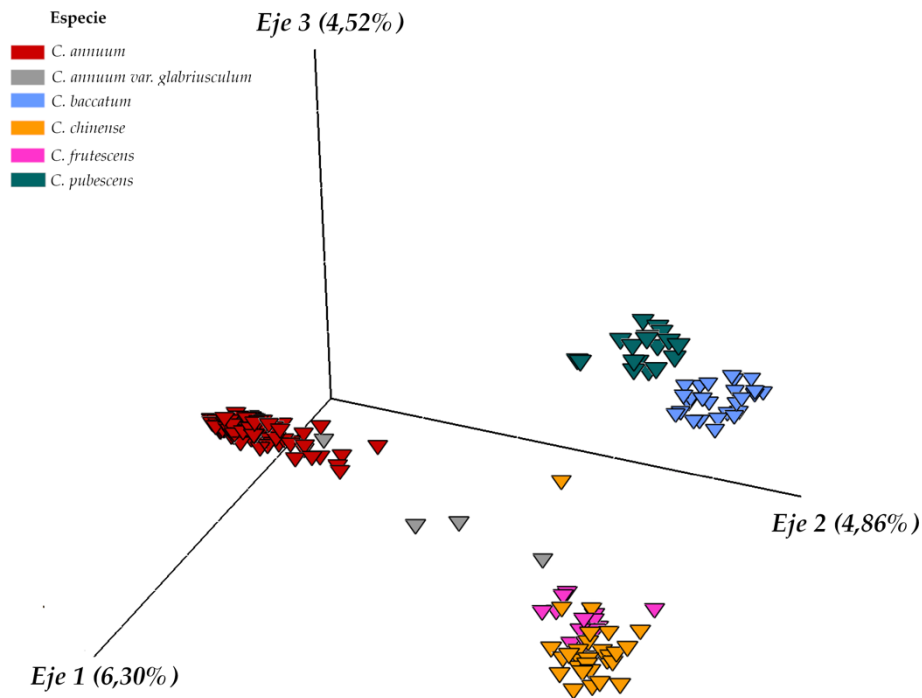


Figura 1.5. Análisis factorial de correspondencia (AFC) para las 230 accesiones y empleando 28 microsátelites.

Se realizaron además sucesivos AFCs para cada uno de los grupos encontrados (Figura A.2), visualizándolos según el país de origen para intentar encontrar relaciones entre las accesiones, pero en ningún caso se encontraron patrones claros. Con todo, el AFC realizado para el conjunto *C. chinense* + *C. frutescens* permitió separar claramente las dos especies (Figura A.2C). En este caso el eje X explica el 100% de la variabilidad y la separación de las accesiones en base a la especie es manifiesto.

Diversidad genética en los grupos encontrados en el análisis de estructura poblacional

Los grupos encontrados en los segundos niveles jerárquicos del análisis “multistep” de STRUCTURE se caracterizaron con ciertos descriptores genéticos. Para ello estos grupos se denominaron *C. annuum*-ibérica, *C. annuum*-mediterránea, *C. annuum*-andina, *C. chinense*, *C. frutescens*, *C. baccatum* y *C. pubescens*. No se incluyó en los análisis el grupo *C. annuum* var. *glabriusculum* al contar con tan solo cuatro accesiones.

Tras la corrección secuencial de Bonferroni, 162 de 196 combinaciones *locus/grupo* (82,7%) mostraron desviación significativa del equilibrio Hardy-Weinberg. Del mismo modo, tras aplicar la corrección secuencial de Bonferroni se detectó desequilibrio de ligamiento entre siete pares de *loci* en total en todos los grupos estudiados pero no se repitió ningún par de *loci* entre grupos.

Teniendo en cuenta que en el cálculo de la frecuencia de alelos con los programas FreeNa y ML-Null Freq para el conjunto global el nivel fue similar en todos los *loci* (Tabla A.1) se optó por emplear solamente ML-Null Freq para calcular los alelos nulos dentro de cada grupo jerárquico. Dabrowski et al. (2015) compararon varios métodos empleados para calcular alelos nulos y demostraron que este era el que mejor rendimiento presentaba. La frecuencia de alelos nulos fue elevada en la mayoría de marcadores y en todos los clústeres estudiados, si bien en ciertos casos la frecuencia fue baja e incluso cero. Destaca el caso del marcador *Hpms2-45*, para el que solo se encontraron alelos nulos en el clúster *C. baccatum* (Tabla 1.5).

Tabla 1.5. Frecuencia de alelos nulos para cada marcador microsatélite y en cada uno de los grupos encontrados en el segundo nivel jerárquico de STRUCTURE calculada con ML-Null Freq.

	<i>C. annuum-</i> ibérica	<i>C. annuum-</i> andina	<i>C. annuum-</i> mediterránea	<i>C. chinense</i>	<i>C. frutescens</i>	<i>C. baccatum</i>	<i>C. pubescens</i>
<i>Hpms1-139</i>	0,35	0,347	0,306	0,313	0,158	0,427	0,157
<i>Hpms1-281</i>	0	0,259	0	0,413	0,32	0,182	0,351
<i>Hpms1-143</i>	0,322	0,406	0,309	0,416	0	0,902	0,915
<i>Hpms1-173</i>	0,338	0,328	0,389	0,404	0,36	0,63	0,325
<i>CAMS405</i>	0,422	0,443	0,42	0,443	0,304	0,349	0,279
<i>GPMS197</i>	0,468	0,466	0,539	0,178	0,366	0,429	0,376
<i>CAMS351</i>	0,46	0,472	0,376	0,274	0	0,378	0,38
<i>Hpms1-5</i>	0,475	0,501	0,631	0,522	0,796	-	0,556
<i>Hpms2-45</i>	0	0	0	0	0	0,287	0
<i>CAMS336</i>	0,356	0,437	0,398	0,423	0,259	0,132	0,213
<i>HpmsE013</i>	0,332	0,403	0,286	0,315	0	0,224	0,333
<i>Hpms2-2</i>	0,353	0,316	0,278	0	0,182	0,255	0,491
<i>EPMS310</i>	0,223	0,381	0,266	0,41	0	0,236	0,325
<i>Hpms2-24</i>	0,33	0,399	0,286	0,253	0,371	0,362	0,298
<i>Hpms1-3</i>	0	0,289	0,22	0,435	0,54	0,861	0,424
<i>CaSIG19</i>	0,302	0,318	0,315	0,305	0,371	0,411	0
<i>Hpms1-6</i>	0,096	0,383	0,249	0,013	0	0,29	0,605
<i>Hpms1-155</i>	0,309	0,4	0,339	0,341	0,406	0,216	0,371
<i>HpmsE015</i>	0,028	0,064	0,216	0	0,044	0	0
<i>HpmsE016</i>	0,412	0,375	0,464	0,33	0,243	0,299	0,213
<i>HpmsE120</i>	0,314	0,332	0,124	0,207	0,182	0	0,296
<i>HpmsAT2-14</i>	0,21	0,263	0,22	0,413	0,289	0,442	0,288
<i>EPMS426</i>	0,207	0,483	0,41	0,309	0,164	0,256	0,425
<i>EPMS472</i>	0,423	0,344	0,396	0,332	0,265	0,441	0
<i>Hpms1-148</i>	0,298	0,314	0,285	0,417	0,538	0,371	0
<i>GPMS178</i>	0,037	0,22	0,066	0,032	0,075	0,341	0
<i>HpmsE075</i>	0,262	0,312	0	0,337	0	0,318	0,296
<i>GPMS159</i>	0,416	0,382	0,466	0,423	0,33	0,445	0,36

Los valores del coeficiente de endogamia (F_{IS}) resultaron ser elevados en los siete clústeres (Tabla 1.6), apuntando a un déficit de heterocigosis en todos los casos. Para evaluar el posible efecto de los alelos nulos en la estima de F_{IS} , cada clúster se analizó con el modelo de endogamia individual bayesiana (*bayesian individual inbreeding model*, IIM) implementado en INest, que permite calcular el coeficiente de endogamia para datos *multiloci* corregido para la presencia de alelos nulos ($F_{IS-INest}$). En los siete casos se observó que los valores de $F_{IS-INest}$ fueron inferiores a los niveles de F_{IS} (Tabla 1.6), sugiriendo que los alelos nulos añaden sesgo al cálculo de F_{IS} . El análisis evalúa además el efecto de los alelos nulos (n), la endogamia (f) y los errores de genotipado (b) en los niveles de homocigosis empleando el modelo completo (nfb) y lo compara con el modelo que excluye el efecto de los alelos nulos (fb) a través del criterio de información de la desviación (*deviance information criterion*, DIC). Que el modelo nfb presente un DIC más bajo que el modelo fb , indica que los alelos nulos tienen una influencia importante en el cálculo del coeficiente de endogamia (Cabrera-Toledo et al., 2020). En todos los casos excepto para el clúster *C. chinense* se observó un DIC más bajo en el modelo nfb (Tabla 1.6), lo que confirma que las elevadas frecuencias de alelos nulos tienen un peso importante en el cálculo del coeficiente de endogamia.

Tabla 1.6. Valores de los coeficientes de endogamia (F_{IS}) para cada clúster obtenidos con GENEPOP, del coeficiente de endogamia corregido para la presencia de alelos nulos ($F_{IS-INest}$) y del criterio de información de la desviación (DIC) tanto con el modelo nfb como con el modelo fb calculados con INest.

Clúster	F_{IS}	$F_{IS-INest}$	DIC_{nfb}	DIC_{fb}
<i>C. annuum</i> - ibérica	0,7945	0,778	5975,231	5992,429
<i>C. annuum</i> - andina	0,8285	0,8165	3419,367	3424,786
<i>C. annuum</i> - mediterránea	0,8963	0,8939	1921,064	1920,8
<i>C. chinense</i>	0,7812	0,7733	2617,991	2617,202
<i>C. frutescens</i>	0,7417	0,7114	874,908	878,567
<i>C. baccatum</i>	0,8348	0,8263	1770,81	1791,53
<i>C. pubescens</i>	0,7964	0,7914	1167,09	1174,391

Discusión

Diversidad genética

En el presente capítulo se estudiaron 230 líneas de las cinco especies cultivadas de pimiento con 28 marcadores microsatélite que ya habían sido empleados previamente en otros trabajos. En nuestro caso, se observó un mayor número de alelos medio por *locus* que en otros análisis previos del género *Capsicum* (González-Pérez et al., 2014; Silvar & García-González, 2016), aumentando por tanto la información genotípica de los recursos fitogenéticos estudiados. A pesar de que Ibiza et al. (2012) y Nicolai et al. (2013) encontraron un número medio de alelos ligeramente mayor, en ambos casos contaban con un número mayor de accesiones y además incluían varias especies silvestres.

Trabajos previos también encontraron una mayor diversidad alélica de *C. annuum* con respecto a las otras especies (Ibiza et al., 2012; Nicolai et al., 2013; González-Pérez et al., 2014). Algunos de estos autores atribuyen este resultado al mayor número de accesiones estudiadas de esta especie, a su diverso origen geográfico y también a que los marcadores empleados han sido diseñados en *C. annuum* y seleccionados en base al polimorfismo encontrado en esta especie (Ibiza et al., 2012; González-Pérez et al., 2014). Sin embargo, en nuestro estudio se calculó la riqueza alélica, que informa del número medio de alelos corregido por rarefacción en función del tamaño muestral del grupo minoritario y que no había sido incluida en ninguno de los trabajos citados anteriormente. La riqueza alélica mostró valores similares para todas las especies, evidenciando que *C. annuum* no es en realidad más diversa que las demás, refutando lo publicado en los artículos citados.

Se sabe que las especies silvestres de *Capsicum* presentan gran variabilidad aún con pocas accesiones estudiadas (González-Pérez et al., 2014). En nuestro trabajo solo fue posible analizar cuatro accesiones de *C. annuum* var. *glabriusculum*. No obstante, se encontraron 92 alelos y 15 de ellos privados, lo que indica que esta especie mantiene gran reservorio de variabilidad genética.

El alto grado de homocigosis que hemos encontrado en todas las especies está en línea con otros estudios realizados en el género *Capsicum* (Ibiza et al., 2012; Yumnam et al., 2012; Nicolai et al., 2013; González-Pérez et al., 2014; Silvar & García-González, 2016) y guarda relación con el proceso de propagación a través de autopolinización en las variedades cultivadas (Ibiza et al., 2012). Curiosamente, *C. frutescens* y *C. chinense*

presentaron las mayores heterocigosidades observadas pese a tener menor número medio de alelos que *C. annuum*, lo que podría sugerir un mayor nivel de alogamia de estas especies en nuestra colección al ser las que presentan mayor diversidad genética. De hecho, ambas especies son las que presentaron un menor coeficiente de endogamia F_{IS} .

Estructura genética

La separación entre *C. annuum* y no-*C. annuum* encontrada con el análisis de inferencia bayesiana fue observada también por González-Pérez et al. (2014) con una colección de accesiones mayoritariamente españolas y por Rai et al. (2013) con una colección de pimientos de la India, por lo que se descarta que la división esté causada por el origen geográfico de las muestras analizadas en el presente trabajo.

Es destacable que tan solo una accesión de *C. annuum* var. *glabriusculum* aparezca en el clúster de *C. annuum*. Probablemente *C. annuum* fue domesticado a partir de *C. annuum* var. *glabriusculum* en México (Eshbaugh, 1993; Perry et al., 2007; Kraft et al., 2014), lo que explicaría la inclusión en el clúster de *C. annuum* de la accesión PM647, originaria de este país. Hoy en día, en México aún se pueden encontrar variedades silvestres de pimiento genéticamente similares a las cultivadas (Kraft et al., 2014). Que las otras tres accesiones aparezcan con pertenencia al grupo no-*C. annuum* guardaría relación con que en este clúster aparecen las accesiones de *C. frutescens* y de *C. chinense*, pertenecientes al complejo *C. annuum* y por lo tanto próximas genéticamente a *C. annuum* var. *glabriusculum*. Algunos autores hipotetizan que un ancestro común

habría dado lugar a las tres especies domesticadas del complejo *C. annuum* a través de procesos de domesticación en diferentes lugares y éste podría ser *C. annuum* var. *glabriusculum* (Moscone et al., 2007; Nicolai et al., 2013; Pereira-Dias et al., 2019), lo que explicaría que apareciera en ambos clústeres.

Con el objetivo de observar subestructura de manera precisa se siguió una aproximación “multistep”, que consiste en reanalizar con STRUCTURE las accesiones contenidas en cada uno de los dos grupos encontrados por separado. En el análisis con STRUCTURE incluyendo solo las accesiones que aparecían en el clúster *C. annuum* se identificó una estructura en 3 grupos que se corresponde en muy alto grado con el origen de las variedades, apareciendo por un lado agrupadas las muestras andinas, por otro las ibéricas y por último las de la zona mediterránea. La diferenciación de las accesiones ibéricas de las no ibéricas fue observada por González-Pérez et al. (2014) y según apuntan dichas autoras guardaría relación con eventos evolutivos divergentes tras la llegada del pimiento a esta región y que daría lugar a un acervo genético distinto y bien definido. La explicación de que aparezcan accesiones de países americanos, principalmente de México, agrupadas con las de la Península Ibérica residiría en el origen de la especie y posterior introducción en España. Como ya se ha comentado anteriormente, *C. annuum* tendría su origen en México y fue introducida en Europa a través de España tras los viajes de Cristóbal Colón por América (Nuez et al., 1996). Los resultados encontrados lo confirman, evidenciando que las accesiones de *C. annuum* de la Península Ibérica guardan similitudes todavía en la actualidad con las originarias de México.

En otro clúster diferenciado se agruparon las accesiones de la región mediterránea. Se sabe que tras la entrada del pimiento en Europa a través de España, este se diversificó por la cuenca mediterránea (Andrews, 1995; Bosland & Votava, 2000; Krishna De, 2003). Pero además esta zona habría tenido incorporación de pimientos procedentes de otro origen. Según Andrews (1993), tras las incursiones de los ejércitos turcos en la zona del estrecho de Ormuz (que delimita el Golfo Pérsico y el Golfo de Omán) y conseguir abrir rutas más al Sur, éstos habrían importado pimientos mediante las rutas comerciales a través de Asia Menor hacia el mar Negro y lo que hoy se conoce como Hungría, zona que conquistarían en el año 1526. Los primeros registros de pimientos en la zona de Europa central datan de poco tiempo después del primer asedio turco en lo que hoy es Austria, que tuvo lugar en el mismo año 1526. Una de las zonas a las que habrían llegado los turcos y de la que podrían haber importado pimientos era la India (Andrews, 1995). Se sabe que en la India los pimientos fueron introducidos por los portugueses, que recientemente habían descubierto Brasil y podrían importar pimientos de esta región (Boxer, 1984). Hacia 1542 ya existían en la India tres linajes diferentes de pimientos (Purseglove, 1968), constituyendo esta zona un importante centro de diversificación. De este modo, la introducción por vía turca del pimiento en Europa llevaría variedades probablemente distintas a las introducidas a través de la Península Ibérica. Esta hipótesis se vería respaldada por nuestros resultados al encontrar un clúster específico para las accesiones de la región mediterránea, que mantienen diferenciación con las de la Península Ibérica. Si bien las colecciones ibérica y andina han sido sujeto de estudio en trabajos previos, la diversidad genética de la colección mediterránea es analizada en este trabajo por primera vez, contribuyendo a un mayor conocimiento de los recursos fitogenéticos de esta zona, que harán posible un uso más eficiente de los mismos.

El tercer clúster está compuesto por la mayoría de accesiones de Perú, Ecuador, Bolivia y México. En la zona andina tiene su origen el género *Capsicum*, concretamente en la zona tropical de Sudamérica que hoy sería Bolivia (Eshbaugh, 1993). La domesticación de *C. annuum* habría tenido lugar en México (probablemente en la zona central del país) a mediados del Holoceno, hace unos 6.500 años (Pickersgill, 1971; Kraft et al., 2014). Así pues, que todas las accesiones americanas aparezcan próximas entre sí atendería a su origen cercano. La región andina es una importante fuente de diversidad genética del género al ser su punto de origen (Perry et al., 2007; Silvar & García-González, 2016). Del mismo modo, México es un gran centro de variabilidad al haber sido la zona de domesticación de *C. annuum* (Pickersgill, 1971), por lo que las accesiones de este clúster guardarían una amplia diversidad que sería interesante explorar en programas de mejora vegetal. Destaca que en este clúster aparezcan las dos accesiones de las Islas Canarias (BGHZ0082 y BGHZ1114). Tras la vuelta de Cristóbal Colón, el pimiento habría recalado en las islas, ya que como documenta Andrews (1995) la expansión se produjo por la Península Ibérica y las islas oceánicas. Estas variedades de pimiento, recién llegadas de América habrían permanecido confinadas en Canarias, manteniendo muy poco flujo con las de la Península, lo que explicaría que en la actualidad se asemejen más genéticamente a las del continente americano, del cual provienen.

Por otra parte se realizó un nuevo análisis con STRUCTURE de las accesiones que aparecían en el clúster no-*C. annuum* en el análisis global. En este caso se encontraron cinco clústeres que agrupan las líneas en función de la especie a la que pertenecen. González-Pérez et al. (2014) también encontraron una división según la especie cuando realizaron un análisis STRUCTURE para las accesiones de no-*C. annuum*, lo que

demuestra que realizar sucesivos análisis es una buena aproximación cuando se quieren encontrar relaciones más precisas de las accesiones. Con este análisis STRUCTURE se pudieron reclasificar siete accesiones que en los bancos de germoplasma se presentaba como *C. annuum* pero que no lo eran en realidad y que se suman a las cuatro reclasificadas como *C. annuum*. La clasificación errónea en los bancos de germoplasma es un hecho común que se ha encontrado también en otros estudios (González-Pérez et al., 2014; Silvar & García-González, 2016). Las clasificaciones en los bancos de germoplasma se realizan siguiendo criterios fenotípicos que en ocasiones hacen difícil diferenciar especies (González-Pérez et al., 2014), especialmente para las especies del complejo *C. annuum*, que presentan homología morfológica (Walsh & Hoot, 2001; van Zonneveld et al., 2015). Por lo tanto, estudios de genotipado como el aquí llevado a cabo proporcionan resultados exactos sobre las líneas que permitirán un uso más eficaz de las mismas. El correcto uso de los recursos fitogenéticos requiere una completa caracterización previa de los mismos y determinar con exactitud la especie a la que pertenecen es la primera etapa.

Se realizó también un análisis factorial de correspondencia (AFC) para observar las relaciones entre accesiones. Este análisis reveló agrupaciones de las accesiones en función de la especie, apuntando a que esta debería ser la primera distinción encontrada en análisis de diversidad dentro del género *Capsicum*. El hecho de que no se observe la distinción *C. annuum*-no *C. annuum* encontrada en STRUCTURE, sino la separación de todas las especies, podría indicar un sesgo en los resultados de STRUCTURE. En el AFC las especies *C. chinense* y *C. frutescens* aparecen agrupadas, resultado en concordancia con la hipótesis de que ambas serían en realidad la misma especie, según apuntan

algunos autores (McLeod et al., 1979; Walsh & Hoot, 2001). Sin embargo, el AFC realizado solo para las muestras de *C. chinense* y *C. frutescens* mostró a ambas especies separadas claramente, demostrando que son en realidad dos especies aunque próximas genéticamente. La distribución encontrada en el AFC para las cuatro accesiones de *C. annuum* var. *glabriusculum* (una agrupada con las *C. annuum*, otra en el grupo de *C. chinense* y *C. frutescens* y las otras dos en posiciones intermedias entre estos grupos) apoyaría la hipótesis de que este es el pariente silvestre de las tres especies (Moscone et al., 2007; Nicolai et al., 2013; Pereira-Dias et al., 2019), presentándose como el nexo de las tres especies domesticadas del complejo *C. annuum*.

Consideraciones metodológicas

En este estudio se han seguido las tres principales metodologías que permiten investigar la diversidad genética: los estadísticos descriptivos, los métodos exploratorios (AFCs) y los *model-based clustering* (STRUCTURE). De todos ellos, el que ha permitido obtener más información sobre la estructura genética y las relaciones entre accesiones de pimiento fue el llevado a cabo con STRUCTURE. Sin embargo, el hecho de que el resultado global de STRUCTURE fuera una marcada distinción *C. annuum*-no *C. annuum* en vez de la división por especie, más esperable y encontrada en el AFC, hace pensar que este resultado pueda estar sesgado. Ninguno de los estudios que encuentra una distinción *C. annuum* y no-*C. annuum* propone una explicación para esta diferenciación tan clara. La partición en estos dos grupos no es congruente con la filogenia de las cinco especies implicadas calculada por Carrizo García et al. (2016) a partir de las secuencias de ADN de dos regiones plásticas y una nuclear. Estos autores encuentran que *C. annuum* está

próxima a las otras dos especies del complejo *C. annuum* (*C. frutescens* y *C. chinense*), pero más alejada de *C. baccatum* y *C. pubescens*. Sin embargo, el grado de diferenciación de las agrupaciones biológicas reveladas por STRUCTURE a partir de marcadores microsatélite no tiene por qué ser congruente con la diferenciación filogenética obtenida a partir de secuencias de ADN. De hecho, cuando el tiempo de divergencia de los distintos linajes evolutivos es largo, los clústeres creados por STRUCTURE suelen no coincidir con la historia evolutiva de las poblaciones/agrupaciones (Kalinowski, 2011). Aún así, la hipotética diferenciación entre *C. annuum* y no-*C. annuum* podría ser un artefacto analítico. Kalinowski (2011) observó que los clústeres revelados por STRUCTURE pueden estar sesgados por variaciones en el tamaño muestral, de tal modo que el programa incluiría todas las muestras del grupo con mayor tamaño muestral en el mismo clúster. Esta asunción sería congruente con los resultados de nuestro estudio, en donde el clúster mayoritario está conformado por todas las accesiones de *C. annuum*, la especie con mayor tamaño muestral, con diferencia.

Para caracterizar los grupos obtenidos con STRUCTURE en el segundo nivel jerárquico se calcularon determinadas estimas de diversidad genética. Pisa et al. (2015) y Rutten et al. (2019) siguieron una metodología similar, evaluando las posibles desviaciones del equilibrio Hardy-Weinberg, el desequilibrio gamético y los coeficientes de endogamia dentro de cada uno de los clústeres definidos. Según Waples (2015) el desequilibrio de ligamiento entre pares de marcadores también puede influir sobre los resultados de estructura genética al producirse asociaciones no aleatorias entre alelos de distintos *loci*. En nuestro estudio, tan solo siete combinaciones de dos marcadores mostraron desequilibrio de ligamiento en alguno de los siete grupos por lo que este fenómeno no

estaría influyendo de manera sustancial en la estructura. Los microsatélites empleados se seleccionaron de tal modo que cubrieran la mayor parte del genoma, minimizando así la posibilidad de ligamiento físico entre ellos. Pese a que no haya ligamiento físico, dos *loci* pueden encontrarse en situación de desequilibrio de ligamiento si están adyacentes a genes con fuertes presiones selectivas, incluso epistáticas (van der Loo et al., 1987). La migración y la deriva también pueden provocar desequilibrio entre *loci*. Estas podrían ser las causas de encontrarlo en algunas parejas de marcadores.

Además del desequilibrio de ligamiento, las desviaciones del equilibrio Hardy-Weinberg también pueden tener un impacto considerable en los análisis de estructura (Waples, 2015), ya que el algoritmo implementado en STRUCTURE está diseñado para agrupar individuos de poblaciones en equilibrio (Kalinowski, 2011). En nuestro caso se observa una desviación del equilibrio Hardy-Weinberg para el 82,7% de las combinaciones *locus/grupo*. El equilibrio Hardy-Weinberg implica que existiendo apareamiento aleatorio y en ausencia de fuerzas evolutivas (selección, deriva genética, mutación y migración) las frecuencias alélicas y genotípicas se mantienen invariables a lo largo de las generaciones (Hardy, 1908; Weinberg, 1908). En plantas cultivadas esta asunción no se cumple, ya que es evidente que las nuevas generaciones se producen mediante cruces o autocruces dirigidos a partir de variables de interés seleccionadas. Así, estarían actuando las fuerzas de selección (artificial) y migración, además de la deriva por reducirse el tamaño efectivo, disminuyendo la variabilidad genética. El resultado del proceso de propagación a partir de solo unos pocos individuos trae como resultado una fuerte homocigosis (Doebly et al., 2006; Rešetnik et al., 2016) que se traduciría en desequilibrio Hardy-Weinberg. El gran defecto de heterocigosis se

evidencia en los resultados obtenidos, al observarse para todos los grupos un coeficiente de endogamia F_{IS} elevado.

Según Gao et al. (2007) dos implicaciones en el mecanismo reproductivo de especies cultivadas como el pimiento, como son la endogamia y la dispersión limitada de los gametos (que crearía una subestructura de la población), producirían sesgos en los clústeres revelados por STRUCTURE. Estos sesgos vendrían dados porque el apareamiento no aleatorio induce correlaciones en los estados alélicos dentro y entre *loci* (Wright, 1965). Así, en poblaciones no panmícticas, las agrupaciones encontradas por STRUCTURE pueden verse sesgadas. Este sesgo de STRUCTURE podría evitarse con el uso del programa InStruct (Gao et al., 2007). Los autores compararon los resultados de ambos programas analizando 16 accesiones de *Oryza rufipogon*, el pariente silvestre del arroz, genotipadas con 111 SNPs y demostraron que InStruct sí clasifica a los individuos en las subpoblaciones adecuadas sin la suposición de equilibrio Hardy-Weinberg dentro de ellas y que no fusiona subpoblaciones en el mismo clúster de manera artificial.

Por otra parte, todos los marcadores microsatélites empleados fueron desarrollados para *C. annuum* (Lee et al., 2004; Minamiyama et al., 2006; Yi et al., 2006; Nagy et al., 2007). La amplificación cruzada en especies distintas (*cross-species*) puede no ser recomendable porque podrían surgir problemas inherentes como homoplasia en el tamaño de los alelos, presencia de alelos nulos, sesgos en el grado de polimorfismo, motivos de repetición partidos o amplificación de *loci* no ortólogos (Primmer et al., 2005; Barbará et al., 2007; Rutkowski et al., 2009; Yue et al., 2010).

La presencia o no de alelos nulos en estos microsatélites no fue tratada en los correspondientes trabajos de desarrollo de estos marcadores (Lee et al., 2004; Minamiyama et al., 2006; Nagy et al., 2007; Yi et al., 2006) ni en otros trabajos que los apliquen en *Capsicum* spp. (i. e. González-Pérez et al., 2014; Ibiza et al., 2012; Silvar & García-González, 2016). Los alelos nulos están causados por mutaciones en la región de anillamiento del cebador que impiden la amplificación del alelo (Pemberton et al., 1995). Están presentes en la mayoría de taxones (Dakin & Avise, 2004) y uno de las consecuencias principales es la sobreestimación del número de homocigotos, que produciría la desviación del equilibrio Hardy-Weinberg (Carlsson, 2008). Se sabe que el programa STRUCTURE es muy sensible a la presencia de alelos nulos (Falush et al., 2007). El propio manual del programa indica que “cuando hay datos perdidos de manera sistemática surge un serio problema, como en el caso de alelos nulos” (Pritchard et al., 2000). En nuestro estudio, 26 de los 28 marcadores mostraron altas frecuencias de alelos nulos ($>0,2$) (Chapuis & Estoup, 2007). Para contrastar que las altas frecuencias de alelos nulos encontradas para todos los marcadores eran un resultado robusto que podría influir en la estructura poblacional encontrada, se calculó la frecuencia de alelos nulos también con el programa ML-Null Freq, que según Dabrowski et al. (2015) es el método de estima de alelos nulos que mejor rendimiento proporciona. ML-Null Freq también determinó que las frecuencias de alelos nulos eran altas en casi todos los marcadores, con resultados muy similares a los proporcionados con FreeNA, corroborando que este es un resultado real y no un artificio producido por el *software*. Dentro de cada grupo, todos los marcadores mostraron frecuencias de alelos nulos elevadas, por lo que este hecho estaría aumentando artificialmente el nivel de homocigosis en unos organismos que ya de por sí tienen defecto de heterocigotos por la selección humana y los cruces

dirigidos a los que son sometidos. El hecho de que muchos estudios que encuentran una marcada distinción *C. annuum*-no *C. annuum* con STRUCTURE empleen muchos de los microsatélites aquí usados y que en ningún caso evalúan el potencial efecto de altas frecuencias de alelos nulos, nos hace apuntar a que este resultado podría estar sesgado.

Los grupos se evaluaron además con el programa INest, que permite obtener una estima de cuánto pueden estar influyendo los alelos nulos en cada conjunto. El criterio de información de la desviación (DIC) fue más bajo en todos los grupos para el modelo nfb , aquel que tiene en cuenta los alelos nulos, que para el fb , que no los tiene en cuenta; indicando que los alelos nulos influyen en el cálculo de la homocigosis (Chybicki & Burczyk, 2009; Cabrera-Toledo et al., 2020). Todos los grupos mostraron unos coeficientes de endogamia próximos a uno, que disminuyeron ligeramente cuando se calculan corregidos para la presencia de alelos nulos ($F_{IS-INest}$). Esto vendría a indicar también que los alelos nulos incrementan artificialmente el grado de homocigosis, en unas plantas con un nivel ya elevado debido a su carácter endogámico.

En conclusión, los resultados de estructura genética aquí presentados y en otros trabajos similares deben ser tomados con cautela. Al tratarse de especies cultivadas, las accesiones estudiadas presentan un déficit de heterocigotos por su carácter endogámico. El nivel de homocigosis se ve incrementado además por la elevada frecuencia de alelos nulos de todos los marcadores y todo ello hace que se produzca el desvío del equilibrio Hardy-Weinberg. Aún así, postulamos que de todos los descritos, el hecho que mayor influencia tendría en el sesgo de los resultados de STRUCTURE sería el del tamaño desigual de los grupos; ya que en el segundo nivel jerárquico seguirían existiendo alelos

nulos, elevada homocigosis y desvíos del equilibrio Hardy-Weinberg pero los grupos encontrados sí presentan congruencia. Este es el primer estudio que explica las limitaciones que puede presentar el análisis poblacional llevado a cabo con STRUCTURE en recursos fitogenéticos de pimiento empleando marcadores moleculares de tipo microsatélite.

Capítulo 1.2

Evaluación de marcadores moleculares ligados a enfermedades y pungencia

Introducción

Los marcadores moleculares también resultan cruciales en la búsqueda de variedades resistentes a diferentes tipos de estreses a los que se enfrentan las plantas y que en ocasiones afectan a su desarrollo y supervivencia. Concretamente, las plagas producidas por patógenos constituyen en la actualidad la principal limitación en los cultivos y producen pérdidas millonarias en todo el mundo. A pesar de que para muchos patógenos que afectan al pimiento existen fitosanitarios químicos que ayudan a controlar las plagas, su uso está siendo cada vez más restringido ya que pueden resultar tóxicos y causar importantes problemas medioambientales (Djian-Caporalino et al., 2007). Además, algunos patógenos pueden desarrollar tolerancia a los tratamientos químicos (Römer et al., 2010), por lo que es necesario buscar nuevas fuentes de resistencia. En este sentido, los marcadores moleculares resultan útiles en la caracterización de regiones genómicas con relación conocida con genes de resistencia, así como en la identificación posterior de variedades que presenten los genotipos resistentes.

En un principio, para identificar regiones de interés pueden resultar útiles los marcadores moleculares de tipo RAPD (*random amplified polymorphic DNA*, ADN polimórfico amplificado al azar) porque se ha observado que pueden ser muy sensibles en regiones adyacentes a *loci* de interés (Michelmore et al., 1991). Sin embargo, los RAPDs presentan serias desventajas como la baja repetibilidad y que los marcadores diseñados en una población puedan no ser polimórficos en otra, lo que disminuye su fiabilidad (Moury et al., 2000). Para subsanar este problema se desarrollaron los

marcadores SCAR (*sequence characterized amplified region*, región amplificadas caracterizada por secuencia), marcadores generados tras secuenciar las regiones RAPD y desarrollar cebadores más específicos (Paran & Michelmore, 1993). Si aún los SCARs resultasen monomórficos, los marcadores de tipo CAPS (*cleaved amplified polymorphic sequence*, secuencia polimórfica amplificada y digerida) pueden subsanar este inconveniente al emplear una digestión con enzimas de restricción tras el proceso de amplificación (Konieczny & Ausubel, 1993).

La agricultura del pimiento se ve afectada por enfermedades causadas por bacterias, virus, nematodos, fitoplasmas, plantas parásitas, hongos y oomicetes (Agrios, 2005; Babu et al., 2011). A continuación se describen los principales patógenos que afectan a los cultivos de pimiento y que se estudian en este capítulo.

El oomicete *Phytophthora capsici* es uno de los patógenos que más afecta al cultivo de pimiento en todo el mundo (Hausbeck & Lamour, 2004; Babu et al., 2011). Los síntomas que produce varían según el hospedador pero incluyen el deterioro de las raíces, marchitamiento foliar, lesiones en el tallo, marchitamiento de los frutos y necrosis de la planta (Babu et al., 2011; Parada-Rojas & Quesada-Ocampo, 2018).

Junto con *P. capsici*, el nematodo *Meloidogyne incognita* es la otra enfermedad transmitida por el suelo que más afecta a las plantaciones de pimiento en todo el mundo (Babu et al., 2011) y puede llegar a causar la pérdida del 100% de la producción en grandes cultivares (Perry et al., 2009; Onkendi et al., 2014). Cuando el parásito infecta a la planta, las células vegetales son modificadas, convirtiéndose en células metabólicamente muy activas de

las que tomar nutrientes, restándole capacidad de crecimiento a la planta y dando lugar a un cultivo deficiente (Perry et al., 2009; Jones et al., 2011). La alteración de la fisiología y el crecimiento de la planta puede favorecer además la infección por otros patógenos oportunistas del suelo (Djian-Caporalino et al., 2007), por lo que los daños causados por *M. incognita* pueden ser aún mayores.

Xanthomonas campestris es una bacteria que produce la enfermedad conocida como “mancha bacteriana del pimiento”, una de las más devastadoras en los cultivos de esta hortícola, tanto en invernadero como en campo abierto (Bouzar et al., 1994; Kousik & Ritchie, 1996). El patógeno produce necrosis en las hojas, tallos, flores y frutos (Potnis et al., 2015) y la infección se ve incrementada con condiciones de humedad o precipitaciones elevadas (Römer et al., 2010).

En cuanto a las patologías causadas por virus, los principales virus que afectan al pimiento son los potyvirus, el virus del bronceado del tomate (*tomato spotted wilt virus*, TSWV), el virus del mosaico del pepino (*Cucumovirus*, CMV) y los tobamovirus. *Potyvirus* es un género que pertenece a la familia *Potyviridae*, que representa el mayor grupo de virus en vegetales. Dentro de este grupo destacan el virus Y de la patata (*potato virus Y*, PVY), el virus del grabado del tabaco (*tobacco etch virus*, TEV) y el virus del moteado del pimiento, (*pepper mottle virus*, PepMov) (Kim et al., 2015). Los potyvirus se transmiten por áfidos (insectos de la familia *Aphididae*) y su control es especialmente difícil porque cada virus puede ser transmitido por muchos grupos de áfidos distintos y una misma especie de áfido puede transmitir a su vez muchos tipos de potyvirus (Gibbs et al., 2008; Ivanov et al., 2014). El virus del bronceado del tomate (TSWV) pertenece al

grupo de los tospovirus, que son transmitidos por trips (insectos de la familia Thripidae) y constituyen la mayor amenaza vírica en pimiento en la actualidad, al estar surgiendo numerosas cepas resistentes (Turina et al., 2012). Se ha documentado que el TSWV puede infectar hasta a 1100 especies distintas, incluyendo otras plantas de interés comercial como el tomate y la lechuga (Parrella et al., 2003). El TSWV produce en el pimiento atrofia en el crecimiento de las plantas jóvenes, manchas cloróticas en las hojas y necrosis en hojas y frutos (Turina et al., 2012). El virus del mosaico del pepino (CMV) afecta a más de 1000 especies de plantas y es el único cucumovirus capaz de infectar al pimiento (Moury & Verdin, 2012). Produce pérdidas que pueden llegar al 80% del total del cultivo ya que afecta al crecimiento vegetativo de la planta y al fruto, con síntomas que incluyen crecimiento anormal de las hojas, frutos deformes, necrosis y muerte de la planta (Avilla et al., 1997; Moury & Verdin, 2012). Por último, los tobamovirus causan importantes pérdidas en las producciones de pimiento al producir clorosis y alteraciones en las hojas y formas, colores y tamaños anormales de los frutos. Son un grupo de virus especialmente difícil de controlar al transmitirse a través de las semillas y dada su alta estabilidad también pueden permanecer en el suelo y en el agua (Moury & Verdin, 2012).

La caracterización del perfil de resistencia de las accesiones conservadas en los bancos de germoplasma a los patógenos descritos previamente permitirá identificar las mejores candidatas como donantes de material genético en programas de mejora vegetal y emplearlas para obtener híbridos resistentes según las necesidades de cultivo (Babu et al., 2011). Así, en este capítulo se analizan 96 accesiones de *C. annuum* de la Península Ibérica y la región mediterránea con diez marcadores moleculares asociados a genes de resistencia a enfermedades, de tipo SCAR, CAPS y tetra-ARMS (*amplification refractory*

mutation system, sistema de amplificación de mutaciones refractarias), con el objetivo de detectar genotipos resistentes de interés.

Además de buscar accesiones potencialmente resistentes, los marcadores moleculares también permiten identificar variedades con otras características de interés. En este capítulo, se ha analizado un marcador SCAR asociado al carácter pungente. La pungencia es una de las principales características que aportan calidad a los cultivos de pimiento (Rodríguez-Maza et al., 2012) y viene dada por la síntesis de unos metabolitos secundarios denominados capsaicinoides en la placenta de los frutos (Suzuki et al., 1980). Estos compuestos poseen capacidad antiinflamatoria, analgésica y antioxidante (The Capsaicin Study Group, 1991; Kirschbaum-Titze et al., 2002; Ochi et al., 2003; Pino et al., 2007) por lo que presentan un gran interés en salud humana, siendo un valor añadido al consumo de pimiento.

Material y métodos

Material vegetal

La potencial resistencia a enfermedades causadas por patógenos que afectan al pimiento se realizó en 96 accesiones de *C. annuum* originarias de la Península Ibérica y la zona mediterránea (Tabla 1.1, columna “Marcadores enfermedad”).

Extracción de ADN

El proceso de siembra de las plantas, extracción y cuantificación de ADN se desarrolló según lo descrito previamente en el Capítulo 1.1 para el análisis con marcadores microsatélite. Las alícuotas de ADN se prepararon a una concentración final de 25ng/μl.

Análisis molecular

Se evaluó la resistencia potencial a siete patógenos del pimiento: *Potyvirus*, virus del bronceado del tomate (TSWV), virus del mosaico del pepino (CMV), *Tobamovirus*, *Phytophthora capsici*, *Meloidogyne incognita* y *Xanthomonas campestris*. Para ello, se analizaron diez marcadores moleculares de tipo SCAR, CAPS y tetra-ARMS asociados a genes. El carácter pungente se evaluó con el marcador SCAR *MAP1* (Tabla 1.7).

Tabla 1.7. Marcadores moleculares ligados a enfermedades y pungencia analizados en las dos colecciones de *Capsicum* spp.

Marcador	Tipo	Enzima de restricción	Carácter asociado	Tipo de patógeno	Locus	Fuente
<i>OpD04.717</i>	SCAR	-	<i>P. capsici</i>	Oomicete	<i>Phyto.5.2</i>	Quirin et al. (2005)
<i>060I2END</i>	SCAR	-	<i>Tobamovirus</i>	Virus	<i>L⁴</i>	Yang et al. (2009)
<i>SCAR_CD</i>	SCAR	-	<i>M. incognita</i>	Nematodo	<i>Me1, Me7</i>	Djian-Caporalino et al. (2007)
<i>PR-Bs3</i>	SCAR	-	<i>X. campestris</i>	Bacteria	<i>Bs3</i>	Römer et al. (2010)
<i>CaTm-int1</i>	CAPS	<i>HinfI</i>	CMV	Virus	<i>Cmr1</i>	Kang et al. (2010)
<i>SCAC568</i>	CAPS	<i>TaqI</i>	TSWV	Virus	<i>Tsw</i>	Moury et al. (2000)
<i>Pvr1-S/Pvr1-R1</i>	CAPS	<i>BsrI/Fnu4HI</i>	<i>Potyvirus</i>	Virus	<i>Pvr1</i>	Yeam et al. (2005)
<i>Pvr1-R2</i>	CAPS	<i>HindIII</i>	<i>Potyvirus</i>	Virus	<i>Pvr1</i>	Yeam et al. (2005)
<i>T200A</i>	Tetra-ARMS	-	<i>Potyvirus</i>	Virus	<i>Pvr2-eIF4E</i>	Rubio et al. (2008)
<i>CSO</i>	CAPS	<i>AlwNI</i>	<i>Potyvirus</i>	Virus	<i>Pvr4</i>	Caranta et al. (1999)
<i>MAP1</i>	SCAR	-	Pungencia	-	<i>Pun1</i>	Rodríguez-Maza et al. (2012)

Las condiciones de PCR se estandarizaron para los marcadores *Pvr1-S/Pvr1-R1*, *Pvr1-R2*, *SCAC568*, *CaTm-int1*, *060I2END*, *OpD04.717* y *SCAR_CD* empleando los protocolos

descritos por sus respectivos autores (Tabla 1.7) con ciertas modificaciones. Las amplificaciones se realizaron en reacciones de PCR en un volumen final de 15 µl que contenían 50 ng de ADN, 1X buffer PCR (NZYTech), 2,5 mM MgCl₂ (NZYTech), dNTPs (NZYTech) a una concentración de 0,2 mM cada uno, 0,2 µM de *primer forward*, 0,2 µM de *primer reverse* y 0,5 U *Taq* polimerasa (NZYTaq DNA Polymerase, NZYTech). El programa de PCR consistió en una desnaturalización inicial a 94°C 3 minutos, 35 ciclos de 94°C 45s, 55°C 45s y 72°C 1 min y una extensión final a 72°C durante 8 min.

Para el marcador *T200A*, la PCR se llevó a cabo en un volumen total de 15 µl con 50 ng de ADN, 1X buffer PCR (NZYTech), 2,5 mM MgCl₂ (NZYTech), dNTPs (NZYTech) a una concentración de 0,2 mM cada uno, 0,1 µM de *primer forward-inner*, 0,1 µM de *primer reverse-inner*, 0,05 µM de *primer forward-outer*, 0,05 µM de *primer reverse-outer* y 0,5 U *Taq* polimerasa (NZYTaq DNA Polymerase, NZYTech). El programa de PCR empleado fue el descrito por Rubio et al. (2008) y consistió en una desnaturalización a 94°C durante 3 min; 35 ciclos de 94°C durante 30s, 54°C durante 30 s y 72°C durante 1 min y con una extensión final a 72°C durante 1 min. En este caso se utilizan dos parejas de *primers* para la amplificación específica de dos alelos con tamaños diferentes que pueden distinguirse mediante electroforesis en gel de agarosa y que por lo tanto permite resolver de manera sencilla un marcador codominante.

Para los marcadores *CSO* y *PR-Bs3* la reacción se efectuó en un volumen total de 15 µl conteniendo 50 ng de ADN, 1X buffer PCR (NZYTech), 2,5 mM MgCl₂ (NZYTech), dNTPs (NZYTech) a una concentración de 0,2 mM cada uno, 0,2 µM de *primer forward*, 0,2 µM de *primer reverse* y 0,5 U *Taq* polimerasa (NZYTaq DNA Polymerase, NZYTech).

El programa de PCR para CSO consistió en una desnaturalización inicial a 93°C durante 1 min; 35 ciclos a 93°C durante 45 s, 47°C durante 1 min y 72°C durante 2 min y una extensión a 72°C durante 8 min. Para amplificar el marcador *PR-Bs3* se siguió un programa de PCR que consta de una desnaturalización a 94°C durante 2 min; 35 ciclos a 94°C durante 10s, 58°C durante 15 s y 72°C durante 45s y extensión final a 72°C durante 5 min.

Los marcadores de tipo *SCAR* y *Tetra-ARMS* fueron revelados directamente mediante electroforesis en gel de agarosa 2-3% en 1X TAE (Tris-acetato 40 mM pH 8, EDTA 1 mM) para observar el tamaño de los fragmentos amplificados. Los productos de PCR de los marcadores tipo CAPS se digirieron con las correspondientes enzimas (Tabla 1.7). Las reacciones de digestión se llevaron a cabo en un volumen final de 15 µl que contenían 10 µl del producto de PCR, 0,5 U de la enzima correspondiente (ThermoFisher Scientific) y 1X buffer de reacción recomendado por la casa comercial, durante 3 h y a la temperatura óptima para cada enzima indicada por el fabricante. Los productos resultantes se visualizaron mediante electroforesis en gel de agarosa 2% en 1X TAE.

Las accesiones se analizaron también para el marcador de pungencia *MAP1* (Rodríguez-Maza et al., 2012). La reacción de PCR para este marcador se realizó en un volumen total de 15 µl que contenían 100 ng de ADN genómico, 1X buffer PCR (NZYTech), 2,5 mM MgCl₂ (NZYTech), dNTPs (NZYTech) a una concentración de 0,2 mM cada uno, 0,02 µM de *primer forward*, 0,2 µM de *primer reverse*, 0,18 µM *primer universal M13* y 0,5 U *Taq* polimerasa (NZYTaq DNA Polymerase, NZYTech). El programa de PCR consistió en una desnaturalización inicial a 94°C durante 5 min; 12 ciclos de 94°C durante 30 s, 62-

56°C (descendiendo a un ritmo de 0,5°C por ciclo) durante 30s, 72°C durante 30s; 35 ciclos 94°C durante 30 s, 56°C durante 30s, 72°C durante 30s; y una extensión final a 72°C durante 10 min. La detección de los productos de amplificación se realizó mediante electroforesis en gel de agarosa al 2% en 1X TAE teñido con *GreenSafe Premium* (NZYTech) y revelado por medio de luz ultravioleta en un fotodocumentador GelDoc XR+ (Bio-Rad). El análisis de los fragmentos amplificados se realizó en los SAI-UDC empleando el mismo protocolo que el descrito previamente para los marcadores microsatélites. La asignación de los alelos se realizó con el programa Geneious 5.6.7.

Resultados

Noventa y seis accesiones de pimiento fueron analizadas con 10 marcadores que aportan información sobre su resistencia potencial ante los principales patógenos que afectan al pimiento. Los genotipos de cada accesión para cada marcador se muestran en la Tabla A.3. Para el marcador *OpD04.717*, la resistencia potencial frente a *P. capsici* viene determinada por el alelo dominante de 700 pb. El marcador *060I2END* informa de resistencia potencial a *Tobamovirus* y presenta una banda de 751 pb para el alelo susceptible, dominante sobre el resistente (banda ausente). Para ninguno de estos dos marcadores se encontró genotipo resistente en las accesiones estudiadas. *SCAR_CD* es un marcador que informa de resistencia potencial frente a *M. incognita*. El alelo resistente es dominante y viene dado por una banda de 107 pb y una de 170 pb, mientras que en el susceptible solo aparece la de 170 pb. Se halló resistencia potencial para *M. incognita* en las accesiones CGN23244 y CAP913 de la región mediterránea (5,88% del total de accesiones de la colección con genotipo). Por su parte, la resistencia potencial a

X. campestris se estudió con el marcador *PR-Bs3*, que puede presentar amplicones de 87 pb asociados al alelo resistente o de 100 pb asociados a la susceptibilidad. En este caso se encontró resistencia potencial para la accesión de *C. annuum* ibéricas BGHZ1756 (1,72%). En el caso del marcador *CaTm-int1*, tras la digestión con la enzima se encuentra una banda de 450 pb para las accesiones que tienen el alelo susceptible, mientras que en las potencialmente resistentes se observarán dos, de 150 y 300 pb. Para este marcador no se encontró resistencia potencial en ninguna línea. El marcador *SCAC568* indica resistencia potencial frente al virus del bronceado del tomate cuando aparece una única banda de 568 pb mientras que el alelo susceptible viene indicado por la presencia de bandas de 220 y 320 pb. Dos de las accesiones de la colección mediterránea mostraron el genotipo resistente para este marcador (5,88%), la CGN17193 y la CGN22825 (Tabla 1.8).

Tabla 1.8. Porcentaje de accesiones de cada colección con genotipo resistente para cada marcador de resistencia y con genotipo pungente para el marcador *MAP1*.

Colección	<i>OpD04.717</i> (<i>P. capsici</i>)	<i>060I2END</i> (<i>Tobamovirus</i>)	<i>SCAR_CD</i> (<i>M. incognita</i>)	<i>PR-Bs3</i> (<i>X. campestris</i>)	<i>CaTm-int1</i> (<i>CMV</i>)	<i>SCAC568</i> (<i>TSWV</i>)	<i>MAP1</i> (<i>pungencia</i>)
<i>C. annuum</i> P. Ibérica	0,00	0,00	0,00	1,72	0,00	0,00	15,00
<i>C. annuum</i> Mediterránea	0,00	0,00	5,88	0,00	0,00	5,88	48,48

La resistencia potencial frente a *Potyvirus* se evaluó mediante la combinación de los cinco marcadores *Pvr1-S*, *Pvr1-R1*, *Pvr1-R2*, *T200A* y *CSO* (Tabla 1.9). El alelo susceptible *Pvr1*⁺ fue el mayoritario en ambas colecciones con frecuencias similares (93,55% en la Península Ibérica y 94,12% en la región mediterránea). El alelo de resistencia *Pvr1*¹ también se encontró a su vez en proporciones semejantes en ambas colecciones, en un 6,45% de las ibéricas y en un 5,88% de las mediterráneas. Ni el alelo *Pvr1* ni el *Pvr1*² se

halló en ninguna accesión. El marcador *T200A* (Rubio et al., 2008) permite identificar variedades resistentes a PepMov y los patotipos 0, 1 y 2 de PVY a través de una reacción PCR que emplea cuatro *primers*. El patrón electroforético determinará el genotipo según el número de bandas visualizadas: el homocigoto susceptible presentará una banda de 402 pb que aparece en todos los casos y una banda de 199 pb correspondiente al alelo susceptible *Pvr2⁺*; para líneas heterocigotas (susceptibles) se observará la banda de 402 pb, la de 199 pb (*Pvr2⁺*) y una de 258 pb correspondiente a uno de los tres alelos de resistencia (*Pvr2¹*, *Pvr2²* o *Pvr2³*); por último, un genotipo resistente presentará la banda de 402 pb y la de 258 pb, perteneciente a alguno de los alelos *Pvr2¹*, *Pvr2²* o *Pvr2³*. Los resultados mostraron genotipo resistente en cuatro accesiones ibéricas (6,90% del total de accesiones con amplificación para el marcador) y en dos mediterráneas (6,25%). El marcador *CSO* está asociado al gen *Pvr4* y también informa de potencial resistencia frente a PepMov y patotipos 0, 1 y 2 de PVY. El alelo resistente viene indicado por la presencia de una banda de 444 pb y el susceptible por una de 458 pb. Se encontró resistencia potencial en ambas colecciones, siendo mayor en la mediterránea, con el 23,53% frente al 6,56% de la ibérica (Tabla 1.9).

Tabla 1.9. Porcentaje de accesiones de cada colección con cada uno de los alelos de los marcadores *Pvr1-S*, *Pvr1-R1* y *Pvr1-R2* y con genotipos resistentes para los loci *Pvr2* y *Pvr4*.

Colección	<i>Pvr1⁺</i>	<i>Pvr1</i>	<i>Pvr1¹</i>	<i>Pvr1²</i>	<i>Pvr2</i>	<i>Pvr4</i>
<i>C. annuum</i> P. Ibérica	93,55	0,00	6,45	0,00	6,90	6,56
<i>C. annuum</i> Mediterránea	94,12	0,00	5,88	0,00	6,25	23,53

La presencia de los alelos *Pvr1⁺*, *Pvr1*, *Pvr1¹* y *Pvr1²* se infiere de la utilización de los marcadores *Pvr1-S*, *Pvr1-R1* y *Pvr1-R2*. El genotipo resistente en el gen *Pvr2* se infiere con el marcador *T200A* y en el gen *Pvr4* con el *CSO*.

La pungencia se estudió mediante el marcador *MAP1*, que presenta una banda de 494 pb asociado a la pungencia y de 479 pb asociado a la ausencia de picante. La frecuencia del alelo pungente varió entre las colecciones de *C. annuum* analizadas, encontrándose en el 15% de las ibéricas y el 48,48% de las mediterráneas (Tabla 1.8).

Discusión

Las plagas en los cultivos de pimiento son el principal factor que afecta a su productividad, por lo que el uso de los recursos fitogenéticos en programas de mejora a través de cruces o ingeniería genética para conseguir variedades resistentes es esencial para paliar las pérdidas provocadas en el cultivo. El correcto uso de los recursos fitogenéticos requiere una precisa y completa caracterización, por lo que los estudios con marcadores moleculares como el que aquí se presenta añaden valor al material empleado, ya que permitirán identificar variedades de pimiento que presenten genotipos que le puedan conferir resistencia a patógenos y presentarlas como candidatas a su inclusión en los programas de mejora. Además se evaluó el potencial carácter picante con un marcador asociado a la generación de capsaicinoides, importantes compuestos nutraceuticos del pimiento.

Resistencia frente a *Phytophthora capsici* (OpD04.717)

OpD04.717 es un marcador asociado al *locus Phyto.5.2*, uno de los seis *loci* de caracteres cuantitativos (*quantitative trait loci*, QTLs) que contribuyen a crear resistencia al oomicete *Phytophthora capsici* (Thabuis et al., 2003; Quirin et al., 2005). Di Dato et al. (2015) y Silvar

& García-González (2017) encontraron resistencia para este marcador en *C. chinense* y *C. baccatum* y apuntaron a las accesiones de *C. annuum* de México como fuente de resistencia. El hecho de no encontrar resistencia potencial en nuestro estudio en ninguna muestra originaria de centros de diversificación posteriores, pero sí en la zona de origen de *C. annuum* en otros trabajos, haría apuntar a accesiones de esta región como las más aptas para hallar resistencia ante el patógeno.

Resistencia frente a *Tobamovirus* (060I2END)

El marcador *060I2END* está ligado al locus *L⁴*, que confiere resistencia a los patotipos *P_{1,2,3}*, los más agresivos y comunes de todos los tobamovirus (Yang et al., 2009). En nuestro caso no se encontró resistencia potencial en ninguna línea de *C. annuum*. Estos resultados están en consonancia con estudios previos que encontraron resistencia en todas las especies cultivadas de pimiento excepto en *C. annuum* (Di Dato et al., 2015; Silvar & García-González, 2017). Este resultado apuntaría a que la resistencia se ha perdido en la mayoría de líneas de *C. annuum*, probablemente ligado a los procesos de expansión y mejora a los que se vio sometida la especie.

Resistencia frente a *Meloidogyne incognita* (SCAR_CD)

El QTL *SCAR_CD* está asociado a diferentes genes (*Me1*, *Me7*) relacionados con resistencia a nematodos, concretamente se ha comprobado su implicación en la defensa frente a nematodos del género *Meloidogyne* (Djian-Caporalino et al., 2007). No se encontró resistencia potencial ante nematodos en las accesiones de la Península Ibérica.

Silvar & García-González (2017) encontraron genotipo resistente en accesiones españolas pero en baja proporción (9,7%), indicando que los pimientos de esta región presentarían poca resistencia potencial a este patógeno. El hecho de que en nuestro estudio sí se encontrase genotipo resistente en dos accesiones de la colección mediterránea apuntaría a que en esta zona la resistencia también se mantendría en bajo nivel. La correcta caracterización de las accesiones con marcadores en los que las frecuencias de resistencia son pequeñas es, si cabe, más importante; ya que el hecho de identificar algunas líneas potencialmente resistentes las presenta como muy buenas candidatas en programas de mejora.

Resistencia frente a *Xanthomonas campestris* (PR-Bs3)

Para el QTL *PR-Bs3*, asociado a resistencia ante la bacteria *X. campestris*, se encontró el alelo resistente en una única accesión de *C. annuum* ibéricas, la BGHZ1756. Silvar & García-González (2017) encontraron resistencia potencial con este marcador en una accesión de *C. annuum* de Ecuador y una de India. Los resultados de estos autores, junto con los aquí presentados, indicarían que para este patógeno existe cierta resistencia en la zona de origen del pimiento así como en sus centros de diversificación secundarios; pero que se está perdiendo esa resistencia en otras zonas de diversificación posteriores como la región mediterránea. El marcador está asociado al gen *Bs3*, implicado en la resistencia a distintas cepas de *X. campestris* (Römer et al., 2010). Otros genes implicados en resistencia a *Xanthomonas* tradicionalmente empleados en programas de mejora vegetal como el *Bs1* y el *Bs2* ya no son útiles porque se han encontrado cepas inmunes a la resistencia conferida por estos genes (Jordan et al., 2006). Por lo tanto, el estudio de

Bs3 es muy útil al conferir resistencia frente a las cepas 0, 1, 4, 7 y 9 de *Xanthomonas* (Römer et al., 2010) e identificar variedades de pimiento que la presenten es vital de cara a su selección para el cultivo.

Resistencia frente al virus del mosaico del pepino (*CaTm-int1*)

La potencial resistencia al virus del mosaico del pepino (*Cucumovirus*, CMV) se estudió a través del marcador *CaTm-int1*, que está ligado el gen de resistencia *Cmr1*, el único identificado con efecto real sobre la resistencia a este patógeno (Kang et al., 2010). No se encontró genotipo resistente para este marcador en ninguna de las accesiones de *C. annuum* empleadas y estudios previos empleando este marcador tampoco encontraron resistencia en esta especie aunque sí en otras como *C. chinense* y *C. frutescens* (Di Dato et al., 2015; Silvar & García-González, 2017). El hecho de no encontrar el alelo resistente en *C. annuum* puede deberse a que esta especie no presente resistencia frente a CMV o a que la resistencia venga dada por un mecanismo genético diferente al de las otras especies.

Resistencia frente al virus del bronceado del tomate (*SCAC568*)

Para estudiar la potencial resistencia al virus del bronceado del tomate (*tomato spotted wilt virus*, TSWV) se empleó el marcador *SCAC568*, asociado al locus *Tsw* (Moury et al., 2000). Pese a que *C. annuum* sería una especie que mayoritariamente muestra susceptibilidad frente al virus del bronceado del tomate (Di Dato et al., 2015), en contraposición a otras especies de pimiento como *C. baccatum*, *C. chinense* y *C. frutescens*

que serían buenos reservorios de resistencia frente a este patógeno (Cebolla-Cornejo et al., 2003); los resultados aquí encontrados presentan dos líneas de esta especie potencialmente resistentes que podrían ser empleadas como donantes de resistencia en programas de mejora vegetal. Estas líneas son CGN17193 y CGN22825. El hecho de que no se haya encontrado el genotipo resistente en ninguna accesión de *C. annuum* de la Península Ibérica y sí en las de la región mediterránea apuntaría a una pérdida de la resistencia en la Península Ibérica tras su introducción, debida a la selección a lo largo de siglos de cultivo, pero que se mantiene en la cuenca mediterránea, zona esta en donde los pimientos podrían tener un origen distinto al de España y Portugal y, por lo tanto, mantendrían un reservorio genético distinto, tal como se apuntó en los resultados obtenidos con los marcadores microsatélite.

Resistencia frente a *Potyvirus* (*Pvr1-S*, *Pvr1-R1*, *Pvr1-R2*, *CSO* y *T200A*)

La resistencia potencial ante *Potyvirus* ha sido estudiada con distintos marcadores ligados a los genes *Pvr1*, *Pvr2* y *Pvr4*. Además del alelo susceptible *Pvr1*⁺, el gen *Pvr1* puede presentar tres alelos que le confieren diferentes niveles de resistencia ante el virus del grabado del tabaco (*tobacco etch virus*, TEV) (alelos *Pvr1*, *Pvr1*²); la cepa 0 del virus Y de la patata (*potato virus Y-0*, PVY-0,) (alelos *Pvr1*, *Pvr1*¹, *Pvr1*²); y virus del moteado del pimiento (*pepper mottle virus*, PepMov,) (alelo *Pvr1*) (Yeam et al., 2005). La presencia de cada alelo viene revelada por el análisis con tres marcadores ligados al gen *Pvr1* desarrollados por Yeam et al. (2005): *Pvr1-S*, *Pvr1-R1* y *Pvr1-R2*. En primer lugar, el marcador *Pvr1-S* informa de la presencia del alelo *Pvr1*⁺, siendo entonces la accesión susceptible, o de uno de los tres alelos resistentes, siendo entonces la accesión resistente

a al menos uno de los tres patógenos. Si la accesión resulta ser resistente se aplica seguidamente el marcador *Pvr1-R1*. Para este marcador, la accesión puede presentar el alelo *Pvr1* y, por lo tanto, ser resistente a los tres potyvirus, o bien ser susceptible a alguno de ellos al portar alguno de los otros dos alelos de resistencia. En este último caso se analiza el marcador *Pvr1-R2* que revela la presencia del alelo *Pvr1¹* o del alelo *Pvr1²*.

De los tres alelos del gen *Pvr1* que confieren resistencia frente a *Potyvirus*, tan solo se encontró el *Pvr1¹* en seis accesiones en total, por lo que atendiendo a los resultados de estos marcadores, en ambas colecciones no se mantendría resistencia potencial frente a TEV y PepMov. Las bajas frecuencias de genotipos resistentes encontradas en las zonas de estudio, evidencian la poca diversidad genética y por lo tanto la baja capacidad de adaptación que tendrían accesiones de centros de diversificación, en donde se han ido seleccionando las variedades más aptas a lo largo del año pero tendiendo a la homogeneidad genética.

El gen *Pvr2-eIF4E* también está implicado en el control de la resistencia a distintos tipos de *Potyvirus* (Ruffel et al., 2002; Robaglia & Caranta, 2006). El efecto del gen puede estar codificado por diez alelos, siendo *Pvr2¹*, *Pvr2²* y *Pvr2³* los más empleados en los programas de mejora en cultivares comerciales ya que proporcionan resistencia frente a PVY-0, PVY-1 y las cepas más comunes de TEV; todos ellos son recesivos con respecto al alelo susceptible *Pvr2⁺* (Kyle & Palloix, 1997; Ruffel et al., 2006). Se encontró la presencia de algún alelo resistente en las mismas accesiones en las que apareció el alelo *Pvr1¹*. Otros autores observaron que la resistencia potencial para este marcador solo aparecía en *C. annuum* (Di Dato et al., 2015; Silvar & García-González, 2017) por lo que

detectar accesiones con genotipo resistente es determinante. El hecho de que las seis accesiones presenten ambos alelos de resistencia las destaca frente a otras al poseer gran nivel de resistencia potencial.

El gen *Pvr4* proporciona resistencia frente a los patotipos 0, 1 y 2 de PVY y PepMov ya que promueve la inhibición de la replicación de los virus en las células huésped del pimiento (Dogimont et al., 1996). Asociado a este *locus* se desarrolló el marcador CSO (Caranta et al., 1999) que aporta información sobre resistencia potencial a estos patógenos. Se encontró genotipo resistente en cuatro accesiones de *C. annuum* ibéricas y en ocho mediterráneas. Los resultados aportados por este marcador son muy relevantes, pues, al no encontrar el alelo *Pvr1* con los marcadores ligados al gen *Pvr1* (*Pvr1-S*, *Pvr1-R1* y *Pvr1-R2*) solamente las accesiones con genotipo resistente para CSO tendrían resistencia potencial a PepMov.

Tras analizar todos los marcadores asociados a resistencia frente a *Potyvirus*, las accesiones de *C. annuum* BGHZ0102, BGHZ0224, BGHZ2546, BGHZ5912, CAP1661, 13027-BPGV, 13113-BPGV, 07149-BPGV, CAP242, CGN16800, CGN16833, CGN16835, CGN17191, CGN17193, CGN22154, CGN22183 y CGN24358 destacan con genotipos resistentes para algún *Potyvirus*, lo que las convierte en excelentes candidatas de cara al cultivo o a actuar como donantes de material genético en programas de mejora genética vegetal. Además, atendiendo a los marcadores para los demás patógenos, también presentan interés las accesiones BGHZ1756, CAP913, CGN22825 y CGN23244, además de la ya citada CGN17193.

Carácter pungente (MAP1)

El carácter picante o no en el pimiento es una de las características más importantes en la selección de variedades para el cultivo. La pungencia viene determinada por distintos genes, entre los que se encuentra *Pun1* (Stellari et al., 2010; Di Dato et al., 2015). El marcador *MAP1* se ha desarrollado a partir de una secuencia relacionada con el *locus Pun1* (Rodríguez-Maza et al., 2012).

Los resultados muestran que la pungencia en *C. annuum* se ha mantenido en las diferentes zonas de cultivo pero con distinta frecuencia, lo que vendría determinado por la selección específica de esta característica según los requerimientos de cultivo de cada región. Además, la selección de otras propiedades de los frutos para su cultivo podría estar interfiriendo en las frecuencias alélicas de *Pun1* por ligamiento genético con otras regiones del cromosoma 2 (Ben-Chaim et al., 2001, 2006). Se ha estudiado además que la ruta de biosíntesis de capsaicinoides puede estar vinculada a otros rasgos agronómicos que se encuentren bajo fuertes presiones selectivas, como la acumulación de biomasa (Han & Lincoln, 1994). Así, este carácter estaría sometido a fuertes presiones directas e indirectas que conllevarían una frecuencia diferente de los alelos según la zona de cultivo.

Capítulo 2

Análisis de la diversidad, estructura genética y relaciones entre accesiones del género *Capsicum* mediante *genotyping-by-sequencing* (GBS)

Introducción

El análisis de marcadores microsatélite en recursos fitogenéticos de *Capsicum* spp. ha permitido comprender mejor los patrones de diversidad genética dentro del género (González-Pérez et al., 2014). Sin embargo, los *loci* microsatélite presentan algunas limitaciones: (1) el empleo de un número bajo de marcadores puede dar lugar a coberturas desiguales en un genoma grande como es el del pimiento, con un tamaño de 3,48 GB (Qin et al., 2014; Taitano et al., 2019); (2) pueden generarse errores humanos durante el genotipado causados por la baja automatización del proceso (Taranto et al., 2016) ; o (3) se pueden generar sesgos debido a la frecuente aplicación de *cross-species amplification* que incrementarían artificialmente la frecuencia de homocigotos (Barbará et al., 2007), entre otros. Estas limitaciones hicieron necesaria la búsqueda de otros marcadores, como los polimorfismos de un único nucleótido (*single nucleotide polymorphisms*, SNPs). Los SNPs representan la forma más abundante de variación genética en los genomas (Bush & Moore, 2012) y sus variantes pueden dar lugar a diferencias funcionales al causar cambios aminoacídicos, cambios en la estabilidad de la transcripción del ARNm o cambios en la afinidad del factor de transcripción (Griffith et al., 2008). Por todo ello son muy útiles en estudios de asociación con caracteres, además de los de diversidad genética.

Con la irrupción de las técnicas de ultrasecuenciación o *next-generation sequencing* (NGS) se reducen los problemas asociados a los microsatélites, al poder generarse una gran cantidad de SNPs de manera automatizada y a un coste cada vez más bajo. La gran cantidad de datos generados proporciona una caracterización genómica de los recursos

fitogenéticos más completa, permitiendo una evaluación más exacta y, por lo tanto, un uso adecuado de los mismos (Varshney et al., 2014; Acquadro et al., 2017). Así, las nuevas técnicas de NGS han hecho que los polimorfismos de un solo nucleótido se convirtieran en los marcadores genéticos predilectos para múltiples aplicaciones de la biología vegetal, entre ellas la conservación y la mejora genética (Mammadov et al., 2012). Una de las metodologías de NGS más empleadas para generar un gran número de SNPs es el *genotyping-by-sequencing* (GBS). Esta técnica se basa en reducir la complejidad del genoma mediante el corte enzimas de restricción. A continuación se ligan adaptadores específicos a los sitios de corte, que contienen una secuencia *barcode* de identificación de cada individuo analizado. Posteriormente se realiza una amplificación de las regiones de interés y los fragmentos generados se secuencian en una plataforma de NGS (Elshire et al., 2011). El análisis de las secuencias incluye etapas de control de calidad, recogida de las secuencias en ficheros, mapeo e identificación de SNVs (*single nucleotide variation*, variantes de un solo nucleótido); aunque estas etapas pueden variar según la aplicación del experimento (Figura 2.1) (Elshire et al., 2011; Poland et al., 2012; Lateef, 2015).

El GBS presenta las siguientes ventajas: (1) genera marcadores fiables independientemente de la especie o población de estudio, (2) no se necesita un genoma de referencia previo y (3) se puede acceder a zonas de interés del genoma que no se podrían alcanzar con los experimentos basados en captura de las secuencias (Elshire et al., 2011; Poland & Rife, 2012; He et al., 2014). Además, con la variante de GBS denominada DArTseq™ (*Diversity Arrays Technology*) (Killian et al., 2012) empleada en nuestro estudio, se seleccionan solamente las fracciones del genoma asociadas a genes

activos al emplear en la digestión enzimas de restricción sensibles a metilación y que por lo tanto no digieren zonas repetitivas del genoma.

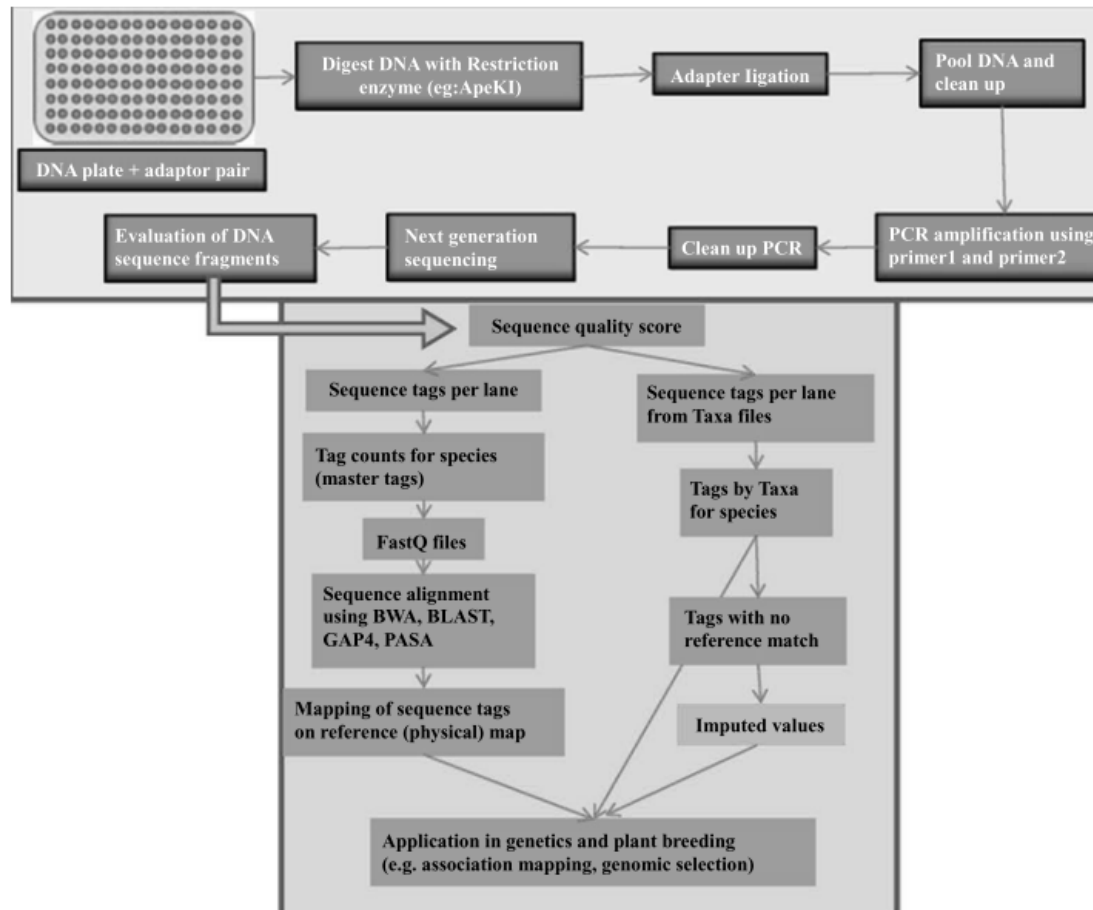


Figura 2.1. Etapas en el desarrollo experimental del GBS. Modificada de Lateef (2015).

En pimiento se han llevado a cabo estudios de GBS para mejorar la comprensión de la evolución del género (Nimmakayala et al., 2016a), para estimar el grado de desequilibrio de ligamiento a lo largo del genoma (Nimmakayala et al., 2016b) y para analizar la diversidad y estructura poblacional (Taranto et al., 2016; Colonna et al., 2019; Pereira-Dias et al., 2019; Taitano et al., 2019) desarrollando en todos los casos miles de SNPs. Empleando SNPs se comprobó que la estructura poblacional viene definida principalmente por el origen y caracteres del fruto (Taranto et al., 2016; Pereira-Dias

et al., 2019) pero también por el tipo de cultivar y condiciones de cultivo (Taitano et al., 2019). La asignación de SNPs que habían sido mapeados inicialmente en el genoma de *C. annuum* en accesiones de *C. baccatum* permitió identificar SNPs homólogos entre especies, determinando además el grado de diferenciación entre genomas (Nimmakayala et al., 2016b). Este hecho confirma la utilidad de los SNPs con diferentes especies, compensando una importante limitación que tenían los marcadores microsatélite. Por lo tanto, se hace necesario identificar nuevos sets de SNPs útiles en distintas especies del género *Capsicum* que permitan establecer comparaciones entre ellas; especialmente si esos SNPs están asociados a regiones codificantes. En este capítulo, se evaluaron las colecciones de pimiento anteriormente descritas mediante la tecnología *genotyping-by-sequencing* (GBS) con el objetivo de determinar la variabilidad existente, la estructura genética y las relaciones entre accesiones.

Material y métodos

Material vegetal

El análisis *genotyping-by-sequencing* (GBS) se realizó mediante la tecnología DArTseq™ (Killian et al., 2012) en 234 líneas de pimiento de las cinco especies cultivadas (Indicadas con una "X" en la columna "GBS" de la Tabla 1.1). Los datos de GBS de 142 líneas de pimiento han sido cedidos por la Dra. Cristina Silvar para la elaboración de la presente Tesis y se incluyeron junto con los generados *de novo*. De las 234 accesiones, 223 también fueron analizadas con marcadores microsatélites en el Capítulo 1.1; no se genotiparon con GBS pero sí con microsatélites las accesiones PI 257144 (Perú); PI 260436 (Bolivia); PI

593925, PI 595908, PI 585265, PI 585267 y PI 585264 (Ecuador), al no disponer del suficiente material vegetal de las mismas. Se incorporaron al ensayo con GBS las accesiones CGN16807, CGN17227, CGN22127, CGN17147 y CGN16817 (Hungría); CGN16808, CGN17214, CAP491 y CAP1254 (Bulgaria); CAP607 (Libia) y CGN17190 (Turquía) no genotipadas con marcadores microsatélite.

De las 234 líneas, 160 son *Capsicum annuum* (de las cuales 4 son *C. annuum* var. *glabriusculum*), 26 *C. chinense*, 11 *C. frutescens*, 24 *C. baccatum* y 13 *C. pubescens*. Las plantas de cada accesión se cultivaron en bandejas de alveolos con una mezcla de sustrato orgánico y perlita en relación 2:1. Las bandejas y posteriormente las plantas se mantuvieron en cámara de cultivo con fotoperiodo de 16 horas luz/ 8 de oscuridad, temperaturas de 25 °C /18 °C (día/noche) y riego periódico.

Extracción de ADN, preparación de las librerías y secuenciación

La extracción de ADN se realizó a partir de aproximadamente 1 g de hojas jóvenes usando el método del bromuro de cetiltrimetilamonio (CTAB)/cloroformo/isoamil alcohol (Doyle & Doyle, 1987). La calidad y concentración del ADN extraído se determinó con un espectrofotómetro NANODROP ND-1000 (NanoDrop Technologies, Inc.). Para algunas muestras no se consiguió obtener suficiente concentración de ADN ni una calidad aceptable, en estos casos se repitió la extracción con el *kit* NZY Plant/Fungi gDNA Isolation kit (NZYTech) a partir de nuevo material vegetal.

Se prepararon alícuotas de las muestras de ADN con calidad adecuada y ratios de absorbancia 230/260 entre 1,8-2,0 y 260/280 entre 1,8-2,2 en 1 X TE (Tris-HCl 10mM, EDTA 1mM) y se enviaron a la empresa *Diversity Arrays Technology Pty Ltd* (Canberra, Australia), donde se generaron las librerías según lo descrito por Killian et al. (2012) y se realizó la secuenciación según el método DArTseq™ empleando la tecnología *Illumina* (Illumina, Inc., San Diego, California, Estados Unidos). Las secuencias generadas, de tipo *single end*, se recogieron en ficheros FASTQ, individuales para cada muestra.

Mapeo y SNV calling

El tratamiento de las secuencias de GBS y los análisis posteriores se realizaron durante una estancia de investigación en el Grupo de Bioinformática y Genómica del Instituto de Conservación y Mejora de la Agrodiversidad Valenciana (COMAV) de la *Universitat Politècnica de València* (Valencia, España). Todos los *scripts* empleados para los análisis están disponibles previa solicitud al Dr. José Blanca (jblanca@upv.es).

La calidad de las secuencias crudas se evaluó con FASTQC 0.11.8 (Andrews et al., 2010). Para eliminar los *primers* y adaptadores de Illumina, así como las secuencias con baja calidad o longitud inesperada se utilizó la herramienta Trimmomatic 0.38 (Bolger et al., 2014) empleando los parámetros ILLUMINACLIP:TruSeq2-SE.fa:2:30:10 (que corta los *primers*, adaptadores y secuencias específicas de Illumina), LEADING:3 y TRAILING:3 (eliminan bases con calidad menor de 3 al principio y al final de la secuencia, respectivamente); SLIDINGWINDOW:4:15 (elimina el fragmento en el que la calidad

media por base es menor de 15) y MINLEN:40 (elimina las secuencias con longitud menor a 40 pb). Las secuencias resultantes se mapearon contra el genoma de referencia de *Capsicum annuum* UCD-10X-F₁ (Hulse-Kemp et al., 2018) usando el mapeador Burrows-Wheeler Aligner (BWA) (Li & Durbin, 2009) con los parámetros por defecto, implementado en un *script* de Python desarrollado en el Grupo de Bioinformática y Genómica del COMAV. Con un nuevo *script* se obtuvieron las estadísticas de mapeo (número de secuencias mapeadas y calidad de mapeo de las secuencias) y se representaron gráficamente los datos por muestra.

La identificación de SNVs o SNV *calling* se realizó con FreeBayes 1.2 (Garrison & Marth, 2012) empleando los parámetros por defecto. Posteriormente se llevó a cabo el filtrado de los SNPs con un nuevo *script* propio, eliminando aquellos con calidad mínima de SNP ≤ 20 ; con profundidad de genotipo (*depth genotype*, DP-GT) < 2 ; recuento mínimo de alelos (*minor allele count*, MAC) ≥ 3 ; heterocigosidad observada $\leq 0,1$ y proporción de datos perdidos $\leq 0,7$.

Estructura poblacional

Para comprender la estructura genética y explorar los patrones de variación genética en el conjunto de accesiones, se realizaron análisis de coordenadas principales (*principal coordinates analysis*, PCoA) mediante un *script* de Python desarrollado en el Grupo de Bioinformática y Genómica del COMAV y utilizando las distancias genéticas de Kosman & Leonard (2005). Previamente a construir los PCoAs, se realizó un filtrado por accesión aplicando criterios de heterocigosidad observada $\leq 0,1$ y proporción de datos

perdidos $\geq 0,97$. Así, se utilizó un esquema de clasificación jerárquico en tres niveles para definir los grupos y subgrupos genéticos dentro de las especies.

Para inferir la estructura poblacional se realizó un análisis clúster implementado en fastSTRUCTURE 1.0 (Raj et al., 2014) con número de clústeres (K) entre 1 y 8, usando como *prior* el método “simple” y *cross-validation*=0. Se escogió aquel valor de K más probable donde la probabilidad marginal de los datos (*marginal likelihood*) es mínima.

Diversidad genética

A partir de los grupos revelados por los PCoAs, se empleó un nuevo *script* de Python para calcular los siguientes descriptores de diversidad genética: heterocigosidad observada (H_o), heterocigosidad esperada (H_e), número medio de alelos por *locus* (N_a), número medio de alelos privados por *locus* (N_{ap}) y porcentaje de marcadores en los que el alelo mayoritario presentó una frecuencia menor de 0,95 (P_{95}), los cuales se considerarán polimórficos. Para calcular cada estimador se requieren al menos cinco muestras por grupo.

Resultados

Secuenciación, pre-procesado de las secuencias y mapeo

En el presente trabajo, se genotiparon con éxito 234 accesiones de *Capsicum* spp. con la tecnología DArTseq™, obteniendo un total de 583.159.174 secuencias *single end* que

fueron almacenadas en ficheros FASTQ. El número de lecturas por muestra varió entre 307.024 (CGN21475) y 5.814.379 (PI 241670), siendo el número medio 2.492.133 (Tabla A.4). Tras la eliminación de los *primers* y adaptadores de secuenciación con Trimmomatic, se mantuvieron un 89,3% de las secuencias iniciales, de las cuales un 96,4% mapearon correctamente contra el genoma de referencia. El mapeador BWA proporciona un estadístico de calidad de mapeo propio (Mapq) para cada secuencia. La mayoría de las secuencias mapearon con calidad alta (>57), pero también un cierto número mapeó con calidad baja (Figura 2.2).

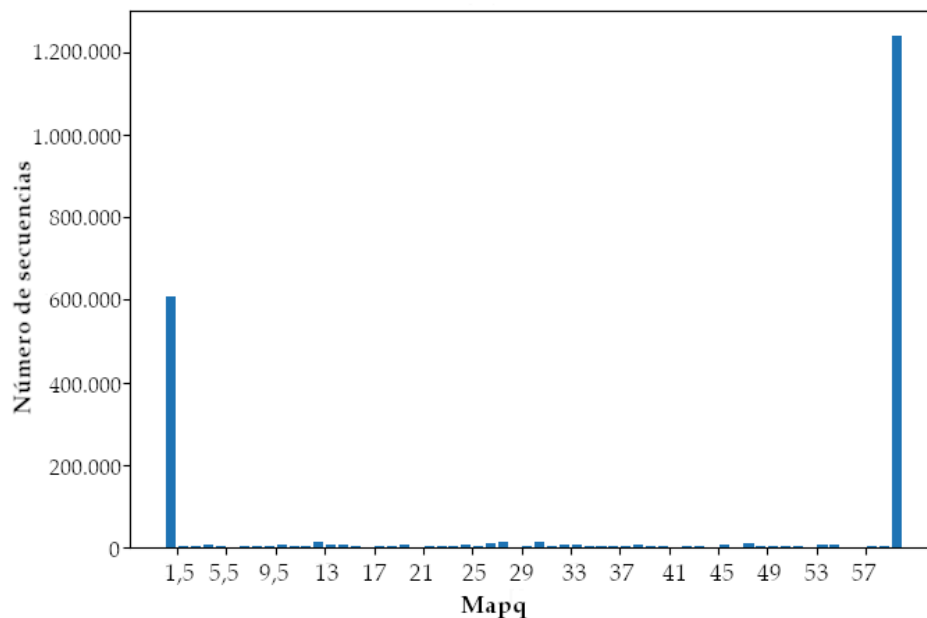


Figura 2.2. Número de secuencias mapeadas para cada valor de Mapq en la muestra BGHZ0099.

Seleccionando solo las secuencias con Mapq mayor de 57 se mantienen 321.847.311 secuencias, un 64,2% del total de lecturas que mapearon (Figura 2.3, Tabla A.4). Mapq es un estimador de que las secuencias estén incorrectamente mapeadas (Li et al., 2009). Valores bajos de Mapq indican que la secuencia podría mapear en otra zona del genoma y valores altos indican que solo mapearía en esa zona, por lo tanto, aunque con este

filtrado se elimine un alto porcentaje de secuencias se asegura que todas las secuencias con las que se trabaje sean secuencias de copia única. Así, se retienen en torno a un 45% de las secuencias iniciales.

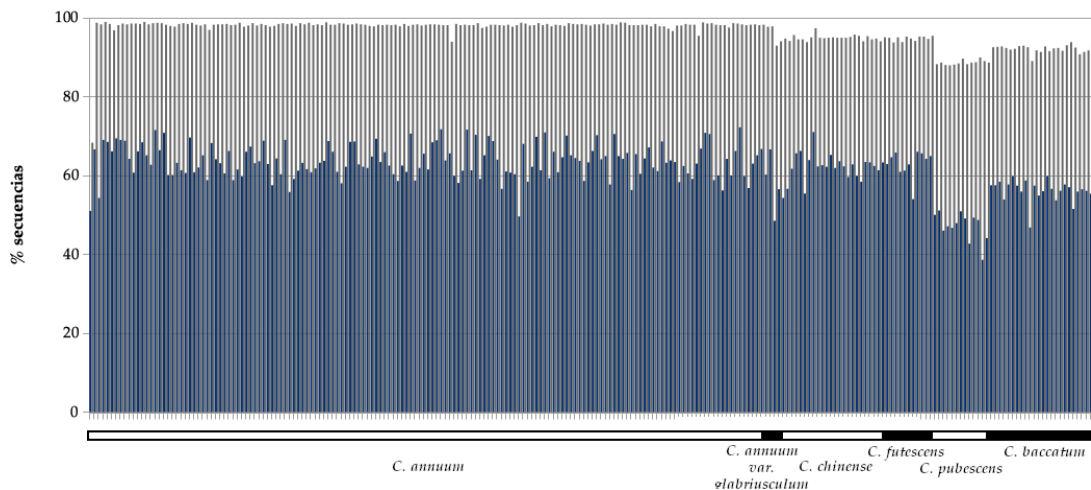


Figura 2.3. Porcentaje de secuencias que mapean (gris) y porcentaje de secuencias que mapean con Mapq mayor de 57 (azul) para cada accesión agrupadas por la especie a la que pertenecen.

Identificación de SNVs

Tras el SNV *calling* se identificaron 245.307 SNVs. Con el filtrado se eliminaron 135.233 SNVs con el criterio de $MAC \geq 3$, 54.712 con el criterio de datos perdidos $\leq 0,7$ y 7.253 con el criterio de heterocigosidad observada $\leq 0,1$; retenándose así 48.109 variantes de alta calidad para los estudios posteriores. El número de SNVs filtrados por cromosoma, varió entre 2.800 en el cromosoma 5 y 6.673 en el cromosoma 3 (Figura 2.4).

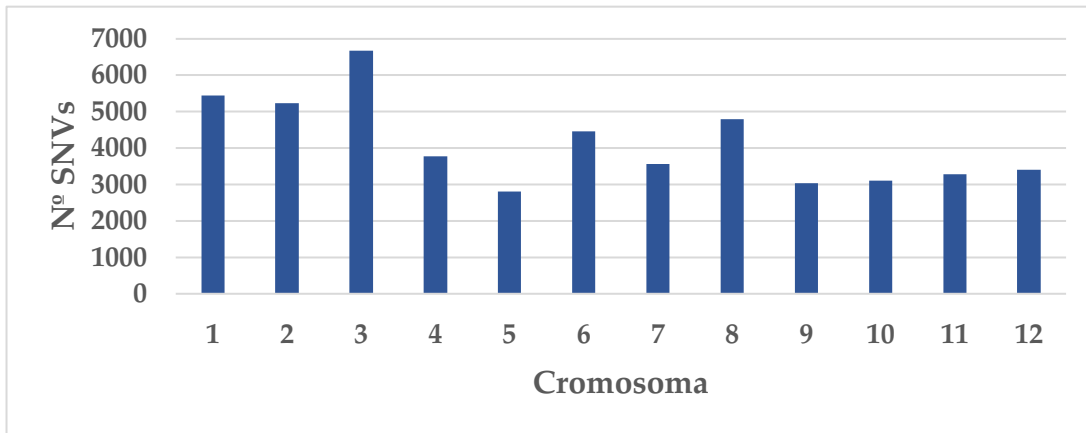


Figura 2.4. Recuento de SNVs tras el filtrado para cada uno de los doce cromosomas de *C. annuum*.

Del total de SNVs, el 88,6% se corresponde a polimorfismos de un solo nucleótido (SNPs), el 10,8% a inserciones/delecciones (*indels*) y un 0,6% son variaciones estructurales. La proporción de transiciones (C↔T o A↔G) encontrada fue mayor (55%) que la de transversiones (45%) (A↔C, C↔G, A↔T o G↔T) (Figura 2.5).

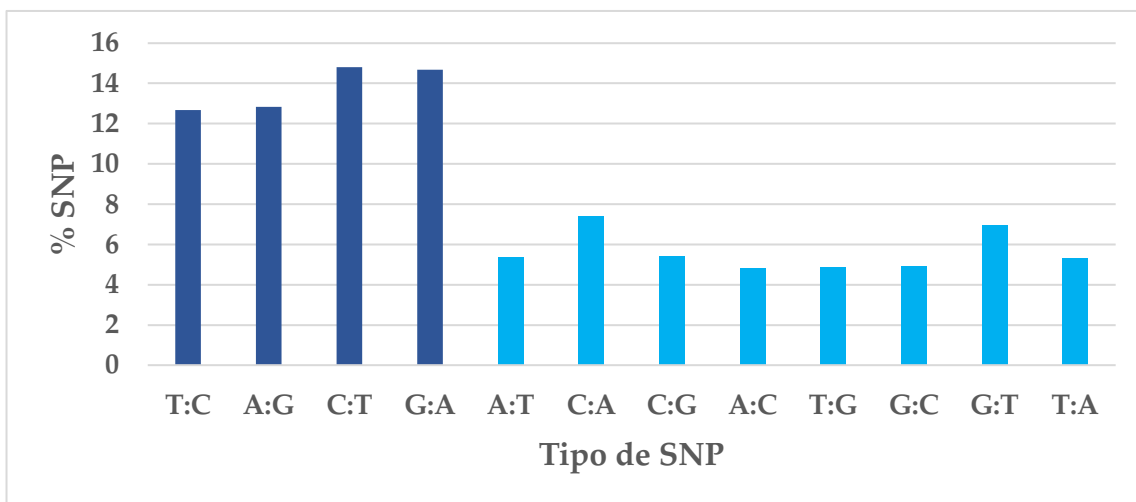


Figura 2.5. Porcentaje de cada tipo de SNP representado según la sustitución nucleotídica en transiciones (azul oscuro) y transversiones (azul claro).

Tras un segundo filtrado por accesión, se eliminaron las muestras CGN23768 y CGN21475 de los análisis por presentar una elevada proporción de datos perdidos.

Estructura genética

Para identificar la estructura genética de la colección completa se realizó un primer PCoA con todas las accesiones y usando los 48.109 SNVs obtenidos después del filtrado (Figura 2.6). Todos los PCoAs se representaron en base a los datos de origen y fenotipo disponibles, pero solo se muestran aquellos que revelaron estructura en base a alguno de esos caracteres.

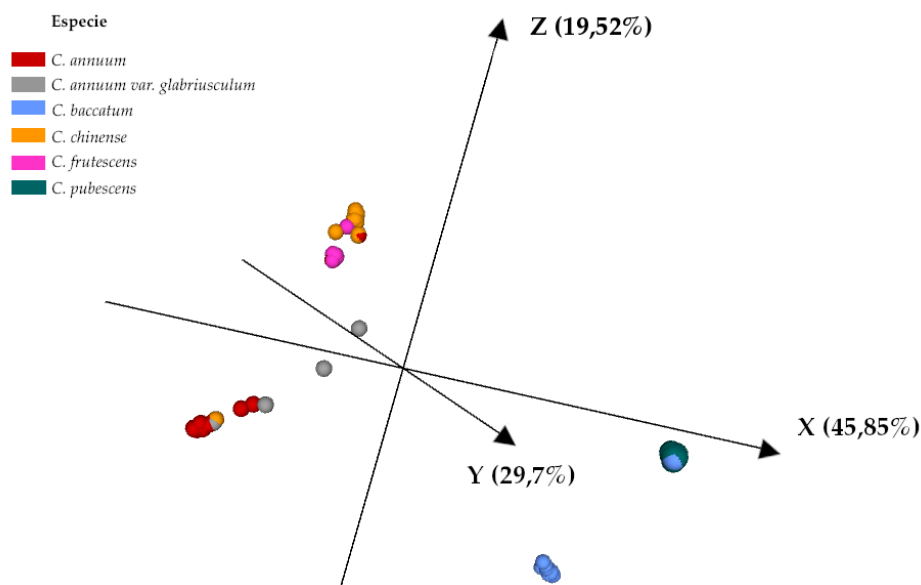


Figura 2.6. Análisis de coordenadas principales global, incluyendo las 232 accesiones y empleando los 48.109 marcadores.

El primer PCoA incluyó las 232 accesiones y mostró un patrón de estructura genética claramente diferenciado por especies, donde los dos primeros componentes explican la

mayor parte de la varianza (45,85% y 29,7%, respectivamente). Alejadas del resto de las accesiones y formando un grupo muy homogéneo se encuentran las 12 accesiones de *C. pubescens*. Aparece también en este grupo la accesión PI 585244 catalogada según el banco de germoplasma como *C. baccatum*. Formando otro grupo claramente diferenciado se encuentran las 23 accesiones de *C. baccatum*, excepto PI 585244. Otra agrupación está formada por las accesiones de las especies *C. chinense* y *C. frutescens*, que aparecen todas ellas próximas entre sí. En este grupo aparecen dos líneas (BGV006008 y PI 573337) catalogadas como *C. annuum*. El cuarto y último grupo está constituido por las restantes accesiones de *C. annuum*, dos accesiones de *C. annuum* var. *glabriusculum* (PM663 y PM647). Finalmente, las líneas PM670 y PM669 (*C. annuum* var. *glabriusculum*) se sitúan en una posición intermedia en el PCoA.

Para realizar los sucesivos PCoAs jerárquicos se seleccionaron los cuatro grupos descritos, es decir, *C. pubescens*, *C. baccatum*, *C. chinense* + *C. frutescens* y *C. annuum*. El segundo PCoA se realizó para las 13 muestras que se agrupaban como *C. pubescens* y que aparecían como el grupo más diferenciado en el PCoA global (Figura 2.6).

El PCoA construido solo con estas 13 accesiones de *C. pubescens* reveló subestructura dentro de este grupo, separándose por un lado las muestras de Ecuador y por otro las de Perú (Figura 2.7A). Esta separación según el país de origen también se corresponde con el nivel de pubescencia de tallo (Figura 2.7B). Las muestras con pubescencia 3 (escasa) y 5 (intermedia) se corresponden con las de Ecuador y aparecen separadas de las de Perú, que presentan pubescencia igual a 7 (densa), según las categorías descritas en IPGRI (1995), lo que parece indicar una separación adaptativa en base a dicho

carácter. Separadas del resto y muy próximas entre sí aparecen las líneas PI 585244 y BGV013702 (de Ecuador y Perú, respectivamente, señaladas en la Figura 2.7A con una flecha).

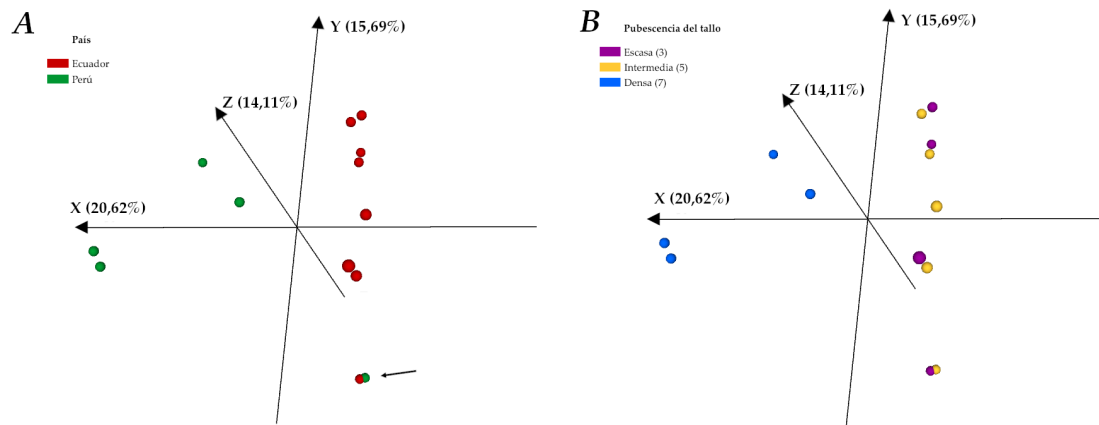


Figura 2.7. Análisis de coordenadas principales incluyendo solo las 13 muestras agrupadas como *C. pubescens* en el PCoA global (Figura 2.6). A. Visualizadas según el país de origen. B. Visualizadas según el grado de pubescencia del tallo. La flecha señala las muestras PI 585244 y BGV013702.

Se representó el subconjunto de las 23 muestras agrupadas como *C. baccatum* (Figura 2.8). En este caso, se observa una agrupación clara por país de origen. Por una parte aparecen las accesiones de Bolivia, por otra las de Perú y por otra las de Ecuador, que son las más próximas entre sí. Dentro de las líneas de Ecuador, destaca la PI 257133 (señalada con una flecha en la Figura 2.8) que aparece bastante alejada del resto, siendo fenotípicamente la única con forma del fruto=5 (acampanulado y en bloque) y forma del fruto en la unión con el pedicelo=3.

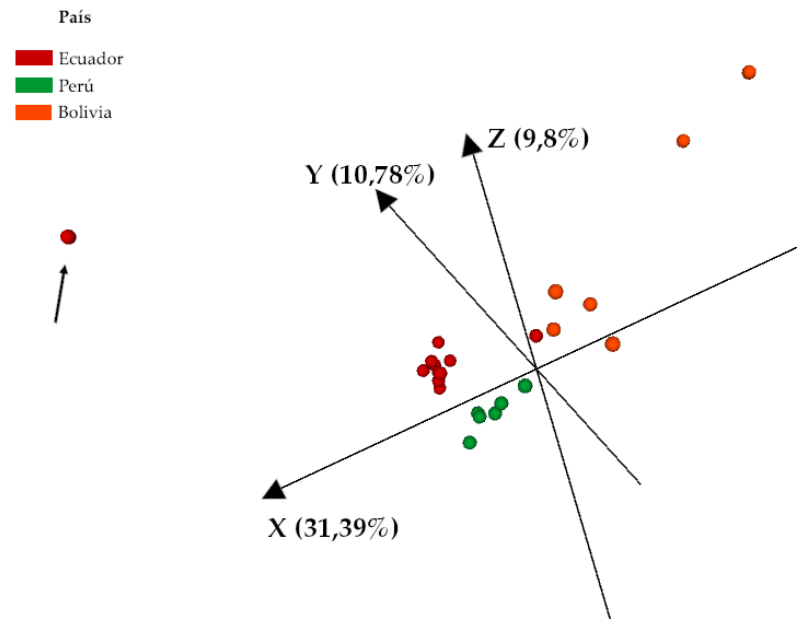


Figura 2.8. Análisis de coordenadas principales de las 23 muestras clasificadas como *C. baccatum* a partir de su agrupación en el PCoA global (Figura 2.6). La flecha señala la accesión PI 257133.

Se representó el conjunto formado por las accesiones de *C. chinense* y *C. frutescens* (Figura 2.9). Se diferenciaron claramente estas dos especies que aparecían próximas en el PCoA global (Figura 2.6). Por un lado aparecen agrupadas las 10 accesiones de *C. frutescens* y la BGV006008 catalogada como *C. annuum*; y por otra parte aparecen las 24 accesiones de *C. chinense* junto con la PI 573337, también clasificada como *C. annuum* en el banco de germoplasma. Además, se consiguió reclasificar las accesiones BGV006008 y PI 573337 que se presentaban como *C. annuum* en los datos del banco de germoplasma, como *C. frutescens* y *C. chinense*, respectivamente.

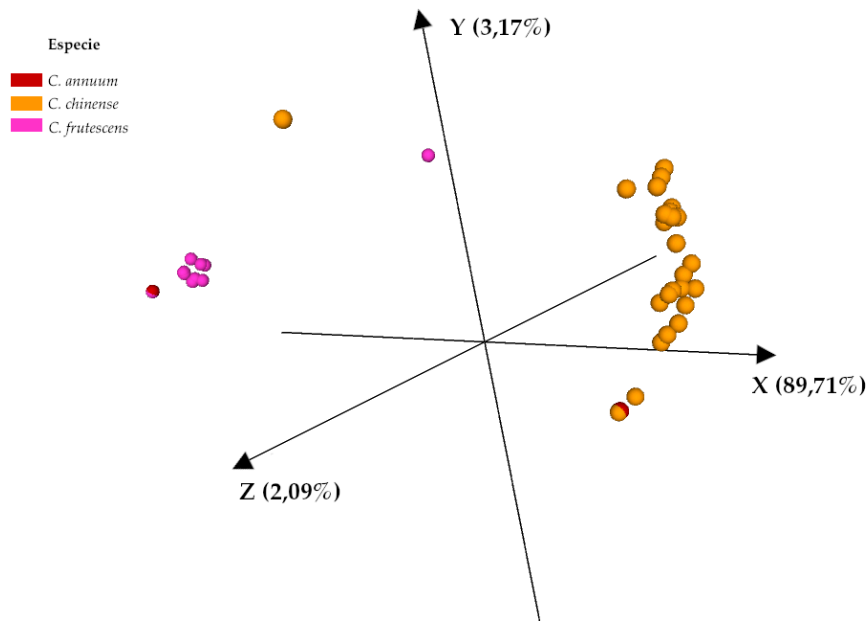


Figura 2.9. Análisis de coordenadas principales de las 38 muestras que conforman el conjunto *C. chinense* + *C. frutescens* en el PCoA global (Figura 2.6).

Se realizaron sucesivos PCoAs jerárquicos, uno para *C. chinense* (Figura 2.10A) y otro para *C. frutescens* (Figura 2.10B), con el fin de comprobar si hay estructura intraespecífica. En el grupo *C. chinense* (Figura 2.10A), se aprecia cierta subestructura en base al país de origen, pero sin estar claramente definida. Destaca que algunas accesiones aparecen muy alejadas de las demás, PI 573337, PI 543203 y CAP472 (señaladas con una flecha en la Figura 2.10A). A su vez, para el grupo de *C. frutescens*, (Figura 2.10B) no se encontró asociación en base al país de origen ni a ningún carácter fenotípico, probablemente debido a que solo se estudiaron siete muestras de Ecuador y cuatro de Bolivia y Perú.

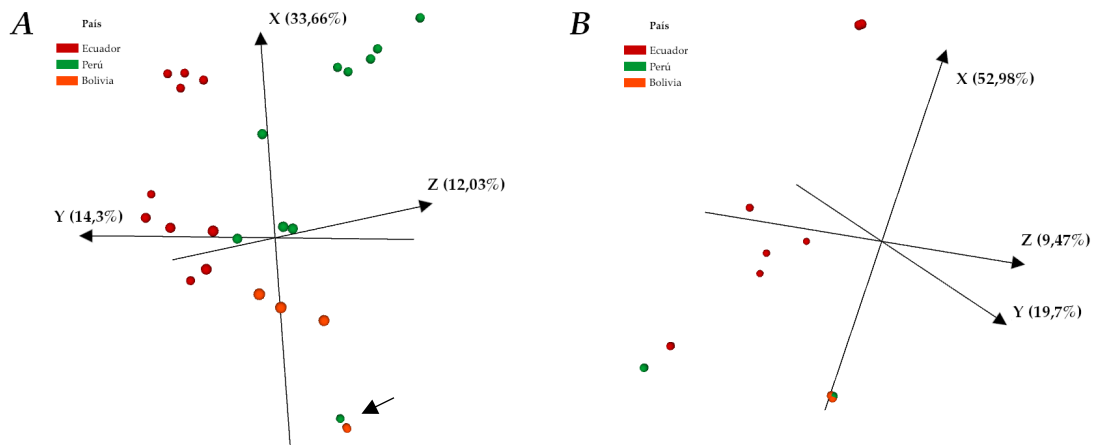


Figura 2.10. Análisis de coordenadas principales para el segundo nivel jerárquico a partir de las agrupaciones observadas en el PCoA del conjunto *C. chinense* + *C. frutescens* (Figura 2.9). A. PCoA para las 25 líneas del grupo *C. chinense*. La flecha negra señala las accesiones PI 573337, PI 543203 y CAP472. B. PCoA para las 11 líneas del grupo *C. frutescens*.

El siguiente PCoA se realizó para las muestras clasificadas como *C. annuum*, 156 en total (Figura 2.11). En este PCoA se observó que aparecen algunas muestras muy alejadas de las demás, correspondiéndose con seis líneas de México, una de Perú y una de Bolivia. Pese a que el resto de las accesiones aparecen más próximas entre sí, es posible diferenciar dos grupos en base a la proximidad de las accesiones, señalados en la imagen como Grupo 1 y Grupo 2. La cercanía entre las líneas del Grupo 2 es mucho más marcada que en el Grupo 1, evidenciando una mayor proximidad genética entre ellas. El Grupo 1 está conformado por cuatro accesiones de Bolivia, tres de Ecuador, seis de México, cuatro de España, dos de Portugal, una de Túnez, una de Israel, una de Bulgaria y las cuatro de Libia. Por su parte, el Grupo 2 incluiría todas las demás accesiones de las colecciones ibérica y mediterránea, así como tres de Ecuador, dos de Perú y dos de México. Cabe destacar que aparecen alejadas y no formando parte de ninguno de los dos grupos una accesión de Bolivia, una de Perú y seis de México.

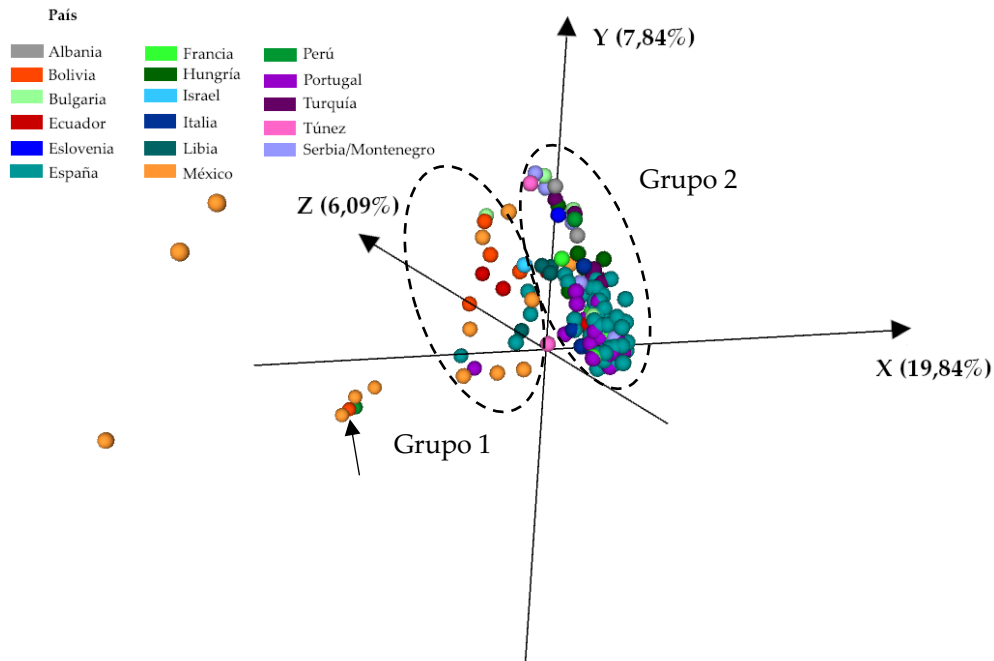


Figura 2.11. Análisis de coordenadas principales para las líneas que aparecen formando el grupo C. *annuum* en el PCoA global (Figura 2.6). La flecha negra señala la accesión CGN20808.

Se realizaron dos PCoAs jerárquicos para cada uno de estos dos grupos. Mientras que en el Grupo 1 no fue posible identificar ninguna relación entre las muestras, en el Grupo 2 sí aparecen agrupaciones que se corresponden en mayor medida con el origen geográfico de las accesiones, observándose mayoritariamente una distinción Oeste-Este de Europa (Figura 2.12). En el subgrupo Oeste Europa se sitúan 58 líneas de España, 19 de Portugal, seis de Italia, cinco de Francia, tres de Israel, una de Bulgaria, una de Túnez y una de Serbia/Montenegro. En este subgrupo también se engloban tres accesiones de Ecuador y una de Perú. En el subgrupo Este Europa se incluyen seis accesiones de Turquía, cinco de Hungría, cuatro de Serbia/Montenegro, dos de Bulgaria, dos de Albania, una de Eslovenia y una de Túnez. Al igual que el subgrupo anterior, también aparece en esta agrupación dos accesiones de México y una de Perú.

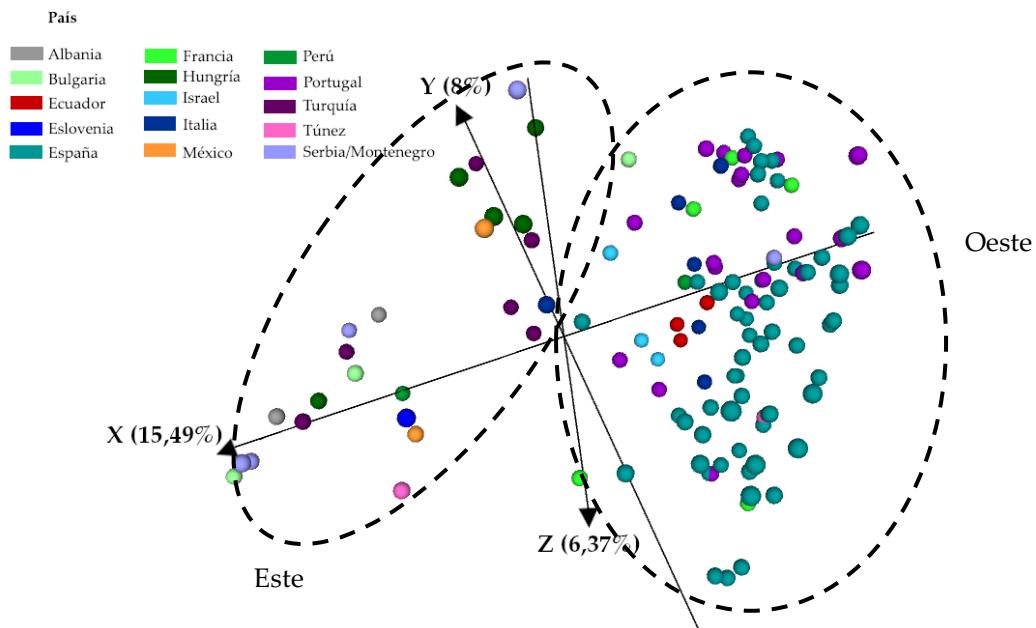


Figura 2.12. Análisis de coordenadas principales para las 122 líneas encuadradas en el “Grupo 2” (segundo nivel jerárquico) a partir de lo observado en el PCoA de *C. annuum* (Figura 2.11).

Para comprobar si existían relaciones dentro de cada uno de estos grupos se realizó una última subdivisión con PCoAs para los subgrupos Oeste Europa y Este Europa. A este nivel no fue posible identificar relaciones claras en ninguno de los dos PCoAs realizados.

Con el mismo objetivo de identificar la estructura genética dentro del conjunto de accesiones, se utilizó también el programa fastSTRUCTURE empleando el conjunto de SNPs filtrados y probando posibles valores de K entre 1 y 8. El valor de K que minimiza la probabilidad marginal fue $K=4$ (Figura 2.13).

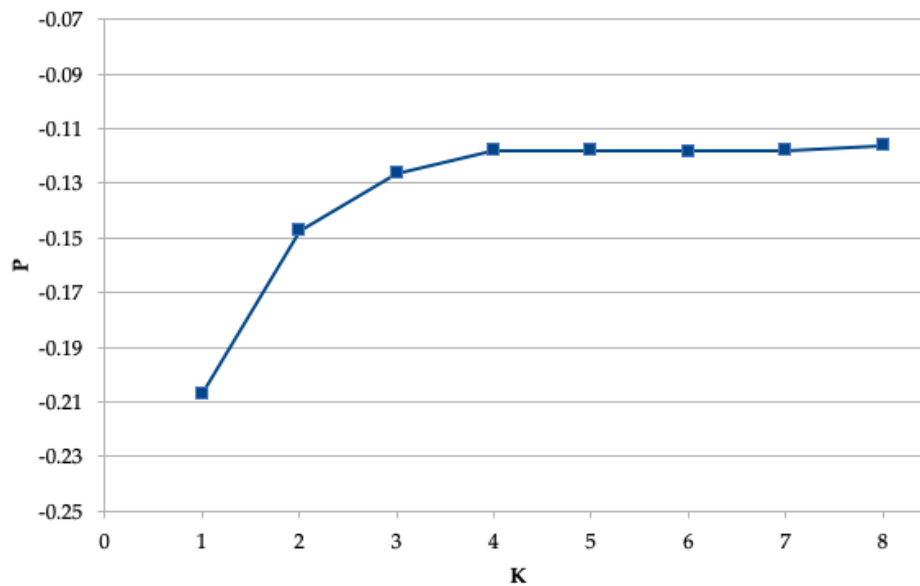


Figura 2.13. Probabilidad marginal (P) para cada clúster (K) testado con el programa fastSTRUCTURE.

Se representaron los coeficientes de pertenencia a cada uno de los cuatro clústeres (Figura 2.14A). Los cuatro grupos se corresponden con los encontrados en el PCoA global (Figura 2.6) corroborando los resultados de los Análisis de Coordenadas Principales. El primer clúster lo componen las 12 líneas de *C. pubescens* de la colección, así como la PI 585244 (*C. baccatum* según los datos iniciales de pasaporte), que al igual que en el PCoA (Figura 2.6) se agrupaba con las demás líneas de *C. pubescens*. Las 23 accesiones de *C. baccatum* también aparecen diferenciadas de las otras especies y agrupadas en un clúster independiente, con coeficientes de pertenencia cercanos al 100% para este grupo. El tercer clúster está compuesto por las accesiones de *C. chinense* y *C. frutescens*, al igual que en el PCoA global (Figura 2.6) en el que también aparecían juntas. En este clúster se incluyen también dos líneas categorizadas en los datos de pasaporte como *C. annuum* (BGV006008 y PI 573337) y una línea de *C. annuum* var. *glabriusculum* (PM670). El último clúster está compuesto por todas las líneas de *C. annuum*,

incluyéndose también aquí una línea *C. annuum* var. *glabriusculum* (PM663). En una posición intermedia entre el clúster de *C. annuum* y el de *C. chinense* + *C. frutescens* se encuentra la accesión de *C. annuum* var. *glabriusculum* PM669, con aproximadamente una pertenencia del 50% a cada uno de los dos clústeres.

Con el objetivo de identificar subestructura dentro de los grupos analizados, se continuó la representación de fastSTRUCTURE para valores superiores de K que pudieran tener sentido biológico. De este modo, para $K=8$ (Figura 2.14B) se consiguió observar subestructura dentro de alguno de los clústeres anteriormente descritos. Para este valor de K , los clústeres de *C. baccatum*, *C. pubescens* y *C. annuum* se mantienen invariables. El clúster antes formado por *C. chinense* y *C. frutescens* ahora se divide en dos, formado cada uno por las líneas respectivas de cada especie. Con esta nueva representación la accesión *C. annuum* var. *glabriusculum* (PM670) aparece con un 50% de pertenencia a cada uno de los dos nuevos clústeres. Por su parte, la accesión PM669 (*C. annuum* var. *glabriusculum*) que para $K=4$ tenía pertenencias de aproximadamente el 50% al clúster de *C. chinense* + *C. frutescens* y 50% al de *C. annuum*, muestra ahora aproximadamente un 50% de pertenencia al clúster de *C. chinense* y otro 50% al de *C. annuum*. Para los restantes tres grupos de $K=8$, que no se corresponden con especies, ninguna línea muestra una pertenencia mayor al 4%. Los resultados de fastSTRUCTURE también corroboran la posición intermedia de *C. annuum* var. *glabriusculum* entre *C. chinense*, *C. frutescens* y *C. annuum*, un hecho congruente con que esta especie sería el pariente silvestre de las otras tres.

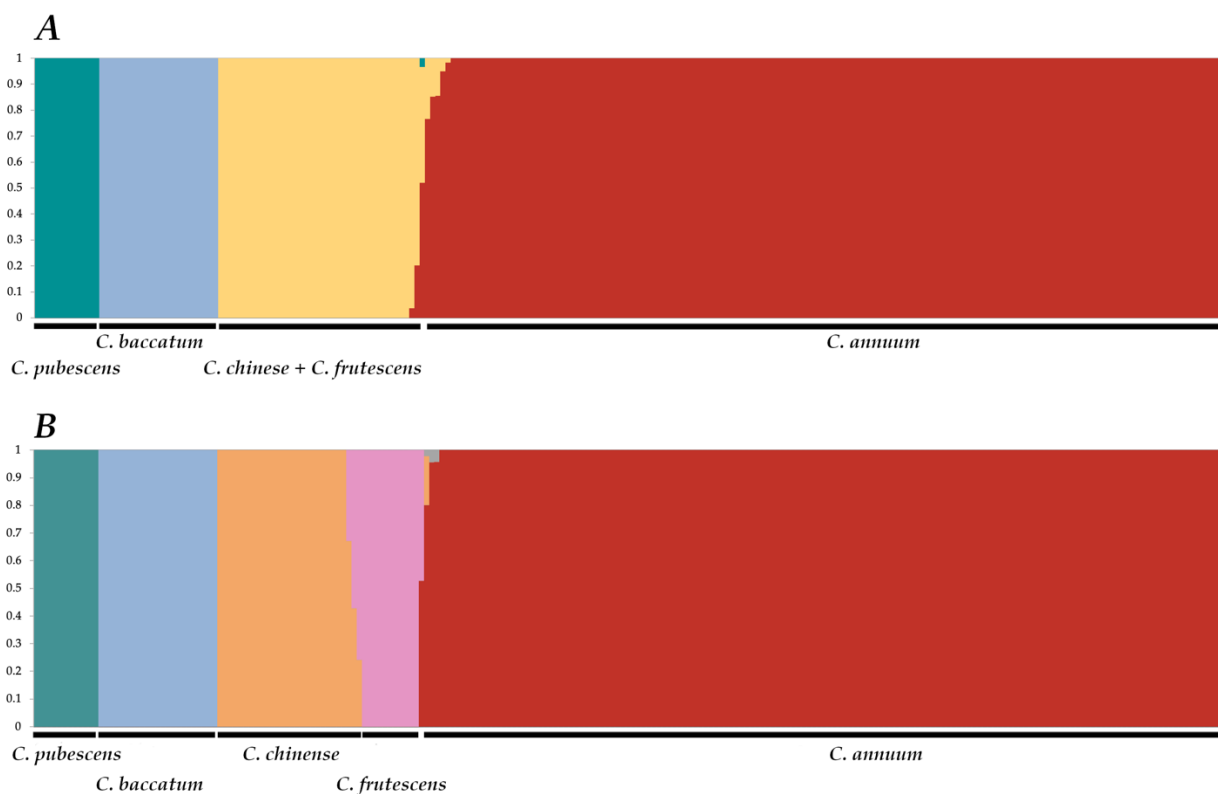


Figura 2.14. Estructura poblacional para las 232 accesiones inferida por el programa fastSTRUCTURE. A. Representación para una estructura en $K=4$ clústeres B. Representación para $K=8$. Cada columna se corresponde con una accesión, cada color indica los distintos clústeres asignados y en el eje Y se representa el coeficiente de pertenencia a cada grupo.

Diversidad genética

Los estadísticos de diversidad genética (Tabla 2.1) se calcularon para los grupos encontrados en los PCoAs, clasificando las líneas según se agrupaban en los gráficos y no según los datos iniciales de pasaporte.

Tabla 2.1. Estadísticas de diversidad para cada grupo de la clasificación jerárquica establecido a partir de los PCoAs. *N*, número de individuos; *Na*, número medio de alelos por *locus*; *Nap*, número medio de alelos privados por *locus*; *Ho*, heterocigosidad observada; *He*, heterocigosidad esperada; *P95*, porcentaje de marcadores en los que la frecuencia del alelo mayoritario es menor de 0,95. No se calcularon los parámetros para la especie *C. annuum* var. *glabriusculum* al ser un grupo compuesto por menos de cinco muestras.

Especies	Grupo	Subgrupo	<i>N</i>	<i>Na</i>	<i>Nap</i>	<i>Ho</i>	<i>He</i>	<i>P95</i>
<i>C. pubescens</i>			13	0,894	0,239	0,012	0,048	0,965
<i>C. baccatum</i>			23	1,021	0,160	0,008	0,032	0,978
<i>C. chinense</i> + <i>C. frutescens</i>			38	1,370	0,203	0,007	0,114	0,919
		<i>C. chinense</i>	25	1,165	0,052	0,005	0,054	0,963
		<i>C. frutescens</i>	11	1,044	0,028	0,005	0,043	0,970
<i>C. annuum</i>			156	1,349	0,152	0,004	0,033	0,978
	Grupo 1		26	1,175	0,005	0,01	0,041	0,385
	Grupo 2		122	1,241	0,008	0,003	0,026	0,5
		Oeste Europa	98	1,200	0,004	0,003	0,024	0,984
		Este Europa	24	1,132	0,001	0,003	0,027	0,981
<i>C. annuum</i> var. <i>glabriusculum</i>			2	-	-	-	-	-

La especie que mostró los valores más altos de heterocigosidad observada fue *C. pubescens* (0,012) en contraposición a *C. annuum* (0,04). La agrupación de las especies *C. chinense* y *C. frutescens* fue la que presentó mayor índice de heterocigosidad esperada, sin embargo, al separar las especies no se observan valores más altos que en las demás especies, por lo que este resultado podría deberse a la unión de dos especies distintas. El número de alelos medio resultó ser mayor en *C. annuum* (1,349) y más bajo en *C. pubescens* (0,849). Pese a que fue la especie que presentó menor número de alelos medio y con un bajo número de muestras analizadas (13), en *C. pubescens* el valor de alelos privados por *locus* fue el máximo. El nivel de polimorfismo (expresado como P95) fue siempre alto y el Grupo 1 resultó ser el caso en el que más bajo se presentó este índice (0,385); que la mayoría de SNPs sean monomórficos puede ser la causa de que no se

encuentren asociaciones ni relaciones entre muestras dentro de este grupo porque todas las líneas serían muy similares genéticamente.

Discusión

El desarrollo de las plataformas de NGS y el progresivo abaratamiento de los costes permitieron en los últimos años su uso en estudios de diversidad y estructura genética como el aquí presentado. Sin embargo, dada la gran cantidad de datos generados por las herramientas de ultrasecuenciación, se debe seguir una metodología que asegure la calidad y rigurosidad de los datos finalmente empleados. La aparición de secuencias con valores de Mapq bajos podría ser debida a (1) la ausencia de posiciones en el genoma de referencia contra las que mapeen las secuencias, por ejemplo por grandes divergencias entre *Capsicum annuum* (especie a la que pertenece el genoma de referencia empleado) y las demás especies estudiadas; (2) secuencias de ADN procedentes del genoma mitocondrial o cloroplástico o (3) repetitividad en el genoma que hace que una lectura pueda mapear en más de una posición (Pereira-Dias et al., 2019). Sin embargo, nuestro estudio retiene un mayor número de secuencias iniciales respecto a otros trabajos con colecciones y número de accesiones similares. Taranto et al., (2016) mantuvieron un 43,4% de las secuencias iniciales y Pereira-Dias et al., (2019) un 40,8% para una colección de 190 muestras de las cinco especies cultivadas del género *Capsicum*. El mayor número de secuencias que se mantiene en este caso puede deberse a que este es el primer estudio que emplea el genoma *UCD-10X-F1* como genoma de referencia. En este caso, al emplear un ensamblaje de secuencias por fases se generó un genoma más ordenado y el más continuo de todos los publicados (Hulse-Kemp et al., 2018).

El número final de SNVs identificadas fue mayor que el generado en otros estudios del género *Capsicum*, en los que no se sobrepasaban los 37.000 (Nimmakayala et al., 2016b; Taranto et al., 2016; Taitano et al., 2019). Esta diferencia en el número de SNVs que se conservan puede ser debida a los distintos criterios de filtrado empleados o al genoma de referencia contra el que se mapea. El genoma *UCD-10X-F₁* procede de la secuenciación de una línea de pimiento híbrida y por lo tanto heterocigota en muchas posiciones (Hulse-Kemp et al., 2018), al contrario de lo empleado en otros genomas de referencia basados en líneas puras y variedades silvestres autopolinizadas (Kim et al., 2014; Qin et al., 2014).

Al igual que en otros estudios, el cromosoma 3 es el que presenta más polimorfismos (Nimmakayala et al., 2016b; Taranto et al., 2016; Pereira-Dias et al., 2019); pero en dichos trabajos aparecían el 8 o el 9 como los que tenían contenían menos SNVs y nunca el 5 como en nuestro caso. La mayor proporción de transiciones respecto a transversiones en el total de SNVs aisladas, coincide con lo encontrado en otros estudios (Taranto et al., 2016; Pereira-Dias et al., 2019) y puede ser explicada como una ventaja evolutiva en caso de errores de emparejamiento, ya que es más probable que se conserve la estructura de la proteína y por esa razón las transiciones se ven favorecidas frente a las transversiones en situaciones de mutación espontánea (Wakeley, 1996). Esto solo resultaría significativo en el caso de que el SNP estuviera en una región codificante. La tecnología DArTseq™ garantiza en gran medida que esto ocurra, ya que, al contrario de otras metodologías GBS, se reduce la complejidad del genoma de tal modo que se seleccionan solamente las regiones de bajo número de copias y asociadas predominantemente a genes activos.

La estructura genética del conjunto de datos fue analizada mediante análisis de coordenadas principales (PCoA), en lugar de emplear los análisis de componentes principales (PCA) típicamente aplicados en estudios de diversidad en recursos fitogenéticos. La utilización de PCoAs viene a subsanar el problema del sesgo que introduce el PCA cuando se tienen datos perdidos. El PCA completa los datos perdidos con la media de los otros datos obtenidos, sesgando de manera importante los datos cuando se trabaja por ejemplo con distintas especies. Por su parte, el PCoA calcula las distancias entre muestras dos a dos, de tal modo que no se tienen en cuenta los datos perdidos (Rohlf, 1972).

La subestructura hallada en el PCoA global que incluyó todas las accesiones analizadas concuerda con lo encontrado en estudios previos. Al igual que en nuestro caso, Pereira-Dias et al. (2019) mediante análisis de componentes principales (PCA) también revelaron una separación clara de la especie *C. baccatum*, la agrupación de *C. chinense* y *C. frutescens* y *C. annuum* por separado; con líneas de *C. annuum* var. *glabriusculum* posicionadas tanto en el grupo de *C. annuum* como en el de *C. chinense* y *C. frutescens*.

Capsicum annuum var. *glabriusculum* se considera el pariente silvestre de *C. annuum* (Bosland & Votava, 2000), por lo que estaría relacionada evolutivamente con todas las especies del complejo *C. annuum* (*C. annuum*, *C. chinense* y *C. frutescens*). Los diferentes orígenes de las accesiones de la especie y que mantenga una gran reserva de variabilidad genética dan lugar a que se agrupen con otras especies diferentes o incluso se posicionen

de manera intermedia, al igual que se observó en otros estudios en los que se incluyó este pariente silvestre (Taranto et al., 2016).

Las accesiones de *C. pubescens* aparecieron en el PCoA global como el grupo más alejado respecto al resto de accesiones. Esta diferenciación es clara ya a nivel fenotípico, pues *C. pubescens* es la única especie domesticada que presenta flores con corola de color púrpura (al igual que las especies silvestres *C. eximium* y *C. cardenasii*) cuando en las demás especies son blancas, semillas de color negro mientras que son claras en las otras especies y elevado grado de pubescencia en el tallo y en las hojas. La marcada separación de *C. pubescens* ya fue observada por otros autores con diferentes marcadores moleculares (Ibiza et al., 2012; Silvar & García-González, 2016). Realizando un PCoA jerárquico para las líneas de *C. pubescens* se alcanzó mayor profundidad en los resultados que si se incluyen todas las especies al conseguirse observar diferencias entre accesiones muy próximas genéticamente, en este caso en función del país de origen (Ecuador y Perú) y del nivel de pubescencia (escasa e intermedia respecto a densa).

La accesión PI 585244 se categorizó en el banco de germoplasma como *C. baccatum*. Sin embargo, en el PCoA global aparecía junto con las *C. pubescens*, lo que podría deberse a una clasificación errónea en el banco, posible pero improbable dada la clara diferenciación morfológica entre ambas especies. Por lo tanto, se apunta a una posible contaminación con ADN de otra especie.

Las muestras catalogadas como *C. baccatum* también aparecieron agrupadas y separadas de las otras especies en el PCoA global, resultado esperable puesto que esta especie

conforma un complejo independiente dentro del género *Capsicum*. *Capsicum baccatum* fue domesticada independientemente de otras especies en los valles de Bolivia, al sur de la cuenca del Amazonas (Pickersgill, 2007), manteniéndose su cultivo en áreas de montaña y, por lo tanto, dificultando el flujo génico (McLeod et al., 1982; Ibiza et al., 2012), lo que explicaría su diferenciación genética respecto a las demás especies del género. Nimmakayala et al. (2016b) ya observaron una marcada separación entre muestras de las especies *C. baccatum* y *C. annuum*, así como Pereira-Dias et al. (2019) donde las muestras de *C. baccatum* se agrupaban exclusivamente en un clúster. En este trabajo detectamos una separación geográfica total en *C. baccatum* en función del país de origen (Bolivia, Perú y Ecuador). Ibiza et al. (2012) ya observaron diferenciación geográfica en *C. baccatum* mediante marcadores microsatélite, si bien se separaban las accesiones de Bolivia de las de Perú-Ecuador, pero no las de estos dos últimos países. Así, se confirma la mayor precisión de los resultados que se obtienen con SNPs que con microsatélites.

Respecto a las accesiones de *C. chinense* y *C. frutescens*, su proximidad observada en el PCoA global y en análisis de inferencia bayesiana fastSTRUCTURE concuerda con los resultados de otros estudios que han encontrado una estrecha relación entre ambas especies utilizando diferentes marcadores (Nicolai et al., 2013; González-Pérez et al., 2014; Pereira-Dias et al., 2019). Algunos autores apuntan a que ambas especies podrían ser consideradas una sola, ya que son morfológicamente muy similares y se pueden cruzar entre sí fácilmente (Pickersgill, 1971; McLeod et al., 1979; Walsh & Hoot, 2001). Sin embargo, en el PCoA realizado solamente con las accesiones de estas dos especies, sí se observó diferenciación clara por especie. Por lo tanto, con nuestro enfoque de realizar sucesivos PCoAs jerárquicos es posible separar las dos especies, que se reflejan próximas

cuando se incluyen con más especies (y por lo tanto se añade variabilidad al conjunto de datos) pero que en realidad se separan perfectamente cuando se trabaja con ellas independientemente, constatando que se trata de especies diferentes si bien muy próximas genéticamente. La separación de ambas especies también se observó con fastSTRUCTURE trabajando con valores más altos de agrupación, siendo la única subestructura hallada dentro de alguno de los clústeres. Estos resultados junto con los del PCoA apoyan que sí se trata de especies diferentes pero su grado de diferenciación con las demás es menor, lo que indicaría que su domesticación fue posterior, al menos, a la de *C. annuum*; Pickersgill (2007) ya apuntó a una fecha posterior de domesticación para *C. chinense*, hace unos 4.000 años, mientras que la de *C. annuum* habría sido hace unos 5.000. Se cree que un ancestro común, a través de la domesticación en lugares independientes dio lugar a *C. chinense*, *C. frutescens* y *C. annuum* (Pereira-Dias et al., 2019). La posición intermedia de las accesiones de *C. annuum* var. *glabriusculum* encontrada así como en otros estudios (Nicolai et al., 2013; Pereira-Dias et al., 2019) apuntarían a que éste podría ser el vínculo común. Sin embargo, el mayor éxito comercial de *C. annuum* provocaría una evolución diferencial de esta especie al aumentar su área de distribución y el número de variedades generadas. Por el contrario, el cultivo de *C. chinense* y *C. frutescens* permanecería restringido a América Latina, permaneciendo confinadas y próximas entre sí (Pickersgill, 1997).

El PCoA de *C. frutescens* y *C. chinense* permitió reclasificar las accesiones BGV006008 y PI 573337, categorizadas como *C. annuum* en el banco de germoplasma, como *C. frutescens* y *C. chinense*, respectivamente. Pese a la clara asignación a *C. frutescens* en los PCoAs, la accesión BGV006008 presenta caracteres fenotípicos marcadamente de *C.*

annuum, por lo que el resultado debe tomarse con precaución. En estudios previos se apuntaba a que las accesiones PI 573337 y PI 543203 podrían pertenecer a la especie *C. frutescens* (Silvar & García-González, 2016), si bien esto parece descartado según nuestros datos al aparecer en el grupo de *C. chinense*. Aún así, clasificaciones erróneas como las aquí encontradas también se han observado en estudios con marcadores microsatélites (González-Pérez et al., 2014; Silvar & García-González, 2016). Esta discrepancia entre los datos de los bancos de germoplasma y la clasificación real no es extraña, ya que la clasificación en los bancos generalmente se lleva a cabo basándose en caracteres morfológicos como las flores y los frutos y que por lo tanto puede llevar a errores, especialmente en especies del complejo *C. annum*, muy similares morfológicamente entre sí. Ambas líneas se presentaban como *C. annum* en los resultados obtenidos en el Capítulo 1 de la presente Tesis Doctoral con marcadores de tipo microsatélite, al contrario de lo encontrado con GBS en este capítulo. Con el empleo de GBS para el análisis de SNPs se aumenta sustancialmente el número de marcadores (más de 48.000 SNPs frente a 28 *loci* microsatélites) y por lo tanto se aumenta la cobertura del genoma que se estudia. Las reclasificaciones taxonómicas basadas en caracteres moleculares evidencian la necesidad de llevar a cabo estudios genotípicos en recursos fitogenéticos, con el fin de caracterizar y preservar correctamente las distintas variedades para poder hacer un uso de ellas más efectivo. El uso de técnicas de NGS para el análisis de marcadores moleculares en recursos fitogenéticos como el que en este trabajo se presenta abre las puertas a una clasificación más exacta de las líneas conservadas en los bancos de germoplasma.

No se detectó subestructura clara para las accesiones de *C. chinense* y *C. frutescens* por separado. En el caso de *C. chinense*, dada la lejanía de las accesiones PI 573337, PI 543203 y CAP472 respecto al resto de accesiones de la especie, se apunta a que éstas podrían ser híbridos entre *C. chinense* y *C. frutescens*, hecho bastante probable debido a la compatibilidad reproductiva existente entre ellas (Onus & Pickersgill, 2004). En cuanto a *C. frutescens*, otros estudios también detectaron que las líneas de esta especie no presentan un patrón de agrupación geográfico (Ibiza et al., 2012) y que es difícil encontrar asociaciones cuando el número de accesiones es bajo (Silvar & García-González, 2016).

Por último, en el PCoA correspondiente a las accesiones de *C. annuum* se detectó un patrón filogeográfico claro: el Grupo 1 (principalmente colección andina) y el Grupo 2 (formado por la mayoría de accesiones de la colección ibérica y colección mediterránea). No formaron parte de ningún grupo y aparecieron alejadas una accesión de Bolivia, una de Perú y seis de México. Esto indica que las accesiones de Bolivia (donde se originó el género *Capsicum* (Eshbaugh, 1993)) y de México (donde se domesticó *C. annuum* (Pickersgill, 2007)), si bien están relacionadas con otros centros de diversificación, guardan cierta identidad genética que las caracteriza.

Dentro del Grupo 1 no fue posible detectar ningún patrón de agrupación entre las muestras, mientras que en el Grupo 2 se observó diferenciación geográfica entre accesiones correspondiente al Oeste-Este de Europa. Además, las accesiones del subgrupo Este Europa presentan mayoritariamente frutos de forma elongada o triangular, mientras que en el Oeste Europa la mayoría tienen frutos en bloque (IPGRI,

1995). La agrupación de las accesiones del Centro-Este de Europa, así como su morfología diferente a la del resto de Europa ya fue observada por Nicolai et al. (2013). Dichos autores apuntaron a que se trataría de un *pool* genético diferente y que los cultivares de esta zona estarían adaptados a condiciones climáticas diferenciales. Los resultados indican una combinación de parámetros de fruto seleccionados según su interés en el cultivo y el origen geográfico (probablemente por confinamientos territoriales) como explicación de la diversificación de *C. annuum*, al igual que sugieren otros autores (Nicolai et al., 2013; Taranto et al., 2016; Colonna et al., 2019;). Ambos hechos estarían directamente relacionados con la mejora de variedades encaminada a seleccionar las más productivas, bien por cruces entre líneas de interés o más recientemente por programas de mejora genética vegetal.

La separación Oeste-Este de Europa encontrada guarda relación con el proceso de diversificación de *C. annuum* en Europa. El pimiento (*C. annuum*) fue introducido en Europa hacia finales del siglo XV a través de la Península Ibérica, a partir de la primera expedición a América de Cristóbal Colón (Andrews, 1995). De este modo, la Península Ibérica se convirtió en un centro de diversificación secundaria, donde los agricultores locales fueron seleccionando durante siglos las variedades de pimiento mejor adaptadas y que mayores rendimientos les podían aportar (Nuez et al., 1996; Bosland & Votava, 2000), lo que explicaría la proximidad encontrada en este estudio entre las accesiones de España, Portugal y países próximos como Italia y Francia. Posteriormente el pimiento se diversificó a través de la región mediterránea y norte de África, hecho reflejado en nuestros resultados con la separación de las accesiones del Este de Europa (a las que se suma una línea de Túnez), de las del Oeste. Algunos autores apuntan a que en la parte

oriental de Europa, el pimiento habría sido introducido a través de Turquía. Estos pimientos provendrían de la India, zona en donde habría sido introducido a su vez por los portugueses tras su importación de Brasil (Andrews, 1995). De este modo en los pimientos de esta zona de Europa se mezclarían dos orígenes, a partir de la Península Ibérica tras introducirlo Cristóbal Colón y a partir de pimientos de la India, originarios de Brasil, lo que explicaría que en nuestros resultados se marque la distinción Oeste-Este. Estos resultados concuerdan con lo encontrado por Taranto et al. (2016), donde se separaban las accesiones mediterráneas y del Mar Negro (Bulgaria y Turquía) pero guardando relación con las españolas y del Norte de África. Destaca además que junto con las accesiones del Grupo 2 aparezcan dos de México, zona de domesticación de *C. annuum* (Pickersgill, 2007); evidenciando el origen de la especie en esta zona y que todavía se mantienen similitudes genéticas con variedades de este país.

La dificultad para separar las accesiones europeas y observar grupos claros radica en el origen común de todas ellas a partir de unas pocas introducidas desde América, que se fueron extendiendo a los demás países, seleccionando variedades adaptadas a las condiciones locales que generaron un nuevo abanico de formas pero que retenían muy poca variabilidad genética (Andrews, 1995; Nuez et al., 1996). Además, a esta selección de pocas variedades se sumó la introducción de las mismas líneas comerciales en diferentes áreas, modificando la estructura genética mediante polinización cruzada con las variedades locales (Ibiza et al., 2012).

En el PCoA para todas las líneas *C. annuum* también se ha podido comprobar que una línea clasificada como *C. annuum* en el banco de germoplasma pero que Silvar & García-

González (2016) reclasificaron como *C. chinense*, la CGN20808, se corresponde en realidad con una *C. annuum* (señalada con una flecha negra en la Figura 2.11). Se incluyó esta línea porque en el PCoA global (Figura 2.6) aparecía con las demás *C. annuum* y de este modo se comprobó que cuando se estudia solo las de esta especie, aparece igualmente agrupada con las demás *C. annuum*. Este resultado también evidencia la idoneidad de combinar distintos tipos de marcadores para obtener conclusiones más robustas.

Los resultados de estructura genética obtenidos mediante inferencia bayesiana con el programa fastSTRUCTURE concuerdan con los hallados con los PCoA, si bien la profundidad de los mismos es mucho menor, ya que en el caso del análisis clúster es imposible ver relaciones intraespecíficas.

Finalmente, la baja diversidad genética detectada en todas las especies es lo esperable en especies del género *Capsicum* y se atribuye al alto nivel de endogamia de las especies (Taranto et al., 2016). Los bajos valores de heterocigosidad coinciden con otros estudios en *Capsicum* empleando SNPs (Nimmakayala, et al., 2016a; Taranto et al., 2016) y microsatélites (Nicolai et al., 2013). De hecho, empleando microsatélites, González-Pérez et al. (2014) encontraron las mismas especies con mayor y menor heterocigosidad observada que en este estudio, *C. pubescens* y *C. annuum*, respectivamente. En contraposición, en el Capítulo 1.1 de esta Tesis Doctoral encontramos que eran *C. chinense* y *C. baccatum* las que mayor y menor nivel de heterocigosidad observada presentaban, respectivamente; por lo que los resultados pueden variar según el estudio aún empleando el mismo tipo de marcadores. La especie que presentó mayor número

medio de alelos fue *C. annuum*. Lo mismo sucedía en los resultados con microsátélites presentados en el Capítulo 1.1 de la presente Tesis Doctoral, pero cuando se calculaba la riqueza alélica, esta era similar en todas las especies. Por ello es necesario resaltar que la estima del número medio de alelos empleando SNPs podría estar sesgado porque se analizaron muchas más accesiones de *C. annuum* que de las demás especies. Por su parte, en nuestro estudio, *C. pubescens* mostró el mayor número medio de alelos privados y la heterocigosidad observada más elevada, lo que indicaría que mantiene un gran reservorio genético, interesante de cara a estudios de mejora vegetal.

Capítulo 3

Caracterización morfológica en cinco especies del género *Capsicum* y análisis de asociación *genome-wide association study* (GWAS) entre datos fenotípicos y genotípicos en *C. annuum*

Introducción

Establecer la relación entre las variantes fenotípicas y los polimorfismos genéticos asociados a ellas es crucial en especies comerciales. Concretamente, en el caso de las especies cultivadas, predecir características agronómicas de interés requiere conocer los *loci* específicos que subyacen al fenotipo así como la arquitectura genética asociada (Korte & Farlow, 2013).

Tradicionalmente, la búsqueda de la relación entre fenotipo y genotipo se llevaba a cabo mediante el mapeo de *loci* de rasgos cuantitativos (*quantitative trait loci*, QTLs; Glazier et al., 2002). Sin embargo, el análisis clásico de los QTLs presenta ciertas limitaciones, ya que requiere del establecimiento de líneas puras que maximicen y minimicen el valor fenotípico, solo es posible evaluar la diversidad genética que segrega entre esas líneas puras parentales (Borevitz & Nordborg, 2003) y la resolución del mapeo puede verse afectada por el grado de recombinación que ha tenido lugar al generar dichas líneas, bien de manera natural o por cruces dirigidos (Balasubramanian et al., 2009; Korte & Farlow, 2013). Los análisis de asociación a nivel del genoma completo (*genome-wide association study*, GWAS) sortean estos inconvenientes al proporcionar información sobre las regiones genómicas que subyacen a ese rasgo, permitiendo inferir los genotipos parentales (Korte & Farlow, 2013). Así, los GWAS más utilizados actualmente identifican la posible asociación entre marcadores genotípicos de tipo SNP (*single nucleotide polymorphisms*, polimorfismo de un solo nucleótido) y la variación un carácter fenotípico de interés analizando un gran número de individuos (Hirschhorn & Daly, 2005). Técnicas estadísticas evalúan esta relación midiendo el grado de asociación entre la

variación fenotípica y la de cada una de las variantes genómicas analizadas mediante modelos lineales (Burghardt et al., 2017). Estudios GWAS previos en pimiento permitieron la identificación de SNPs asociados al contenido de capsaicinoides (Nimmakayala et al., 2016; Han et al., 2018), a la longitud del pedicelo del fruto (Nimmakayala et al., 2016b) y a ciertos caracteres morfológicos del fruto (Colonna et al., 2019).

El éxito de un estudio GWAS viene dado por una selección adecuada de las muestras, así como por el número de los SNPs generados y la calidad de los mismos. Al incluir el mayor número de variantes posible es más probable que alguna de ellas esté asociada a la región genómica que controla el fenotipo de interés (Korte & Farlow, 2013). Asimismo, un análisis fenotípico lo más exhaustivo posible permite encontrar asociaciones más precisas. Herramientas de cuantificación automatizadas como Tomato Analyzer (Rodríguez et al., 2010), posibilitan un examen fenotípico muy completo ya que toman medidas cuantitativas a partir de imágenes escaneadas de las secciones de los frutos, identificando ciertos rasgos casi imposibles de cuantificar manualmente y minimizando el componente subjetivo (Tripodi & Greco, 2018). A pesar de que el programa fue desarrollado originalmente para analizar variaciones en la morfología de tomate (Brewer et al., 2007; Gonzalo & van der Knaap, 2008), su uso también se ha extendido en estudios de berenjena (Kaushik et al., 2016) o melón (Pereira et al., 2018). En pimiento, Tripodi & Greco (2018) realizaron un extenso estudio de fenotipado utilizando Tomato Analyzer con el objetivo de describir la diversidad morfológica de los frutos de las cinco especies cultivadas de pimiento y sus variedades silvestres. Posteriormente, Naegele et al., (2016) establecieron relaciones entre variedades de *C. annuum* en base a datos

fenotípicos de sus frutos obtenidos con esta herramienta. El fenotipado proporcionado por Tomato Analyzer ha permitido además estudiar la base genética de la diversidad de los frutos mediante estudios de asociación de tipo GWAS en *C. annuum* (Colonna et al., 2019; Solomon et al., 2019). Los autores encontraron SNPs asociados significativamente a parámetros que definen la forma rectangular del fruto, la asimetría, el bloqueo y los extremos distales y proximales (Colonna et al., 2019) y al tamaño del fruto (Solomon et al., 2019).

Por lo tanto, la combinación de un gran número de marcadores genómicos con un fenotipado completo es la base para los estudios GWAS. En el presente trabajo se realizará un análisis GWAS a partir de SNPs generados mediante GBS y datos obtenidos de un completo fenotipado de la planta y los frutos, tanto con descriptores típicos como con parámetros medidos con Tomato Analyzer. El GWAS se llevará a cabo para las accesiones de *C. annuum*, así como para todas las accesiones en las del complejo *C. annuum* (*C. annuum*, *C. frutescens* y *C. chinense*) conjuntamente con el fin de comprobar si este tipo de análisis es susceptible de ser aplicado cuando se emplean especies distintas, aunque estén relacionadas. Los resultados obtenidos permitirán establecer qué variantes alélicas están asociadas a fenotipos concretos de caracteres con interés agronómico. La inclusión de dichas variantes, o la región genómica en la que se encuentran, en programas de mejora hará posible obtener nuevas líneas con el fenotipo seleccionado. Además, el genotipado de los marcadores asociados al rasgo posibilitará predecir el fenotipo de una línea de manera temprana, sin necesidad de esperar a un desarrollo completo de las plantas que implicaría una mayor inversión de recursos.

Material y métodos

Datos fenotípicos

Ciento noventa y seis accesiones de las cinco especies domesticadas de *Capsicum* se cultivaron en invernadero en el Centro de Investigaciones Agrarias de Mabegondo (Mabegondo, A Coruña) ($43^{\circ}14'46,8''\text{N}$ $8^{\circ}16'21,3''\text{O}$) en los periodos de abril a septiembre de 2017 y 2018. El ensayo incluyó 3-4 plantas de cada línea y se realizó como un diseño en bloque al azar. Cuando se observó más de un fenotipo por accesión se consideró solo aquel presente en la mayoría de plantas, y cuando no fue posible se descartó la línea, para así garantizar que no se introdujesen errores debidos al envío de semillas no correspondientes con la línea sembrada.

En el invernadero se tomaron medidas de “altura de la planta”, “anchura de la planta” y “longitud del tallo” según lo descrito en IPGRI (1995) así como del diámetro del tallo (Tabla 3.1). Además se realizó el fenotipado de la planta para los rasgos cualitativos “color del tallo”, “antocianina en el nudo”, “forma del tallo”, “pubescencia del tallo” y “hábito de crecimiento de la planta” (Tabla 3.2) según IPGRI (1995). Los frutos se recogieron en estado de maduración completa y de manera individualizada para cada línea. Tras su recogida, se registraron datos fenotípicos para “color del fruto en estado maduro”, “forma del fruto”, “forma del fruto en la unión con el pedicelo”, “cuello en la base del fruto”, “forma del ápice del fruto”, “apéndice en el fruto” y “tipo de epidermis del fruto” (IPGRI, 1995) (Tabla 3.2) y se registró el peso del fruto y la longitud del pedicelo del fruto (Tabla 3.1) tomando 20 réplicas de cada accesión siempre que fue

posible. Tras el fenotipado, los frutos se cortaron en secciones transversales y longitudinales y se escanearon en un escáner ScanJet G3110 (HP, California, Estados Unidos). Las imágenes escaneadas se procesaron con el programa Tomato Analyzer 3.0 (Rodríguez et al., 2010) con el que se realizaron 36 medidas en los frutos (Tabla 3.3), 24 de ellas en sección longitudinal del fruto y 12 en sección transversal. Siempre que fue posible, de cada accesión se procesaron 20 frutos en sección longitudinal y 20 en sección transversal.

Tabla 3.1. Descripción de los caracteres fenotípicos cuantitativos medidos para planta y fruto.

Parámetro	Abreviatura	Descripción
Altura de la planta	AIP	Altura (cm) cuando comienza a madurar el primer fruto en el 50% de las plantas
Anchura de la planta	AnP	Anchura (cm) medida inmediatamente después de la primera cosecha, en el punto más ancho
Longitud del tallo	LT	Altura (cm) hasta la primera bifurcación, inmediatamente después de la primera cosecha
Diámetro del tallo	DT	Diámetro (cm) extrapolado a partir del perímetro del tallo, medido en la parte del medio hasta la primera bifurcación, inmediatamente después de la primera cosecha
Peso del fruto	Pes	Peso (g) del fruto maduro
Longitud del pedicelo del fruto	LonPe	Longitud (cm) del pedicelo del fruto maduro

Tabla 3.2. Descripción de los caracteres fenotípicos cualitativos estudiados para planta y fruto.

Parámetro	Abreviatura	Descripción	Valores según IPGRI (1995)
Color del tallo	CT	Registrado en las plantas jóvenes	1=verde; 2= verde con rayas púrpura; 3=morado; 4=otro (especificar)
Antocianina en el nudo	AN	Observado cuando la planta está madura	1=verde; 3=morado claro; 5=morad; 7=morado oscuro
Forma del tallo	FT	Observado cuando la planta está madura	1=cilíndrico; 2=angular; 3=achatado (aplastado)
Pubescencia del tallo	PBT	Observado cuando la planta está madura, excluyendo los dos primeros nudos debajo del brote	3=escasa; 5=intermedia; 7=densa
Hábito de crecimiento de la planta	HC	Observado cuando ha empezado a madurar el primer fruto en el 50% de las plantas	3=postrada; 5=intermedia (compacta); 7=erecta; 9=otro (especificar)
Color del fruto en estado maduro	CM	Observado en frutos maduros (a la primera cosecha)	1=blanco; 2=amarillo-limón; 3=amarillo-naranja pálido; 4=amarillo-naranja; 5=naranja pálido; 6=naranja; 7=rojo claro; 8=rojo; 9=rojo oscuro; 10=morado; 11=marrón; 12=negro; 13=otro (especificar)
Forma del fruto	FF	Observado en frutos maduros (a la primera cosecha)	1=elongado; 2=casi redondo; 3=triangular; 4=acampanulado; 5=acampanulado y en bloque; 6=otro (especificar)
Forma del fruto en la unión con el pedicelo	FFU	Observado en frutos maduros (a la primera cosecha)	1=agudo; 2=obtusos; 3=truncado; 4=cordado; 5=lobulado
Cuello en la base del fruto	CBF	Observado en frutos maduros (a la primera cosecha)	0=ausente; 1=presente
Forma del ápice del fruto	FAF	Observado en frutos maduros (a la primera cosecha)	1=puntudo; 2=romo; 3=hundido; 4=hundido y puntudo; 5= otro (especificar)
Apéndice en el fruto	AF	Vestigio de la floración. Observado en frutos maduros (a la primera cosecha)	0=ausente; 1=presente
Tipo de epidermis del fruto	TE	Observado en frutos maduros (a la primera cosecha)	1=lisa; 2=semirrugosa; 3=rugosa

Tabla 3.2. Descripción de los caracteres fenotípicos cualitativos estudiados para planta y fruto. (continuación)

Parámetro	Abreviatura	Descripción	Valores según IPGRI (1995)
Color del tallo	CT	Registrado en las plantas jóvenes	1=verde; 2= verde con rayas púrpura; 3=morado; 4=otro (especificar)
Antocianina en el nudo	AN	Observado cuando la planta está madura	1=verde; 3=morado claro; 5=morad; 7=morado oscuro
Forma del tallo	FT	Observado cuando la planta está madura	1=cilíndrico; 2=angular; 3=achatado (aplastado)
Pubescencia del tallo	PBT	Observado cuando la planta está madura, excluyendo los dos primeros nudos debajo del brote	3=escasa; 5=intermedia; 7=densa
Hábito de crecimiento de la planta	HC	Observado cuando ha empezado a madurar el primer fruto en el 50% de las plantas	3=postrada; 5=intermedia (compacta); 7=erecta; 9=otro (especificar)
Color del fruto en estado maduro	CM	Observado en frutos maduros (a la primera cosecha)	1=blanco; 2=amarillo-limón; 3=amarillo-naranja pálido; 4=amarillo-naranja; 5=naranja pálido; 6=naranja; 7=rojo claro; 8=rojo; 9=rojo oscuro; 10=morado; 11=marrón; 12=negro; 13=otro (especificar)
Forma del fruto	FF	Observado en frutos maduros (a la primera cosecha)	1=elongado; 2=casi redondo; 3=triangular; 4=acampanulado; 5=acampanulado y en bloque; 6=otro (especificar)
Forma del fruto en la unión con el pedicelo	FFU	Observado en frutos maduros (a la primera cosecha)	1=agudo; 2=obtuso; 3=truncado; 4=cordado; 5=lobulado
Cuello en la base del fruto	CBF	Observado en frutos maduros (a la primera cosecha)	0=ausente; 1=presente
Forma del ápice del fruto	FAF	Observado en frutos maduros (a la primera cosecha)	1=puntudo; 2=romo; 3=hundido; 4=hundido y puntudo; 5= otro (especificar)
Apéndice en el fruto	AF	Vestigio de la floración. Observado en frutos maduros (a la primera cosecha)	0=ausente; 1=presente
Tipo de epidermis del fruto	TE	Observado en frutos maduros (a la primera cosecha)	1=lisa; 2=semirrugosa; 3=rugosa

Tabla 3.3. Descripción de los caracteres fenotípicos medidos con el programa Tomato Analyzer.

Categoría	Parámetro	Abreviatura	Descripción	
Medidas básicas (Basic Measurements)	Perímetro (<i>Perimeter</i>)	P	Longitud del perímetro (mm)	
	Área (<i>Area</i>)	A	Área del fruto (mm ²)	
	Anchura en la mitad de la altura (<i>Width Mid-height</i>)	WMH	Anchura medida a la mitad de la altura del fruto (mm)	
	Anchura máxima (<i>Maximum Width</i>)	MW	Mayor distancia horizontal del fruto (mm)	
	Altura en la mitad de la anchura (<i>Height Mid-width</i>)	HMW	Altura medida a la mitad de la altura del fruto (mm)	
	Altura máxima (<i>Maximum Height</i>)	MH	Mayor distancia vertical del fruto (mm)	
	Altura en la curva (<i>Curved Height</i>)	CH	Altura medida a lo largo de una línea curva que atraviesa el fruto (mm)	
	Índice de la forma del fruto (<i>Fruit Shape Index</i>)	Índice externo de la forma del fruto I (<i>Fruit Shape Index External I</i>)	FSIEI	Relación entre la altura máxima y la anchura máxima
		Índice externo de la forma del fruto II (<i>Fruit Shape Index External II</i>)	FSIEII	Relación entre la altura en la mitad de la anchura y la anchura en la mitad de la altura
		Índice de la forma curva del fruto (<i>Curved Fruit Shape Index</i>)	CFSI	Relación entre la altura en la curva y la anchura del fruto en la mitad de la altura de la curva, medida perpendicularmente a la línea de la altura de la curva
Bloqueo (<i>Blockiness</i>)	Bloqueo proximal del fruto (<i>Proximal Fruit Blockiness</i>)	PFB	Relación entre la anchura en el punto de bloqueo superior y la anchura en la mitad de la altura	
	Bloqueo distal del fruto (<i>Distal Fruit Blockiness</i>)	DFB	Relación entre la anchura en el punto de bloqueo inferior y la anchura en la mitad de la altura	
	Forma triangular del fruto (<i>Fruit Shape Triangle</i>)	FST	Relación entre la anchura en el punto de bloqueo superior y la anchura en el punto de bloqueo inferior	

Tabla 3.3. Descripción de los caracteres fenotípicos medidos con el programa Tomato Analyzer. (continuación)

Categoría	Parámetro	Abreviatura	Descripción
Homogeneidad (Homogeneity)	Elipsoide (<i>Ellipsoid</i>)	ELL	Relación entre el error resultante de la elipse que mejor se ajusta y el área del fruto. Indica la precisión de ajustar la forma real del fruto a una elipse (valores pequeños indican una forma real elipsoide)
	Circular (<i>Circular</i>)	CIR	Relación entre el error resultante del círculo que mejor se ajusta y el área del fruto. Indica la precisión de ajustar la forma real del fruto a un círculo (valores pequeños indican una forma real circular)
	Rectangular (<i>Rectangular</i>)	REC	Relación entre el área del rectángulo que delimita el fruto con el área del rectángulo delimitado por el fruto
Forma del extremo proximal del fruto (Proximal Fruit End Shape)	Altura en los hombros (<i>Shoulder Height</i>)	SH	Relación entre la altura media de los puntos que quedan por encima del punto final proximal y la altura máxima
	Área de la hendidura proximal (<i>Proximal Indentation Area</i>)	PIA	Área de la hendidura proximal dividida por el área total del fruto, multiplicada por 10
Asimetría (<i>Asymmetry</i>)	Obovoide (<i>Obovoid</i>)	OBO	Relación entre el área del fruto por encima y por debajo de la mitad de la altura (valores más altos indican un área mayor por debajo de la mitad de la altura)
	Ovoide (<i>Ovoid</i>)	OVO	Relación entre el área del fruto por debajo y por encima de la mitad de la altura (valores más altos indican un área mayor por encima de la mitad de la altura)
	V. Asimetría (<i>V. Asymmetry</i>)	VA	Distancia media entre una línea vertical que atraviesa el fruto en la mitad de la anchura y el punto medio de la anchura del fruto a cada altura
	H. Asimetría. Ob (<i>H. Asymmetry.ob</i>)	HAB	Si el área del fruto es mayor por debajo de la mitad de la altura que por arriba, entonces es la distancia media entre una línea horizontal que atraviesa el fruto a la mitad de la altura y el punto medio de la altura del fruto a cada anchura, en caso contrario es 0.
	H. Asimetría. Ov (<i>H. Asymmetry.ov</i>)	HAV	Si el área del fruto es mayor por encima de la mitad de la altura que por debajo, entonces es la distancia media entre una línea horizontal que atraviesa el fruto a la mitad de la altura y el punto medio de la altura del fruto a cada anchura, en caso contrario es 0.
	Posición de la anchura máxima (<i>Width Widest Pos</i>)	WWP	Relación entre la altura a la que se produce la anchura máxima y la altura máxima

Tabla 3.3. Descripción de los caracteres fenotípicos medidos con el programa Tomato Analyzer. (continuación)

Categoría	Parámetro	Abreviatura	Descripción
Sección latitudinal (<i>Latitudinal Section</i>)	Grado de lóbulos (<i>Lobedness Degree</i>)	LD	Desviación estándar de las distancias desde el centro del fruto al perímetro, multiplicada por 100
	Área del pericarpio (<i>Pericarp Area</i>)	PA	Relación entre el área dentro del pericarpio y el área del fruto
	Grosor del pericarpio (<i>Pericarp Thickness</i>)	PT	Relación entre el grosor medio del pericarpio en las líneas verticales y horizontales que pasan por el centro del fruto y la media de la altura máxima y la anchura máxima
Valores promedio de color (<i>Average Color Values</i>)	Rojo promedio (<i>Average Red</i>)	AR	Media del valor R (0-255) en el sistema de color RGB a través de todos los píxeles del fruto
	Verde promedio (<i>Average Green</i>)	AG	Media del valor G (0-255) en el sistema de color RGB a través de todos los píxeles del fruto
	Azul promedio (<i>Average Blue</i>)	AB	Media del valor B (0-255) en el sistema de color RGB a través de todos los píxeles del fruto
	Luminosidad media (<i>Average Luminosity</i>)	AL	Media de la luminosidad (0-240) a través de todos los píxeles del fruto
	Media del valor L (<i>Average L Value</i>)	ALV	Media del valor L* (0-100) en el sistema de color L*a*b* a través de todos los píxeles del fruto
	Media del valor a (<i>Average a Value</i>)	AAV	Media del valor a* (0-100) en el sistema de color L*a*b* a través de todos los píxeles del fruto
	Media del valor b (<i>Average b Value</i>)	ABV	Media del valor b* (0-100) en el sistema de color L*a*b* a través de todos los píxeles del fruto
	Tono promedio (<i>Average Hue</i>)	HA	Tono (0°-360°) representado por la media del valor a y la media del valor b
	Croma promedio (<i>Average Chroma</i>)	CHR	Croma representado por la media del valor a y la media del valor b

Para el GWAS se seleccionaron 117 accesiones de *C. annuum* y 154 pertenecientes al complejo *C. annuum* (Tabla 1.1, columna GWAS), de las que se disponía de información fenotípica.

El GWAS se llevó a cabo para los 36 caracteres medidos con Tomato Analyzer así como para el peso del fruto, longitud del pedicelo, altura de la planta, anchura de la planta, longitud del tallo y diámetro del tallo, tanto en *C. annuum* como en el complejo *C. annuum*.

Para cada carácter se testó si la distribución fenotípica se ajustaba a una distribución normal a través del test de Lilliefors (Kolmogorov-Smirnov corregido) (Lilliefors, 1967) y se visualizó la distribución de los datos a través de gráficos de normalidad (histogramas y gráficos Q-Q) mediante el entorno R 3.6.3 (R Core Team, 2013). A las variables que no se ajustaron a la normalidad se les aplicó la transformación Box-Cox (Box & Cox, 1982) empleando R 3.6.3. En las variables transformadas se volvió a testar la normalidad con el test Lilliefors y aquellas que se ajustaron a la normalidad se incluyeron posteriormente en el análisis GWAS.

Datos genotípicos

Tanto para el grupo *C. annuum* como para el complejo *C. annuum*, el archivo de datos genotípicos se generó a partir del fichero VCF de 245.307 SNPs descrito en el Capítulo 2. Una vez seleccionadas las 117 accesiones de *C. annuum* y las 154 del complejo *C. annuum*, se procedió a eliminar los *indels* y los SNPs no asignados a cromosomas con VCFtools

0.1.17 (Danecek et al., 2011) y Tassel 5.0 (Bradbury et al., 2007), respectivamente. A continuación, aquellos SNPs con $DP-GT < 2$, se transformaron a dato perdido, eliminándose posteriormente todas las variantes con una proporción de datos perdidos $\geq 0,3$. Además, se realizó un filtrado de las accesiones con una proporción de datos perdidos $\geq 0,3$. Tras el filtrado se realizó la imputación de genotipos faltantes con Beagle 5.1 (Browning et al., 2018) según el método *phasing* descrito en Browning & Browning (2007), con 6 iteraciones de *burn-in* y 12 iteraciones de *phasing*. Finalmente, se realizó un filtrado con Tassel para eliminar los SNPs con una frecuencia mínima de alelo (*minor allele frequency*, MAF) $\leq 0,05$ y heterocigosidad observada $\geq 0,1$. Además, se eliminaron aquellas accesiones con una proporción de SNPs heterocigotos $\geq 0,4$.

Cálculo de la *Q-matrix* y la *K-matrix*

La estructura genética del conjunto de accesiones se estimó con el *software* ADMIXTURE 1.3.0 (Alexander et al., 2009) con valores de K que variaron entre 1 y 10; realizando 200 réplicas de *bootstrap* y una validación cruzada (*cross-validation*) de $CV=10$ para estimar el valor de K más probable. Las *Q-matrix* se construyeron con los coeficientes de pertenencia de cada accesión a cada grupo establecido por el análisis poblacional. La *K-matrix* (matriz *kinship*), se calculó usando el *software* Tassel 5.0 con el método “centered IBS” según lo descrito en Endelman & Jannink (2012)

Estudio de asociación a nivel del genoma completo (*genome-wide association study*, GWAS)

El estudio de asociación se llevó a cabo con el software Tassel 5.0, empleando tres aproximaciones a partir de los dos modelos que tiene implementados: (1) GLM+Q, Modelo Lineal Generalizado condicionado con estimaciones de la estructura poblacional empleando la *Q-matrix*; (2) MLM+K, Modelo Lineal Mixto (Yu et al., 2006) condicionado con la *K-matrix* y (3) MLM+Q+K, Modelo Lineal Mixto que incorpora como covariantes la *Q-matrix* y la *K-matrix*. Para calcular los modelos MLM, Tassel implementa el algoritmo P3D (Zhang et al., 2010) para reducir el tiempo de computación.

Para determinar el grado de ajuste de los modelos en cada variable y en cada uno de los archivos genotípicos, se observaron los gráficos Q-Q de $-\log_{10}(\text{p-valor})$. Con el fin de establecer a partir de que p-valor se considera una asociación significativa se empleó el paquete “simpleM” de R (Gao et al., 2008), que se basa en una modificación de la corrección de Bonferroni para pruebas múltiples. Las representaciones de los gráficos Manhattan y Q-Q se realizaron con el paquete “qqman” de R (Turner, 2018).

Desequilibrio de ligamiento (*linkage disequilibrium*, LD)

El desequilibrio de ligamiento entre los SNPs se determinó calculando el coeficiente de correlación de las frecuencias alélicas al cuadrado (R^2) y el valor absoluto de D' para los pares de sitios polimórficos con el programa Tassel 5.0, empleando el método “site by all” para estudiar posiciones concretas. Se fijó $R^2=0,2$ como el umbral por encima del cual

se consideró que una pareja de SNPs se encuentra en desequilibrio de ligamiento (Pasam et al., 2012).

Mapeo contra el genoma anotado de *C. annuum*

Para comprobar la posible implicación biológica de los SNPs significativamente asociados con algún carácter fenotípico, se realizó un *blast* de las secuencias que contenían el SNP contra el genoma anotado de *C. annuum* CM334 (Kim et al., 2014) a través de la base de datos de Solgenomics (<https://solgenomics.net/to-ols/blast>; Genoma CDS de *Capsicum annuum* cv CM334 v1.55). Además, se realizó una comparación adicional realizando otro *blast* general mediante la herramienta BlastN (Zhang et al., 2000) del NCBI (<https://blast.ncbi.nlm.nih.gov/>).

Resultados

Variabilidad morfológica

De las 196 accesiones de *Capsicum* sembradas se obtuvieron datos para 191 en al menos alguno de los doce caracteres cualitativos de planta y fruto (Tabla A.5), ocho cuantitativos de planta y fruto (Tabla A.6) y 34 cuantitativos de fruto obtenidos con el *software* Tomato Analyzer (Tabla A.7). Cuando se observó más de un fenotipo por accesión se consideró solo aquel presente en la mayoría de plantas, y cuando no fue posible se descartó la línea, para así garantizar que no se introdujesen errores debidos al envío de semillas no correspondientes con la línea sembrada.

La menor variabilidad se encontró en aquellos parámetros que definen la arquitectura de la planta, con valores similares en ellos pero destacando las accesiones de *C. frutescens* y *C. baccatum* con los niveles más elevados de altura y ancho de la planta (Tabla A.6). La mayoría de accesiones de *C. pubescens*, así como algunas de *C. annuum* mostraron presencia de antocianina en el nudo. Un parámetro determinante en la identificación de accesiones de *C. pubescens* es la pubescencia del tallo, mostrando estas accesiones valores altos en dicho carácter. Varias accesiones de *C. annuum* de México también presentaron los mismos valores de pubescencia (Tabla A.5).

Los parámetros más variables fueron aquellos relacionados con la forma, el tamaño y el color del fruto, encontrándose una gran diversidad morfológica (Figura 3.1). Fue dentro de las especies *C. chinense* y *C. annuum* donde se encontró la mayor variabilidad. Los

frutos más grandes se encontraron en *C. annuum*, con un peso medio de 53,3 g pero con un amplio rango que varió entre los 0,1 g y los 202,4 g, lo que explicaría que el coeficiente de variación fuera del 95%.

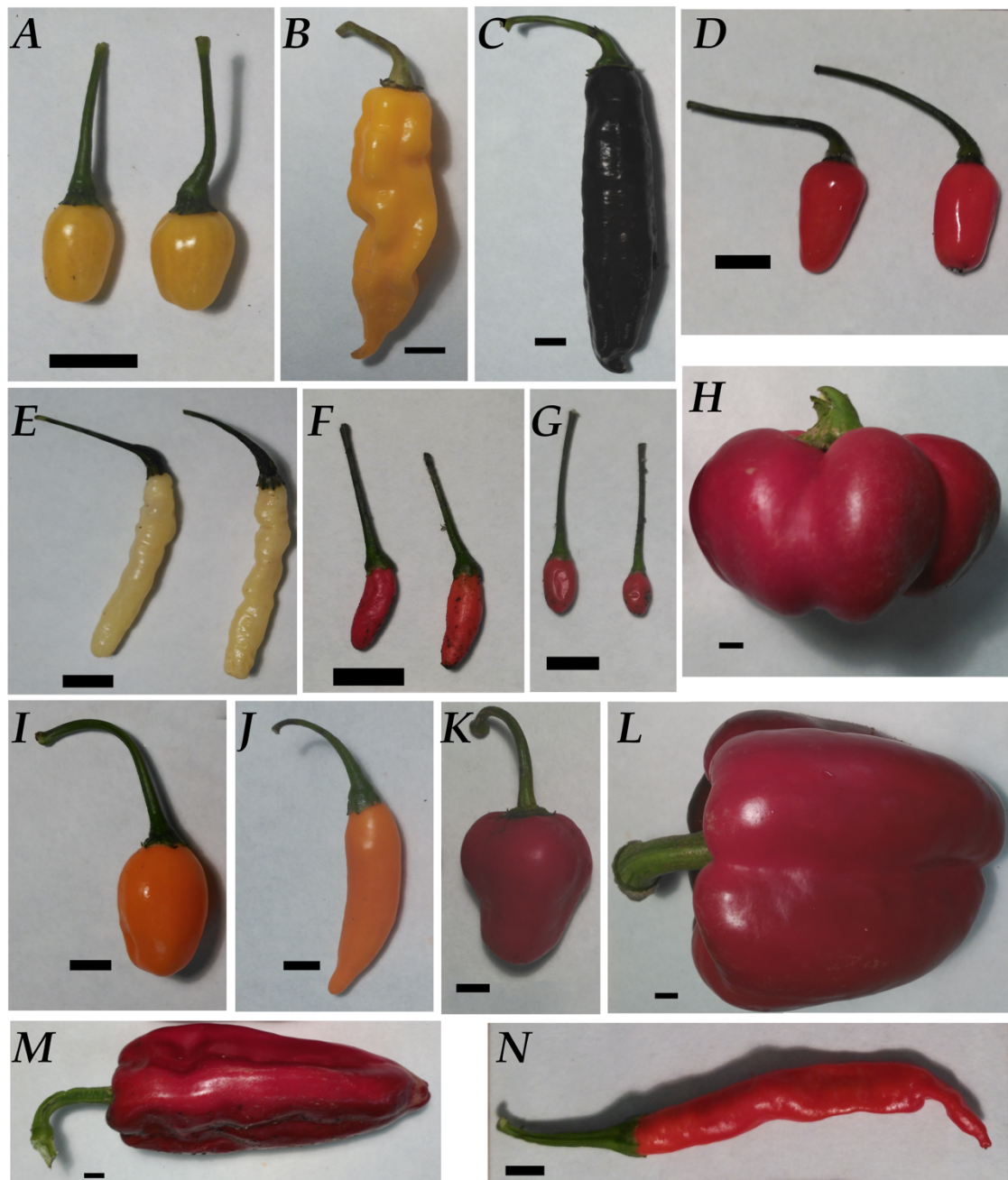


Figura 3.1. Diversidad morfológica encontrada en los frutos de las accesiones estudiadas. A. CAP472. **B.** CAP868. **C.** PI 315015. **D.** PI 595907. **E.** PI 573337. **F.** PM647. **G.** PM670. **H.** CGN17214. **I.** PI 585244. **J.** PI 590505. **K.** PI 585269. **L.** BGHZ0775. **M.** BGHZ0113. **N.** PI 241670. Las barras de tamaño indican en todos los casos 1 cm. Fuente: imágenes propias.

Los resultados obtenidos en el fenotipado con descriptores tradicionales tuvieron su correspondencia con los resultados del programa Tomato Analyzer. Así, parámetros indicadores del tamaño del fruto como perímetro, área, anchura máxima o altura máxima fueron mayores en las líneas de *C. annuum*. Sin embargo, los coeficientes de variación también fueron mayores en esta especie, lo que explicaría la variabilidad intraespecífica existente. Dentro de los parámetros obtenidos con el programa Tomato Analyzer, Obovoide y H. Asimetría Ob. mostraron resultados de cero para todos los frutos estudiados, por lo que no se tuvieron en cuenta en los análisis GWAS.

Análisis de asociación en *C. annuum*

Identificación de SNPs

Se estudió un conjunto de 117 accesiones de *C. annuum* seleccionadas a fin de lograr una completa representación de la diversidad genética, morfológica y geográfica de la especie, así como del pariente silvestre *C. annuum* var. *glabriusculum*. Todas las accesiones se eligieron según los datos de especie proporcionados por los diferentes bancos de germoplasma y según los resultados de reclasificación taxonómica mostrados en los anteriores capítulos de la presente Tesis. Cuarenta y cinco eran accesiones originarias de España, diez de México, siete de Portugal, seis de Ecuador, seis de Turquía, cinco de Italia, cinco de Bolivia, cinco de Hungría, cuatro de Bulgaria, cuatro de Israel, tres de Francia, tres de Libia, tres de Perú, tres de Túnez, tres de

Serbia/Montenegro, dos de Albania y una de Eslovenia; así como dos accesiones de *C. annuum* var. *glabriusculum* (una de México y una de Colombia).

El fichero VCF original obtenido tras el *SNV calling* para todas las accesiones analizadas con GBS se filtró para seleccionar solo aquellas que pertenecían a la especie *C. annuum* y *C. annuum* var. *glabriusculum*, obteniendo un nuevo fichero VCF con 117 líneas y 245.307 SNVs. Tras eliminar los *indels* y los SNPs no asignados a cromosomas en el *SNV calling* se conservaron 177.935 SNPs. Después de filtrarlos eliminando aquellos con una proporción de datos perdidos $\geq 0,3$ y las accesiones con proporción de datos perdidos $\geq 0,3$ se conservaron 112 accesiones con 68.694 SNPs. Posteriormente se realizó la imputación de genotipos faltantes con Beagle 5.1. Seguidamente, el archivo resultante se filtró para eliminar SNPs con $MAF \leq 0,05$ y heterocigosidad observada $\geq 0,1$ así como aquellas accesiones con proporción de SNPs heterocigotos $\geq 0,4$, obteniendo un fichero con 111 líneas y 4.787 SNPs de alta calidad que se empleó para el análisis GWAS.

Estructura poblacional

El análisis con ADMIXTURE reveló que la estructura óptima del conjunto de datos es $K=3$ grupos (Figura A.3A), sin existir una correspondencia entre los grupos y la procedencia geográfica de las accesiones. El grupo mayoritario engloba a una variedad de accesiones con morfologías y orígenes diversos, si bien los dos grupos minoritarios sí se agrupan en cierto modo según datos fenotípicos (Figura 3.2). El segundo clúster reúne accesiones con frutos pequeños (peso medio de 1,91 g/fruto), de color en estado maduro mayoritariamente rojo (8 según IPGRI, 1995) y forma del fruto elongada (1 según IPGRI,

1995). Las accesiones de este clúster presentan además valores más elevados de altura y anchura de planta y de longitud del tallo, que se corresponden con plantas de mayor porte que permiten dar mayor número de frutos para suplir el pequeño tamaño de estos. El tercer clúster está formado por accesiones con frutos predominantemente de color rojo o rojo oscuro en estado maduro (8 y 9 respectivamente en IPGRI, 1995), de forma triangular o acampanulada y en bloque (3 y 5 respectivamente según IPGRI, 1995) y de tamaño grande, con un peso medio de 38,79 g/fruto.

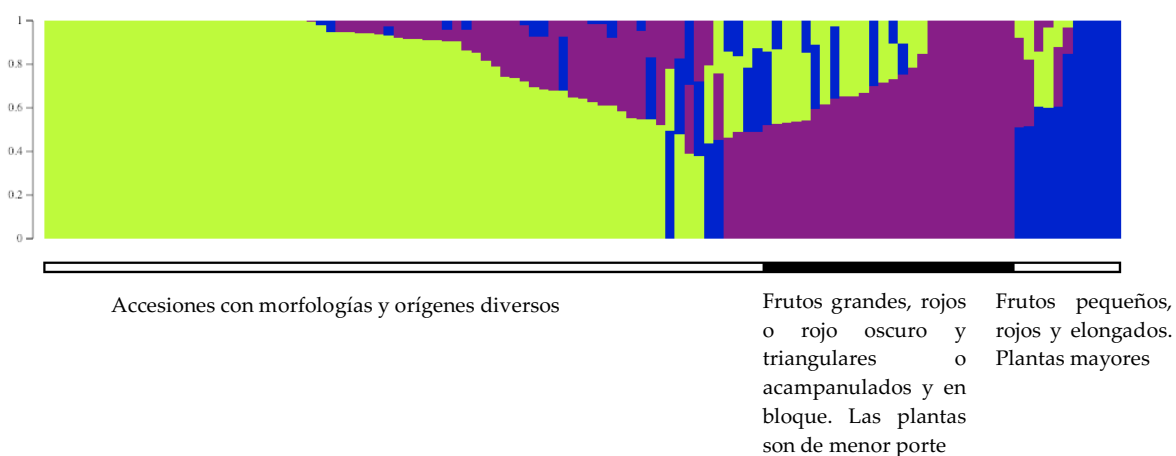


Figura 3.2. Estructura poblacional inferida con el programa ADMIXTURE para el conjunto de accesiones de *C. annuum* analizadas en el GWAS para $K=3$ clústeres. Cada línea vertical representa una accesión y cada color representa el coeficiente de pertenencia a cada clúster.

GWAS

El número de marcadores independientes se calculó mediante el método *simpleM*. Este método estima el número efectivo de tests independientes entre datos altamente correlacionados, como los SNPs, sujetos frecuentemente a un fuerte desequilibrio de ligamiento, y aplica una corrección de Bonferroni basada en esa estimación (Gao et al, 2008). El número de marcadores independientes resultó ser 1184 y el p -valor para

considerar una asociación significativa se fijó en $4,2 \times 10^{-5}$ ($p=0,05/1184$). El GWAS reveló asociaciones significativas con alguno de los tres métodos (GLM+Q, MLM+Q+K o MLM+K) entre 157 SNPs diferentes y 40 parámetros. Tras examinar los gráficos Q-Q se mantuvieron tan solo las asociaciones con buen ajuste del modelo (Figura A.4). Así se encontraron los SNPs *S1_157644336*, *S8_130837307* y *S2_147083777* significativamente asociados a la longitud del tallo en los modelos MLM+Q+K (Figura 3.3A) y MLM+K (Figura 3.3B); los SNPs *S11_4165433*, *S1_201025316*, *S1_19470319* y *S4_210105544* asociados significativamente al verde promedio para los modelos MLM+Q+K (Figura 3.3C) y MLM+K (Figura 3.3D); el SNP *S11_4165433* con asociación también al parámetro media del valor L, obtenido con el Tomato Analyzer, tanto con el modelo MLM+Q+K (Figura 3.3E) como con el MLM+K (Figura 3.3F) y los SNPs *S9_213716963* y *S2_147083777* asociados al área de la hendidura proximal según el modelo GLM+Q (Figura 3.3G).

Para la variable longitud de tallo se observó un SNP, el *S2_138891298*, que se acercó al umbral de significación pese a no llegar a él, con p -valor= $7,25 \times 10^{-5}$ en el modelo MLM+Q+K (Figura 3.3A) y p -valor= $7,76 \times 10^{-5}$ en el modelo MLM+K (Figura 3.3B). Dada su proximidad con el SNP significativo *S2_147083777* se estudió el posible desequilibrio de ligamiento entre ambos con el programa Tassel y se encontró que ambos SNPs no se hallaban en desequilibrio de ligamiento ($R^2=0,005$; $D'=0,455$) por lo que se descartó su posible efecto en el carácter.

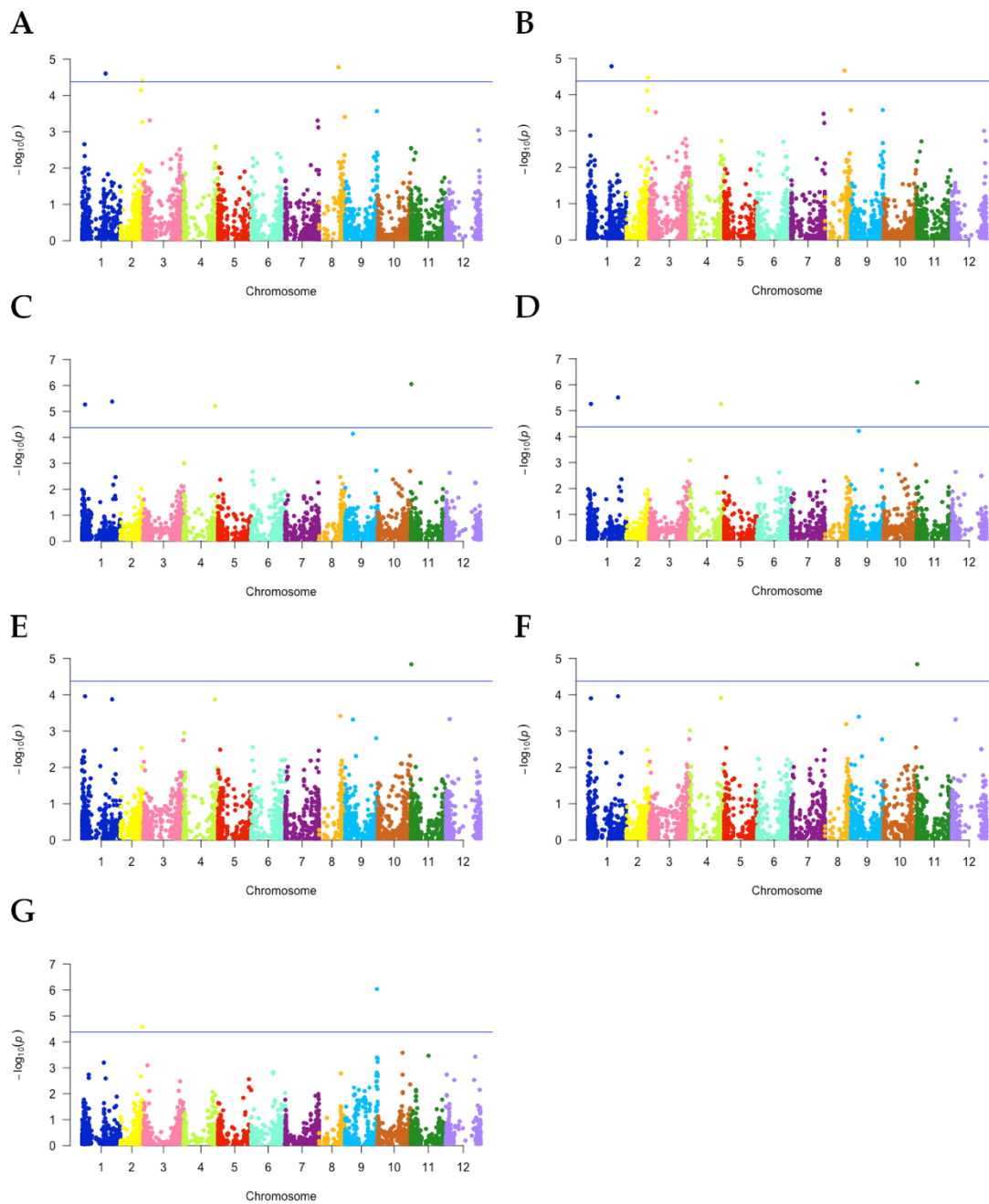


Figura 3.3. Gráficos Manhattan para los parámetros con asociaciones significativas. A. Gráfico para el parámetro longitud del tallo según el modelo MLM+Q+K. B. Gráfico para el parámetro longitud del tallo según el modelo MLM+K. C. Gráfico para el parámetro verde promedio según el modelo MLM+Q+K. D. Gráfico para el parámetro verde promedio según el modelo MLM+K. E. Gráfico para el parámetro media del valor L según el modelo MLM+Q+K. F. Gráfico para el parámetro media del valor L según el modelo MLM+K. G. Gráfico para el parámetro área de hendidura proximal según el modelo GLM+Q. En todos casos, la línea horizontal azul indica el valor umbral para considerar una asociación significativa.

Marcadores con asociación

S1_157644336

El SNP *S1_157644336* del cromosoma 1 mostró asociación significativa con la longitud del tallo con los modelos MLM+Q+K (p-valor= $2,48 \times 10^{-5}$) y MLM+K (p-valor= $1,64 \times 10^{-5}$). El BLASTn de las lecturas que contienen el SNP contra el genoma anotado de *C. annuum* reveló que dicho SNP se localiza en el locus *CA01g22700*, correspondiente al gen codificante de una presunta poligalacturonasa, con una identidad del 98,56% (Tabla 3.4), con similitud a la codificada por el locus *At1g48100* de *Arabidopsis thaliana*. El SNP se corresponde con una transición A/G. Las accesiones que presentan el alelo A tienen una longitud del tallo menor (promedio de 27,3 cm) que las que presentan G (45,8 cm).

S8_130837307

El marcador *S8_130837307* también mostró asociación significativa con la longitud del tallo con MLM+Q+K (p-valor= $1,66 \times 10^{-5}$) y MLM+K (p-valor= $2,16 \times 10^{-5}$). Ochenta y ocho accesiones presentaron el alelo A mientras que 26 mostraron el alternativo T y dos de ellas A/T. Ninguno de los *blast* realizados con las secuencias que contienen el SNP mostró identidad con ninguna secuencia anotada de *C. annuum* (Tabla 3.4).

S2_147083777

El SNP S2_147083777 fue el tercero que mostró asociación significativa con la longitud de tallo con ambos modelos, MLM+Q+K ($p\text{-valor}=3,97\times 10^{-5}$) y MLM+K ($p\text{-valor}=3,38\times 10^{-5}$); y con el área de la hendidura proximal con el modelo GLM+Q ($p\text{-valor}=2,67\times 10^{-5}$). Puede presentar el alelo G o el alternativo A, que se encontró en 20 de las muestras estudiadas. El *blast* realizado a través de la base de datos Solgenomics contra el genoma anotado de pimiento mostró una identidad del 97,1% y un *E-value* de 8×10^{-27} con el locus CA02g26060 de *C. annuum*, que codificaría para una proteína de función desconocida (*detected protein of unknown function*) (Tabla 3.4). El segundo *blast*, realizado en la web del NCBI mostró el mismo resultado.

S11_4165433

El SNP S11_4165433 se corresponde con una transición C→T y mostró asociación significativa con dos parámetros obtenidos con el Tomato Analyzer, el verde promedio y la media del valor L (Tabla 3.4). Con el verde promedio según el modelo MLM+Q+K presentó un $p\text{-valor}=8,84\times 10^{-7}$ y con MLM+K $p\text{-valor}=2,48\times 10^{-5}$. Por su parte, para la media del valor L con MLM+Q+K $p\text{-valor}=1,45\times 10^{-5}$ y con MLM+K $p\text{-valor}=8,08\times 10^{-7}$. No se encontró correspondencia entre las secuencias que lo contienen y las anotaciones del genoma de *C. annuum* con ninguno de los dos *blast* empleados.

S1_201025316

El SNP localizado en la posición 201025316 del cromosoma 1 mostró un *p*-valor por encima del umbral significativo en el parámetro verde promedio, tanto con el modelo MLM+Q+K (*p*-valor= $8,16 \times 10^{-6}$) como con MLM+K (*p*-valor= $3,09 \times 10^{-6}$). Puede presentar el alelo T o el alternativo A y el blast realizado con las secuencias que lo contienen mostró con una identidad del 90,91% que se corresponden con el gen *CA01g27080*, que codifica para una proteína ECERIFERUM, similar a la ECERIFERUM-1 de *Solanum tuberosum* (Tabla 3.4). Siete accesiones de las estudiadas mostraron el alelo alternativo. Estas muestras presentaron un valor medio para el parámetro mayor que el de las accesiones con el alelo de referencia, 112,97 frente a 83,12.

S1_19470319

Con el promedio del valor verde obtenido con el programa Tomato Analyzer, también mostró asociación significativa el SNP *S1_19470319* tanto con el modelo MLM+Q+K (*p*-valor= $5,39 \times 10^{-6}$) como con el MLM+K (*p*-valor= $5,48 \times 10^{-6}$). Tras realizar los *blast* con las secuencias que lo contiene, tanto con Solgenomics como con BlastN del NCBI, no se encontró ninguna coincidencia con la secuencia del genoma anotado de *C. annuum*.

S4_210105544

El cuarto SNP asociado significativamente al promedio del valor verde fue el *S4_210105544* tanto mediante MLM+Q+K (*p*-valor= $6,17 \times 10^{-6}$) como MLM+K (*p*-

valor= $5,54 \times 10^{-6}$). Al realizar el *blast* con las lecturas que lo contienen contra el genoma anotado se encontró identidad con ocho regiones del genoma anotado de *C. annuum*, todas ellas correspondientes a genes que codifican glucosiltransferasas. (Tabla 3.4).

S9_213716963

El SNP S9_213716963 mostró asociación significativa con el área de la hendidura proximal con el modelo GLM+Q (p -valor= $9,21 \times 10^{-7}$), parámetro obtenido con el Tomato Analyzer que informa de la forma del fruto en su extremo proximal o superior, es decir, la zona de inserción del pedicelo. Tras realizar el *blast* no se encontró similitud con ninguna secuencia del genoma de *C. annuum* CM334 pero sí con una región del cromosoma 9 del genoma de *C. annuum* *Zunla-1* (Qin et al., 2014) codificante para una proteína alfa soluble de unión a NFS, con una identidad del 90% (Tabla 3.4).

Tabla 3.4. SNPs que mostraron asociaciones significativas con parámetros fenotípicos. El nombre del SNP identifica el cromosoma y la posición en la que se encuentra. Para cada uno se indica alelo de referencia/alternativo de la variante, el parámetro al que se asocia, el *locus* con anotación con el que presenta similitud, el porcentaje de identidad y *E-value* para el *blast* y la proteína codificada.

SNP	Variante	Parámetro	Correspondencia con la anotación	% Identidad	<i>E-value</i>	Proteína
S1_157644336	A/G	Longitud del tallo	CA01g22700	98,56%	2×10 ⁻²⁸	Poligalacturonasa At1g48100-like
S8_130837307	A/T	Longitud del tallo	-	-	-	-
S2_147083777	G/A	Longitud del tallo y Área de la hendidura proximal	CA02g26060	97,1%	8×10 ⁻²⁷	Desconocida
S11_4165433	C/T	Verde promedio y Media del valor L	-	-	-	-
S1_201025316	T/A	Verde promedio	CA01g27080	90,91%	2×10 ⁻²⁴	ECERIFERUM ECERIFERUM-1-like (<i>S. tuberosum</i>)
S1_19470319	T/G	Verde promedio	-	-	-	-
S4_210105544	T/C	Verde promedio	CA00g36460	97,14 %	2×10 ⁻²⁷	Glucosiltransferasa
			CA04g15380	90,91 %	2×10 ⁻¹⁸	Glucosiltransferasa
			CA07g06320	89,39 %	8×10 ⁻¹⁷	Glucosiltransferasa
			CA07g06310	89,39 %	8×10 ⁻¹⁷	Glucosiltransferasa
			CA07g06300	89,39 %	8×10 ⁻¹⁷	Glucosiltransferasa
			CA04g15390	87,69 %	1×10 ⁻¹⁴	Glucosiltransferasa
			CA07g06280	86,57 %	5×10 ⁻¹⁴	Glucosiltransferasa
CA07g06270	85,71 %	8×10 ⁻¹²	Glucosiltransferasa			
S9_213716963	A/G	Área de la hendidura proximal	LOC107841155	90,0%	3×10 ⁻²⁵	Proteína alfa soluble de unión a NFS

Análisis de asociación en el complejo *C. annuum*

Identificación de SNPs

A partir del archivo VCF generado inicialmente tras el SNV *calling* se seleccionaron las 154 accesiones pertenecientes a las especies del complejo *C. annuum* (*C. annuum*, *C. chinense*, *C. frutescens* y *C. annuum* var. *glabriusculum*) y se creó un nuevo fichero VCF, que se filtró para eliminar los *indels* y las variantes no asignadas a cromosomas, así como los SNPs que aparecen como dato perdido en más del 30% de las muestras y las accesiones con más del 30% de datos perdidos, obteniendo un nuevo fichero con 149 accesiones y 68.565 SNPs. Sobre este fichero se realizó la imputación de genotipos faltantes y se filtró de nuevo aplicando los criterios de $MAF \leq 0,05$, heterocigosidad observada $\geq 0,1$ y heterocigosidad observada por accesión $\geq 0,4$. Se obtuvo así un fichero con 18.532 SNPs y 149 accesiones de las especies del complejo *C. annuum*.

Estructura poblacional

Los valores de *cross-validation* obtenidos con ADMIXTURE revelaron una estructura genética con mayor sentido para las accesiones del complejo *C. annuum* de $K=7$ grupos (Figura A.3B). Se encontró un grupo conformado por las accesiones de *C. chinense*, otro con las de *C. frutescens* y cinco para las de *C. annuum* (Figura 3.4). Además, 28 accesiones de *C. annuum* no se pudieron asignar a ningún clúster al tener coeficientes de pertenencia inferiores al 50% para todos ellos, formando un grupo de pertenencias mixto. El clúster *C. annuum* I engloba accesiones con orígenes diversos (región andina, Península Ibérica

y región mediterránea) y morfologías también dispares. En el clúster *C. annuum* II aparecen accesiones con frutos de color rojo y con forma elongada o triangular (1 o 3, respectivamente, según IPGRI, 1995). En *C. annuum* III se engloban accesiones con origen en su mayoría en el este de Europa (principalmente Bulgaria, Turquía y Hungría) y con frutos también predominantemente de color rojo y forma triangular o elongada. El clúster *C. annuum* IV está formado principalmente por accesiones de España y Portugal y que tienen frutos de color rojo y forma elongada, triangular o acampanulada y en bloque (1, 3 y 5 respectivamente en IPGRI, 1995). Asimismo, las medidas y fenotipado de planta resultaron ser muy similares en los tres grupos. Por lo tanto, las accesiones de estos tres clústeres (II, III y IV) presentarían gran similitud en la morfología y la diferencia entre ellas vendría dada por su origen geográfico. El último clúster con accesiones de *C. annuum* es *C. annuum* V y en él, cuatro de las seis accesiones presentes tienen origen en México siendo una de ellas *C. annuum* var. *glabriusculum*. Además, en este clúster todos los frutos son rojos y elongados, pero de pequeño tamaño (peso medio de 0,6 g).

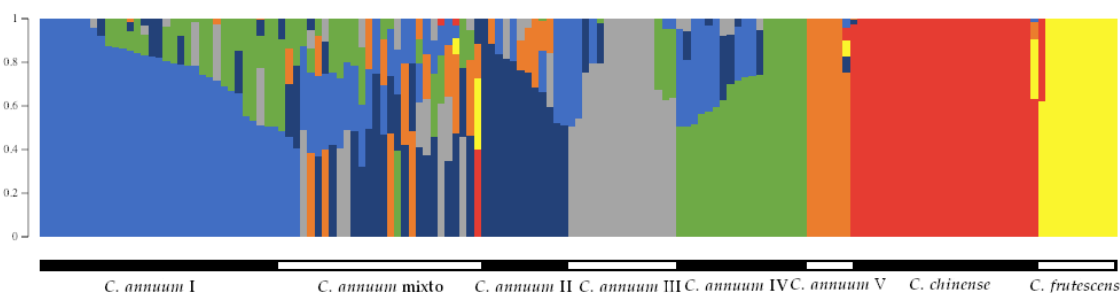


Figura 3.4. Estructura poblacional inferida con el programa ADMIXTURE para el conjunto de accesiones del complejo *C. annuum* analizadas en el GWAS para $K=7$ clústeres. Cada línea vertical representa una accesión y cada color representa el coeficiente de pertenencia a cada clúster.

GWAS

Tras estudiar todas las variables con los modelos GLM+Q, MLM+Q+K y MLM+K se obtuvieron los gráficos Q-Q. En ninguno de los gráficos obtenidos se observó el ajuste ideal del p-valor observado con el p-valor esperado, obteniendo únicamente gráficos en los que el modelo no se ajustaba bien, con desvíos tanto hacia arriba como hacia abajo de los p-valores observados. Estos resultados indican que el modelo presenta sobrerrepresentación de falsos positivos y falsos negativos (se muestran ejemplos de ambas situaciones en las Figuras 3.5A y 3.5B, respectivamente). Al no poder asegurar que los resultados obtenidos fueran asociaciones reales, se detuvo el análisis para el complejo *C. annuum* en este punto.

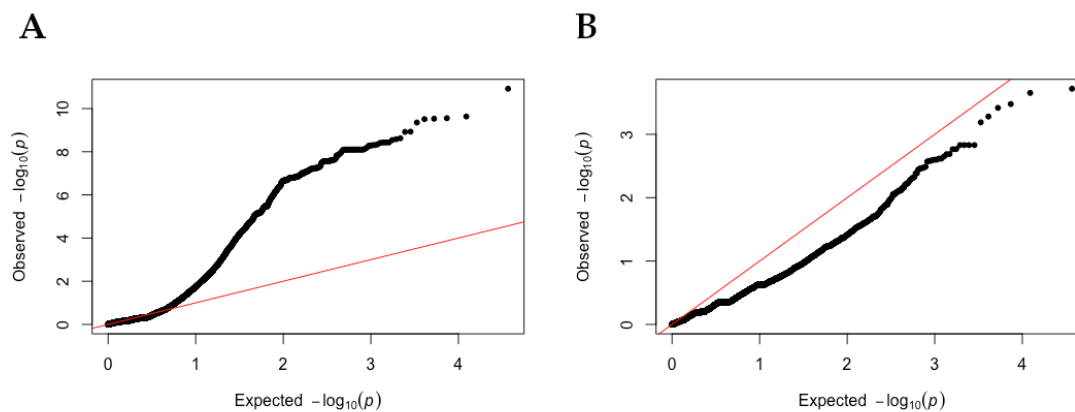


Figura 3.5. Ejemplos de gráficos Q-Q que muestran modelos mal ajustados para el complejo *C. annuum*. A. Gráfico Q-Q para el parámetro verde promedio con el modelo GLM+Q, inflado hacia arriba y por tanto indicando que los resultados están sesgados por falsos positivos. B. Gráfico Q-Q para el parámetro bloqueo proximal del fruto con el modelo MLM+K, abombado hacia abajo, indicando presencia de falsos negativos. En todos los casos, la línea roja indica el $-\log_{10}(\text{p-valor})$ esperado y los puntos negros el $-\log_{10}(\text{p-valor})$ observado para cada SNP.

Discusión

Variabilidad morfológica

La descripción de la diversidad fenotípica alcanzada en los centros de diversificación primarios y secundarios permitirá conocer cómo fue la evolución del pimiento y realizar una gestión más eficaz de los recursos fitogenéticos. Los resultados del presente estudio revelaron gran diversidad para los caracteres analizados en accesiones de las cinco especies cultivadas en la actualidad y que proceden de distintas áreas geográficas. Aún así, los resultados presentados deben ser tomados con cautela ya que la mayoría de líneas estudiadas son de la especie *C. annuum*.

Es destacable la presencia antocianina en *C. pubescens* y alguna accesión de *C. annuum* ya que se ha visto que dicho pigmento está asociado a resistencia a patógenos y tolerancia a distintos estreses (Gould et al., 2010), por lo que puede ser un fenotipo de interés a localizar cuando se quieren encontrar variedades con resistencia y tolerancia.

En otras especies de solanáceas se ha visto que la pubescencia tendría implicación defensiva frente a herbívoros e insectos (Simmons & Gurr, 2005), por lo que las accesiones de *C. pubescens* y de *C. annuum* de México que presentaban valores elevados de pubescencia del tallo deberían tenerse en cuenta en la búsqueda de variedades resistentes.

Los frutos más variables y de mayor tamaño se encontraron en *C. annuum* pero también fue la especie de la que más accesiones se incluyeron. El aumento de tamaño y la variedad morfológica serían resultado de la selección efectuada en las variedades de las especies cultivadas a lo largo de los años (Paran & Van Der Knaap, 2007). La gran variación encontrada en la forma y color de fruto es una interesante fuente de información, ya que se trata de los principales parámetros a seleccionar en programas de mejora vegetal (Tripodi & Greco, 2018). Las accesiones de *C. annuum* var. *glabriusculum* presentaron frutos de pequeño tamaño, de forma alargada y color rojo. Se sabe que este sería el pariente silvestre de *C. annuum* (Bosland & Votava, 2000) y que el proceso de domesticación para dar lugar a *C. annuum* habría tenido lugar en México (Eshbaugh, 1993), lo que explicaría la gran similitud encontrada entre los frutos de México y los de *C. annuum* var. *glabriusculum*, ya que aunque en los pimientos mexicanos se han observado varios tamaños, predominan los pequeños y todos son rojos y de forma triangular. Sería posteriormente en los centros de diversificación secundarios donde se habría producido la mayor variación en formas y colores de frutos (Nuez et al., 1996; Djian-Caporalino et al., 2007;), hecho plasmado en nuestros resultados a través de la amplia gama fenotípica que muestran las accesiones europeas de *C. annuum*.

Análisis de asociación en *C. annuum*

Tras las sucesivas etapas de filtrado se mantuvo un conjunto de datos final con 111 accesiones de *C. annuum* y 4.787 SNPs. El número de SNPs empleado es similar al de otros estudios de asociación para *C. annuum* (Nimmakayala et al., 2016a) mientras que el número de accesiones empleadas es mayor en nuestro caso. Para generar estos datos

se realizó además la imputación de genotipos faltantes, una etapa fundamental en el GWAS ya que, con datos procedentes de secuenciación masiva, la cobertura desigual en las distintas muestras daría lugar a que para muchos SNPs no presenten datos en todas las accesiones, reduciendo la robustez estadística en la búsqueda de asociaciones (Burghardt et al., 2017). Con la imputación, se reconstruyen los datos faltantes a partir de un panel con muestras sí genotipadas (Browning et al., 2018), obteniendo así un conjunto de datos más completo.

Previamente al GWAS se calculó la estructura genética del conjunto final de accesiones incluidos ya que la estructura poblacional puede producir desviaciones entre los p-valores observados y esperados (Vilhjálmsón & Nordborg, 2012) e incrementar el número de resultados con falsos positivos en estudios de asociación (Kang et al., 2008). El primer motivo por el que la estructura poblacional influye en la búsqueda de asociaciones es que si un fenotipo varía de manera conjunta con la estructura poblacional, muchos SNPs no asociados también van a covariar con el fenotipo simplemente debido a la relación entre accesiones (Burghardt et al., 2017). Otra razón es que accesiones genéticamente próximas también tendrían fenotipos similares porque crecen en ambientes parecidos (Burghardt et al., 2017), sin embargo, este hecho es más fácil de controlar si las plantas de todas las variedades se hacen crecer en el mismo ambiente. Por dichos motivos, la estructura genética puede ser un condicionante en estudios GWAS y el posible efecto de la estructura poblacional debe incluirse como covariante si es posible.

Marcadores con asociación*S1_157644336*

El SNP *S1_157644336* mostró asociación significativa con la longitud de tallo y el *blast* realizado contra el genoma de referencia mostró que dicho SNP se localiza en la región génica perteneciente a una presunta poligalacturonasa similar a la codificada por el *locus At1g48100* en *Arabidopsis thaliana*.

Las poligalacturonasas son enzimas hidrolíticas implicadas en el metabolismo de los glúcidos y en la organización de la pared celular a través de la regulación de las pectinas (González-Carranza et al., 2007; Khan et al., 2019). Se han identificado numerosos genes que codifican para presuntas poligalacturonasas en diferentes especies como *Arabidopsis*, melón y arroz (Hadfield et al., 1998; Kim et al., 2006; González-Carranza et al., 2007). Dada su función, las poligalacturonasas están implicadas en diversos procesos fisiológicos del desarrollo de plantas, como la separación de las micrósporas durante el desarrollo del polen (Rhee et al., 2003), la abscisión del órgano floral (González-Carranza et al., 2007), la abscisión del fruto (Peng et al., 2013) y la morfología de la hoja (Atkinson et al., 2002). Ogawa et al. (2009) también comprobaron la implicación de las poligalacturonasas en el crecimiento de la planta, ya que el incremento de la expresión de genes que codifican estas proteínas tendría como resultado una hidrólisis mayor de los residuos galacturonanos de la pectina, inhibiendo el crecimiento celular.

Los descriptores para *Capsicum* (IPGRI, 1995) definen la longitud del tallo como la altura en centímetros hasta la primera bifurcación, inmediatamente después de la primera cosecha, y se ha visto que este es un parámetro importante para definir la arquitectura de la planta (Solomon et al., 2019). Nuestros resultados muestran que el SNP *S1_157644336* se corresponde con una transición A→G. En nuestro estudio, las accesiones con el alelo A tienen una longitud del tallo menor que las que presentan el alelo G. Nuestra hipótesis es que las variedades modernas habrían sido seleccionadas tras evolucionar hasta conseguir plantas con mayor longitud del tallo, lo que les conferiría un mayor porte, teniendo en cuenta que: (i) el alelo A está presente en el mayor número de accesiones estudiadas; (ii) que casi todas las catalogadas como silvestres en los bancos de germoplasma lo tienen; y (iii) que las que presentan el alelo alternativo son principalmente originarias de países de la cuenca mediterránea (Bulgaria, Eslovenia, Albania y Serbia/Montenegro), centro de diversificación secundario de la especie. Así, el alelo ancestral sería A y el alelo reciente G. Esta evolución vendría favorecida por la transición A→G en el SNP *S1_157644336*, que produciría menor expresión de la poligalacturonasa y por lo tanto menor hidrólisis de pectina, lo que se traduciría en un mayor crecimiento de la planta.

S1_201025316

El SNP *S1_201025316* mostró asociación significativa con el color verde promedio, parámetro calculado al incorporar la aplicación “Color Test” en el programa Tomato Analyzer (Rodríguez et al., 2010). El programa calcula índices de color de los frutos con distintas opciones, entre ellas el espacio RGB, midiendo el valor de los colores rojo, verde

y azul en cada píxel de la imagen de los frutos escaneados y devolviendo la media de todas las medidas por fruto (Rodríguez et al., 2010). Para el verde promedio medido con el Tomato Analyzer, las accesiones que presentaron el alelo alternativo (A) del SNP *S1_201025316* mostraron en media un mayor valor que las que tenían el alelo de referencia (T), lo que indica que sus frutos presentan una coloración más verdosa. Dicho SNP se localiza en una región génica que codificaría para una proteína de tipo ECERIFERUM. En *Arabidopsis thaliana* se ha observado que las proteínas ECERIFERUM están implicadas en la biosíntesis de ceras de la cutícula y en la fertilidad del grano de polen (Aarts et al., 1995; Bourdenx et al., 2011). Plantas mutantes para los genes que codifican para las ECERIFERUM, los genes *Cer*, y que por lo tanto tienen alterada la ruta de biosíntesis de ceras, presentan una coloración más verdosa y más brillante que los silvestres al tener ceras sin alcanos en su estructura, uno de los compuestos bioquímicos esenciales en su conformación (Bianchi et al., 1985; Koornneef et al., 1989; Aarts et al., 1995; Bourdenx et al., 2011). La expresión de los genes *Cer* también se ha observado en los órganos reproductivos, participando en la biosíntesis de ceras en flores y frutos (Bourdenx et al., 2011). De este modo, las proteínas ECERIFERUM también estarían implicadas en la producción de ceras en los frutos, determinando por tanto su color. Así, en el presente estudio se plantea la implicación del SNP *S1_201025316* en la generación de color verde en frutos al estar localizado en una región que codificaría para una proteína ECERIFERUM que determinaría el color verde de manera similar a como lo hacen en *A. thaliana*. La variante alélica A en el SNP produciría o bien una reducción en la expresión de las proteínas ECERIFERUM o bien su ausencia total si forma parte de un gen ortólogo de *Cer*. A su vez, la reducción en proteínas ECERIFERUM conllevaría la biosíntesis de ceras de la cutícula sin alcanos, aumentando la coloración verde de los

frutos. El color es una de las características principales que se tiene en cuenta cuando se seleccionan variedades de pimiento para su consumo (Bianchi & Lo Scalzo, 2018) así como para otras aplicaciones como el uso ornamental (Virga et al., 2020), por lo que el conocimiento de las variaciones genéticas y los mecanismos moleculares que lo determinan es fundamental para seleccionar las variedades de mayor interés o mejorar otras mediante ingeniería genética para adaptarlas a las necesidades del mercado.

S4_21010544

El SNP localizado en la posición 210205544 del cromosoma 4 también se mostró asociado significativamente al color verde promedio medido con el Tomato Analyzer. Los *blast* realizados con las secuencias que contienen el SNP mostraron identidad con ocho regiones génicas que codifican para glucosiltransferasas.

Las glucosiltransferasas son enzimas encargadas de la formación de glucósidos mediante la transferencia de azúcares activados por nucleótidos-difosfato (generalmente UDP-glucosa) a sustratos aceptores de bajo peso molecular, entre los que se incluyen la mayoría de metabolitos secundarios (terpenoides, sustratos fenólicos, alcaloides) (Vogt & Jones, 2000). Estas enzimas están implicadas en una amplia variedad de procesos tales como el metabolismo de las auxinas (Grubb et al., 2004), la estabilización de fenoles y terpenos volátiles involucrados en el sabor en frutos como el limón o la mandarina (Frydman et al., 2004) o la detoxificación de toxinas bióticas y xenobióticas como herbicidas y pesticidas (Bowles et al., 2005). Otro de los procesos en los que están implicadas es la formación de la antocianina, pigmento responsable de conferir color

rojo, púrpura o azul a las hojas, frutos y flores de las plantas (Hoshino & Tamura, 1999; Vogt & Jones, 2000).

La similitud de las lecturas que contienen este SNP con ocho *loci* correspondientes a glucosiltransferasas probablemente se deba a que las lecturas son cortas (≤ 77 nucleótidos) y cubren solo el algún dominio común a estas enzimas. Por lo tanto, no es posible identificar qué glucosiltransferasa concreta es. Además, el *blast* es un algoritmo que busca similitudes entre las secuencias problema y las secuencias contenidas en la base de datos, indicando cuán diferentes o no son, pero no asegura que se correspondan exactamente unas con otras (Zhang et al., 2000). Por estas razones, podemos afirmar que el SNP *S4_210105544* se encuentra en una región génica que codifica para una glucosiltransferasa relacionada con el grado de verde del fruto a través de su implicación en la formación de antocianina. En cualquier caso, los resultados aquí presentados son variantes génicas candidatas a tener un efecto causal en el fenotipo, pero la constatación real de su efecto requiere de experimentos funcionales independientes (Burghardt et al., 2017). Por otra parte, el resultado de un análisis GWAS no siempre es identificar un factor causal o predictivo para un carácter fenotípico, sino acotar zonas del genoma con contribución al fenotipo estudiado (Korte & Farlow, 2013), como en este caso.

S9_213716963

Los descriptores para *Capsicum* (IPGRI, 1995) establecen cinco categorías para clasificar la forma del fruto en la unión con el pedicelo; agudo, obtuso, truncado, cordado o

lobulado. Pese a que estas categorías son útiles para realizar una clasificación y describir las accesiones, no están adaptadas a realizar medidas cuantitativas precisas en los frutos, por lo que el empleo del software Tomato Analyzer es una buena estrategia para suplir esta ausencia y obtener un fenotipado de los frutos más completo (Brewer et al., 2006; Rodríguez et al., 2010). El programa Tomato Analyzer calcula el parámetro “área de la hendidura proximal del fruto” como la relación del área de la hendidura (definida por los hombros a ambos extremos del fruto) y el área total del fruto multiplicado por 10 (Brewer et al., 2006; Rodríguez et al., 2010), por lo que vendría a dar información de la morfología del fruto, concretamente de la zona del fruto en la que se inserta el pedicelo. Asociado significativamente a este parámetro se encontraron el SNP *S2_14708377*, para el que no se encontró identidad entre las secuencias que lo contienen y el genoma anotado de *C. annuum*; y el *S9_213716963*, que el *blast* mostró que se encuentra en una región génica que codificaría para proteína alfa soluble de unión a NFS.

Las proteínas de tipo alfa soluble de unión a factores sensibles a N-etilmaleimida (NFS) son proteínas de transporte intracelular implicadas en eventos de fusión de vesículas en la membrana plasmática (Südhof & Rothman, 2009; Zhao et al., 2015; Bayless et al., 2019). Este rol fundamental en el tráfico de proteínas y lípidos a través de la membrana hace que estas proteínas sean esenciales en el desarrollo y que participen en multitud de procesos (Sanderfoot et al., 2000). Se ha investigado ampliamente su implicación en la resistencia a patógenos (Cook et al., 2012; Matsye et al., 2012; Pant et al., 2014) y en las respuestas a estrés osmótico, por salinidad o por bajas temperaturas (Mazel et al., 2004; Leshem et al., 2006; Lee et al., 2009).

No se ha encontrado en la bibliografía referencias a la posible implicación de las proteínas alfa solubles de unión a NFS en la forma de fruto. En base a nuestros resultados, las accesiones con el alelo A en el SNP *S9_213716963* tienen un valor de hendidura proximal mayor que las que tienen el alelo alternativo G (0,071 frente a 0,042). Esto parece indicar que la variación en dicho SNP podría tener influencia en el fenotipo a través de la variación en la expresión de la proteína pero no se puede precisar más porque el papel biológico de la proteína es muy amplio. Por lo tanto, se requieren futuros experimentos para validar la función de si esta proteína tiene influencia sobre los frutos.

S8_130837307, S2_147083777, S11_4165433 y S1_19470319

Los SNPs *S8_130837307, S11_4165433 y S1_19470319* mostraron asociación significativa con algún parámetro fenotípico pero al realizar el *blast* entre las secuencias que los contienen y diferentes genomas anotados de *C. annuum* no se encontró ninguna identidad. Un resultado de falta de coincidencia en el *blast* puede deberse a que las secuencias empleadas sean muy cortas, de menos de 20 nucleótidos (Zhang et al., 2000) pero se descarta esta causa porque se seleccionaron secuencias con una longitud mayor. Así, apuntamos a que las secuencias a testar no se encontraban anotadas en los genomas de referencia empleados. En el genoma de pimiento *Zunla-1* se asignaron a cromosomas el 79% de los *scaffolds* secuenciados (Qin et al., 2014) y en el genoma *CM334* el 86% (Kim et al., 2014) por lo que es probable que algunas de las secuencias que contienen los SNPs de interés no se encuentren entre esas y por lo tanto no se cuente con información para ellas. Pese a que la aproximación de realizar *blast* contra distintos genomas puede paliar

el sesgo de secuencias de un genoma u otro, aún no se estaría cubriendo el genoma completo de *C. annuum*.

Las secuencias que contienen el SNP *S2_147083777* mostraron identidad con una región génica que codificaría para una proteína de función desconocida. En muchos casos, los genomas presentan regiones no anotadas pero que, en base a su secuencia, presentan similitudes con otras regiones que codifican para proteínas sí caracterizadas (Shumilin et al., 2012). Pese a que no se encuentren completamente caracterizadas su estudio es muy importante ya que, en muchos casos, proteínas altamente conservadas en distintas especies (y que por lo tanto podrían tener una importante función biológica) presentan homología en la secuencia nucleotídica que las codificaría, de tal modo que se puede inferir su presencia pese a no estar anotada (Galperin & Koonin, 2004).

Análisis de asociación en el complejo *C. annuum*

En el GWAS para las especies del complejo *C. annuum* se llevó a cabo con un conjunto de datos final de 18.532 SNPs y 149 accesiones de las tres especies. El número de SNPs finales es mayor que el obtenido en el caso del análisis con *C. annuum* porque se incluyen más especies. Del mismo modo, en Nimmakayala et al. (2016b) se identificaron 36.621 SNPs en una colección de *C. annuum* y *C. baccatum* pero al estudiarse solo las accesiones de *C. annuum* el número de SNPs se redujo a una quinta parte (Nimmakayala et al., 2016a). El programa ADMIXTURE determinó que en nuestro caso la estructura poblacional está determinada por la especie, el origen geográfico y la morfología del fruto. Pereira-Dias et al., (2019) encontraron una estructura similar para las especies del

complejo *C. annuum* y agrupaciones en base a la morfología del fruto, apareciendo en los mismos clústeres aquellas accesiones de *C. annuum* con frutos rojos, grandes y en bloque. Estos autores también encontraron agrupadas y claramente diferenciadas las accesiones originarias de México con morfologías tipo cayena e incluyéndose con ellas accesiones *C. annuum* var. *glabriusculum*. Se cree que *C. annuum* var. *glabriusculum* es el ancestro de *C. annuum* en México (Moscone et al., 2007; Hayano-Kanashiro et al., 2016), lo que explicaría la proximidad con las accesiones de dicho país debido al flujo génico existente entre formas cultivadas y silvestres (Pereira-Dias et al., 2019). Comparando estos resultados con los obtenidos anteriormente solo para las accesiones de *C. annuum* se puede concluir que los clústeres de agrupación en esta especie vienen definidos en primer lugar por la morfología y de manera secundaria por el origen geográfico, observándose el efecto de éste cuando se aumenta el número de grupos. Las otras dos especies incluidas, *C. frutescens* y *C. chinense* presentarían un origen común (Moscone et al., 2007; Nicolai et al., 2013) que determinaría que aparezcan próximas en muchos estudios y agrupadas en el mismo clúster (Pereira-Dias et al., 2019; Taitano et al., 2019). Sin embargo, cuando se obtienen resultados de subestructuras más precisas (bien aumentando el número de subgrupos o el tamaño muestral), ambas especies se separan claramente, como sucede en nuestro caso en los resultados encontrados en este capítulo con ADMIXTURE, en los resultados mostrados en el Capítulo 2 o en otros trabajos (Colonna et al., 2019).

El GWAS para las 149 accesiones del complejo *C. annuum* se llevó a cabo del mismo modo que el expuesto anteriormente para *C. annuum*. Sin embargo, en este caso ninguna de las variables fenotípicas estudiadas mostró buen ajuste en el gráfico Q-Q con ninguno

de los modelos empleados. En los análisis GWAS, el gráfico Q-Q muestra el $-\log_{10}(p\text{-valor})$ esperado (eje X) frente al $-\log_{10}(p\text{-valor})$ observado de la probabilidad de asociación (eje Y). Si el Q-Q muestra un ajuste perfecto, con una línea recta de la probabilidad observada sobre la esperada, significa que la hipótesis nula es cierta y por lo tanto no existe asociación significativa o polimorfismo causal. Una desviación de esta línea recta indicaría que la hipótesis nula no es cierta y por lo tanto se encontrarían asociaciones. Sin embargo, si el gráfico presenta una línea recta cerca del cruce de los ejes pero después se dirige hacia arriba, estaría indicando que hay falsos positivos, y si se dispersa hacia abajo la conclusión es que existen falsos negativos. La situación deseada es una línea recta con una pequeña cola final desviada hacia arriba, lo que indicaría que tanto los falsos positivos como los falsos negativos están controlados (Kaler et al., 2020). En nuestro estudio los datos no se ajustaron bien en ningún caso, observándose desviaciones hacia arriba y hacia abajo de los p -valores observados. Con modelos claramente inflados por falsos positivos y negativos no se puede asegurar que los resultados sean fiables. Uno de los aspectos que más influye a la hora de generar falsos positivos y falsos negativos en el GWAS es la fuerte estructura poblacional (Myles et al., 2009; Vilhjálmsson & Nordborg, 2012; Korte & Farlow, 2013; Huang & Han, 2014) ya que puede causar fuertes asociaciones entre los alelos a lo largo del genoma (desequilibrio de ligamiento), produciendo un resultado espurio (Rosenberg & Nordborg, 2006; Platt et al., 2010). Para paliar el sesgo producido por la estructura poblacional, esta se debe inferir del conjunto de datos e incluir las estimas de pertenencia como covariable, a través de la matriz Q (Myles et al., 2009). Pese a que en este estudio se probaron distintos modelos de asociación y se incluyó la matriz Q como covariable, la estructura poblacional es tan fuerte (porque se incluyen especies diferentes) que es

imposible corregir los falsos positivos y falsos negativos que crea. Ante esta situación, no se puede afirmar que los resultados obtenidos sean fiables y se detuvo el ensayo.

Epílogo

En el Capítulo 1.1 de esta Tesis se genotiparon 28 *loci* microsatélite en tres colecciones de recursos fitogenéticos de pimiento. La especie con mayor número de alelos y de alelos privados fue *C. annuum* si bien se contaba con muchas más muestras que de las demás especies y todos los marcadores fueron desarrollados en esta especie. Los niveles de riqueza alélica sí fueron similares en todas las especies. En general la heterocigosidad observada fue baja, resultado del proceso de autopolinización en especies cultivadas. *Capsicum chinense* y *C. frutescens* fueron las especies para las que se encontró un mayor nivel de heterocigosidad observada así como un coeficiente de endogamia más bajo, lo que demuestra que estas serían las especies con más diversidad.

El análisis de inferencia bayesiana realizado con STRUCTURE reveló una estructura de la colección completa en dos grupos que se correspondieron con la distinción *C. annuum*-no *C. annuum*. Esta separación fue comúnmente encontrada con anterioridad pero es difícil de explicar, pues sería esperable que se separaran especies más alejadas y aparecieran juntas otras más próximas como por ejemplo las tres del complejo *C. annuum*. Para profundizar en las relaciones dentro de cada clúster se realizaron sucesivos análisis STRUCTURE para cada uno de ellos por separado. Así, se encontró una agrupación en tres grupos para las accesiones de *C. annuum* según la región de origen. El primer clúster agrupaba a las líneas de la región andina, zona de origen del pimiento; el segundo a las de la Península Ibérica, centro de diversificación secundaria y punto de entrada en Europa; y el tercero a las de la cuenca mediterránea, zona de propagación posterior del pimiento. Para las accesiones no-*C. annuum* se observó una agrupación en cinco grupos que se corresponde con la especie. El análisis factorial de correspondencia realizado para todo el conjunto de datos mostró agrupaciones según la

especie a la que pertenecen las accesiones, con la excepción de las especies *C. chinense* y *C. frutescens* que aparecían próximas, demostrando su proximidad genética. Esta metodología da como resultado una agrupación más esperable, en base a la especie, lo que podría indicar un sesgo en los resultados de STRUCTURE. Dicho sesgo estaría causado por un número desigual de muestras en los distintos grupos/especies; por la limitación de STRUCTURE cuando se trabaja con especies sin cruzamiento libre, que presentan una alta endogamia y por lo tanto se apartan del equilibrio Hardy-Weinberg; y por emplear unos marcadores con elevadas frecuencias de alelos nulos, que sobreestiman aún más el nivel de homocigosis y contribuyen por lo tanto a romper el equilibrio Hardy-Weinberg. Pese a que los resultados obtenidos con marcadores microsatélite aportan información genética valiosa sobre los recursos fitogenéticos, los posibles sesgos derivados de la metodología empleada hacen necesario que se contrasten con otros marcadores moleculares. Esto se expuso en el Capítulo 2 empleando SNPs desarrollados mediante *next-generation sequencing*.

En el Capítulo 1.2 las accesiones de la Península Ibérica y la cuenca mediterránea se analizaron con diez marcadores moleculares asociados a genes de resistencia a patógenos del pimiento y un marcador asociado a un gen que codifica para la pungencia. Los resultados proporcionan información sobre resistencia potencial, lo que permitiría identificar líneas provechosas para el cultivo o como donantes de material genético en programas de mejora sin realizar ensayos de infección, más laboriosos y costosos. Tan solo se encontraron accesiones con genotipos resistentes para los marcadores *SCAR_CD*, que informó de resistencia potencial a nematodos del género *Meloidogyne*; *PR-Bs3*, marcador de resistencia ante *X. campestris* y *SCAC568*, que da información de potencial

resistencia al virus del bronceado del tomate. Las bajas frecuencias de genotipos resistentes serían resultado del proceso de selección al que se vieron sometidas las variedades de pimiento en sus centros de diversificación, como son la Península Ibérica y la región mediterránea. Durante años los agricultores seleccionaron variedades adaptadas a las necesidades de cultivo pero muy homogéneas genéticamente, perdiendo mucha diversidad y potencial capacidad de respuesta a patógenos y otros estreses. La resistencia frente a distintos tipos de *Potyvirus* se estudió a través de marcadores ligados a los genes *Pvr1*, *Pvr2* y *Pvr4*. Con respecto al gen *Pvr1*, solamente se encontró el alelo *Pvr1¹* que confiere resistencia a PVY-0, en frecuencias bajas y similares en ambas colecciones. Para el marcador *T200A*, ligado al gen *Pvr2*, que controla la resistencia a PVY-0, PVY-1 y TEV también se encontró resistencia potencial en ambas colecciones de *C. annuum* en frecuencias semejantes. Por su parte, el marcador *CSO* ligado al gen *Pvr4* informó de resistencia potencial a PVY-0, PVY-1, PVY-2 y PepMov encontrándose el genotipo resistente en mayor proporción de líneas mediterráneas que ibéricas. Además, se evaluó el posible carácter pungente de las accesiones a través del marcador *MAP1*, ligado al gen de pungencia *Pun1*. En la zona mediterránea se encontró posible pungencia en mayor porcentaje en la Península Ibérica. La cualidad pungente vendría seleccionada por los intereses en el cultivo de cada zona y podría estar ligada además a otras características agronómicas.

El Capítulo 2 confirma la gran utilidad de las herramientas de secuenciación masiva como el GBS para el estudio de las relaciones genéticas en recursos fitogenéticos del género *Capsicum*. En dicho capítulo, presentamos los resultados de estructura y diversidad genética mediante la identificación de 48.109 SNVs altamente informativos

distribuidos a lo largo de todo el genoma para 234 accesiones procedentes de tres áreas de gran interés: una de la Península Ibérica, centro de diversificación secundario del pimiento al ser la zona de entrada en Europa; otra de la cuenca mediterránea, región a través de la que se distribuyó el pimiento a lo largo de Europa; y otra de la región andina, zona de origen del género. Siguiendo una metodología basada en PCoAs jerárquicos, que permitió obtener resultados en mayor profundidad que utilizando el conjunto de datos al completo, se consiguió comprobar que la estructura genética de *Capsicum* viene definida por la especie de la que se trata, su origen geográfico y caracteres morfológicos. El primer PCoA separó claramente tres de las especies estudiadas (*C. baccatum*, *C. pubescens* y *C. annuum*,) pero las especies *C. frutescens* y *C. chinense* aparecieron unidas. A menor escala sí se consiguieron separar claramente *C. frutescens* y *C. chinense*, indicando que son dos especies diferentes si bien próximas genéticamente. *Capsicum baccatum* y *C. pubescens* mostraron cierta subestructura según el país de origen, si bien en el caso de *C. pubescens* esta subestructura también guarda relación con el grado de pubescencia del tallo, evidenciando la selección de las líneas conforme a datos fenotípicos concretos según la zona de cultivo para esta especie. Para *C. annuum* se encontró una marcada estructura genética en base al origen de las accesiones, huella de su diversificación a lo largo del mundo. En primer lugar se diferenciaron la mayoría de líneas de la región andina de las europeas. A escala europea, aparecieron próximas las accesiones de la Península Ibérica con las de países cercanos y por último se agruparon las accesiones de los demás países europeos, evidenciando además la selección de las accesiones en base a características de interés del cultivo. Además, esta fue la especie que presentó mayor número de alelos medio. Así, este estudio profundiza en las relaciones entre las variedades locales de pimiento e indica qué grupos mantendrían mayor

variabilidad genética, resultados importantes de cara a la selección de accesiones en programas de mejora vegetal.

Por lo tanto en esta Tesis Doctoral se emplean dos tipos de marcadores para evaluar la diversidad y estructura genética de una colección de recursos fitogenéticos. Los resultados presentan un patrón de estructura genética similar pero con marcadas diferencias. Con los marcadores microsatélite el primer resultado que se encuentra es la distinción de las accesiones *C. annuum* de las no-*C. annuum* en STRUCTURE, resultado encontrado previamente por otros autores pero que sin embargo se plantea como espurio. El PCoA generado a partir de los datos de GBS muestra una estructura de las accesiones según la especie de pertenencia, resultado más esperado que el de *C. annuum*-no *C. annuum* y que concuerda con lo observado en el AFC realizado con los microsatélites. Sucesivos AFCs realizados en los subgrupos encontrados no permitieron dilucidar relaciones dentro de ellos pero sí los sucesivos PCoAs jerárquicos empleando GBS. Destaca el nivel de precisión de los resultados de GBS en *C. annuum*, que permitieron inferir el patrón de dispersión de la especie: se observan por una parte las accesiones de la región andina, zona de origen del género y de México, donde se domesticó la especie. Por otra parte se observa la agrupación de las accesiones europeas, que si se estudian por separado se separan por una parte las de la Península Ibérica y países próximos y por otro las de la zona oriental de Europa, que habrían recibido introducciones de distintas vías. Si bien este patrón también se observa en los resultados de STRUCTURE con microsatélites, estos aportan mucha menos información pues solo se observa la separación andina-Península Ibérica-mediterránea.

Los resultados confirman que el empleo de GBS permite obtener resultados más precisos eliminando además sesgos de los marcadores microsatélites, ya que estos marcadores no presentarían la limitación de los alelos nulos. El hecho de que no se encuentre la marcada separación *C. annuum*-no *C. annuum* con estos marcadores, confirma que se trataría de un resultado artificial causado por STRUCTURE cuando se emplean marcadores microsatélite.

En el Capítulo 3 se ha estudiado la variabilidad fenotípica de 191 accesiones del género *Capsicum* a través de 54 parámetros descriptores. Se encontró gran variabilidad en las especies cultivadas, especialmente en aquellos caracteres relacionados con la forma, el tamaño y el color del fruto, relacionados con el proceso de expansión del pimiento desde su domesticación. Los resultados obtenidos le confieren a la colección estudiada una gran importancia como reservorio de características de interés en programas de mejora vegetal y aportan un gran valor al conocimiento de los recursos fitogenéticos. Dentro de la especie *C. annuum* se ha estudiado la asociación entre los datos fenotípicos y los genotípicos obtenidos con GBS. Tras el proceso de filtrado se obtuvieron 4.787 SNPs para los que se estudió la asociación con los caracteres fenotípicos mediante un análisis GWAS empleando diferentes modelos de asociación. Tras seleccionar los SNPs que rindieron asociación significativa mediante el método *simpleM*, se encontraron ocho marcadores asociados a cuatro caracteres. Para el SNP *S1_157644336* se encontró asociación significativa con la longitud del tallo. Una variación A/G en este SNP alteraría la expresión de una proteína poligalacturonasa teniendo como resultado una menor hidrólisis de pectina y, por lo tanto, mayor longitud del tallo. El SNP *S1_201025316* se encontraría en una región génica codificante para proteínas de tipo ECERIFERUM, con

implicación en la biosíntesis de ceras y que por ello tendría implicación en la coloración verde del fruto. El SNP *S4_210105544* también se encontró asociado con el valor promedio del color verde y su implicación vendría a través de la alteración de la expresión de alguna glucosiltransferasa, probablemente afectando a la ruta de biosíntesis de antocianina en la que están involucradas dichas enzimas. Para el SNP *S9_213716963* se encontró asociación significativa con el área de la hendidura proximal. Este SNP se localizaría en una región génica que codifica una proteína alfa soluble de unión a NFS. Este tipo de proteínas participan principalmente en el transporte transmembrana, estando involucradas en multitud de procesos biológicos. Pese a que no se ha podido identificar la vía concreta en la que la variación del SNP se relaciona con la variación de este parámetro del fruto. El resultado es atractivo de cara a continuar investigando sobre la implicación de esta variante y de las proteínas alfa soluble de unión a NFS en la forma del fruto, concretamente en su extremo proximal, y por lo tanto arrojar luz sobre un parámetro de potencial interés a la hora de seleccionar variedades de pimiento en base a su forma. A mayores, otros cuatro SNPs mostraron asociaciones significativas con algún parámetro fenotípico pero las secuencias que los contienen no mostraron identidad con ninguna región anotada del genoma o lo hicieron con una secuencia que codificaría para una proteína de función desconocida.

Por otra parte se realizó el análisis GWAS para las accesiones que pertenecen a las tres especies del complejo *C. annuum*. En este caso, después de realizar los filtrados correspondientes se retuvieron 149 líneas con 18.532 SNPs. En ellas se procedió de igual modo que con las muestras de *C. annuum*, analizando la posible asociación mediante los modelos GLM+Q, MLM+Q+K y MLM+K, pero desafortunadamente ninguno de ellos

permitió establecer asociación con ningún carácter ya que tras explorar los gráficos Q-Q se concluyó que los ajustes de los modelos no eran buenos. A tenor de esto, lo más probable es que los SNPs que aparecían como asociados significativamente fueran en realidad falsos positivos y falsos negativos. La aparición de fuertes desviaciones en los gráficos Q-Q que indican la existencia de falsos positivos y falsos positivos podría deberse al efecto de una fuerte estructura poblacional al incluir distintas especies. Pese a incluir la estructura poblacional como covariante en los modelos empleados (a través de la matriz Q) en ningún caso se consiguió contrarrestar estos resultados.

Conclusiones

Capítulo 1. Análisis de la diversidad genética de recursos fitogenéticos de pimiento (*Capsicum* spp.)

Capítulo 1.1. Estudio de la variabilidad y estructura genética mediante marcadores microsatélite.

- I. Se analizó la variabilidad y estructura genética de 230 accesiones de las cinco especies cultivadas de pimiento empleando 28 *loci* microsatélite. Todas las especies mostraron valores bajos de heterocigosidad como consecuencia del alto grado de endogamia existente en especies cultivadas y una riqueza alélica similar.
- II. La estructura genética para el conjunto de accesiones encontrada con el análisis bayesiano en dos grupos se corresponde con la distinción *C. annuum*-no *C. annuum*, mientras que con el análisis factorial de correspondencia se encontró una agrupación más esperable en base a la especie. Esta diferencia entre ambos resultados apuntaría a un agrupamiento espurio generado por STRUCTURE.
- III. El análisis de la estructura genética de las accesiones de *C. annuum* reveló diferenciación según el origen geográfico en tres clústeres: un clúster conformado por accesiones de la región andina (zona de origen del pimiento), otro por las de la Península Ibérica (centro de diversificación secundario) y el último por las de la región mediterránea (zona de expansión posterior y con pimientos de distintos orígenes).

- IV. Las accesiones del conjunto no-*C. annuum* se agruparon en base a la clasificación taxonómica de las muestras, lo que apoya la división de las líneas de pimiento en función de la especie.
- V. Los grupos encontrados en el segundo nivel jerárquico mostraron desequilibrio de ligamiento y elevadas frecuencias de alelos nulos. Las altas tasas de alelos nulos sobreestiman los coeficientes de endogamia al incrementar falsamente el número de homocigotos.
- VI. La agrupación encontrada para el conjunto de datos en dos clústeres, *C. annuum*-no *C. annuum*, se debería a un sesgo causado por STRUCTURE debido al tamaño desigual de los grupos/especies, al gran defecto de heterocigosis en plantas cultivadas que produce desvíos del equilibrio Hardy-Weinberg, a la elevada frecuencia de alelos nulos que sobreestimaría el nivel de homocigosis y a otras limitaciones derivadas del *cross-species*.

Capítulo 1.2. Evaluación de marcadores moleculares ligados a enfermedades y pungencia.

- VII. La resistencia potencial encontrada con los marcadores de enfermedad fue muy baja, tanto en la Península Ibérica como en la región mediterránea. Este resultado vendría dado por la pérdida de diversidad genética causada por el proceso de selección de variedades en las zonas de cultivo.
- VIII. El análisis del marcador de pungencia MAP1 detectó pungencia potencial con frecuencias muy diferentes en la Península Ibérica y la región

mediterránea, resultado de la selección desigual de características de interés agronómico según el área de cultivo.

Capítulo 2. Análisis de la diversidad, estructura genética y relaciones entre accesiones del género *Capsicum* mediante *genotyping-by-sequencing* (GBS).

- IX. Empleando marcadores de tipo SNP también se detectaron niveles bajos de diversidad genética en todas las especies, al igual que con microsatélites. *Capsicum pubescens* fue la especie que presentó la mayor variabilidad, indicando que es la especie que mantiene un mayor reservorio genético.
- X. El análisis de coordenadas principales agrupó a las accesiones en base a la especie, confirmando que las especies *C. frutescens* y *C. chinense* si bien están muy próximas genéticamente, sí se corresponden realmente con especies diferentes.
- XI. A nivel intraespecífico, las accesiones de *C. baccatum* y las de *C. pubescens* mostraron agrupaciones claras en base al origen geográfico y a caracteres fenotípicos.
- XII. Dentro de *C. annum* se identificó una distinción en base a la zona de origen. Se encontró una separación Este-Oeste de Europa que guardaría relación con el proceso de expansión del pimiento tras su introducción en el continente, por la Península Ibérica pero también por vía turca de pimientos originarios de la India.

Capítulo 3. Caracterización morfológica en cinco especies del género *Capsicum* y análisis de asociación *genome-wide association study* (GWAS) entre datos fenotípicos y genotípicos en *C. annuum*.

- XIII. El análisis fenotípico de las accesiones mostró una elevada diversidad morfológica en los frutos de pimiento, especialmente dentro de *C. annuum* y *C. chinense*.
- XIV. Mediante un análisis de asociación a nivel del genoma completo entre los datos fenotípicos y los SNPs desarrollados se identificaron nueve asociaciones significativas entre 8 SNPs y 4 parámetros morfológicos relacionados con rasgos de interés y para cuatro SNPs se identificaron posibles mecanismos a través de los que estaría implicado en la determinación del fenotipo.
- XV. El SNP *S1_157644336* asociado significativamente a la longitud del tallo se encuentra en una región génica que codifica para una poligalacturonasa, proteínas implicadas en el crecimiento de la planta al inhibir el crecimiento celular.
- XVI. El SNP *S1_201025316* mostró asociación significativa con el verde promedio. Se encontraría en una región que codifica para una proteína ECERIFERUM, proteínas que determinan el color verde al estar implicadas en la ruta de biosíntesis de ceras.
- XVII. El SNP *S4_210105544* también se encontró asociado al color verde promedio. La secuencia que lo contiene se corresponde con un dominio común a

glucosiltransferasas, enzimas implicadas en la generación de color al participar en la producción de antocianina.

- XVIII. Para el SNP *S9_213716963* se encontró asociación con el área de la hendidura proximal. Se encuentra en una región codificante para una proteína alfa soluble de unión a NFS, un tipo de proteínas implicadas en multitud de procesos biológicos que hace imposible determinar la forma concreta en la que influirían en la variabilidad fenotípica del parámetro.

Referencias

- Aarts, M. G., Keijzer, C. J., Stiekema, W. J., & Pereira, A. (1995). Molecular characterization of the CER1 gene of *Arabidopsis* involved in epicuticular wax biosynthesis and pollen fertility. *Plant Cell*, 7(12), 2115-2127.
- Acquadro, A., Barchi, L., Gramazio, P., Portis, E., Vilanova, S., Comino, C., ... Lanteri, S. (2017). Coding SNPs analysis highlights genetic relationships and evolution pattern in eggplant complexes. *PLoS ONE*, 12(7), e0180774.
- Adetula, O. A. (2006). Genetic diversity of *Capsicum* using random amplified polymorphic DNAs. *African Journal of Biotechnology*, 5, 120-122.
- Agrios, G. N. (2005). *Plant Pathology* (5ª). Ámsterdam, Países Bajos: Academic Press.
- Alexander, D. H., Novembre, J., & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research*, 19(9), 1655-1664.
- Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., & Lander, E. S. (2000). An SNP map of the human genome generated by reduced representation shotgun sequencing. *Nature*, 407(6803), 513-516.
- Andrews, J. (1993). Diffusion of Mesoamerican Food Complex to Southeastern Europe. *Geographical Review*, 83(2), 194-204.
- Andrews, J. (1995). *Peppers: The domesticated Capsicums*. Austin, Estados Unidos: University of Texas Press.
- Andrews, S., Krueger, F., Seaman-Pichon, A., Biggins, F., & Wingett, S. (2010). FastQC: A quality control tool for high throughput sequence data. Recuperado 7 de febrero de 2019, de <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>
- Atkinson, R. G., Schröder, R., Hallett, I. C., Cohen, D., & MacRae, E. A. (2002). Overexpression of polygalacturonase in transgenic apple trees leads to a range of novel phenotypes involving changes in cell adhesion. *Plant Physiology*, 129(1), 122-133.
- Avilla, C., Collar, J. L., Duque, M., Pérez, P., & Fereres, A. (1997). Impact of floating rowcovers on bell pepper yield and virus incidence. *Hortscience*, 32, 882-883.
- Babu, B. S., Pandravada, S. R., Prasada Rao, R. D. V. J., Anitha, K., Chakrabarty, S. K., & Varaprasad, K. S. (2011). Global sources of pepper genetic resources against arthropods, nematodes and pathogens. *Crop Protection*, 30(4), 389-400.

- Baenas, N., Belović, M., Ilic, N., Moreno, D. A., & García-Viguera, C. (2019). Industrial use of pepper (*Capsicum annum* L.) derived products: Technological benefits and biological advantages. *Food Chemistry*, 274, 872-885.
- Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., ... Johnson, E. A. (2008). Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLOS ONE*, 3(10), e3376.
- Balasubramanian, S., Schwartz, C., Singh, A., Warthmann, N., Kim, M. C., Maloof, J. N., ... Weigel, D. (2009). QTL Mapping in New *Arabidopsis thaliana* Advanced Intercross-Recombinant Inbred Lines. *PLOS ONE*, 4(2), e4318.
- Barbará, T., Palma-Silva, C., Paggi, G. M., Bered, F., Fay, M. F., & Lexer, C. (2007). Cross-species transfer of nuclear microsatellite markers: Potential and limitations. *Molecular Ecology*, 16(18), 3759-3767.
- Barboza, G. E., Carrizo García, C., Leiva González, S., Scaldaferrro, M., & Reyes, X. (2019). Four new species of *Capsicum* (Solanaceae) from the tropical Andes and an update on the phylogeny of the genus. *PloS One*, 14(1), e0209792.
- Barboza, G. E., de Bem, L., & Stehmann, J. R. (2020). *Capsicum carassense* (Solanaceae), a new species from the Brazilian Atlantic Forest. *PhytoKeys*, 140, 125-138.
- Bayless, A. M., Zapotocny, R. W., Han, S., Grunwald, D. J., Amundson, K. K., & Bent, A. F. (2019). The *rhg1-a* (*Rhg1* low-copy) nematode resistance source harbors a copia-family retrotransposon within the *Rhg1*-encoded α -SNAP gene. *Plant Direct*, 3(8), e00164.
- Belkhir, K., Borsa, P., Chikni, L., Raufaste, N., & Bonhomme, F. (2004). GENETIX4. 05, logiciel sous Windows TM pour la génétique des populations. *Laboratoire génome, populations, interactions, CNRS UMR*. Université de Montpellier II: Montpellier, France.
- Ben-Chaim, A., Borovsky, Y., Falise, M., Mazourek, M., Kang, B. C., Paran, I., & Jahn, M. (2006). QTL analysis for capsaicinoid content in *Capsicum*. *Theoretical and Applied Genetics*, 113, 1481-1490.
- Ben-Chaim, A., Paran, I., Grube, R. C., Jahn, M., van Wijk, R., & Peleman, J. (2001). QTL mapping of fruit-related traits in pepper (*Capsicum annum*). *Theoretical and Applied Genetics*, 102, 1016-1028.

- Bianchi, A., Bianchi, G., Avato, P., & Salamini, F. (1985). Biosynthetic pathways of epicuticular wax of maize as assessed by mutation, light, plant age and inhibitor studies. *Maydica*, 30, 179-198.
- Bianchi, G., & Lo Scalzo, R. (2018). Characterization of hot pepper spice phytochemicals, taste compounds content and volatile profiles in relation to the drying temperature. *Journal of Food Biochemistry*, 42, e12675.
- Bley, K., Boorman, G., Mohammad, B., McKenzie, D., & Babbar, S. (2012). A comprehensive review of the carcinogenic and anticarcinogenic potential of capsaicin. *Toxicologic Pathology*, 40(6), 847-873.
- Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics*, 30(15), 2114-2120.
- Borevitz, J. O., & Nordborg, M. (2003). The Impact of Genomics on the Study of Natural Variation in *Arabidopsis*. *Plant Physiology*, 132(2), 718-725.
- Bosland, P. W. (1996). *Capsicums: innovative uses of an ancient crop*. En J. Janick (Ed.), *Progress in new crops* (pp. 479-487). Arlington, Estados Unidos: ASHS Press.
- Bosland, P. W., & Votava, E. (2000). *Peppers: vegetable and spice Capsicums* (1.^a ed.). Wallingford, Reino Unido: CABI publishing.
- Bourdenx, B., Bernard, A., Domergue, F., Pascal, S., Léger, A., Roby, D., ... Joubès, J. (2011). Overexpression of *Arabidopsis* ECERIFERUM1 promotes wax very-long-chain alkane biosynthesis and influences plant response to biotic and abiotic stresses. *Plant Physiology*, 156(1), 29-45.
- Bouzar, H., Jones, J. B., Stall, R. E., Hodge, N. C., Minsavage, G. V., Benedict, A. A., & Alvarez, A. M. (1994). Physiological, chemical, serological, and pathogenic analyses of a worldwide collection of *Xanthomonas campestris* pv. *vesicatoria* strains. *Phytopathology*, 84, 663-671.
- Bowles, D., Isayenkova, J., Lim, E. K., & Poppenberger, B. (2005). Glycosyltransferases: Managers of small molecules. *Current Opinion in Plant Biology*, 8(3 SPEC. ISS.), 254-263.
- Box, G. E., & Cox, D. R. (1982). An analysis of transformations revisited, rebutted. *Journal of the American Statistical Association*, 77(377), 209-210.
- Boxer, C. R. (1984). *From Lisbon to Goa 1500-1750. Studies in Portuguese maritime enterprise*.

- Londres, Reino Unido: Variorum Reprints.
- Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y., & Buckler, E. (2007). TASSEL: Software for Association Mapping of Complex Traits in Diverse Samples. *Bioinformatics*, 23(19), 2633-2635.
- Brederson, J.-D., Kym, P. R., & Szallasi, A. (2013). Targeting TRP channels for pain relief. *European Journal of Pharmacology*, 716(1-3), 61-76.
- Brewer, M. T., Lang, L., Fujimura, K., Dujmovic, N., Gray, S., & Van Der Knaap, E. (2006). Development of a controlled vocabulary and software application to analyze fruit shape variation in tomato and other plant species. *Plant Physiology*, 141(1), 15-25.
- Brewer, M. T., Moyseenko, J. B., Monforte, A. J., & van der Knaap, E. (2007). Morphological variation in tomato: a comprehensive study of quantitative trait loci controlling fruit shape and development. *Journal of Experimental Botany*, 58(6), 1339-1349.
- Browning, B. L., Zhou, Y., & Browning, S. R. (2018). A One-Penny Imputed Genome from Next-Generation Reference Panels. *American Journal of Human Genetics*, 103(3), 338-348.
- Browning, S. R., & Browning, B. L. (2007). Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *American Journal of Human Genetics*, 81(5), 1084-1097.
- Burghardt, L. T., Young, N. D., & Tiffin, P. (2017). A Guide to Genome-Wide Association Mapping in Plants. *Current protocols in plant biology*, 2(1), 22-38.
- Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, 8(12), e1002822.
- Cabrera-Toledo, D., Vargas-Ponce, O., Ascencio-Ramírez, S., Valadez-Sandoval, L. M., Pérez-Alquicira, J., Morales-Saavedra, J., & Huerta-Galván, O. F. (2020). Morphological and Genetic Variation in Monocultures, Forestry Systems and Wild Populations of *Agave maximiliana* of Western Mexico: Implications for Its Conservation. *Frontiers in Plant Science*, 11, 817.
- Caranta, C., Thabuis, A., & Palloix, A. (1999). Development of a CAPS marker for the *Pvr4* locus: a tool for pyramiding potyvirus resistance genes in pepper. *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada*,

- 42(6), 1111-1116.
- Carlsson, J. (2008). Effects of microsatellite null alleles on assignment testing. *Journal of Heredity*, 99(6), 616-623.
- Carrizo García, C., Barfuss, M. H. J., Sehr, E. M., Barboza, G. E., Samuel, R., Moscone, E. A., & Ehrendorfer, F. (2016). Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae). *Annals of Botany*, 118(1), 35-51.
- Carvalho, S. I. C., Bianchetti, L. B., Ragassi, C. F., Ribeiro, C. S. C., Reifschneider, F. J. B., Buso, G. S. C., & Faleiro, F. G. (2017). Genetic variability of a Brazilian *Capsicum frutescens* germplasm collection using morphological characteristics and SSR markers. *Genetics and Molecular Research*, 16(3), 1-18.
- Cebolla-Cornejo, J., Soler, S., Gomar, B., Soria, M. D., & Nuez, F. (2003). Screening *Capsicum* germplasm for resistance to tomato spotted wilt virus (TSWV). *Annals of Applied Biology*, 143(2), 143-152.
- Chapa-Oliver, A. M., & Mejía-Teniente, L. (2016). Capsaicin: From plants to a cancer-suppressing agent. *Molecules*, 21(8), 931.
- Chapuis, M. P., & Estoup, A. (2007). Microsatellite null alleles and estimation of population differentiation. *Molecular Biology and Evolution*, 24(3), 621-631.
- Cheng, G. X., Li, R. J., Wang, M., Huang, L. J., Khan, A., Ali, M., & Gong, Z. H. (2018). Variation in leaf color and combine effect of pigments on physiology and resistance to whitefly of pepper (*Capsicum annuum* L.). *Scientia Horticulturae*, 229, 215-225.
- Chile Pepper Institute (CPI). (1997). Introducing Peruvian aji chiles. *Chile Pepper Institute Newsletter*, 6(3), 1-4.
- Chybicki, I. J., & Burczyk, J. (2009). Simultaneous estimation of null alleles and inbreeding coefficients. *Journal of Heredity*, 100(1), 106-113.
- Clark, R., & Lee, S.-H. (2016). Anticancer Properties of Capsaicin Against Human Cancer. *Anticancer Research*, 36(3), 837-843.
- Colonna, V., D'Agostino, N., Garrison, E., Albrechtsen, A., Meisner, J., Facchiano, A., ... Tripodi, P. (2019). Genomic diversity and novel genome-wide association with fruit morphology in *Capsicum*, from 746k polymorphic sites. *Scientific Reports*, 9, 10067.
- Cook, D. E., Lee, T. G., Guo, X., Melito, S., Wang, K., Bayless, A. M., ... Bent, A. F. (2012). Copy number variation of multiple genes at *Rhg1* mediates nematode resistance in

- soybean. *Science*, 338(6111), 1206-1209.
- Craufurd, P. Q., & Wheeler, T. R. (2009). Climate change and the flowering time of annual crops. *Journal of Experimental Botany*, 60(9), 2529-2539.
- Dabrowski, M. J., Bornelöv, S., Kruczyk, M., Baltzer, N., & Komorowski, J. (2015). «True» null allele detection in microsatellite loci: A comparison of methods, assessment of difficulties and survey of possible improvements. *Molecular Ecology Resources*, 15(3), 477-488.
- Dakin, E. E., & Avise, J. C. (2004). Microsatellite null alleles in parentage analysis. *Heredity*, 93(5), 504-509.
- Danecek, P., Auton, A., Abecasis, G., Albers, C. A., Banks, E., DePristo, M. A., ... Durbin, R. (2011). The variant call format and VCFtools. *Bioinformatics*, 27(15), 2156-2158.
- Davey, J. W., & Blaxter, M. L. (2010). RADSeq: next-generation population genetics. *Briefings in Functional Genomics*, 9(5-6), 416-423.
- Davey, J. W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., & Blaxter, M. L. (2013). Special features of RAD Sequencing data: Implications for genotyping. *Molecular Ecology*, 22(11), 3151-3164.
- 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. *Nature Reviews Genetics*, 12(7), 499-510.
- Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1), 1-38.
- DeWitt, D., & Bosland, P. W. (1996). *Peppers of the world. An identification guide*. Berkeley, Estados Unidos: Ten Speed Press.
- Di Dato, F., Parisi, M., Cardi, T., & Tripodi, P. (2015). Genetic diversity and assessment of markers linked to resistance and pungency genes in *Capsicum* germplasm. *Euphytica*, 204(1), 103-119.
- Di Sotto, A., Vecchiato, M., Abete, L., Toniolo, C., Giusti, A., Mannina, L., ... Di Giacomo, S. (2018). *Capsicum annuum* L. var. Cornetto di Pontecorvo PDO: Polyphenolic profile and in vitro biological activities. *Journal of Functional Foods*, 40, 679-691.
- Diversity Arrays Technology Pty Ltd. (2020). Diversity Arrays Technology. Recuperado

- 14 de agosto de 2020, de <https://www.diversityarrays.com/technology-and-resources/dartseq/>
- Djian-Caporalino, C., Fazari, A., Arguel, M. J., Vernie, T., VandeCastele, C., Faure, I., ... Abad, P. (2007). Root-knot nematode (*Meloidogyne* spp.) *Me* resistance genes in pepper (*Capsicum annuum* L.) are clustered on the P9 chromosome. *Theoretical and Applied Genetics*, 114(3), 473-486.
- Djian-Caporalino, C., Lefebvre, V., Sage-Daubèze, A., & Palloix, A. (2007). *Capsicum*. En R. Singh (Ed.), *Genetic resources, chromosome engineering and crop improvement series, Vol 3, Vegetable crops* (3.^a ed., pp. 185-243). Florida, Estados Unidos: CRC Press.
- Doebley, J. F. (1989). Isozymic evidence and the evolution of crop plants. En D. Soltis & P. Soltis (Eds.), *Isozymes in Plant Biology* (pp. 165-191). Portland, Oregon, Estados Unidos: Dioscorides Press.
- Doebley, J. F., Gaut, B. S., & Smith, B. D. (2006). The Molecular Genetics of Crop Domestication. *Cell*, 127, 1309-1321.
- Dogimont, C., Palloix, A., Daubze, A. M., Marchoux, G., Selassie, K. G., & Pochard, E. (1996). Genetic analysis of broad spectrum resistance to potyviruses using doubled haploid lines of pepper (*Capsicum annuum* L.). *Euphytica*, 88, 231-239.
- Doyle, J. J., & Doyle, J. L. (1987). A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin*, 19(1), 11-15.
- Earl, D. A., & vonHoldt, B. M. (2012). STRUCTURE HARVESTER: A website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources*, 4(2), 359-361.
- 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.
- Endelman, J. B., & Jannink, J. L. (2012). Shrinkage estimation of the realized relationship matrix. *G3: Genes, Genomes, Genetics*, 2(11), 1405-1413.
- Eshbaugh, W. H. (1968). A nomenclatural note on the genus *Capsicum*. *Taxon*, 17(1), 51-52.
- Eshbaugh, W. H. (1970). A Biosystematic and Evolutionary Study of *Capsicum baccatum* (Solanaceae). *Brittonia*, 22(1), 31-43.

- Eshbaugh, W. H. (1979). Biosystematic and evolutionary study of the *Capsicum pubescens* complex. En *National Geographic Society Reports* (pp. 143-162). Washington, Estados Unidos: National Geographic Society.
- Eshbaugh, W. H. (1982). Variation and evolution in *Capsicum eximium* Hunz. *Baileya*, 21, 193-198.
- Eshbaugh, W. H. (1993). Peppers: History and Exploitation of a Serendipitous New Crop Discover. En J. Janick & J. Simon (Eds.), *New crops* (pp. 132-139). New York, Estados Unidos: John Wiley & Sons Ltd.
- Esquinas-Alcázar, J. T. (1993). La diversidad genética como material básico para el desarrollo agrícola. En J. I. Cubero & M. T. Moreno (Eds.), *La Agricultura del Siglo XXI* (pp. 79-102). Madrid, España: Mundi-Prensa.
- Eun, M. H., Han, J. H., Yoon, J. B., & Lee, J. (2016). QTL mapping of resistance to the Cucumber mosaic virus P1 strain in pepper using a genotyping-by-sequencing analysis. *Horticulture Environment and Biotechnology*, 57(6), 589-597.
- Evanno, G., Regnaut, S., & Goudet, J. (2005). Detecting the number of clusters of individuals using the software STRUCTURE: A simulation study. *Molecular Ecology*, 14(8), 2611-2620.
- Falush, D., Stephens, M., & Pritchard, J. K. (2007). Inference of population structure using multilocus genotype data: Dominant markers and null alleles. *Molecular Ecology Notes*, 7(4), 574-578.
- FAO. (2001). *Tratado Internacional sobre los Recursos Fitogenéticos para la Alimentación y la Agricultura*.
- FAOSTAT. (2018). FAOSTAT, Base de datos estadísticos corporativos de la Organización de las Naciones Unidas para la Alimentación y la Agricultura. Datos sobre alimentación y agricultura. Recuperado 28 de agosto de 2020, de <http://www.fao.org/faostat/es/#data/QC>
- Frydman, A., Weisshaus, O., Bar-Peled, M., Huhman, D. V., Sumner, L. W., Marin, F. R., ... Eyal, Y. (2004). Citrus fruit bitter flavors: Isolation and functional characterization of the gene *Cm1,2RhaT* encoding a 1,2 rhamnosyltransferase, a key enzyme in the biosynthesis of the bitter flavonoids of citrus. *Plant Journal*, 40(1), 88-100.

- Galano, A., & Martínez, A. (2012). Capsaicin, a tasty free radical scavenger: mechanism of action and kinetics. *The Journal of Physical Chemistry. B*, 116(3), 1200-1208.
- Galperin, M. Y., & Koonin, E. V. (2004). «Conserved hypothetical» proteins: Prioritization of targets for experimental study. *Nucleic Acids Research*, 32(18), 5452-5463.
- Gao, H., Williamson, S., & Bustamante, C. D. (2007). A Markov chain Monte Carlo approach for joint inference of population structure and inbreeding rates from multilocus genotype data. *Genetics*, 176(3), 1635-1651.
- Gao, X., Starmer, J., & Martin, E. R. (2008). A multiple testing correction method for genetic association studies using correlated single nucleotide polymorphisms. *Genetic Epidemiology*, 32(4), 361-369.
- Garrison, E., & Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv preprint*, arXiv:12073907.
- Genesys. (2020). Genesys Global Portal on Plant Genetic Resources. Recuperado 13 de agosto de 2020, de <https://www.genesys-pgr.org/>.
- Gibbs, A. J., Ohshima, K., Phillips, M. J., & Gibbs, M. J. (2008). The prehistory of potyviruses: their initial radiation was during the dawn of agriculture. *PLoS ONE*, 3, e2523.
- Glazier, A. M., Nadeau, J. H., & Aitman, T. J. (2002). Finding genes that underlie complex traits. *Science*, 298, 2345-2349.
- González-Carranza, Z. H., Elliott, K. A., & Roberts, J. A. (2007). Expression of polygalacturonases and evidence to support their role during cell separation processes in *Arabidopsis thaliana*. *Journal of Experimental Botany*, 58(13), 3719-3730.
- González-Pérez, S., Garcés-Claver, A., Mallor, C., Sáenz De Miera, L. E., Fayos, O., Pomar, F., ... Silvar, C. (2014). New insights into *Capsicum* spp relatedness and the diversification process of *Capsicum annuum* in Spain. *PLoS ONE*, 9(12), e116276.
- Gonzalo, M. J., & van der Knaap, E. (2008). A comparative analysis into the genetic bases of morphology in tomato varieties exhibiting elongated fruit shape. *Theoretical and Applied Genetics*, 116(5), 647-656.
- Goudet, J. (2003). FSTAT version 2. 9. 4: a program to estimate and test population genetics parameters. Updated from Goudet (1995).

- Gould, K. S., Dudle, D. A., & Neufeld, H. S. (2010). Why some stems are red: Cauline anthocyanins shield photosystem II against high light stress. *Journal of Experimental Botany*, *61*(10), 2707-2717.
- Govindarajan, V. S., & Salzer, U. J. (1986). *Capsicum* — production, technology, chemistry, and quality — part II. Processed products, standards, world production and trade. *C R C Critical Reviews in Food Science and Nutrition*, *23*(3), 207-288.
- Griffith, O. L., Montgomery, S. B., Bernier, B., Chu, B., Kasaian, K., Aerts, S., ... Consortium, O. R. A. (2008). ORegAnno: an open-access community-driven resource for regulatory annotation. *Nucleic Acids Research*, *36*(Database issue), D107-D113.
- Grubb, C. D., Zipp, B. J., Ludwig-Müller, J., Masuno, M. N., Molinski, T. F., & Abel, S. (2004). *Arabidopsis* glucosyltransferase UGT74B1 functions in glucosinolate biosynthesis and auxin homeostasis. *Plant Journal*, *40*(6), 893-908.
- Grube, R. C., Zhang, Y., Murphy, J. F., Loaiza-Figueroa, F., Lackney, V. K., Provvidenti, R., & Jahn, M. K. (2000). New source of resistance to *Cucumber mosaic virus* in *Capsicum frutescens*. *Plant Disease*, *84*(8), 885-891.
- Guzmán, F. A., Ayala, H., Azurdia, C., Duque, M. C., & de Vicen, M. C. (2005). AFLP assessment of genetic diversity of *Capsicum* genetic resources in Guatemala: Home gardens as an option for conservation. *Crop Science*, *45*, 363-370.
- Hadfield, K. A., Rose, J. K. C., Yaver, D. S., Berka, R. M., & Bennett, A. B. (1998). Polygalacturonase gene expression in ripe melon fruit supports a role for polygalacturonase in ripening-associated pectin disassembly. *Plant Physiology*, *117*(2), 363-373.
- Hammer, K., Arrowsmith, N., & Gladis, T. (2003). Agrobiodiversity with emphasis on plant genetic resources. *Naturwissenschaften*, *90*(6), 241-150.
- Han, K., Lee, H., Ro, N., Hur, O., Lee, J., Kwon, J., & Kang, B. (2018). QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotechnology Journal*, *16*(9), 1546-1558.
- Han, K. P., & Lincoln, D. E. (1994). The evolution of carbon allocation to plant secondary metabolites—a genetic analysis of cost in *Diplacus aurantiacus*. *Evolution*, *48*, 1550-1563.

- Hanacek, P., Vyhnanek, T., Rohrer, M., Cieslarova, J., & Stavelikova, H. (2009). DNA polymorphism in genetic resources of red pepper using microsatellite markers. *Horticultural Science*, 36, 127-132.
- Hardy, G. H. (1908). Mendelian proportions in a mixed population. *Science*, 28, 49-50.
- Hausbeck, M. K., & Lamour, K. H. (2004). *Phytophthora capsici* on vegetable crops: research progress and management challenges. *Plant Disease*, 88(12), 1292-1303.
- Hayano-Kanashiro, C., Gámez-Meza, N., & Medina-Juárez, L. Á. (2016). Wild Pepper *Capsicum annuum* L. var. *glabriusculum*: Taxonomy, plant morphology, distribution, genetic diversity, genome sequencing, and phytochemical compounds. *Crop Science*, 56(1), 1-11.
- He, J., Zhao, X., Laroche, A., Lu, Z., Liu, H., & Li, Z. (2014). Genotyping-by-sequencing (GBS), an ultimate marker-assisted selection (MAS) tool to accelerate plant breeding. *Frontiers in Plant Science*, 5, 484.
- Hirschhorn, J. N., & Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nature Reviews Genetics*, 6(2), 95-108.
- Holm, S. (1979). A Simple Sequentially Rejective Multiple Test Procedure. *Scandinavian Journal of Statistics*, 6(2), 65-70.
- Hornero-Méndez, D., Gómez-Ladrón de Guevara, R., & Mínguez-Mosquera, M. I. (2000). Carotenoid Biosynthesis Changes in Five Red Pepper (*Capsicum annuum* L.) Cultivars during Ripening. Cultivar Selection for Breeding. *Journal of Agricultural and Food Chemistry*, 48(9), 3857-3864.
- Hoshino, T., & Tamura, H. (1999). Anthocyanidin glycosides: color variation and color stability. En R. Ikan (Ed.), *Naturally Occurring Glycosides* (pp. 43-82). John Wiley & Sons.
- Huang, X., & Han, B. (2014). Natural variations and genome-wide association studies in crop plants. *Annual Review of Plant Biology*, 65, 531-551.
- Hulse-Kemp, A. M., Maheshwari, S., Stoffel, K., Hill, T. A., Jaffe, D., Williams, S. R., ... Van Deynze, A. (2018). Reference quality assembly of the 3.5-Gb genome of *Capsicum annuum* from a single linked-read library. *Horticulture Research*, 5, 4.
- Ibiza, V. P., Blanca, J., Cañizares, J., & Nuez, F. (2012). Taxonomy and genetic diversity of domesticated *Capsicum* species in the Andean region. *Genetic Resources and Crop*

- Evolution*, 59(6), 1077-1088.
- Ince, A. G., Karaca, M., & Onus, A. N. (2009). Development and utilization of diagnostic DAMD-PCR markers for *Capsicum* accessions. *Genetic Resources and Crop Evolution*, 56, 211-221.
- Ince, A. G., Karaca, M., & Onus, A. N. (2010). Genetic relationships within and between capsicum species. *Biochemical Genetics*, 48(1-2), 83-95.
- IPGRI. (1995). *Descriptors for Capsicum (Capsicum spp.)*. Roma, Italia: Int. Plant Genetic Resources Inst.
- Iquebal, M. A., Sarika, Arora, V., Verma, N., Rai, A., & Kumar, D. (2013). First whole genome based microsatellite DNA marker database of tomato for mapping and variety identification. *BMC Plant Biology*, 13(1), 197.
- Ittah, Y., Kanner, J., & Granit, R. (1993). Hydrolysis study of carotenoid pigments of paprika (*Capsicum annuum* L. variety Lehava) by HPLC/photodiode array detection. *Journal of Agricultural and Food Chemistry*, 41(6), 899-901.
- Ivanov, K. I., Eskelin, K., Lohmus, A., & Makinen, K. (2014). Molecular and cellular mechanisms underlying potyvirus infection. *Journal of General Virology*, 95, 1415-1429.
- Jakobsson, M., & Rosenberg, N. A. (2007). CLUMPP: A cluster matching and permutation program for dealing with label switching and multimodality in analysis of population structure. *Bioinformatics*, 23(14), 1801-1806.
- Jensen, R. J., McLeod, M. J., Eshbaugh, W. H., & Guttman, S. I. (1979). Numerical Taxonomic Analyses of Allozymic Variation in *Capsicum* (Solanaceae). *Taxon*, 28(4), 315-327.
- Jeong, W. Y., Jin, J. S., Cho, Y. A., Lee, J. H., Park, S., Jeong, S. W., ... Shin, S. C. (2011). Determination of polyphenols in three *Capsicum annuum* L. (bell pepper) varieties using high-performance liquid chromatography-tandem mass spectrometry: their contribution to overall antioxidant and anticancer activity. *Journal of Separation Science*, 34(21), 2967-2974.
- Jones, J., Gheysen, G., & Fenoll, C. (2011). *Genomics and Molecular Genetics of Plant-Nematode Interactions*. Dordrecht, Países Bajos: Springer.
- Jordan, T., Römer, P., Meyer, A., Szczesny, R., Pierre, M., Piffanelli, P., ... Lahaye, T.

- (2006). Physical delimitation of the pepper *Bs3* resistance gene specifying recognition of the AvrBs3 protein from *Xanthomonas campestris* pv. *vesicatoria*. *Theoretical and Applied Genetics*, 113(5), 895-905.
- Kaler, A. S., Gillman, J. D., Beissinger, T., & Purcell, L. C. (2020). Comparing Different Statistical Models and Multiple Testing Corrections for Association Mapping in Soybean and Maize. *Frontiers in Plant Science*, 10, 1794.
- Kalinowski, S. T. (2011). The computer program STRUCTURE does not reliably identify the main genetic clusters within species: Simulations and implications for human population structure. *Heredity*, 106(4), 625-632.
- Kalinowski, S. T., & Taper, M. L. (2006). Maximum likelihood estimation of the frequency of null alleles at microsatellite loci. *Conservation Genetics*, 7(6), 991-995.
- Kang, H., Zaitlen, N. A., Wade, C. M., Kirby, A., Heckerman, D., Daly, M. J., & Eskin, E. (2008). Efficient control of population structure in model organism association mapping. *Genetics*, 178(3), 1709-1723.
- Kang, J.-H., Kim, C.-S., Han, I.-S., Kawada, T., & Yu, R. (2007). Capsaicin, a spicy component of hot peppers, modulates adipokine gene expression and protein release from obese-mouse adipose tissues and isolated adipocytes, and suppresses the inflammatory responses of adipose tissue macrophages. *FEBS Letters*, 581(23), 4389-4396.
- Kang, W. H., Hoang, N. H., Yang, H. B., Kwon, J. K., Jo, S. H., Seo, J. K., ... Kang, B. C. (2010). Molecular mapping and characterization of a single dominant gene controlling CMV resistance in peppers (*Capsicum annuum* L.). *Theoretical and Applied Genetics*, 120(8), 1587-1596.
- Kaushik, P., Prohens, J., Vilanova, S., Gramazio, P., & Plazas, M. (2016). Phenotyping of Eggplant Wild Relatives and Interspecific Hybrids with Conventional and Phenomics Descriptors Provides Insight for Their Potential Utilization in Breeding. *Frontiers in Plant Science*, 7, 677.
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., ... Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.

- Khan, N., Fatima, F., Haider, M. S., Shazadee, H., Liu, Z., Zheng, T., & Fang, J. (2019). Genome-wide identification and expression profiling of the polygalacturonase (PG) and pectin methylesterase (PME) genes in grapevine (*Vitis vinifera* L.). *International Journal of Molecular Sciences*, 20(13), 3180.
- Killian, A., Wenzl, P., Huttner, E., Carling, J., Xia, L., Blois, H., ... Uszynski, G. (2012). Diversity Arrays Technology: A Generic Genome Profiling Technology on Open Platforms. En Pompanon, F., & Bonin, A. (Eds), *Data Production and Analysis in Population Genomics. Methods in Molecular Biology (Methods and Protocols)*, vol 888 (pp. 177-196). Totowa, Nueva Jersey, Estados Unidos: Humana Press.
- Kim, C.-S., Kawada, T., Kim, B.-S., Han, I.-S., Choe, S.-Y., Kurata, T., & Yu, R. (2003). Capsaicin exhibits anti-inflammatory property by inhibiting I κ B- α degradation in LPS-stimulated peritoneal macrophages. *Cellular Signalling*, 15(3), 299-306.
- Kim, J., Shiu, S. H., Thoma, S., Li, W. H., & Patterson, S. E. (2006). Patterns of expansion and expression divergence in the plant polygalacturonase gene family. *Genome Biology*, 7(9), R87.
- Kim, S., Lee, H., Seo, S., Lee, J. H., & Choi, D. (2015). RNA-Dependent RNA Polymerase (NIb) of the Potyviruses Is an Avirulence Factor for the Broad-Spectrum Resistance Gene *Pvr4* in *Capsicum annuum* cv. CM334. *PLoS ONE*, 10(3), e0119639.
- Kim, S., Park, J., & Hwang, I. K. (2004). Composition of Main Carotenoids in Korean Red Pepper (*Capsicum annuum*, L) and Changes of Pigment Stability During the Drying and Storage Process. *Journal of Food Science*, 69(1), 37-44.
- Kim, S., Park, M., Yeom, S. I., Kim, Y. M., Lee, J. M., Lee, H. A., ... Choi, D. (2014). Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nature Genetics*, 46(3), 270-278.
- Kirschbaum-Titze, P., Hiepler, C., Mueller-Seitz, E., & Petz, M. (2002). Pungency in paprika (*Capsicum annuum*). 1. Decrease of capsaicinoid content following cellular disruption. *Journal of Agricultural and Food Chemistry*, 50, 1260-1263.
- Kobata, K., Sutoh, K., Todo, T., Yazawa, S., Iwai, K., & Watanabe, T. (1999). Nordihydrocapsiate, a new capsinoid from the fruits of a nonpungent pepper, *Capsicum annuum*. *Journal of Natural Products*, 62(2), 335-336.
- Konieczny, A., & Ausubel, F. M. (1993). A procedure for mapping *Arabidopsis* mutations

- using codominant ecotype-specific PCR-based markers. *Plant Journal*, 4, 403-410.
- Koornneef, M., Hanhart, C. J., & Thiel, F. (1989). A genetic and phenotypic description of Eceriferum (cer) mutants in *Arabidopsis thaliana*. *Journal of Heredity*, 80(2), 118-122.
- Korte, A., & Farlow, A. (2013). The advantages and limitations of trait analysis with GWAS: A review. *Plant Methods*, 9, 29.
- Kosman, E., & Leonard, K. J. (2005). Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Molecular Ecology*, 14(2), 415-424.
- Kousik, C. S., & Ritchie, D. F. (1996). Disease potential of pepper bacterial spot pathogen races that overcome the *Bs2* gene for resistance. *Phytopathology*, 86(12), 1336-1343.
- Kraft, K. H., Brown, C. H., Nabhan, G. P., Luedeling, E., De Jesús Luna Ruiz, J., D'Eeckenbrugge, G. C., ... Gepts, P. (2014). Multiple lines of evidence for the origin of domesticated chili pepper, *Capsicum annuum*, in Mexico. *Proceedings of the National Academy of Sciences of the United States of America*, 111(17), 6165-6170.
- Krishna De, A. (Ed.). (2003). *Capsicum. The genus Capsicum*. Londres, Reino Unido: CRC Press.
- Kyle, M. M., & Palloix, A. (1997). Proposed revision of nomenclature for potyvirus resistance genes in *Capsicum*. *Euphytica*, 97, 183-188.
- Lateef, D. D. (2015). DNA Marker Technologies in Plants and Applications for Crop Improvements. *Journal of Biosciences and Medicines*, 3, 7-18.
- Lee, D. G., Ahsan, N., Lee, S. H., Lee, J. J., Bahk, J. D., Kang, K. Y., & Lee, B. H. (2009). Chilling stress-induced proteomic changes in rice roots. *Journal of Plant Physiology*, 166(1), 1-11.
- Lee, J. M., Nahm, S. H., Kim, Y. M., & Kim, B. D. (2004). Characterization and molecular genetic mapping of microsatellite loci in pepper. *Theoretical and Applied Genetics*, 108(4), 619-627.
- Leshem, Y., Melamed-Book, N., Cagnac, O., Ronen, G., Nishri, Y., Solomon, M., ... Levine, A. (2006). Suppression of *Arabidopsis* vesicle-SNARE expression inhibited fusion of H₂O₂-containing vesicles with tonoplast and increased salt tolerance. *Proceedings of the National Academy of Sciences of the United States of America*, 103(47), 18008-18013.

- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*, 25(14), 1754-1760.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), 2078-2079.
- Lilliefors, H. W. (1967). On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown. *Journal of the American Statistical Association*, 62, 399-402.
- Liu, F. M., Zhang, N. N., Liu, X. J., Yang, Z. J., Jia, H. Y., & Xu, D. P. (2019). Genetic Diversity and Population structure analysis of *Dalbergia odorifera* Germplasm and Development of a core collection using Microsatellite Markers. *Genes*, 10(4), 281.
- Liu, W. Y., Kang, W., & Kang, B. (2013). Basic Information on Pepper. En B. C. Kang & C. Kole (Eds.), *Genetics, Genomics and Breeding of Peppers and Eggplants* (pp. 1-15). Boca Raton, Florida, Estados Unidos: CRC Press.
- Loaiza-Figueroa, F., Ritland, K., Cancino, J. A. L., & Tanksley, S. D. (1989). Patterns of genetic variation of the genus *Capsicum* (Solanaceae) in Mexico. *Plant Systematics and Evolution*, 165(3-4), 159-188.
- Mammadov, J., Aggarwal, R., Buyyarapu, R., & Kumpatla, S. (2012). SNP Markers and Their Impact on Plant Breeding. *International Journal of Plant Genomics*, 2012, 728398.
- Martín, I. (2015). *Conservación de recursos fitogenéticos*. Madrid, España: Ministerio de Agricultura, Alimentación y Medio Ambiente.
- Martins, T. R., & Barkman, T. J. (2005). Reconstruction of Solanaceae Phylogeny Using the Nuclear Gene SAMT. *Systematic Botany*, 30(2), 435-447.
- Matsye, P. D., Lawrence, G. W., Youssef, R. M., Kim, K. H., Lawrence, K. S., Matthews, B. F., & Klink, V. P. (2012). The expression of a naturally occurring, truncated allele of an α -SNAP gene suppresses plant parasitic nematode infection. *Plant Molecular Biology*, 80(2), 131-155.
- Mazel, A., Leshem, Y., Tiwari, B. S., & Levine, A. (2004). Induction of Salt and Osmotic Stress Tolerance by Overexpression of an Intracellular Vesicle Trafficking Protein AtRab7 (AtRabG3e). *Plant Physiology*, 134(1), 118-128.
- McLeod, M., Eshbaugh, W. H., & Guttman, S. I. (1979). A preliminary biochemical systematic study of the genus *Capsicum*-Solanaceae. En J. Hawkes, R. Lester, & A.

- Skelding (Eds.), *The biology and taxonomy of the Solanaceae* (pp. 701-714). New York: Academic Press.
- McLeod, M., Guttman, S. I., & Eshbaugh, W. H. (1982). Early evolution of chili peppers (*Capsicum*). *Economic Botany*, 36(4), 361-368.
- McLeod, M. J., Guttman, S. I., Eshbaugh, W. H., & Rayle, R. E. (1983). An Electrophoretic Study of Evolution in *Capsicum* (Solanaceae). *Evolution; international journal of organic evolution*, 37(3), 562–574.
- Michelmore, R. W., Paran, I., & Kesseli, R. V. (1991). Identification of markers linked to disease resistance genes by bulked segregant analysis: A rapid method to detect markers in specific regions using segregating populations. *Proceedings of the National Academy of Sciences of the United States of America*, 88, 9828-9832.
- Minamiyama, Y., Tsuru, M., & Hirai, M. (2006). An SSR-based linkage map of *Capsicum annuum*. *Molecular Breeding*, 18(2), 157-169.
- Minguez-Mosquera, M. I., & Hornero-Mendez, D. (1994). Formation and transformation of pigments during the fruit ripening of *Capsicum annuum* Cv. Bola and Agridulce. *Journal of Agricultural and Food Chemistry*, 42(1), 38-44.
- Monma, S., & Sakata, Y. (1997). Screening of *Capsicum* accessions for resistance to cucumber mosaic virus. *Journal of the Japanese Society for Horticultural Science*, 65(4), 769-776.
- Moscone, E. A., Scaldaferrro, M. A., Grabiele, M., Cecchini, N. M., García, Y. S., Jarret, R., ... Ehrendorfer, F. (2007). The evolution of chili peppers (*Capsicum*-Solanaceae): A cytogenetic perspective. *Acta Horticulturae*, 745, 137-169.
- Moury, B., Pflieger, S., Blattes, A., Lefebvre, V., & Palloix, A. (2000). A CAPS marker to assist selection of tomato spotted wilt virus (TSWV) resistance in pepper. *Genome*, 43(1), 137-142.
- Moury, B., & Verdin, E. (2012). Viruses of Pepper Crops in the Mediterranean Basin. A Remarkable Stasis. En G. Loebenstein, & H. Lecoq (Eds.), *Viruses and Virus Diseases of Vegetables in the Mediterranean Basin, Vol 84, Advances in Virus Research* (pp. 127-162). San Diego, California, Estados Unidos: Academic Press.
- Myles, S., Peiffer, J., Brown, P. J., Ersoz, E. S., Zhang, Z., Costich, D. E., & Buckler, E. (2009). Association mapping: Critical considerations shift from genotyping to

- experimental design. *Plant Cell*, 21(8), 2194-2202.
- Naegele, R. P., Mitchell, J., & Hausbeck, M. K. (2016). Genetic diversity, population structure, and heritability of fruit traits in *Capsicum annuum*. *PLoS ONE*, 11(7), e0156969.
- Nagy, I., Stágel, A., Sasvári, Z., Röder, M., & Ganai, M. (2007). Development, characterization, and transferability to other Solanaceae of microsatellite markers in pepper (*Capsicum annuum* L.). *Genome*, 50(7), 668-688.
- Navarro, J., Flores, P., Garrido, C., & Martínez, V. (2006). Changes in the contents of antioxidant compounds in pepper fruits at different ripening stages, as affected by salinity. *Food Chemistry*, 96(1), 66-73.
- Nicolai, M., Cantet, M., Lefebvre, V., Sage-Palloix, A. M., & Palloix, A. (2013). Genotyping a large collection of pepper (*Capsicum* spp.) with SSR loci brings new evidence for the wild origin of cultivated *C. annuum* and the structuring of genetic diversity by human selection of cultivar types. *Genetic Resources and Crop Evolution*, 60(8), 2375-2390.
- Nimmakayala, P., Abburi, V. L., Saminathan, T., Alaparathi, S. B., Almeida, A., Davenport, B., ... Reddy, U. K. (2016a). Genome-wide Diversity and Association Mapping for Capsaicinoids and Fruit Weight in *Capsicum annuum* L. *Scientific Reports*, 6, 38081.
- Nimmakayala, P., Abburi, V. L., Saminathan, T., Almeida, A., Davenport, B., Davidson, J., ... Reddy, U. K. (2016b). Genome-Wide Divergence and Linkage Disequilibrium Analyses for *Capsicum baccatum* Revealed by Genome-Anchored Single Nucleotide Polymorphisms. *Frontiers in Plant Science*, 7, 1646.
- Nuez, F., Gil, R., & Costa, J. (1996). *El cultivo de pimientos, chiles y ajíes*. Madrid, España: Mundi-Prensa.
- Ochi, T., Takaishi, Y., Kogure, K., & Yamauti, I. (2003). Antioxidant activity of a new capsaicin derivative from *Capsicum annuum*. *Journal of Natural Products*, 66, 1094-1096.
- Ogawa, M., Kay, P., Wilson, S., & Swain, S. M. (2009). *Arabidopsis* Dehiscence Zone Polygalacturonase1 (ADPG1), ADPG2, and Quartet2 are polygalacturonases required for cell separation during reproductive development in *Arabidopsis*. *Plant*

- Cell*, 21(1), 216-233.
- Olmstead, R. G., Sweere, J. A., Spangler, R. E., Bohs, L., & Palmer, J. D. (1999). Phylogeny and provisional classification of the Solanaceae based on Chloroplast DNA. En M. Nee, D. E. Symon, R. N. Lester, & J. P. Jessop (Eds.), *Solanaceae IV: Advances in Biology and Utilization* (pp. 111-137). Kew, Reino Unido: Royal Botanic Gardens, Kew Publishing.
- Onkendi, E. M., Kariuki, G. M., Marais, M., & Moleleki, L. N. (2014). The threat of root-knot nematodes (*Meloidogyne* spp.) in Africa: a review. *Plant Pathology*, 63, 727-737.
- Onus, A. N., & Pickersgill, B. (2004). Unilateral incompatibility in *Capsicum* (Solanaceae): Occurrence and taxonomic distribution. *Annals of Botany*, 94(2), 289-295.
- Ortiz, R., Delgado, F., Alvarado, G., & Crossa, J. (2010). Classifying vegetable genetic resources—A case study with domesticated *Capsicum* spp. *Scientia Horticulturae*, 126, 186-191.
- Osorio-Guarín, J. A., Quackenbush, C. R., & Cornejo, O. E. (2018). Ancestry informative alleles captured with reduced representation library sequencing in *Theobroma cacao*. *PloS One*, 13(10), e0203973.
- Pant, S. R., Matsye, P. D., McNeece, B. T., Sharma, K., Krishnavajhala, A., Lawrence, G. W., & Klink, V. P. (2014). Syntaxin 31 functions in *Glycine max* resistance to the plant parasitic nematode *Heterodera glycines*. *Plant Molecular Biology*, 85(1-2), 107-121.
- Parada-Rojas, C. H., & Quesada-Ocampo, L. M. (2018). Analysis of microsatellites from transcriptome sequences of *Phytophthora capsici* and applications for population studies. *Scientific Reports*, 8, 5194.
- Paran, I., & Michelmore, R. W. (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. *Theoretical and Applied Genetics*, 85, 985-993.
- Paran, I., & Van Der Knaap, E. (2007). Genetic and molecular regulation of fruit and plant domestication traits in tomato and pepper. *Journal of Experimental Botany*, 58(14), 3841-3852.
- Parrella, G., Gognalons, P., Gebre-Selassie, K., Vovlas, C., & Marchoux, G. (2003). An update of the host range of tomato spotted wilt virus. *Journal of Plant Pathology*, 8, 227-264.

- Pasam, R., Sharma, R., Malosetti, M., van Eeuwijk, F., Haseneyer, G., Kilian, B., & Graner, A. (2012). Genome-wide association studies for agronomical traits in a world wide spring barley collection. *BMC Plant Biology*, *12*, 16.
- Peakall, R., & Smouse, P. E. (2012). GenALEx 6.5: Genetic analysis in Excel. Population genetic software for teaching and research-an update. *Bioinformatics*, *28*(19), 2537-2539.
- Pemberton, J. M., Slate, J., Bancroft, D. R., & Barret, J. A. (1995). Nonamplifying alleles at microsatellite loci: a caution for parentage and population studies. *Molecular Ecology*, *4*(2), 249-252.
- Peng, G., Wu, J., Lu, W., & Li, J. (2013). A polygalacturonase gene clustered into clade E involved in lychee fruitlet abscission. *Scientia Horticulturae*, *150*, 244-250.
- Pereira, L., Ruggieri, V., Pérez, S., Alexiou, K. G., Fernández, M., Jahrmann, T., ... Garcia-Mas, J. (2018). QTL mapping of melon fruit quality traits using a high-density GBS-based genetic map. *BMC Plant Biology*, *18*, 324.
- Pereira-Dias, L., Vilanova, S., Fita, A., Prohens, J., & Rodríguez-Burruezo, A. (2019). Genetic diversity, population structure, and relationships in a collection of pepper (*Capsicum* spp.) landraces from the Spanish centre of diversity revealed by genotyping-by-sequencing (GBS). *Horticulture Research*, *6*, 54.
- Perry, L., Dickau, R., Zarrillo, S., Holst, I., Pearsall, D. M., Piperno, D. R., ... Zeidler, J. A. (2007). Starch fossils and the domestication and dispersal of chili peppers (*Capsicum* spp. L.) in the Americas. *Science*, *315*(5814), 986-988.
- Perry, R. N., Moens, M., & Starr, J. L. (2009). *Root-knot Nematodes*. Wallingford, Reino Unido: CABI International.
- 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.
- Pickersgill, B. (1971). Relationships Between Weedy and Cultivated Forms in Some Species of Chili Peppers (Genus *Capsicum*). *Evolution*, *25*(4), 683-691.
- Pickersgill, B. (1997). Genetic resources and breeding of *Capsicum* spp. *Euphytica*, *96*, 129-133.
- Pickersgill, B. (2007). Domestication of plants in the Americas: Insights from Mendelian

- and molecular genetics. *Annals of Botany*, 100(5), 925-940.
- Pino, J., Gonzalez, M., Ceballo, L., Centurion-Yah, A. R., Trujillo-Aguirre, J., Latournerie-Moreno, L., & Sauri-Duch, E. (2007). Characterization of total capsaicinoids, colour and volatile compounds of Habanero chilli pepper (*Capsicum chinense* Jack.) cultivars grown in Yucatan. *Food Chemistry*, 104 (4), 1682-1686.
- Pisa, G., Orioli, V., Spilotros, G., Fabbri, E., Randi, E., & Bani, L. (2015). Detecting a hierarchical genetic population structure: The case study of the Fire Salamander (*Salamandra salamandra*) in Northern Italy. *Ecology and Evolution*, 5(3), 743-758.
- Platt, A., Vilhjálmsson, B. J., & Nordborg, M. (2010). Conditions under which genome-wide association studies will be positively misleading. *Genetics*, 186(3), 1045-1052.
- Poland, J., & Rife, T. (2012). Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome*, 5, 92-102.
- Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J.-L. (2012). Development of High-Density Genetic Maps for Barley and Wheat Using a Novel Two-Enzyme Genotyping-by-Sequencing Approach. *PLOS ONE*, 7(2), e32253.
- Potnis, N., Timilsina, S., Strayer, A., Shantharaj, D., Barak, J. D., Paret, M. L., ... Jones, J. B. (2015). Bacterial spot of tomato and pepper: diverse *Xanthomonas* species with a wide variety of virulence factors posing a worldwide challenge. *Molecular Plant Pathology*, 16(9), 907-920.
- Primmer, C. R., Painter, J. N., Koskinen, M. T., Palo, J. U., & Merilä, J. (2005). Factors affecting avian cross-species microsatellite amplification. *Journal of Avian Biology*, 36(4), 348-360.
- Prince, J., Loaiza-Figueroa, F., & Tanksley, D. (1992). Restriction fragment length polymorphism and genetic distance among Mexican accessions of *Capsicum*. *Genome*, 35, 726-732.
- Pritchard, J. K., Stephens, M., & Donnelly, P. (2000). Inference of Population Structure Using Multilocus Genotype Data. *Genetics*, 155, 945-959.
- Purseglove, J. M. (1968). *Tropical crops*. Vol. 2. Nueva York, Estados Unidos: John Wiley & Sons.
- Qin, C., Yu, C., Shen, Y., Fang, X., Chen, L., Min, J., ... Zhang, Z. (2014). Whole-genome sequencing of cultivated and wild peppers provides insights into *Capsicum*

- domestication and specialization. *Proceedings of the National Academy of Sciences of the United States of America*, 111(14), 5135-5140.
- Quince, C., Lanzen, A., Davenport, R., & Turnbaugh, P. (2011). Removing Noise From Pyrosequenced Amplicons. *BMC Bioinformatics*, 12, 38.
- Quirin, E. A., Ogundiwin, E. A., Prince, J. P., Mazourek, M., Briggs, M. O., Chlanda, T. S., ... Jahn, M. M. (2005). Development of sequence characterized amplified region (SCAR) primers for the detection of *Phyto.5.2*, a major QTL for resistance to *Phytophthora capsici* Leon. in pepper. *Theoretical and Applied Genetics*, 110(4), 605-612.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Viena, Austria: R Foundation for Statistical Computing.
- Rabinowicz, P. D., Schutz, K., Dedhia, N., Yordan, C., Parnell, L. D., Stein, L., ... Martienssen, R. A. (1999). Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. *Nature Genetics*, 23, 305-308.
- Rai, V. P., Kumar, R., Kumar, S., Rai, A., Kumar, S., Singh, M., ... Paliwal, R. (2013). Genetic diversity in *Capsicum* germplasm based on microsatellite and random amplified microsatellite polymorphism markers. *Physiology and Molecular Biology of Plants*, 19(4), 575-586.
- Raj, A., Stephens, M., & Pritchard, J. K. (2014). FastSTRUCTURE: Variational inference of population structure in large SNP data sets. *Genetics*, 197(2), 573-589.
- Raman, R. (2017). The impact of Genetically Modified (GM) crops in modern agriculture: A review. *GM Crops and Food*, 8(4), 195-208.
- Ramasamy, R. K., Ramasamy, S., Bindroo, B. B., & Naik, V. G. (2014). STRUCTURE PLOT: A program for drawing elegant STRUCTURE bar plots in user friendly interface. *SpringerPlus*, 3(1), 431.
- Rešetnik, I., Baričević, D., Rusu, D. B., Carović-Stanko, K., Chatzopoulou, P., Dajić-Stevanovic, Z., ... Šatović, Z. (2016). Genetic diversity and demographic history of wild and cultivated/naturalised plant populations: Evidence from Dalmatian Sage (*Salvia officinalis* L., Lamiaceae). *PLoS ONE*, 11(7), e0159545.
- Rhee, S. Y., Osborne, E., Poindexter, P. D., & Somerville, C. R. (2003). Microspore Separation in the *quartet* 3 Mutants of Arabidopsis Is Impaired by a Defect in a Developmentally Regulated Polygalacturonase Required for Pollen Mother Cell

- Wall Degradation. *Plant Physiology*. 133, 1170-1180.
- Robaglia, C., & Caranta, C. (2006). Translation initiation factors: a weak link in plant RNA virus infection. *Trends in Plant Science*, 11, 40-45.
- Rodríguez, G. R., Moyseenko, J. B., Robbins, M. D., Morejón, N. H., Francis, D. M., & van der Knaap, E. (2010). Tomato analyzer: A useful software application to collect accurate and detailed morphological and colorimetric data from two-dimensional objects. *Journal of Visualized Experiments*, 37, e1856.
- Rodríguez-Maza, M. J., Garcés-Claver, A., Park, S. W., Kang, B. C., & Arnedo-Andrés, M. S. (2012). A versatile PCR marker for pungency in *Capsicum* spp. *Molecular Breeding*, 30(2), 889-898.
- Rohlf, J. F. (1972). An empirical comparison of three ordination techniques in numerical taxonomy. *Systematic Biology*, 21(3), 271-280.
- Römer, P., Jordan, T., & Lahaye, T. (2010). Identification and application of a DNA-based marker that is diagnostic for the pepper (*Capsicum annuum*) bacterial spot resistance gene *Bs3*. *Plant Breeding*, 129(6), 737-740.
- Rosenberg, N. A., & Nordborg, M. (2006). A general population-genetic model for the production by population structure of spurious genotype-phenotype associations in discrete, admixed or spatially distributed populations. *Genetics*, 173(3), 1665-1678.
- Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, 8(1), 103-106.
- Rubio, M., Caranta, C., & Palloix, A. (2008). Functional markers for selection of potyvirus resistance alleles at the *pvr2-eIF4E* locus in pepper using tetra-primer ARMS-PCR. *Genome*, 51(9), 767-771.
- Ruffel, S., Dussault, M. ., Palloix, A., Moury, B., Bendahmane, A., Robaglia, C., & Caranta, C. (2002). A natural recessive resistance gene against potato virus Y in pepper corresponds to the eukaryotic initiation factor 4E (eIF4E). *Plant Journal*, 32, 1067-1075.
- Ruffel, S., Gallois, J. L., Moury, B., Robaglia, C., Palloix, A., & Caranta, C. (2006). Simultaneous mutations in translation initiation factors eIF4E and eIF(iso)4E are required to prevent pepper veinal mottle virus infection of pepper. *Journal of General*

- Virology*, 87, 2089-2098.
- Rutkowski, R., Sielezniew, M., & Szostak, A. (2009). Contrasting levels of polymorphism in cross-amplified microsatellites in two endangered xerothermophilous, obligatorily myrmecophilous, butterflies of the genus *Phengaris* (Maculinea) (Lepidoptera: Lycaenidae). *European Journal of Entomology*, 106(4), 457-469.
- Rutten, A., Cox, K., Scheppers, T., Broecke, B. Vanden, Leirs, H., & Casaer, J. (2019). Analysing the recolonisation of a highly fragmented landscape by wild boar using a landscape genetic approach. *Wildlife Biology*, 1, 1-11.
- Sanderfoot, A. A., Assaad, F. F., & Raikhel, N. V. (2000). The Arabidopsis genome. An abundance of soluble N-ethylmaleimide-sensitive factor adaptor protein receptors. *Plant Physiology*, 124(4), 1558-1569.
- Schuelke, M. (2000). An economic method for the fluorescent labeling of PCR fragments. *Nature Biotechnology*, 18(2), 233-234.
- Scientific Committee on Food on Capsaicin. (2002). Final Opinion of the Scientific Committee on Food on Capsaicin. *European Commission*, (February), 1-12.
- Shumilin, I. A., Cymborowski, M., Chertihin, O., Jha, K. N., Herr, J. C., Lesley, S. A., ... Minor, W. (2012). Identification of unknown protein function using metabolite cocktail screening. *Structure*, 20(10), 1715-1725.
- Silvar, C., & García-González, C. A. (2016). Deciphering genetic diversity in the origins of pepper (*Capsicum* spp.) and comparison with worldwide variability. *Crop Science*, 56(6), 3100-3111.
- Silvar, C., & García-González, C. A. (2017). Screening old peppers (*Capsicum* spp.) for disease resistance and pungency-related traits. *Scientia Horticulturae*, 218, 249-257.
- Simmons, A. T., & Gurr, G. M. (2005). Trichomes of *Lycopersicon* species and their hybrids: Effects on pests and natural enemies. *Agricultural and Forest Entomology*, 7(4), 265-276.
- Simone, D. A., Baumann, T. K., & LaMotte, R. H. (1989). Dose-dependent pain and mechanical hyperalgesia in humans after intradermal injection of capsaicin. *Pain*, 38(1), 99-107.
- Solomon, A. M., Han, K., Lee, J. H., Lee, H. Y., Jang, S., & Kang, B. C. (2019). Genetic diversity and population structure of Ethiopian *Capsicum* germplasms. *PLoS ONE*,

- 14(5), e0216886.
- Srinivasan, K. (2016). Biological Activities of Red Pepper (*Capsicum annuum*) and Its Pungent Principle Capsaicin: A Review. *Critical Reviews in Food Science and Nutrition*, 56(9), 1488-1500.
- Stellari, G. M., Mazourek, M., & Jahn, M. M. (2010). Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. *Heredity*, 104(5), 460-471.
- Stift, M., Kolář, F., & Meirmans, P. G. (2019). Structure is more robust than other clustering methods in simulated mixed-ploidy populations. *Heredity*, 123(4), 429-441.
- Südhof, T. C., & Rothman, J. E. (2009). Membrane fusion: Grappling with SNARE and SM proteins. *Science*, 323(5913), 474-477.
- Sun, T., Xu, Z., Wu, C.-T., Janes, M., Prinyawiwatkul, W., & No, H. K. (2007). Antioxidant activities of different colored sweet bell peppers (*Capsicum annuum* L.). *Journal of Food Science*, 72(2), S98-102.
- Surh, Y.-J. (2002). More Than Spice: Capsaicin in Hot Chili Peppers Makes Tumor Cells Commit Suicide. *JNCI: Journal of the National Cancer Institute*, 94(17), 1263-1265.
- Suzuki, T., Fujiwake, H., & Iwai, K. (1980). Intracellular localization of capsaicin and its analogues, capsaicinoid, in *Capsicum* fruit 1. Microscopic investigation of the structure of the placenta of *Capsicum annuum* var. *annuum* cv. Karaya-tsubusa. *Plant and Cell Physiology*, 21 (5), 839-853.
- Taitano, N., Bernau, V., Jardón-Barbolla, L., Leckie, B., Mazourek, M., Mercer, K., ... van der Knaap, E. (2019). Genome-wide genotyping of a novel Mexican Chile Pepper collection illuminates the history of landrace differentiation after *Capsicum annuum* L. domestication. *Evolutionary Applications*, 12(1), 78-92.
- Tam, S. M., Lefebvre, V., Palloix, A., Sage-Palloix, A. M., Mhiri, C., & Grandbastien, M.-A. (2009). LTR-retrotransposons Tnt1 and T135 markers reveal genetic diversity and evolutionary relationships of domesticated peppers. *Theoretical and Applied Genetics*, 119(6), 973-989.
- Tamisier, L., Szadkowski, M., Nemouchi, G., Lefebvre, V., Szadkowski, E., Duboscq, R., ... Moury, B. (2020). Genome-wide association mapping of QTLs implied in potato

- virus Y population sizes in pepper: evidence for widespread resistance QTL pyramiding. *Molecular Plant Pathology*, 21(1), 3-16.
- Taranto, F., D'Agostino, N., Greco, B., Cardi, T., & Tripodi, P. (2016). Genome-wide SNP discovery and population structure analysis in pepper (*Capsicum annuum*) using genotyping by sequencing. *BMC Genomics*, 17, 943.
- Tewksbury, J. J., Reagan, K. M., Machnicki, N. J., Carlo, T. A., Haak, D. C., Peñaloza, A. L. C., & Levey, D. J. (2008). Evolutionary ecology of pungency in wild chilies. *Proceedings of the National Academy of Sciences of the United States of America*, 105(33), 11808-11811.
- Thabuis, A., Palloix, A., Pflieger, S., Daubèze, A. M., Caranta, C., & Lefebvre, V. (2003). Comparative mapping of *Phytophthora* resistance loci in pepper germplasm: Evidence for conserved resistance loci across Solanaceae and for a large genetic diversity. *Theoretical and Applied Genetics*, 106(8), 1473-1485.
- The Capsaicin Study Group. (1991). Treatment of painful diabetic neuropathy with topical capsaicin. A multicenter, double-blind, vehicle-controlled study. *Archives of Internal Medicine*, 151(11), 2225-2229.
- Thompson, J. T. (1995). *The Great Hot Sauce Book*. Nueva York, Estados Unidos: Ten Speed Press.
- Tong, N., & Bosland, P. W. (1999). *Capsicum tovarii*, a new member of the *Capsicum baccatum* complex. *Euphytica*, 109(2), 71-77.
- Topuz, A., & Ozdemir, F. (2007). Assessment of carotenoids, capsaicinoids and ascorbic acid composition of some selected pepper cultivars (*Capsicum annuum* L.) grown in Turkey. *Journal of Food Composition and Analysis*, 20(7), 596-602.
- Tripodi, P., & Greco, B. (2018). Large scale phenotyping provides insight into the diversity of vegetative and reproductive organs in a wide collection of wild and domesticated peppers (*Capsicum* spp.). *Plants*, 7, 103.
- Turina, M., Tavella, L., & Ciuffo, M. (2012). Tospoviruses in the Mediterranean Area. En G. Loebenstein, & H. Lecoq (Eds.), *Viruses and Virus Diseases of Vegetables in the Mediterranean Basin, Vol 84, Advances in Virus Research* (pp. 403-437). San Diego, California, Estados Unidos: Academic Press.
- Turner, S. D. (2018). qqman: an R package for visualizing GWAS results using Q-Q and

- manhattan plots. *Journal of Open Source Software*, 3(25), 731-732.
- USDA-ARS, N. G. R. P. (2020). Germplasm Resources Information Network (GRIN). Recuperado 18 de agosto de 2020, de <https://npgsweb.ars-grin.gov/gringlobal/taxon/taxonomysimple.aspx>
- van der Loo, W., Arthur, C. P., Richardson, B. J., Wallage-Drees, M., & Hamers, R. (1987). Nonrandom allele associations between unlinked protein loci: are the polymorphisms of the immunoglobulin constant regions adaptive? *Proceedings of the National Academy of Sciences of the United States of America*, 84(9), 3075-3079.
- Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., ... Sonstegard, T. S. (2008). SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, 5(3), 247-252.
- van Zonneveld, M., Meckelmann, S. W., Riegel, D. W., Ríos, L., Peña, K., Mueller-Seitz, E., & Petz, M. (2015). Capsaicinoids, flavonoids, tocopherols, antioxidant capacity and color attributes in 23 native Peruvian chili peppers (*Capsicum* spp.) grown in three different locations. *European Food Research and Technology*, 240(2), 273-283.
- Varshney, R. K., Graner, A., & Sorrells, M. E. (2005). Genomics-assisted breeding for crop improvement. *Trends in Plant Science*, 10(12), 621-630.
- Varshney, R. K., Terauchi, R., & McCouch, S. R. (2014). Harvesting the Promising Fruits of Genomics: Applying Genome Sequencing Technologies to Crop Breeding. *PLoS Biology*, 12(6), e1001883.
- Vilhjálmsson, B. J., & Nordborg, M. (2012). The nature of confounding in genome-wide association studies. *Nature Reviews Genetics*, 14(1), 1-2.
- Virga, G., Licata, M., Consentino, B. B., Tuttolomondo, T., Sabatino, L., Leto, C., & La Bella, S. (2020). Agro-morphological characterization of sicilian chili pepper accessions for ornamental purposes. *Plants*, 9, 1400.
- Vogt, T., & Jones, P. (2000). Glycosyltransferases in plant-natural product synthesis: Characterization of a supergene family. *Trends in Plant Science*, 5(9), 380-386.
- Votava, E. J., Baral, J. B., & Bosland, P. W. (2005). Genetic diversity of chile (*Capsicum annuum* var. *annuum* L.) landraces from northern New Mexico, Colorado, and Mexico. *Economic Botany*, 59(1), 8-17.

- Wageningen University and Research. (2020). CGN Germplasm Search. Recuperado 20 de agosto de 2020, de <https://cgngenis.wur.nl/>
- Wakeley, J. (1996). The excess of transitions among nucleotide substitutions: New methods of estimating transition bias underscore its significance. *Trends in Ecology and Evolution*, 11(4), 158-162.
- Walsh, B. M., & Hoot, S. B. (2001). Phylogenetic relationships of *Capsicum* (Solanaceae) using DNA sequences from two noncoding regions: The chloroplast atpB-rbcL spacer region and nuclear waxy introns. *International Journal of Plant Sciences*, 162(6), 1409-1418.
- Waples, R. S. (2015). Testing for hardy-weinberg proportions: Have we lost the plot? *Journal of Heredity*, 106(1), 1-19.
- Weinberg, W. (1908). On the demonstration of heredity in man. *Anuarios de la Sociedad de Historia Natural de Württemberg*, 64, 368-382.
- Weisenfeld, N. I., Kumar, V., Shah, P., Church, D. M., & Jaffe, D. B. (2018). Corrigendum: Direct determination of diploid genome sequences. *Genome research*, 28(4), 757-767.
- Wickland, D. P., Battu, G., Hudson, K. A., Diers, B. W., & Hudson, M. E. (2017). A comparison of genotyping-by-sequencing analysis methods on low-coverage crop datasets shows advantages of a new workflow, GB-eaSy. *BMC Bioinformatics*, 18, 586.
- Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, 19, 395-420.
- Wu, F., Eannetta, N. T., Xu, Y., Durrett, R., Mazourek, M., Jahn, M. M., & Tanksley, S. D. (2009). A COSII genetic map of the pepper genome provides a detailed picture of synteny with tomato and new insights into recent chromosome evolution in the genus *Capsicum*. *Theoretical and Applied Genetics*, 118(7), 1279-1293.
- Xing, F., Cheng, G., & Yi, K. (2006). Study on the antimicrobial activities of the capsaicin microcapsules. *Applied Polymer Science*, 102(2), 1318-1321.
- Yang, H. B., Liu, W. Y., Kang, W. H., Jahn, M., & Kang, B. C. (2009). Development of SNP markers linked to the *L* locus in *Capsicum* spp. by a comparative genetic analysis. *Molecular Breeding*, 24(4), 433-446.
- Yeam, I., Kang, B. C., Lindeman, W., Frantz, J. D., Faber, N., & Jahn, M. M. (2005). Allele-

- specific CAPS markers based on point mutations in resistance alleles at the *pvr1* locus encoding eIF4E in *Capsicum*. *Theoretical and Applied Genetics*, 112(1), 178-186.
- Yi, G., Lee, J. M., Lee, S., Choi, D., & Kim, B. D. (2006). Exploitation of pepper EST-SSRs and an SSR-based linkage map. *Theoretical and Applied Genetics*, 114(1), 113-130.
- Yu, J., Pressoir, G., Briggs, W. H., Bi, I. V., Yamasaki, M., Doebley, J. F., ... Buckler, E. S. (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nature Genetics*, 38(2), 203-208.
- Yue, G. H., Kovacs, B., & Orban, L. (2010). A new problem with cross-species amplification of microsatellites: generation of non-homologous products. *Dong wu xue yan jiu = Zoological research*, 31(2), 131-140.
- Yumnam, J. S., Tyagi, W., Pandey, A., Meetei, N. T., & Rai, M. (2012). Evaluation of Genetic Diversity of Chilli Landraces from North Eastern India Based on Morphology, SSR Markers and the *Pun1* Locus. *Plant Molecular Biology Reporter*, 30(6), 1470-1479.
- Zhang, L.-X., Cooney, R. V., & Bertram, J. S. (1991). Carotenoids enhance gap junctional communication and inhibit lipid peroxidation in C3H/10T1/2 cells: relationship to their cancer chemopreventive action. *Carcinogenesis*, 12(11), 2109-2114.
- Zhang, Z., Ersoz, E., Lai, C.-Q., Todhunter, R. J., Tiwari, H. K., Gore, M. A., ... Buckler, E. S. (2010). Mixed linear model approach adapted for genome-wide association studies. *Nature Genetics*, 42(4), 355-360.
- Zhang, Z., Schwartz, S., Wagner, L., & Miller, W. (2000). A greedy algorithm for aligning DNA sequences. *Journal of Computational Biology*, 7(1-2), 203-214.
- Zhao, M., Wu, S., Zhou, Q., Vivona, S., Cipriano, D., Cheng, Y., & Brunger, A. (2015). Mechanistic insights into the recycling machine of the SNARE complex. *Nature*, 518, 61-67.

Anexos

Anexos de tablas

Tabla A.1. Frecuencias de alelos nulos para cada marcador microsatélite calculadas con los programas FreeNa y ML-Null Freq.

<i>Locus</i>	Alelos nulos (FreeNA)	Alelos nulos (ML-Null Freq)
<i>Hpms1-139</i>	0,411	0,426
<i>Hpms1-281</i>	0,383	0,383
<i>Hpms1-143</i>	0,433	0,544
<i>Hpms1-173</i>	0,417	0,452
<i>CAMS405</i>	0,398	0,446
<i>GPMS197</i>	0,436	0,484
<i>CAMS351</i>	0,442	0,482
<i>Hpms1-5</i>	0,443	0,611
<i>Hpms2-45</i>	0,331	0,340
<i>CAMS336</i>	0,421	0,444
<i>HpmsE013</i>	0,399	0,419
<i>Hpms2-2</i>	0,386	0,411
<i>EPMS310</i>	0,409	0,416
<i>Hpms2-24</i>	0,433	0,444
<i>Hpms1-3</i>	0,365	0,464
<i>HpmsCaSIG19</i>	0,385	0,409
<i>Hpms1-6</i>	0,259	0,356
<i>Hpms1-155</i>	0,421	0,437
<i>HpmsE015</i>	0,063	0,081
<i>HpmsE016</i>	0,422	0,442
<i>HpmsE120</i>	0,370	0,403
<i>HpmsAT2-14</i>	0,387	0,435
<i>EPMS426</i>	0,383	0,426
<i>EPMS472</i>	0,406	0,463
<i>Hpms1-148</i>	0,395	0,418
<i>GPMS178</i>	0,162	0,207
<i>HpmsE075</i>	0,365	0,373
<i>GPMS159</i>	0,433	0,455

Tabla A.2. Parámetros de diversidad genética por especie y por locus microsatélite. *N*, número de accesiones con genotipo; *N_a*, número de alelos; *N_{ap}*, número de alelos privados; *H_o*, heterocigosidad observada; *H_e*, heterocigosidad esperada; *F_{is}*, coeficiente de endogamia; Alelos nulos, frecuencia de alelos nulos calculada con el programa FreeNa. *, accesiones reclasificadas taxonómicamente según los resultados de los marcadores microsatélite.

	<i>H_{pms1-139}</i>	<i>H_{pms1-281}</i>	<i>H_{pms1-143}</i>	<i>H_{pms1-173}</i>	<i>CAMS405</i>	<i>GPMS197</i>	<i>CAMS351</i>	<i>H_{pms1-5}</i>	<i>H_{pms2-45}</i>	<i>CAMS336</i>	<i>H_{pmsE013}</i>	<i>H_{pms2-2}</i>	<i>EPMS310</i>	<i>H_{pms2-24}</i>
<i>C. annuum</i>														
<i>N</i>	144	147	143	146	134	136	134	124	145	142	141	143	147	146
<i>N_a</i>	6	4	6	6	9	14	12	14	2	8	7	4	10	6
<i>N_{ap}</i>	0	1	3	0	3	3	6	7	0	3	1	1	4	2
<i>H_o</i>	0.118	0.007	0.021	0.082	0.075	0.044	0.045	0.024	0.007	0.092	0.021	0.042	0.027	0.062
<i>H_e</i>	0.685	0.060	0.664	0.677	0.661	0.884	0.699	0.706	0.007	0.713	0.411	0.464	0.400	0.686
<i>F_{is}</i>	0.829	0.887	0.969	0.879	0.888	0.951	0.936	0.966	0.000	0.873	0.949	0.910	0.932	0.911
Alelos nulos	0.337	0.112	0.390	0.359	0.358	0.446	0.388	0.402	0.000	0.364	0.300	0.296	0.296	0.373
<i>C. chinense</i>														
<i>N</i>	27	28	28	28	27	27	28	25	28	27	28	28	28	27
<i>N_a</i>	4	6	6	7	6	4	6	9	2	7	3	3	6	3
<i>N_{ap}</i>	0	2	0	3	1	1	1	3	0	2	0	1	0	1
<i>H_o</i>	0.000	0.000	0.000	0.033	0.034	0.000	0.000	0.000	0.000	0.071	0.033	0.033	0.033	0.000
<i>H_e</i>	0.404	0.720	0.729	0.718	0.784	0.187	0.351	0.853	0.064	0.690	0.516	0.096	0.767	0.190
<i>F_{is}</i>	1.000	1.000	1.000	0.955	0.958	1.000	1.000	1.000	1.000	0.900	0.938	0.663	0.958	1.000
Alelos nulos	0.395	0.182	0.420	0.331	0.349	0.369	0.378	0.000	0.287	0.132	0.224	0.255	0.236	0.316
<i>C. baccatum</i>														
<i>N</i>	23	24	5	16	24	24	24	0	24	24	24	24	24	23
<i>N_a</i>	5	2	4	5	6	8	6	0	2	2	3	2	3	4
<i>N_{ap}</i>	1	0	1	0	1	2	3	0	0	1	0	0	0	3
<i>H_o</i>	0.040	0.000	0.000	0.125	0.038	0.125	0.038	0.000	0.038	0.000	0.000	0.077	0.038	0.040
<i>H_e</i>	0.714	0.142	0.720	0.664	0.564	0.770	0.661	0.000	0.440	0.074	0.210	0.426	0.297	0.494
<i>F_{is}</i>	0.946	1.000	1.000	0.823	0.934	0.844	0.944	Monomórfico	0.916	1.000	1.000	0.826	0.875	0.922
Alelos nulos	0.157	0.351	0.400	0.325	0.279	0.224	0.380	0.453	0.000	0.213	0.333	0.315	0.325	0.298
<i>C. pubescens</i>														
<i>N</i>	16	16	3	16	16	13	16	14	16	16	16	15	16	16
<i>N_a</i>	2	3	3	2	4	8	4	7	4	2	2	2	2	3
<i>N_{ap}</i>	0	2	0	0	0	1	2	3	1	1	0	2	0	1
<i>H_o</i>	0.000	0.056	0.000	0.000	0.000	0.308	0.000	0.000	0.500	0.000	0.000	0.000	0.000	0.000
<i>H_e</i>	0.105	0.619	0.667	0.475	0.346	0.695	0.599	0.827	0.427	0.198	0.500	0.444	0.475	0.370
<i>F_{is}</i>	1.000	0.915	1.000	1.000	1.000	0.584	1.000	1.000	-0.142	1.000	1.000	1.000	1.000	1.000
Alelos nulos	0.313	0.420	0.422	0.401	0.421	0.213	0.294	0.461	0.124	0.369	0.317	0.114	0.416	0.214
<i>C. frutescens</i>														
<i>N</i>	11	11	11	11	11	11	11	4	11	11	11	11	11	10
<i>N_a</i>	3	4	2	3	3	4	2	2	1	5	2	2	2	3
<i>N_{ap}</i>	1	0	0	0	1	1	0	1	0	1	0	0	0	1
<i>H_o</i>	0.077	0.077	0.077	0.000	0.083	0.000	0.083	0.000	0.000	0.154	0.077	0.000	0.083	0.000
<i>H_e</i>	0.210	0.583	0.074	0.544	0.226	0.556	0.080	0.480	0.000	0.568	0.074	0.142	0.080	0.292
<i>F_{is}</i>	0.657	0.878	0.000	1.000	0.656	1.000	0.000	1.000	Monomórfico	0.747	0.000	1.000	0.000	1.000
Alelos nulos	0.158	0.320	0.000	0.360	0.163	0.366	0.000	0.327	0.001	0.259	0.000	0.182	0.000	0.264
<i>C. annuum</i> var. <i>glabriscautum</i>														
<i>N</i>	4	4	4	3	4	4	3	2	4	3	4	4	3	4
<i>N_a</i>	4	4	4	2	4	4	3	1	2	3	4	3	3	3
<i>N_{ap}</i>	0	0	0	0	0	1	1	1	0	1	0	1	1	0
<i>H_o</i>	0.000	0.000	0.000	0.000	0.000	0.250	0.000	0.000	0.250	0.000	0.000	0.000	0.000	0.000
<i>H_e</i>	0.750	0.750	0.750	0.444	0.750	0.719	0.667	0.000	0.219	0.667	0.750	0.625	0.667	0.625
<i>F_{is}</i>	1.000	1.000	1.000	1.000	1.000	0.727	1.000	Monomórfico	0.000	1.000	1.000	1.000	1.000	1.000
Alelos nulos	0.429	0.429	0.429	0.315	0.429	0.278	0.400	0.001	0.000	0.400	0.429	0.387	0.400	0.387

Tabla A.2. Parámetros de diversidad genética por especie y por locus microsatélite. *N*, número de accesiones con genotipo; *N_a*, número de alelos; *N_{ap}*, número de alelos privados; *H_o*, heterocigosidad observada; *H_e*, heterocigosidad esperada; *F_{is}*, coeficiente de endogamia; Alelos nulos, frecuencia de alelos nulos calculada con el programa FreeNa. *, accesiones reclasificadas taxonómicamente según los resultados de los marcadores microsatélite. (continuación).

	<i>Hpmst-3</i>	<i>HpmstCaSfG19</i>	<i>Hpmst-6</i>	<i>Hpmst-155</i>	<i>HpmstE015</i>	<i>HpmstE016</i>	<i>HpmstE120</i>	<i>HpmstA12-14</i>	<i>EPMS426</i>	<i>EPMS472</i>	<i>Hpmst-148</i>	<i>GPMSt78</i>	<i>HpmstE075</i>	<i>GPMSt59</i>
<i>C. annuum</i>														
<i>N</i>	144	146	140	145	145	142	139	144	142	134	145	138	146	141
<i>N_a</i>	3	7	8	7	4	17	6	2	12	4	4	16	2	13
<i>N_{ap}</i>	1	1	3	3	1	8	2	0	6	2	0	4	1	2
<i>H_o</i>	0,007	0,103	0,200	0,055	0,297	0,063	0,022	0,000	0,056	0,045	0,076	0,659	0,048	0,078
<i>H_e</i>	0,074	0,569	0,341	0,577	0,382	0,782	0,263	0,382	0,523	0,533	0,524	0,766	0,361	0,794
<i>F_{is}</i>	0,907	0,821	0,417	0,905	0,227	0,920	0,919	1,000	0,893	0,917	0,856	0,143	0,868	0,902
Alelos nulos	0,126	0,306	0,137	0,336	0,079	0,404	0,230	0,293	0,321	0,325	0,292	0,085	0,250	0,400
<i>C. ditense</i>														
<i>N</i>	27	27	27	28	28	26	28	26	26	27	25	28	28	28
<i>N_a</i>	8	5	7	4	2	7	3	8	12	2	6	9	3	9
<i>N_{ap}</i>	2	0	1	0	0	1	0	2	6	0	0	0	0	2
<i>H_o</i>	0,000	0,000	0,667	0,000	0,900	0,214	0,000	0,071	0,296	0,034	0,037	0,733	0,000	0,000
<i>H_e</i>	0,766	0,307	0,724	0,504	0,499	0,596	0,238	0,652	0,665	0,441	0,436	0,774	0,524	0,740
<i>F_{is}</i>	1,000	1,000	0,097	1,000	-0,796	0,651	1,000	0,894	0,568	0,924	0,918	0,070	1,000	1,000
Alelos nulos	0,381	0,143	0,290	0,216	0,000	0,299	0,001	0,344	0,256	0,308	0,371	0,341	0,318	0,419
<i>C. baccatum</i>														
<i>N</i>	7	2	4	24	24	24	24	21	24	22	24	24	24	24
<i>N_a</i>	3	2	4	2	2	6	1	8	3	1	6	7	2	9
<i>N_{ap}</i>	0	0	2	0	0	0	0	2	0	0	2	2	0	1
<i>H_o</i>	0,000	0,000	0,038	0,000	0,846	0,077	0,000	0,130	0,000	0,000	0,077	0,154	0,000	0,080
<i>H_e</i>	0,612	0,087	0,412	0,204	0,497	0,493	0,000	0,720	0,272	0,423	0,713	0,748	0,857	0,524
<i>F_{is}</i>	1,000	1,000	0,910	1,000	-0,692	0,850	Monomórfico	0,826	1,000	1,000	0,896	0,801	1,000	0,910
Alelos nulos	0,424	0,001	0,001	0,300	0,000	0,213	0,296	0,161	0,000	0,001	0,000	0,000	0,296	0,299
<i>C. putrescens</i>														
<i>N</i>	16	16	10	15	16	16	16	11	14	16	16	14	16	15
<i>N_a</i>	5	1	1	3	3	2	2	4	7	1	4	3	2	5
<i>N_{ap}</i>	0	0	0	0	0	1	0	0	1	1	1	1	0	0
<i>H_o</i>	0,000	0,000	0,000	0,000	0,167	0,000	0,056	0,333	0,214	0,000	0,167	0,500	0,056	0,118
<i>H_e</i>	0,735	0,000	0,000	0,381	0,156	0,198	0,498	0,642	0,520	0,000	0,205	0,508	0,498	0,573
<i>F_{is}</i>	1,000	Monomórfico	Monomórfico	1,000	-0,041	1,000	0,894	0,514	0,612	Monomórfico	0,215	0,048	0,894	0,806
Alelos nulos	0,435	0,274	0,010	0,348	0,000	0,226	0,238	0,357	0,203	0,290	0,298	0,049	0,350	0,428
<i>C. frutescens</i>														
<i>N</i>	11	11	9	11	11	11	11	11	11	11	11	11	11	11
<i>N_a</i>	5	3	4	3	3	2	2	5	3	2	5	4	2	3
<i>N_{ap}</i>	1	0	1	1	0	2	0	0	0	0	2	0	0	0
<i>H_o</i>	0,000	0,000	0,778	0,182	0,636	0,000	0,000	0,154	0,077	0,077	0,000	0,462	0,077	0,000
<i>H_e</i>	0,777	0,292	0,599	0,492	0,657	0,260	0,142	0,633	0,269	0,453	0,760	0,512	0,074	0,462
<i>F_{is}</i>	1,000	1,000	-0,244	0,658	0,079	1,000	1,000	0,774	0,733	0,842	1,000	0,138	0,000	1,000
Alelos nulos	0,437	0,264	0,000	0,212	0,034	0,243	0,182	0,289	0,164	0,265	0,433	0,075	0,000	0,330
<i>C. annuum</i> var. <i>glabriusculum</i>														
<i>N</i>	3	4	3	4	4	4	4	4	3	3	4	4	3	4
<i>N_a</i>	3	4	2	4	5	3	4	3	3	2	4	5	2	4
<i>N_{ap}</i>	0	0	0	2	1	2	1	0	1	0	2	0	0	0
<i>H_o</i>	0,000	0,000	0,000	0,000	1,000	0,000	0,250	0,000	0,000	0,000	0,250	0,250	0,000	0,000
<i>H_e</i>	0,667	0,750	0,444	0,750	0,688	0,625	0,656	0,625	0,667	0,444	0,719	0,781	0,444	0,750
<i>F_{is}</i>	1,000	1,000	1,000	1,000	-0,333	1,000	0,700	1,000	1,000	1,000	0,727	0,750	1,000	1,000
Alelos nulos	0,400	0,429	0,315	0,429	0,000	0,387	0,236	0,387	0,400	0,315	0,278	0,300	0,315	0,429

Tabla A.3. Genotipos de cada accesión para los diez marcadores de enfermedad y el marcador de pungencia. R: resistente. S: susceptible. P: pungente. NP: no pungente. -: dato perdido.

	OpD04.717	06012END	SCAR_CD	CaTm-int1	PR-Bs3	SCAC568	Pvr1-S	pvr1-R1	pvr1-R2	T200A	CSO	MAP1
<i>C. annuum</i> ibérica												
BGHZ0102	S	S	S	S	-	S	S	S	S	S	R	NP
BGHZ0113	S	S	S	S	S	S	S	-	S	S	S	-
BGHZ0179	S	S	S	S	S	S	S	-	S	S	S	P
BGHZ0206	S	S	S	S	S	-	S	S	S	S	S	NP
BGHZ0213	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0224	S	S	S	S	S	S	S	S	S	S	R	NP
BGHZ0282	S	S	S	S	-	-	S	S	S	S	S	NP
BGHZ0283	S	S	S	S	S	S	S	S	S	-	S	NP
BGHZ0287	S	S	S	S	S	S	S	S	S	S	S	P
BGHZ0331	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0337	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0338	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0413	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0418	S	S	S	S	S	S	S	S	S	-	S	NP
BGHZ0463	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0775	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0782	S	S	S	S	S	S	S	S	S	S	S	P
BGHZ0958	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ0971	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ1109	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ1375	S	S	S	S	S	S	S	S	S	S	S	P
BGHZ1519	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ1521	S	S	S	S	S	S	S	S	S	S	S	P
BGHZ1756	S	S	S	S	R	-	S	S	S	S	S	NP
BGHZ2154	S	S	S	S	S	S	S	S	S	S	S	P
BGHZ2546	S	S	S	S	S	S	R	S	S	R	S	NP
BGHZ2561	S	S	S	S	S	S	S	S	S	S	S	P
BGHZ3337	S	S	S	S	S	S	S	S	S	-	S	NP
BGHZ5388	S	S	S	S	S	S	S	S	S	S	S	NP
BGHZ5912	S	S	S	S	S	S	S	S	S	S	R	NP
PA-145	S	S	S	S	S	S	S	S	S	S	S	NP
Ol27-05	S	S	S	S	S	S	S	S	S	S	S	NP
CO12B	S	S	S	S	S	S	S	S	S	S	S	NP
PI309-05	S	S	S	S	S	S	S	S	S	S	S	NP
BR113-05	S	S	S	S	S	S	S	S	S	S	S	NP
CGN17234	S	S	S	S	S	S	S	S	S	S	S	NP
CGN17235	S	S	S	S	S	S	S	S	S	S	S	P
CGN23293	S	S	S	S	S	S	S	S	S	S	S	NP
CGN20507	S	S	S	S	S	S	S	S	S	S	S	NP
CAP1661	S	S	S	S	S	S	R	S	S	R	S	NP
03038-BPGV	S	S	S	S	-	-	S	S	S	-	-	P
04793-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
05572-BPGV	S	S	S	S	-	-	S	S	S	S	S	-
05740-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
06235-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
07149-BPGV	S	S	S	S	S	S	S	S	S	S	R	NP
07822-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
11355-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
11688-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
11805-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
11849-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
12348-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
13027-BPGV	S	S	S	S	S	S	R	S	S	R	S	NP
13113-BPGV	S	S	S	S	S	S	R	S	S	R	S	NP
12950-BPGV	S	S	S	S	S	S	S	S	S	S	S	NP
CGN22161	S	S	S	S	S	S	S	S	S	S	S	NP
CGN16899	S	S	S	S	S	S	S	S	S	S	S	NP
CAP1371	S	S	S	S	S	S	S	S	S	S	S	NP
CAP1539	S	S	S	S	S	S	S	S	S	S	S	NP
CGN21475	S	S	S	S	S	S	S	S	S	S	S	NP
CGN16805	S	S	S	S	S	S	S	S	S	S	S	NP
CGN16897	S	S	S	S	S	S	S	S	S	S	S	NP

Tabla A.3. Genotipos de cada accesión para los diez marcadores de enfermedad y el marcador de pungencia. R: resistente. S: susceptible. P: pungente. NP: no pungente. -: dato perdido. (continuación).

	<i>OpD04.717</i>	<i>06012END</i>	<i>SCAR_CD</i>	<i>CaTm-int1</i>	<i>PR-Bs3</i>	<i>SCAC568</i>	<i>Per1-S</i>	<i>per1-R1</i>	<i>per1-R2</i>	<i>T200A</i>	<i>CSO</i>	<i>MAP1</i>
<i>C. annuum mediterránea</i>												
CGN16894	S	S	S	S	S	S	S	S	S	S	S	NP
CGN17172	S	S	S	S	S	S	S	S	S	S	S	NP
CGN21486	S	S	S	S	S	S	S	S	S	S	S	NP
CGN23239	S	S	S	S	S	S	S	S	S	S	S	NP
CGN23244	S	S	R	S	S	S	S	S	S	S	S	P
CGN16800	S	S	S	S	S	S	R	S	S	R	S	P
CGN16881	S	S	S	S	S	S	S	S	S	-	S	P
CGN22138	S	S	S	S	S	S	S	S	S	S	S	NP
CAP242	S	S	S	S	S	S	R	S	S	R	R	NP
CAP913	S	S	R	S	S	S	S	S	S	S	S	NP
CAP68	S	S	S	S	S	S	S	S	S	S	S	NP
CGN23235	S	S	S	S	S	S	S	S	S	S	S	NP
CGN17192	S	S	S	S	S	S	S	S	S	-	S	P
CGN16835	S	S	S	S	S	S	S	S	S	S	R	P
CGN17193	S	S	S	S	S	R	S	S	S	S	R	NP
CGN22183	S	S	S	S	S	S	S	S	S	S	R	P
CGN21581	S	S	S	S	S	S	S	S	S	S	S	NP
CGN16916	S	S	S	S	S	S	S	-	S	S	S	NP
CGN21480	S	S	S	S	S	S	S	S	S	S	S	NP
CGN22825	S	S	S	S	S	R	S	S	S	S	S	P
CGN16917	S	S	S	S	S	S	S	-	S	S	S	NP
CGN17191	S	S	S	S	S	S	S	S	S	S	R	NP
CGN17010	S	S	S	S	S	S	S	S	S	S	S	NP
CGN24358	S	S	S	S	S	S	S	S	S	S	R	P
CGN16833	S	S	S	S	S	S	S	S	S	S	R	-
CGN22154	S	S	S	S	S	S	S	S	S	S	R	P
CAP1049	S	S	S	S	S	S	S	S	S	S	S	NP
CAP1051	S	S	S	S	S	S	S	S	S	S	S	P
CAP606	S	S	S	S	S	S	S	S	S	S	S	P
CAP609	S	S	S	S	S	S	S	S	S	S	S	P
CAP614	S	S	S	S	S	S	S	S	S	S	S	P
CAP1538	S	S	S	S	S	S	S	S	S	S	S	P
CAP977	S	S	S	S	S	S	S	S	S	S	S	P
CAP979	S	S	S	S	S	S	S	S	S	S	S	P

Tabla A.4. Estadísticas de las lecturas de GBS empleadas. Lecturas crudas: lecturas obtenidas tras el proceso de secuenciación, sin realizar ningún tratamiento.

Accesión	Especie	Lecturas crudas		Lecturas después de la limpieza con Trimmomatic			Lecturas mapeadas				
		Total	% lecturas con calidad (valor Phred)>30	Total	% lecturas con respecto al total de lecturas crudas	% lecturas con calidad (valor Phred)>30	Total	% lecturas mapeadas con respecto a las conservadas después de la limpieza	Lecturas con calidad Mapq>57	% con calidad>57 mapeadas con respecto a las conservadas después de la limpieza	% mapeadas que se descartan por tener calidad Mapq<57
CGN21475	<i>C. annuum</i>	307024	97,8	300789	98	98,6	205364	68,3	153502	51	17,3
BGHZ0331	<i>C. annuum</i>	658003	95,7	591166	89,8	97,2	583681	98,7	393759	66,6	32,1
CGN16833	<i>C. annuum</i>	708001	95,8	669175	94,5	97,3	658130	98,3	363168	54,3	44
CGN16805	<i>C. annuum</i>	734279	97,7	692353	94,3	98,6	684858	98,9	477740	69	29,9
CGN16800	<i>C. annuum</i>	750188	97,7	306201	40,8	98,6	301268	98,4	209813	68,5	29,9
CGN22783	<i>C. annuum</i>	752718	97,9	689535	91,6	98,7	667778	96,8	455814	66,1	30,7
CGN16881	<i>C. annuum</i>	834791	98,2	800506	95,9	98,8	785277	98,1	556562	69,4	28,7
CGN23239	<i>C. annuum</i>	864487	97,6	841899	97,4	98,5	828857	98,5	581083	69	29,5
CAP977	<i>C. annuum</i>	965111	98	916962	95	98,7	901518	98,3	630497	68,8	29,5
CAP242	<i>C. annuum</i>	1036575	97,8	1001395	96,6	98,6	986320	98,5	643356	64,2	34,3
CGN21486	<i>C. annuum</i>	1039182	98,2	984387	94,7	98,9	969418	98,5	598002	60,7	37,8
CGN16897	<i>C. annuum</i>	1054671	97,7	895974	85	98,6	881730	98,4	592272	66,1	32,3
CGN16894	<i>C. annuum</i>	1081499	97,8	980528	90,7	98,7	969254	98,9	670850	68,4	30,5
CGN23244	<i>C. annuum</i>	1109953	97,5	899242	81	98,4	884208	98,3	585670	65,1	33,2
CGN17172	<i>C. annuum</i>	1112938	97,9	1028923	92,5	98,7	1014487	98,6	645438	62,7	35,9
BGHZ0097	<i>C. annuum</i>	4947473	95,9	2916111	58,9	97,5	2859633	98,1	1897938	65,1	33
PI241670	<i>C. annuum</i>	5814379	96,8	5506973	94,7	97,9	5409769	98,2	3672296	66,7	31,5
CAP979	<i>C. annuum</i>	1328391	97,6	1278361	96,2	98,5	1261686	98,7	914574	71,5	27,2
CAP606	<i>C. annuum</i>	1552567	97,3	1473722	94,9	98,4	1453096	98,6	979053	66,4	32,2
CGN22138	<i>C. annuum</i>	1613566	98,1	1546410	95,8	98,7	1518275	98,2	1095548	70,8	27,4
CGN22154	<i>C. annuum</i>	1635263	97,2	1576248	96,4	98,2	1542669	97,9	947522	60,1	37,8
CGN22825	<i>C. annuum</i>	1640602	97	1540952	93,9	98,1	1505231	97,7	926362	60,1	37,6
CGN24358	<i>C. annuum</i>	1677822	97,1	1375244	82	98,2	1353906	98,4	869276	63,2	35,2
BGHZ5912	<i>C. annuum</i>	1776284	95,9	1500191	84,5	97,4	1479092	98,6	919616	61,3	37,3
CGN22183	<i>C. annuum</i>	1790970	98	1711052	95,5	98,7	1684293	98,4	1037304	60,6	37,8
CGN21480	<i>C. annuum</i>	1884464	97,8	1810395	96,1	98,6	1786764	98,7	1260442	69,6	29,1
CAP1661	<i>C. annuum</i>	1917372	95,7	1642247	85,7	97,3	1611872	98,2	998797	60,8	37,4
BPGV11688	<i>C. annuum</i>	1929187	96,7	1633700	84,7	97,9	1601243	98	1013291	62	36
COI2B	<i>C. annuum</i>	1940160	96	1700118	87,6	97,5	1671231	98,3	1106727	65,1	33,2
CGN17191	<i>C. annuum</i>	1957401	97,1	1528860	78,1	98,4	1482034	96,9	899195	58,8	38,1
BGHZ1519	<i>C. annuum</i>	1959170	96,2	1673000	85,4	97,6	1643242	98,2	1141598	68,2	30
BGHZ3337	<i>C. annuum</i>	1979424	95,6	1441582	72,8	97,4	1416938	98,3	924287	64,1	34,2
BGHZ0775	<i>C. annuum</i>	1979606	96,4	1784224	90,1	97,7	1753835	98,3	1125588	63,1	35,2
BGHZ0337	<i>C. annuum</i>	1988157	96	1751543	88,1	97,4	1723183	98,4	1060078	60,5	37,9
BGHZ1521	<i>C. annuum</i>	1991679	96,4	1782721	89,5	97,7	1749735	98,1	1179534	66,2	31,9
BPGV11355	<i>C. annuum</i>	2002735	96,3	1726904	86,2	97,7	1695670	98,2	1015519	58,8	39,4
BPGV07149	<i>C. annuum</i>	2010272	96,3	1739759	86,5	97,5	1716737	98,7	1070244	61,5	37,2
BGHZ1114	<i>C. annuum</i>	2010948	95,8	1707747	84,9	97,3	1668198	97,7	1019872	59,7	38
BGHZ0969	<i>C. annuum</i>	2014583	95,3	1832425	91	97,1	1795703	98	1209028	66	32
CGN16916	<i>C. annuum</i>	2022758	98	1885928	93,2	98,7	1858851	98,6	1269346	67,3	31,3
BGHZ2561	<i>C. annuum</i>	2025900	96,3	1837835	90,7	97,6	1801950	98	1159010	63,1	34,9
OI2705	<i>C. annuum</i>	2032822	96,4	1814094	89,2	97,7	1784631	98,4	1154025	63,6	34,8
CGN17235	<i>C. annuum</i>	2032877	96,6	1838844	90,5	97,9	1804747	98,1	1265968	68,8	29,3
CGN20507	<i>C. annuum</i>	2033931	96,5	1476819	72,6	97,9	1443231	97,7	929561	62,9	34,8
BPGV04793	<i>C. annuum</i>	2041531	96,4	1877231	92	97,7	1837780	97,9	1078936	57,5	40,4
BGHZ3269	<i>C. annuum</i>	2044726	96,3	1859601	90,9	97,6	1830123	98,4	1195447	64,3	34,1
CGN17010	<i>C. annuum</i>	2044921	96,6	1704829	83,4	98,1	1681559	98,6	1027603	60,3	38,3
CAP614	<i>C. annuum</i>	2047059	97,6	1936556	94,6	98,5	1904706	98,4	1336880	69	29,4
PA145	<i>C. annuum</i>	2048905	96,6	1894944	92,5	97,8	1866503	98,5	1056659	55,8	42,7
BGHZ1375	<i>C. annuum</i>	2049210	96	1768082	86,3	97,5	1730426	97,9	1044541	59,1	38,8
BGHZ2546	<i>C. annuum</i>	2049559	96,2	1766450	86,2	97,6	1741759	98,6	1081833	61,2	37,4
BR11305	<i>C. annuum</i>	2053400	95,9	1832201	89,2	97,4	1800767	98,3	1157568	63,2	35,1

Tabla A.4. Estadísticas de las lecturas de GBS empleadas. Lecturas crudas: lecturas obtenidas tras el proceso de secuenciación, sin realizar ningún tratamiento. (continuación).

Accesión	Especie	Lecturas crudas		Lecturas después de la limpieza con Trimmomatic			Lecturas mapeadas				
		Total	% lecturas con calidad (valor Phred)>30	Total	% lecturas con respecto al total de lecturas crudas	% lecturas con calidad (valor Phred)>30	Total	% lecturas mapeadas con respecto a las conservadas después de la limpieza	Lecturas con calidad Mapq>57	% con calidad >57 mapeadas con respecto a las conservadas después de la limpieza	% mapeadas que se descartan por tener calidad Mapq<57
CGN23293	<i>C. annuum</i>	2059552	96,5	1739215	84,4	97,7	1716831	98,7	1071682	61,6	37,1
BGHZ0782	<i>C. annuum</i>	2060581	96,5	1775434	86,2	97,8	1742278	98,1	1080062	60,8	37,3
C431	<i>C. annuum</i>	2072328	96,2	1825615	88,1	97,6	1794314	98,3	1127547	61,8	36,5
BGHZ1109	<i>C. annuum</i>	2072730	96,2	1895993	91,5	97,6	1859547	98,1	1198790	63,2	34,9
CGN17192	<i>C. annuum</i>	2073990	97,8	1948772	94	98,6	1926247	98,8	1241312	63,7	35,1
CGN17214	<i>C. annuum</i>	2076972	97,1	1965446	94,6	98,2	1930469	98,2	1351068	68,7	29,5
BGHZ0113	<i>C. annuum</i>	2084763	96,7	1936619	92,9	97,9	1901599	98,2	1278702	66	32,2
BPGV07822	<i>C. annuum</i>	2085478	96	1803855	86,5	97,5	1779270	98,6	1097698	60,9	37,7
PI30905	<i>C. annuum</i>	2088980	95,6	1835111	87,8	97,3	1807298	98,5	1063861	58	40,5
C434	<i>C. annuum</i>	2096281	96,8	1833752	87,5	97,9	1800638	98,2	1140378	62,2	36
CAP491	<i>C. annuum</i>	2098401	97,1	1930084	92	98,2	1897358	98,3	1322689	68,5	29,8
CGN16835	<i>C. annuum</i>	2103761	97,8	1952231	92,8	98,6	1922142	98,5	1339676	68,6	29,9
CGN16917	<i>C. annuum</i>	2105760	97,4	1789941	85	98,4	1760277	98,3	1124231	62,8	35,5
BGHZ0284	<i>C. annuum</i>	2106154	96,6	1914621	90,9	97,9	1879565	98,2	1190541	62,2	36
BGHZ0966	<i>C. annuum</i>	2115601	96,3	1910813	90,3	97,6	1871173	97,9	1182748	61,9	36
BPGV06235	<i>C. annuum</i>	2117605	95,9	1770432	83,6	97,5	1731116	97,8	1146830	64,8	33
PI543188	<i>C. annuum</i>	2122825	97,3	1991652	93,8	98,3	1955835	98,2	1380895	69,3	28,9
BGHZ0447	<i>C. annuum</i>	2126277	96,6	1923260	90,5	97,8	1886257	98,1	1219900	63,4	34,7
BGHZ0181	<i>C. annuum</i>	2128908	96,3	1940142	91,1	97,7	1905032	98,2	1277826	65,9	32,3
BGHZ1374	<i>C. annuum</i>	2137134	96	1805525	84,5	97,5	1770759	98,1	1128364	62,5	35,6
BGHZ5388	<i>C. annuum</i>	2139629	96,1	1929806	90,2	97,5	1895227	98,2	1164594	60,3	37,9
BGHZ2169	<i>C. annuum</i>	2145373	96	1889312	88,1	97,5	1848419	97,8	1108044	58,6	39,2
BGHZ0958	<i>C. annuum</i>	2146391	96,3	1791334	83,5	97,6	1763370	98,4	1118860	62,5	35,9
BGHZ0283	<i>C. annuum</i>	2146571	96	1915951	89,3	97,5	1875280	97,9	1167304	60,9	37
CGN21581	<i>C. annuum</i>	2147589	97,9	2022827	94,2	98,6	1985899	98,2	1427948	70,6	27,6
BGHZ0338	<i>C. annuum</i>	2148800	95,4	1859038	86,5	97,1	1822323	98	1150367	61,9	36,1
CGN16808	<i>C. annuum</i>	2149805	96,8	1953396	90,9	98	1920327	98,3	1336670	68,4	29,9
CGN17147	<i>C. annuum</i>	2151887	97	1975932	91,8	98,1	1939504	98,2	1293272	65,5	32,7
BGHZ1756	<i>C. annuum</i>	2153032	96,4	1950359	90,6	97,7	1917987	98,3	1200080	61,5	36,8
PI213915	<i>C. annuum</i>	2158265	96,6	1981479	91,8	98	1946265	98,2	1364714	68,9	29,3
CGN22155	<i>C. annuum</i>	2162451	97	2005848	92,8	98,2	1966813	98,1	1438311	71,7	26,4
BPGV05572	<i>C. annuum</i>	2167970	96,1	1884513	86,9	97,5	1853151	98,3	1105556	58,7	39,6
BGHZ0215	<i>C. annuum</i>	2171777	96,6	1973649	90,9	97,8	1935840	98,1	1259719	63,8	34,3
PI573337	<i>C. annuum</i>	2173354	96,5	2025753	93,2	97,9	1901649	93,9	1329806	65,6	28,3
BGHZ1174	<i>C. annuum</i>	2173902	96,1	1966084	90,4	97,5	1934495	98,4	1177835	59,9	38,5
C430	<i>C. annuum</i>	2177861	96,2	1964727	90,2	97,6	1927249	98,1	1141810	58,1	40
BPGV11805	<i>C. annuum</i>	2178638	96,3	1824479	83,7	97,6	1792087	98,2	1117196	61,2	37
CAP524	<i>C. annuum</i>	2182705	96,7	2034628	93,2	98	1994955	98,1	1457607	71,6	26,5
BGHZ1753	<i>C. annuum</i>	2184285	96,3	1993649	91,3	97,6	1954970	98,1	1222703	61,3	36,8
CGN17190	<i>C. annuum</i>	2196385	96,6	2022390	92,1	97,9	1994390	98,6	1422070	70,3	28,3
BGHZ0224	<i>C. annuum</i>	2199696	96,2	2013031	91,5	97,6	1960280	97,4	1188801	59,1	38,3
BGHZ0463	<i>C. annuum</i>	2199697	96,4	1959014	89,1	97,7	1913674	97,7	1276182	65,1	32,6
CGN17193	<i>C. annuum</i>	2199723	97,7	1961966	89,2	98,6	1927165	98,2	1373988	70	28,2
CGN21527	<i>C. annuum</i>	2205405	96,3	2040698	92,5	97,7	2003403	98,2	1401615	68,7	29,5
CGN17234	<i>C. annuum</i>	2205994	96,3	2011692	91,2	97,6	1973881	98,1	1287030	64	34,1
BGHZ0213	<i>C. annuum</i>	2211403	96,3	1951420	88,2	97,7	1911920	98	1105158	56,6	41,4
BGHZ0287	<i>C. annuum</i>	2211682	96,2	1874038	84,7	97,5	1841121	98,2	1143791	61	37,2
BGHZ0413	<i>C. annuum</i>	2212086	96,6	2002633	90,5	97,8	1956255	97,7	1215396	60,7	37
BGHZ2154	<i>C. annuum</i>	2212778	95,7	1920040	86,8	97,4	1883658	98,1	1158500	60,3	37,8
BPGV03038	<i>C. annuum</i>	2223869	95,8	2039696	91,7	97,3	2012644	98,7	1011231	49,6	49,1
CGN22127	<i>C. annuum</i>	2233182	96,7	2064501	92,4	97,9	2033214	98,5	1403229	68	30,5
BGHZ0971	<i>C. annuum</i>	2243092	96,5	1956568	87,2	97,7	1917395	98	1142599	58,4	39,6

Tabla A.4. Estadísticas de las lecturas de GBS empleadas. Lecturas crudas: lecturas obtenidas tras el proceso de secuenciación, sin realizar ningún tratamiento. (continuación).

Accesión	Especie	Lecturas crudas		Lecturas después de la limpieza con Trimmomatic			Lecturas mapeadas				
		Total	% lecturas con calidad (valor Phred)>30	Total	% lecturas con respecto al total de lecturas crudas	% lecturas con calidad (valor Phred)>30	Total	% lecturas mapeadas con respecto a las conservadas después de la limpieza	Lecturas con calidad Mapq>57	% con calidad>57 mapeadas con respecto a las conservadas después de la limpieza	% mapeadas que se descartan por tener calidad Mapq<57
BPGV05740	<i>C. annuum</i>	2244196	95,1	1236034	55,1	97,5	1212285	98,1	768735	62,2	35,9
PI439212	<i>C. annuum</i>	2247043	96,8	2102394	93,6	98	2073422	98,6	1468311	69,8	28,8
BGHZ0208	<i>C. annuum</i>	2247579	95,8	1965808	87,5	97,3	1927738	98,1	1204244	61,3	36,8
BGV013687	<i>C. annuum</i>	2253251	97	2128136	94,4	98,1	2093407	98,4	1508438	70,9	27,5
BGHZ0282	<i>C. annuum</i>	2265327	96,3	2081133	91,9	97,6	2036101	97,8	1233913	59,3	38,5
CAP607	<i>C. annuum</i>	2271651	96,7	2142943	94,3	98	2104822	98,2	1413648	66	32,2
BGHZ0280	<i>C. annuum</i>	2278610	96,3	2009273	88,2	97,5	1972080	98,1	1221226	60,8	37,3
BGHZ0106	<i>C. annuum</i>	2283515	96,5	2079783	91,1	97,8	2036537	97,9	1343707	64,6	33,3
CGN16807	<i>C. annuum</i>	2304567	96,7	2128520	92,4	97,9	2091819	98,3	1408630	66,2	32,1
CGN16817	<i>C. annuum</i>	2317018	96,7	2147785	92,7	97,9	2117669	98,6	1505984	70,1	28,5
CGN23235	<i>C. annuum</i>	2318065	98,1	2152440	92,9	98,8	2118428	98,4	1401773	65,1	33,3
BGHZ0206	<i>C. annuum</i>	2318290	96,4	1984746	85,6	97,6	1950633	98,3	1278588	64,4	33,9
BGHZ0102	<i>C. annuum</i>	2318615	96	2040884	88	97,4	2008369	98,4	1300150	63,7	34,7
BGHZ0099	<i>C. annuum</i>	2325004	96,5	2122381	91,3	97,7	2084133	98,2	1244203	58,6	39,6
BGHZ0104	<i>C. annuum</i>	2325481	96,1	2002420	86,1	97,5	1961979	98	1267772	63,3	34,7
CGN17227	<i>C. annuum</i>	2333077	96,2	2156435	92,4	97,6	2119816	98,3	1512947	70,2	28,1
CAP1051	<i>C. annuum</i>	2350076	97,3	1933227	82,3	98,3	1903730	98,5	1238273	64,1	34,4
BGHZ0179	<i>C. annuum</i>	2405892	96,5	2179032	90,6	97,8	2140804	98,2	1414167	64,9	33,3
BGHZ0418	<i>C. annuum</i>	2445356	96,3	2239996	91,6	97,6	2203617	98,4	1291848	57,7	40,7
BGV013691	<i>C. annuum</i>	2456455	96,3	2261792	92,1	97,7	2222032	98,2	1595511	70,5	27,7
CAP1254	<i>C. annuum</i>	2548063	96,5	2337647	91,7	97,9	2309318	98,8	1517132	64,9	33,9
CAP1538	<i>C. annuum</i>	2620360	97,3	2488292	95	98,3	2454920	98,7	1596496	64,2	34,5
CAP68	<i>C. annuum</i>	2859658	97,5	2728168	95,4	98,4	2675971	98,1	1791214	65,7	32,4
CAP609	<i>C. annuum</i>	2877229	97,4	2728111	94,8	98,4	2677485	98,1	1535079	56,3	41,8
CAP1049	<i>C. annuum</i>	3002467	97,3	2295647	76,5	98,5	2252331	98,1	1500620	65,4	32,7
CAP913	<i>C. annuum</i>	3013538	97,8	2850644	94,6	98,6	2800453	98,2	1722925	60,4	37,8
C94	<i>C. annuum</i>	3710731	96	3182455	85,8	97,4	3124894	98,2	2046167	64,3	33,9
CAP1371	<i>C. annuum</i>	3975300	95,8	3209528	80,7	97,4	3138696	97,8	2155052	67,1	30,7
C204	<i>C. annuum</i>	4088076	95,8	3272957	80,1	97,4	3220258	98,4	2028060	62	36,4
BPGV12950	<i>C. annuum</i>	4106073	96,2	3397644	82,7	97,6	3321778	97,8	2076205	61,1	36,7
C141	<i>C. annuum</i>	4231480	96,2	3339494	78,9	97,7	3265093	97,8	2291982	68,6	29,2
BGV006008	<i>C. annuum</i>	4329982	97,1	4104603	94,8	98,2	3916654	95,4	2584147	63	32,4
CGN21460	<i>C. annuum</i>	4236162	96	3519504	83,1	97,5	3421030	97,2	2225531	63,2	34
CGN22102	<i>C. annuum</i>	4236202	96,4	3561443	84,1	97,7	3441081	96,6	2271289	63,8	32,8
PI585246	<i>C. annuum</i>	4245879	96,9	3921675	92,4	98,1	3875501	98,8	2618887	66,8	32
CGN23765	<i>C. annuum</i>	4337366	96,1	3531648	81,4	97,5	3461025	98	2237377	63,4	34,6
BPGV12348	<i>C. annuum</i>	4338260	96,1	3788988	87,3	97,6	3712808	98	2210449	58,3	39,7
C87	<i>C. annuum</i>	4341634	96,4	3671409	84,6	97,7	3612344	98,4	2290946	62,4	36
CGN22161	<i>C. annuum</i>	4349949	96,3	3675055	84,5	97,6	3607900	98,2	2223687	60,5	37,7
CGN24355	<i>C. annuum</i>	4357301	95,9	3692231	84,7	97,4	3624555	98,2	2181077	59,1	39,1
CGN23254	<i>C. annuum</i>	4382553	95,8	3706675	84,6	97,3	3636180	98,1	2224578	60	38,1
PI585238	<i>C. annuum</i>	4404427	96,7	4080792	92,7	98	4019681	98,5	2889401	70,8	27,7
BGV005857	<i>C. annuum</i>	4405030	97	4141562	94	98,1	4084673	98,6	2921760	70,5	28,1
BPGV13027	<i>C. annuum</i>	4407176	96,1	3767087	85,5	97,5	3697501	98,2	2213970	58,8	39,4
CGN21526	<i>C. annuum</i>	4415135	96,4	3825049	86,6	97,7	3730717	97,5	2454471	64,2	33,3
BPGV11849	<i>C. annuum</i>	4417023	96,4	3682007	83,4	97,7	3629997	98,6	2207844	60	38,6
BPGV13113	<i>C. annuum</i>	4466794	95,8	3890133	87,1	97,4	3816008	98,1	2187154	56,2	41,9
CGN16899	<i>C. annuum</i>	4471614	95,6	3921664	87,7	97,3	3847289	98,1	2345956	59,8	38,3
BGV006055	<i>C. annuum</i>	4483486	96,9	4168574	93	98	4106297	98,5	2760844	66,2	32,3
BGV005981	<i>C. annuum</i>	4493991	96,8	4212428	93,7	98	4140817	98,3	3040631	72,2	26,1
CAP1539	<i>C. annuum</i>	4640925	96,3	3918413	84,4	97,6	3846922	98,2	2225948	56,8	41,4
BGHZ0082	<i>C. annuum</i>	4680366	95,5	2706871	57,8	97,5	2661217	98,3	1705600	63	35,3
Media	<i>C. annuum</i>	2425718,9	96,6	2127791,4	88,4	97,9	2087479,8	98	1353711,6	63,6	34,3

Tabla A.4. Estadísticas de las lecturas de GBS empleadas. Lecturas crudas: lecturas obtenidas tras el proceso de secuenciación, sin realizar ningún tratamiento. (continuación).

Accesión	Especie	Lecturas crudas			Lecturas después de la limpieza con Trimmomatic			Lecturas mapeadas			
		Total	% lecturas con calidad (valor Phred)>30	% lecturas con respecto al total de lecturas crudas	Total	% lecturas con calidad (valor Phred)>30	Total	% lecturas mapeadas con respecto a las conservadas después de la limpieza	Lecturas con calidad Mapq>57	% con calidades>57 mapeadas con respecto a las conservadas después de la limpieza	% mapeadas que se descartan por tener calidad Mapq<57
PI585252	<i>C. chinense</i>	4506360	96,6	4250468	94,3	97,9	4033506	94,9	2673731	62,9	32
PI585253	<i>C. chinense</i>	4246061	96,5	3986114	93,9	97,8	3783023	94,9	2483339	62,3	32,6
PI593933	<i>C. chinense</i>	4246229	96,9	3952360	93,1	98,1	3759019	95,1	2355770	59,6	35,5
PI360725	<i>C. chinense</i>	4276327	96,7	3932841	92	97,9	3764095	95,7	2471452	62,8	32,9
CGN23259	<i>C. chinense</i>	4312963	96,6	3960833	91,8	97,9	3777528	95,4	2370650	59,9	35,5
CGN17040	<i>C. chinense</i>	4318441	96,8	3996850	92,6	98,1	3757077	94	2334596	58,4	35,6
PI224427	<i>C. chinense</i>	4329413	96,8	3953617	91,3	98	3768356	95,3	2504997	63,4	31,9
BGV005890	<i>C. chinense</i>	4329714	97,1	3987143	92,1	98,2	3767749	94,5	2523972	63,3	31,2
PI585278	<i>C. chinense</i>	4482612	96,7	4239419	94,6	98	4025624	95	2682841	63,3	31,7
PI593922	<i>C. chinense</i>	4424120	96,8	4091360	92,5	98	3844901	94	2507935	61,3	32,7
CAP868	<i>C. chinense</i>	2258484	96,3	2102272	93,1	97,7	1997394	95	1370121	65,2	29,8
CGN17045	<i>C. chinense</i>	1102514	97,8	1042196	94,5	98,6	980450	94,1	589950	56,6	37,5
CGN19211	<i>C. chinense</i>	1058597	97,9	1008321	95,3	98,6	954411	94,7	547337	54,3	40,4
CGN17220	<i>C. chinense</i>	2105528	96,5	1948598	92,5	97,8	1863112	95,6	1202508	61,7	33,9
PI543203	<i>C. chinense</i>	2129862	96,9	2020694	94,9	98	1908819	94,5	1337965	66,2	28,3
PI543184	<i>C. chinense</i>	2130978	97,1	1951465	91,6	98,1	1830087	93,8	1081200	55,4	38,4
PI593929	<i>C. chinense</i>	4380088	97,1	4079269	93,1	98,1	3862893	94,7	2544312	62,4	32,3
CAP472	<i>C. chinense</i>	2227876	97,1	2071460	93	98,2	1964854	94,9	1288675	62,2	32,7
PI315020	<i>C. chinense</i>	2231328	97,2	2111805	94,6	98,3	2004815	94,9	1307271	61,9	33
CGN23565	<i>C. chinense</i>	2218014	96,6	2062974	93	97,9	1957597	94,9	1311935	63,6	31,3
BGV013705	<i>C. chinense</i>	2211427	97,3	2052146	92,8	98,3	1949358	95	1267399	61,8	33,2
CAP1364	<i>C. chinense</i>	2172809	96,8	1984855	91,3	98	1881852	94,8	1241713	62,6	32,2
PI260478	<i>C. chinense</i>	2157239	96,8	1993815	92,4	98	1894493	95	1273587	63,9	31,1
BGV14749	<i>C. chinense</i>	2143455	97,2	1991752	92,9	98,3	1882093	94,5	1306873	65,6	28,9
PI315015	<i>C. chinense</i>	2168534	96,8	2034910	93,8	98	1931810	94,9	1268340	62,3	32,6
CGN20808	<i>C. chinense</i>	2163101	96,4	1948491	90,1	97,9	1895601	97,3	1383770	71	26,3
Media	<i>C. chinense</i>	3012772,1	96,9	2798308,8	93	98,1	2655404,5	94,9	1739701,5	62,1	32,8
BGV011957	<i>C. baccatum</i>	4515654	97,2	4286087	94,9	98,2	3964545	92,5	2463014	57,5	35
BGV13703	<i>C. baccatum</i>	1147947	97,7	1050892	91,5	98,6	976377	92,9	587489	55,9	37
BGV005852	<i>C. baccatum</i>	4395499	96,6	4129645	94	97,9	3824385	92,6	2372744	57,5	35,1
BGV008233	<i>C. baccatum</i>	2344032	97	2149112	91,7	98,1	2003985	93,2	1191639	55,4	37,8
PI585249	<i>C. baccatum</i>	2312009	96,7	2180196	94,3	97,9	1998781	91,7	1222095	56,1	35,6
PI585239	<i>C. baccatum</i>	2184332	96,8	2035894	93,2	98	1866629	91,7	1167773	57,4	34,3
CAP1366	<i>C. baccatum</i>	2188755	96,6	2066668	94,4	97,9	1886218	91,3	1134208	54,9	36,4
PI257135	<i>C. baccatum</i>	2188861	97,2	2021855	92,4	98,2	1874777	92,7	1132928	56	36,7
PI215739	<i>C. baccatum</i>	2189344	97,4	2022235	92,4	98,4	1849583	91,5	1209883	59,8	31,7
BGV013986	<i>C. baccatum</i>	2175725	97,2	2046559	94,1	98,3	1893891	92,5	1200860	58,7	33,8
PI585244	<i>C. baccatum</i>	2177230	96,2	1992524	91,5	97,7	1773727	89	933495	46,8	42,2
PI257133	<i>C. baccatum</i>	2274216	96,6	2125164	93,4	97,9	1927893	90,7	1187094	55,9	34,8
PI355813	<i>C. baccatum</i>	2277195	95,9	2136010	93,8	97,5	1951005	91,3	1206240	56,5	34,8
PI595905	<i>C. baccatum</i>	2203719	96,9	2013355	91,4	98,1	1857019	92,2	1139617	56,6	35,6
PI596061	<i>C. baccatum</i>	2269423	96,7	2100663	92,6	98	1941760	92,4	1081152	51,5	40,9
CGN23205	<i>C. baccatum</i>	2117222	97,1	1982868	93,7	98,2	1839200	92,8	1139110	57,4	35,4
BGV006064	<i>C. baccatum</i>	2260504	96,6	2066016	91,4	97,9	1921267	93	1192560	57,7	35,3
CGN17174	<i>C. baccatum</i>	2255753	96,8	2121724	94,1	98	1942772	91,6	1190538	56,1	35,5
PI290982	<i>C. baccatum</i>	2222171	97,1	2010221	90,5	98,1	1855161	92,3	1079443	53,7	38,6
PI590505	<i>C. baccatum</i>	2215919	96,3	2084429	94,1	97,7	1955239	93,8	1187453	57	36,8
PI585241	<i>C. baccatum</i>	2147717	97	2001133	93,2	98,1	1847711	92,3	1077794	53,9	38,4
PI593932	<i>C. baccatum</i>	2151261	97,2	1969018	91,5	98,3	1825875	92,7	1150115	58,4	34,3
CGN21479	<i>C. baccatum</i>	2121380	96,7	1976520	93,2	97,9	1819970	92,1	1181130	59,8	32,3
CGN17042	<i>C. baccatum</i>	2131751	96,5	1949339	91,4	97,9	1790635	91,9	1125063	57,7	34,2
Media	<i>C. baccatum</i>	2702205,2	96,9	2511224,8	92,9	98,1	2354594,6	93,6	1500497,6	59,3	34,3

Tabla A.4. Estadísticas de las lecturas de GBS empleadas. Lecturas crudas: lecturas obtenidas tras el proceso de secuenciación, sin realizar ningún tratamiento. (continuación).

Accesión	Especie	Lecturas crudas			Lecturas después de la limpieza con Trimmomatic			Lecturas mapeadas			
		Total	% lecturas con calidad (valor Phred)>30	% lecturas con respecto al total de lecturas crudas	Total	% lecturas con calidad (valor Phred)>30	Total	% lecturas mapeadas con respecto a las conservadas después de la limpieza	Lecturas con calidad Mapq>57	% con calidades>57 mapeadas con respecto a las conservadas después de la limpieza	% mapeadas que se descartan por tener calidad Mapq<57
PI585275	<i>C. pubescens</i>	2543799	96,9	2398057	94,3	98,1	2115401	88,2	1198549	50	38,2
CAP217	<i>C. pubescens</i>	2258811	96,4	2113360	93,6	97,8	1859840	88	972835	46	42
BGV013300	<i>C. pubescens</i>	2359383	96,7	2180021	92,4	98	1932444	88,6	1114571	51,1	37,5
PI355394	<i>C. pubescens</i>	2229746	96,4	2044232	91,7	97,8	1831053	89,6	1040926	50,9	38,7
PI585271	<i>C. pubescens</i>	2210985	97	2082733	94,2	98,1	1840243	88,4	998121	47,9	40,5
CAP1486	<i>C. pubescens</i>	2205683	97	2069489	93,8	98,1	1818331	87,9	975367	47,1	40,8
PI585269	<i>C. pubescens</i>	2192897	96,5	2007638	91,6	97,9	1769007	88,1	937769	46,7	41,4
CGN22108	<i>C. pubescens</i>	2181646	96,2	2016911	92,4	97,7	1779744	88,2	991308	49,1	39,1
BGV013702	<i>C. pubescens</i>	2148957	96,6	1939688	90,3	97,9	1718736	88,6	829002	42,7	45,9
CAP357	<i>C. pubescens</i>	1097060	97,8	1000011	91,2	98,7	886180	88,6	440776	44,1	44,5
CGN23768	<i>C. pubescens</i>	657838	97,9	629890	95,8	98,7	560906	89	243148	38,6	50,4
PI585262	<i>C. pubescens</i>	2083624	97	1911449	91,7	98,1	1719026	89,9	931120	48,7	41,2
PI585273	<i>C. pubescens</i>	2117494	96,7	1964183	92,8	98	1741956	88,7	968104	49,3	39,4
Media	<i>C. pubescens</i>	2022147,9	96,9	1873666,3	92,8	98,1	1659451,3	88,6	895507,4	47,1	41,5
PI560941	<i>C. frutescens</i>	2140736	96,8	1997658	93,3	98	1901784	95,2	1318477	66	29,2
PI595907	<i>C. frutescens</i>	2213364	97	2078630	93,9	98,1	1947101	93,7	1343672	64,6	29,1
PI585256	<i>C. frutescens</i>	4332761	96,9	4004864	92,4	98,1	3812203	95,2	2626208	65,6	29,6
PI585254	<i>C. frutescens</i>	4470425	96,6	4137579	92,6	97,9	3946967	95,4	2686811	64,9	30,5
PI585257	<i>C. frutescens</i>	4447869	96,9	4079793	91,7	98,1	3857890	94,6	2620144	64,2	30,4
PI355808	<i>C. frutescens</i>	3905621	96,5	3530293	90,4	97,8	3352262	95	2369390	67,1	27,9
PI593924	<i>C. frutescens</i>	2224751	97,1	2078297	93,4	98,2	1975162	95	1366570	65,8	29,2
PI593920	<i>C. frutescens</i>	1128269	97,7	1046294	92,7	98,6	996311	95,2	640297	61,2	34
BGV13699	<i>C. frutescens</i>	1148324	97,6	943228	82,1	98,5	887469	94,1	509564	54	40,1
PI439488	<i>C. frutescens</i>	2041189	96,8	1884065	92,3	98	1783516	94,7	1183649	62,8	31,9
CGN22792	<i>C. frutescens</i>	2171980	97,2	2041101	94	98,3	1914039	93,8	1243705	60,9	32,9
Media	<i>C. frutescens</i>	2747753,5	97	2529254,7	91,7	98,1	2397700,4	94,7	1628044,3	63,4	31,3
PM670	<i>C. annuum</i> var. <i>glabriusculum</i>	4238768	96,1	3433139	81	97,5	3225655	94	1940103	56,5	37,5
PM669	<i>C. annuum</i> var. <i>glabriusculum</i>	4232991	96,4	3521972	83,2	97,6	3271020	92,9	1707929	48,5	44,4
PM647	<i>C. annuum</i> var. <i>glabriusculum</i>	3857858	95,9	3076658	79,8	97,5	3010385	97,8	2048707	66,6	31,2
PM663	<i>C. annuum</i> var. <i>glabriusculum</i>	1104502	98	1056245	95,6	98,7	1031503	97,7	635798	60,2	37,5
Media	<i>C. annuum</i> var. <i>glabriusculum</i>	3358529,8	96,6	2772003,5	84,9	97,8	2634640,8	95,6	1583134,3	58	37,7

Tabla A.5. Datos de fenotipado de cada accesión para los parámetros cualitativos de planta y fruto.

Accesión	CT	AN	FT	PBT	HC	CM	FF	FFU	CBF	FAF	AF	TE
BGHZ0082	1	5	2	3	7	9	3	4	0	2	0	1
BGHZ0097	1	1	2	3	7	7	3	2	0	3	0	2
BGHZ0102	1	1	2	3	7	8	3	5	0	2	0	2
BGHZ0106	1	1	2	3	7	8	1	2	0	1	0	3
BGHZ0113	1	1	2	3	7	9	3	5	0	3	0	2
BGHZ0179	1	1	2	3	7	8	5	4	0	4	0	2
BGHZ0181	1	5	2	3	7	8	5	3	0	3	0	3
BGHZ0206	1	1	2	3	7	8	5	5	0	3	0	2
BGHZ0213	1	1	2	3	7	8	3	4	0	2	0	1
BGHZ0224	1	1	2	3	7	8	2	3	0	2	0	1
BGHZ0280	1	5	2	3	7	8	2	2	0	2	0	1
BGHZ0282	1	1	2	5	7	8	1	3	0	3	0	2
BGHZ0284	1	1	2	3	5	8	1	2	0	1	0	2
BGHZ0337	1	1	2	3	7	8	5	4	0	3	0	2
BGHZ0338	1	5	2	3	5	8	3	5	0	3	0	3
BGHZ0413	1	7	2	3	7	8	3	5	0	2	0	2
BGHZ0418	-	-	-	-	-	-	-	-	-	-	-	-
BGHZ0447	1	1	2	3	7	8	3	3	0	3	0	2
BGHZ0463	1	5	2	3	7	11	4	3	0	3	0	3
BGHZ0775	1	1	2	3	7	9	5	5	0	3	0	2
BGHZ0782	1	5	2	3	7	6	3	2	0	2	0	1
BGHZ0958	-	-	-	-	-	-	-	-	-	-	-	-
BGHZ0969	1	1	2	3	7	8	3	3	0	2	0	1
BGHZ0971	1	1	2	3	7	8	3	3	0	3	0	2
BGHZ1109	1	1	2	3	7	8	5	5	0	3	0	2
BGHZ1114	1	1	2	3	7	8	1	2	0	1	0	1
BGHZ1174	1	5	2	3	7	8	5	3	0	3	0	2
BGHZ1375	1	1	2	3	7	8	1	1	0	2	0	3
BGHZ1521	1	1	2	3	7	9	1	2	1	1	0	3
BGHZ1753	1	5	2	3	7	8	1	2	0	1	0	3
BGHZ2154	1	1	2	3	7	8	1	3	0	2	0	2
BGHZ2546	1	1	2	3	7	8	3	5	0	3	0	1
BGHZ2561	1	1	2	3	7	8	2	5	0	3	0	3
BGHZ3337	1	5	2	3	7	9	3	3	0	3	0	2
BGHZ5912	1	5	2	3	7	8	5	5	0	3	0	1
BR113-05	1	1	2	3	5	7	5	5	0	4	0	2
C-431	1	5	2	3	7	8	2	4	0	3	0	2
CAP1661	1	5	2	3	7	9	3	5	0	3	0	2

Valores de los descriptores según IPGRI (1995). CT: color del tallo. AN: antocianina del nudo. FT: forma del tallo. PBT: pubescencia del tallo. HC: hábito de crecimiento de la planta. CM: color del fruto en estado maduro. FF: forma del fruto. FFU: forma del fruto en la unión con el pedicelo. CBF: cuello en la base del fruto. FAF: forma del ápice del fruto. AF: apéndice en el fruto. TE: tipo de epidermis del fruto.

Tabla A.5. Datos de fenotipado de cada accesión para los parámetros cualitativos de planta y fruto. (continuación).

Accesión	CT	AN	FT	PBT	HC	CM	FF	FFU	CBF	FAF	AF	TE
CGN17234	1	1	2	3	7	9	1	4	0	2	0	2
CGN17235	1	1	2	3	7	8	1	2	0	1	0	1
CGN20507	1	1	2	3	7	9	5	4	0	3	0	3
CGN23293	1	1	2	3	7	9	5	5	0	3	0	2
OI27-05	1	3	2	3	7	8	3	5	0	2	0	2
PA-145	1	1	2	3	7	8	5	4	0	3	0	2
PI309-05	1	1	2	3	7	8	3	4	0	2	0	1
CAP1371	1	1	2	3	7	9	3	4	0	3	0	3
CAP1539	1	5	2	3	7	9	1	2	0	1	0	2
CGN16805	1	7	2	3	7	8	3	4	0	2	0	1
CGN16897	1	7	2	3	7	8	5	5	0	3	0	2
CGN16899	2	5	2	3	7	9	2	4	0	3	0	3
CGN21475	2	5	2	3	7	9	2	4	0	3	0	2
CGN22161	1	1	2	3	5	8	5	5	0	4	0	2
C-87	1	1	2	3	3	7	3	5	0	2	1	2
C-94	1	7	2	5	7	8	5	3	0	3	0	2
C-204	1	7	2	7	7	8	3	3	0	2	0	1
CGN21460	2	7	1	3	7	8	1	2	0	1	0	1
CGN21526	1	1	2	7	7	8	3	2	0	1	0	1
CGN21527	1	1	2	7	5	7	1	2	0	1	0	2
CGN22102	1	7	2	3	7	8	3	4	0	2	0	2
CGN22783	1	5	2	3	7	8	1	2	0	1	0	1
CGN23254	1	3	1	3	5	9	3	4	0	2	0	2
CGN24355	1	1	2	5	7	8	1	1	0	1	0	2
BGV005857	1	1	1	3	5	8	3	3	0	3	0	2
BGV005981	1	7	1	3	5	9	5	3	0	3	0	2
BGV006008	1	1	2	5	5	8	5	4	0	3	0	2
BGV006055	1	1	2	3	5	8	3	3	1	2	0	2
PI241670	1	7	1	3	5	8	1	2	0	1	0	2
PI585238	1	1	1	3	5	8	1	2	0	1	0	2
PI585246	1	7	1	3	5	8	5	2	0	3	0	1
BGV013687	1	1	1	3	5	8	5	5	0	4	0	1
BGV013691	1	1	1	3	5	8	3	4	0	1	0	1
CAP524	2	7	1	3	7	8	1	2	0	2	0	1
CGN20808	1	7	2	7	7	8	1	1	0	1	0	1
CGN22155	1	7	1	3	5	9	1	2	0	2	0	1
PI213915	1	1	2	3	7	8	1	2	0	1	0	2
PI439212	1	1	1	3	5	8	1	2	0	1	0	1

Valores de los descriptores según IPGRI (1995). CT: color del tallo. AN: antocianina del nudo. FT: forma del tallo. PBT: pubescencia del tallo. HC: hábito de crecimiento de la planta. CM: color del fruto en estado maduro. FF: forma del fruto. FFU: forma del fruto en la unión con el pedicelo. CBF: cuello en la base del fruto. FAF: forma del ápice del fruto. AF: apéndice en el fruto. TE: tipo de epidermis del fruto.

Tabla A.5. Datos de fenotipado de cada accesión para los parámetros cualitativos de planta y fruto. (continuación).

Accesión	CT	AN	FT	PBT	HC	CM	FF	FFU	CBF	FAF	AF	TE
PI543188	1	7	1	3	3	8	3	2	0	2	0	1
CGN17172	1	5	2	3	7	8	3	4	0	3	0	3
CGN21486	1	1	2	3	7	9	5	4	0	4	0	3
CGN23239	1	1	2	3	7	8	3	4	0	3	0	3
CAP68	1	3	2	3	7	8	2	4	0	3	0	2
CAP242	1	5	2	3	7	8	5	5	0	3	0	2
CAP913	1	1	2	3	7	4	3	4	0	2	0	2
CGN16800	1	1	2	3	7	8	2	3	0	3	0	2
CGN16881	1	1	2	3	7	4	5	5	0	2	1	1
CGN16835	1	3	2	3	7	8	1	2	0	1	0	3
CGN17193	1	1	2	3	7	8	3	3	0	2	0	2
CGN23235	1	1	2	3	7	9	5	5	0	3	0	2
CGN21581	1	1	2	3	7	9	3	4	0	3	0	2
CGN16916	1	1	2	3	7	8	5	5	0	4	0	3
CGN16917	1	5	2	3	7	9	5	5	0	3	0	2
CGN21480	1	1	2	3	7	7	3	5	0	3	0	2
CGN22825	1	1	2	3	7	8	1	2	0	1	0	2
CGN16833	1	5	2	3	7	8	5	4	0	3	0	3
CGN17010	1	5	2	3	7	8	5	4	0	3	0	2
CGN17190	2	5	2	3	7	8	5	5	0	4	0	2
CGN17191	1	5	2	3	7	8	3	4	0	1	0	2
CGN22154	1	1	2	3	7	8	5	5	0	3	0	3
CGN24358	1	5	2	3	7	9	1	2	1	1	0	3
CAP1049	1	3	2	3	7	8	3	2	0	3	0	2
CAP1051	1	5	2	3	7	8	1	1	1	1	0	3
CAP491	1	1	2	3	7	8	1	2	1	1	0	3
CAP1254	1	1	2	3	7	2	1	1	0	1	0	2
CGN16808	1	1	2	3	7	9	3	2	0	1	0	2
CGN17214	1	1	2	3	7	9	2	5	0	4	0	3
CGN16807	1	7	2	3	7	8	2	4	0	3	0	2
CGN16817	1	1	2	3	7	8	2	5	0	4	0	3
CGN17147	1	1	2	3	7	9	3	3	0	2	0	1
CGN17227	1	7	2	3	7	8	3	5	0	1	1	1
CGN22127	1	1	2	3	7	9	1	2	0	2	0	2
CAP607	1	3	2	3	7	9	3	2	0	2	0	2
CAP609	1	3	2	3	7	8	5	3	0	3	0	3
CAP614	1	1	2	3	7	8	3	3	0	1	0	2
CAP977	1	3	2	3	7	9	5	3	0	3	0	3

Valores de los descriptores según IPGRI (1995). CT: color del tallo. AN: antocianina del nudo. FT: forma del tallo. PBT: pubescencia del tallo. HC: hábito de crecimiento de la planta. CM: color del fruto en estado maduro. FF: forma del fruto. FFU: forma del fruto en la unión con el pedicelo. CBF: cuello en la base del fruto. FAF: forma del ápice del fruto. AF: apéndice en el fruto. TE: tipo de epidermis del fruto.

Tabla A.5. Datos de fenotipado de cada accesión para los parámetros cualitativos de planta y fruto. (continuación).

Accesión	CT	AN	FT	PBT	HC	CM	FF	FFU	CBF	FAF	AF	TE
CAP979	1	1	2	3	7	8	3	4	0	3	0	3
CAP1538	1	1	2	3	7	8	5	4	0	4	0	2
PM647	2	5	2	5	7	8	1	2	0	1	0	2
PM670	1	5	2	3	7	8	2	2	0	2	0	2
CGN23259	1	1	1	3	7	6	5	2	0	2	0	2
CGN17040	1	1	1	3	7	5	2	2	0	3	0	1
PI593922	1	1	1	3	7	8	3	3	0	1	0	2
PI585252	1	1	2	3	5	9	4	2	0	1	0	2
PI593929	1	1	1	3	7	7	4	3	0	3	1	1
PI585278	1	7	2	3	5	8	3	4	0	3	0	2
PI585253	1	1	1	3	7	8	3	2	0	2	0	2
BGV005890	1	1	1	3	7	7	3	3	0	3	0	2
PI224427	1	1	2	3	7	7	5	2	0	3	0	3
PI360725	1	7	1	3	7	8	1	1	0	1	0	2
PI593933	1	1	2	3	5	4	3	3	0	2	1	2
BGV013705	1	1	2	5	5	8	3	3	0	1	0	2
CAP472	1	7	1	3	5	2	2	1	0	2	0	1
CAP868	1	1	1	5	5	2	3	4	1	1	0	3
CAP1364	2	7	1	3	7	6	1	1	0	1	0	3
CGN17220	1	1	2	5	3	8	5	3	0	3	0	3
CGN19211	1	1	1	3	5	6	4	3	0	3	0	3
CGN23565	1	1	2	5	7	3	1	1	1	1	0	2
PI260478	1	1	1	5	7	8	4	2	0	1	0	2
PI315015	1	1	2	3	7	11	1	2	0	1	0	2
PI315020	1	1	2	3	7	8	3	3	0	3	0	3
BGV014749	1	1	1	3	5	7	2	3	0	3	1	1
CGN17045	1	1	1	3	5	8	3	4	0	3	0	2
PI543184	1	1	1	3	7	7	3	3	1	3	0	2
PI543203	1	7	1	3	5	1	1	2	0	3	0	3
PI355808	1	1	1	3	7	-	-	-	-	-	-	-
PI585254	1	1	1	3	7	-	-	-	-	-	-	-
PI585256	1	1	2	3	7	-	-	-	-	-	-	-
PI585257	1	1	1	3	7	-	-	-	-	-	-	-
PI593920	1	1	1	3	7	8	5	1	0	3	0	3
PI593924	1	1	1	3	7	-	-	-	-	-	-	-
PI595907	1	1	1	3	7	7	3	2	0	2	0	1
BGV013699	1	1	1	3	7	-	-	-	-	-	-	-
CGN22792	1	1	1	3	7	-	-	-	-	-	-	-

Valores de los descriptores según IPGRI (1995). CT: color del tallo. AN: antocianina del nudo. FT: forma del tallo. PBT: pubescencia del tallo. HC: hábito de crecimiento de la planta. CM: color del fruto en estado maduro. FF: forma del fruto. FFU: forma del fruto en la unión con el pedicelo. CBF: cuello en la base del fruto. FAF: forma del ápice del fruto. AF: apéndice en el fruto. TE: tipo de epidermis del fruto.

Tabla A.5. Datos de fenotipado de cada accesión para los parámetros cualitativos de planta y fruto. (continuación).

Accesión	CT	AN	FT	PBT	HC	CM	FF	FFU	CBF	FAF	AF	TE
PI439488	1	1	1	3	7	-	-	-	-	-	-	-
PI560941	1	1	1	3	7	-	-	-	-	-	-	-
PI573337	1	7	1	3	5	1	1	2	1	2	0	3
BGV005852	1	1	1	3	7	8	3	1	0	2	0	1
BGV006064	1	1	2	3	7	8	3	2	0	1	0	2
BGV011957	1	1	1	3	5	8	1	2	0	3	0	3
PI257133	1	1	2	3	7	8	5	3	0	3	0	2
PI257135	1	1	2	3	7	8	1	2	0	1	0	2
PI355813	1	1	2	3	7	8	3	2	0	1	0	2
PI585239	1	1	2	3	7	8	1	2	0	3	0	2
PI585244	1	1	1	3	7	4	2	2	0	2	0	2
PI585249	1	1	2	3	7	8	3	1	0	1	0	2
PI593932	1	1	2	3	7	7	1	2	0	2	0	1
PI595905	1	1	1	3	5	8	1	2	0	3	0	3
BGV013703	1	1	1	3	5	8	1	2	0	1	0	1
CAP1366	1	1	2	3	7	6	1	1	0	3	0	2
CGN17174	1	1	1	3	5	8	4	2	0	2	0	2
CGN21479	1	1	2	3	5	6	1	2	0	2	0	2
PI215739	1	1	1	3	5	8	3	1	0	1	0	1
PI290982	1	1	2	3	7	6	3	2	0	3	0	2
BGV008233	1	1	2	3	7	6	1	2	0	1	0	2
BGV013986	1	1	2	3	5	8	1	2	0	2	0	1
CGN17042	1	1	2	3	7	8	1	2	0	2	0	2
CGN23205	3	7	2	5	5	8	3	2	0	2	0	1
PI590505	1	1	2	3	5	6	3	2	0	1	0	1
PI596061	1	1	2	3	5	8	1	2	0	3	0	2
BGV013300	1	1	1	3	5	9	5	3	0	2	0	1
PI355394	1	7	1	5	5	8	5	4	0	2	0	1
PI585262	1	7	1	3	3	8	4	4	1	3	0	2
PI585269	1	7	2	5	3	9	4	3	0	2	0	1
PI585271	1	7	1	5	3	8	3	3	0	1	0	1
PI585273	1	1	1	3	5	8	5	3	0	2	0	1
PI585275	1	7	1	5	3	8	4	3	1	2	0	1
BGV013702	1	5	1	5	3	-	-	-	-	-	-	-
CAP217	2	7	2	7	3	8	4	4	1	3	0	1
CAP357	1	1	1	7	3	6	4	3	0	3	0	1
CAP1486	3	7	2	7	3	9	5	4	0	3	0	2
CGN22108	1	1	2	7	5	9	4	4	1	2	0	2
CGN23768	1	7	1	7	3	2	5	4	0	3	0	2

Valores de los descriptores según IPGRI (1995). CT: color del tallo. AN: antocianina del nudo. FT: forma del tallo. PBT: pubescencia del tallo. HC: hábito de crecimiento de la planta. CM: color del fruto en estado maduro. FF: forma del fruto. FFU: forma del fruto en la unión con el pedicelo. CBF: cuello en la base del fruto. FAF: forma del ápice del fruto. AF: apéndice en el fruto. TE: tipo de epidermis del fruto.

Tabla A.6. Media de los datos de fenotipado de cada accesión para los parámetros cuantitativos de planta y fruto.

Accesión	Pes	LonPe	AIP	AnP	LT	DT
BGHZ0082	8,57	3,66	101,00	52,33	23,00	1,62
BGHZ0097	38,19	4,31	92,00	51,00	20,50	1,67
BGHZ0102	115,83	4,09	94,67	40,67	25,17	1,69
BGHZ0106	18,90	5,80	122,67	68,33	31,00	1,74
BGHZ0113	69,21	4,62	95,67	48,33	22,67	1,63
BGHZ0179	62,00	5,54	100,50	56,17	31,50	1,29
BGHZ0181	22,61	3,76	72,33	42,67	21,67	1,05
BGHZ0206	110,62	4,93	89,50	57,00	21,50	2,10
BGHZ0213	133,41	3,96	98,67	55,00	19,50	2,20
BGHZ0224	91,05	4,47	95,00	53,00	25,83	2,02
BGHZ0280	2,20	2,20	93,00	36,50	22,00	1,38
BGHZ0282	55,54	5,41	84,67	46,33	21,83	1,86
BGHZ0284	9,83	3,87	81,00	44,33	22,97	1,33
BGHZ0337	83,54	4,02	45,00	86,33	23,67	1,59
BGHZ0338	65,15	4,57	85,67	46,00	30,33	1,76
BGHZ0413	66,53	4,48	92,83	59,67	23,00	1,70
BGHZ0418	-	-	82,00	47,00	29,00	1,19
BGHZ0447	20,00	4,01	104,00	46,00	39,17	1,32
BGHZ0463	27,03	3,46	103,50	49,00	29,00	1,67
BGHZ0775	172,68	4,81	101,33	68,67	31,33	2,01
BGHZ0782	13,10	3,68	128,33	54,33	24,83	1,56
BGHZ0958	-	-	84,00	56,00	20,50	1,46
BGHZ0969	8,58	2,26	107,00	58,50	19,00	2,21
BGHZ0971	76,62	5,54	102,00	56,67	33,00	2,01
BGHZ1109	202,30	3,54	82,00	51,00	19,33	1,92
BGHZ1114	1,63	3,23	128,00	60,00	39,00	1,38
BGHZ1174	112,11	4,32	89,00	49,50	27,00	1,58
BGHZ1375	19,38	6,13	126,00	62,00	22,47	1,76
BGHZ1521	44,14	6,58	122,50	66,00	23,67	1,73
BGHZ1753	14,28	4,13	101,00	49,33	32,17	1,18
BGHZ2154	21,27	5,52	121,50	59,00	38,50	1,89
BGHZ2546	122,84	4,65	80,67	51,67	25,17	1,85
BGHZ2561	117,69	3,40	105,33	47,00	24,50	1,58
BGHZ3337	69,21	4,81	98,00	41,83	26,00	1,72
BGHZ5912	111,46	5,39	87,00	50,00	31,17	1,61
BR113-05	38,26	3,74	63,67	53,67	19,00	1,70
C-431	42,45	3,38	88,50	50,50	22,25	1,78
CAP1661	94,31	6,24	92,33	60,33	27,67	1,74
CGN17234	99,24	4,54	86,67	43,33	23,67	1,21
CGN17235	5,65	4,44	98,33	50,00	16,33	2,24
CGN20507	121,76	4,47	105,33	53,67	31,50	2,12

Pes: peso del fruto (g). LonPe: longitud del pedicelo del fruto (cm). AIP: altura de la planta (cm). AnP: ancho de la plata (cm). LT: longitud del tallo (cm). DT: diámetro del tallo (cm). PorH2O: porcentaje de agua del fruto. PesSe: peso seco por fruto (g). DE: desviación estándar. CV: coeficiente de variación.

Tabla A.6. Media de los datos de fenotipado de cada accesión para los parámetros cuantitativos de planta y fruto. (continuación).

Accesión	Pes	LonPe	AIP	AnP	LT	DT
CGN23293	139,41	5,33	106,00	58,33	27,33	2,05
OI27-05	162,16	5,07	109,33	57,00	26,00	1,97
PA-145	28,02	5,32	131,67	66,00	31,67	1,72
PI309-05	202,40	2,93	95,33	47,00	24,33	1,91
CAP1371	36,03	3,88	100,00	52,33	21,00	1,43
CAP1539	1,92	2,95	98,00	46,33	31,67	1,34
CGN16805	62,16	5,55	78,33	46,67	22,50	1,57
CGN16897	173,85	4,33	81,00	46,67	42,33	1,64
CGN16899	24,87	3,00	78,00	43,33	33,33	1,45
CGN21475	23,61	2,82	86,67	45,67	37,17	1,27
CGN22161	162,70	3,67	74,33	49,67	19,33	1,51
C-87	47,89	3,46	35,67	26,33	22,33	1,01
C-94	14,78	3,70	105,50	50,00	24,25	1,64
C-204	2,55	2,91	103,00	49,00	30,00	1,43
CGN21460	0,38	3,14	178,00	55,25	47,00	1,23
CGN21526	4,05	2,34	154,50	84,50	62,15	0,92
CGN21527	0,75	3,04	122,75	59,75	30,00	0,99
CGN22102	11,57	3,23	87,67	61,00	30,00	1,16
CGN22783	0,10	3,10	68,67	38,67	17,33	0,72
CGN23254	30,96	4,99	110,25	63,75	34,50	1,06
CGN24355	7,38	4,29	99,33	43,00	55,67	1,22
BGV005857	18,89	4,28	58,33	51,00	19,67	0,79
BGV005981	16,48	3,75	81,75	92,50	31,75	1,16
BGV006008	33,63	4,33	81,38	48,88	25,25	1,54
BGV006055	18,63	4,00	94,75	43,00	22,63	1,16
PI241670	3,92	3,33	76,50	45,75	30,00	0,72
PI585238	4,22	2,62	95,50	49,00	16,50	0,76
PI585246	14,23	3,95	86,25	53,25	26,50	0,72
BGV013687	28,30	4,06	79,50	52,75	20,50	1,19
BGV013691	29,07	3,64	75,33	50,67	28,33	1,25
CAP524	0,26	2,84	148,33	54,67	60,67	0,87
CGN20808	0,24	2,28	155,75	47,00	32,75	1,07
CGN22155	9,20	3,41	87,75	58,25	27,25	0,92
PI213915	5,58	3,95	145,33	69,67	58,33	0,85
PI439212	3,61	3,31	87,00	49,00	23,00	0,64
PI543188	23,15	3,35	69,33	56,00	14,33	0,69
CGN17172	106,89	5,08	110,33	37,33	29,33	1,03
CGN21486	155,80	5,28	118,00	60,00	34,00	1,03
CGN23239	17,09	4,90	120,33	50,00	28,67	0,90
CAP68	60,42	3,92	108,67	45,00	22,67	0,93
CAP242	55,26	3,79	73,67	32,67	26,00	0,87

Pes: peso del fruto (g). LonPe: longitud del pedicelo del fruto (cm). AIP: altura de la planta (cm). AnP: ancho de la plata (cm). LT: longitud del tallo (cm). DT: diámetro del tallo (cm). PorH2O: porcentaje de agua del fruto. PesSe: peso seco por fruto (g). DE: desviación estándar. CV: coeficiente de variación.

Tabla A.6. Media de los datos de fenotipado de cada accesión para los parámetros cuantitativos de planta y fruto. (continuación).

Accesión	Pes	LonPe	AIP	AnP	LT	DT
CAP913	160,83	3,92	76,00	35,33	21,00	1,00
CGN16800	14,22	2,15	111,67	56,33	29,33	0,85
CGN16881	116,69	4,42	97,00	46,00	19,00	1,22
CGN16835	34,76	4,78	87,33	40,00	47,33	0,69
CGN17193	69,66	4,98	93,00	39,67	53,33	1,06
CGN23235	150,44	3,73	105,33	35,67	69,67	0,92
CGN21581	31,00	4,20	110,67	46,67	76,00	0,71
CGN16916	59,14	2,76	61,00	38,67	15,67	0,87
CGN16917	85,39	3,54	98,67	40,33	21,67	0,93
CGN21480	48,43	3,21	68,33	36,67	17,00	0,98
CGN22825	25,39	4,81	132,00	53,00	39,00	1,14
CGN16833	14,22	2,15	111,67	56,33	29,33	0,85
CGN17010	65,75	3,98	86,00	49,00	24,33	0,80
CGN17190	76,55	3,40	100,00	43,33	24,67	1,03
CGN17191	44,43	5,03	116,33	36,33	33,67	1,03
CGN22154	74,73	2,75	77,33	35,33	21,00	0,86
CGN24358	22,96	4,72	93,00	54,00	23,00	0,88
CAP1049	57,39	4,76	108,67	42,33	22,67	1,20
CAP1051	19,56	4,38	124,00	50,67	27,67	0,83
CAP491	14,66	4,13	87,67	26,00	28,00	0,65
CAP1254	2,30	3,04	80,00	47,00	21,50	0,48
CGN16808	50,76	5,42	114,50	34,00	35,50	0,70
CGN17214	118,55	3,15	101,67	36,00	26,67	1,20
CGN16807	70,95	2,38	81,33	35,67	19,33	0,90
CGN16817	115,98	3,35	80,33	40,00	23,00	0,94
CGN17147	63,15	2,46	69,00	41,00	16,67	0,80
CGN17227	77,57	3,63	76,33	39,67	16,33	0,93
CGN22127	26,89	5,98	78,00	28,67	26,67	0,65
CAP607	11,44	6,02	134,33	50,67	33,00	1,28
CAP609	14,21	5,38	106,67	42,67	26,67	1,04
CAP614	3,26	3,38	112,33	45,33	36,67	0,81
CAP977	9,62	3,48	101,33	49,33	30,67	1,31
CAP979	38,33	5,50	113,00	40,75	35,00	1,11
CAP1538	34,05	5,65	131,00	53,00	40,33	1,21
Media C. annuum	53,29	4,07	97,55	49,66	28,80	1,29
DE C. annuum	50,69	1,01	22,21	10,76	10,86	0,43
CV C. annuum	0,95	0,25	0,23	0,22	0,38	0,34
PM647	0,20	2,51	91,83	54,50	29,17	0,83
PM670	0,14	2,60	101,17	35,17	60,17	0,95
Media C. annuum var. glabriusculum	0,17	2,56	96,50	44,83	44,67	0,89
DE C. annuum var. glabriusculum	0,04	0,06	6,60	13,67	21,92	0,09
CV C. annuum var. glabriusculum	0,25	0,02	0,07	0,30	0,49	0,10

Pes: peso del fruto (g). LonPe: longitud del pedicelo del fruto (cm). AIP: altura de la planta (cm). AnP: ancho de la plata (cm). LT: longitud del tallo (cm). DT: diámetro del tallo (cm). PorH2O: porcentaje de agua del fruto. PesSe: peso seco por fruto (g). DE: desviación estándar. CV: coeficiente de variación.

Tabla A.6. Media de los datos de fenotipado de cada accesión para los parámetros cuantitativos de planta y fruto. (continuación).

Accesión	Pes	LonPe	AIP	AnP	LT	DT
CGN23259	7,00	4,37	101,50	43,75	48,63	1,19
CGN17040	0,58	2,68	128,00	66,00	39,33	1,04
PI593922	4,78	3,41	88,50	48,50	27,50	0,91
PI585252	16,05	4,78	98,00	46,50	31,00	0,99
PI593929	10,46	4,33	108,25	63,50	38,75	1,19
PI585278	14,17	3,44	96,00	102,00	37,00	1,37
PI585253	7,33	4,50	130,33	71,00	44,67	1,32
BGV005890	19,11	4,03	98,50	55,00	45,50	1,19
PI224427	2,03	6,67	149,00	72,00	48,00	1,27
PI360725	2,27	3,50	159,20	55,00	61,25	1,15
PI593933	2,93	4,49	101,33	74,00	31,67	1,22
BGV013705	10,95	3,73	85,50	70,50	26,50	1,16
CAP472	0,48	2,17	91,75	72,50	30,25	0,76
CAP868	9,85	3,74	107,75	64,00	45,00	1,18
CAP1364	0,64	2,38	124,17	60,33	48,83	1,07
CGN17220	2,57	2,73	71,25	52,00	25,75	0,72
CGN19211	2,90	2,95	107,25	63,50	36,75	1,34
CGN23565	3,41	3,01	123,75	52,25	45,80	1,13
PI260478	3,46	3,19	120,50	64,00	39,50	1,31
PI315015	14,19	4,67	101,67	44,00	30,00	1,20
PI315020	3,87	2,95	162,50	80,00	44,00	1,27
BGV014749	5,95	2,93	98,33	84,00	19,67	1,33
CGN17045	7,16	2,34	72,25	51,75	25,00	1,18
PI543184	12,62	3,49	100,50	48,00	25,75	1,29
PI543203	1,15	3,22	98,27	68,00	38,83	1,06
Media <i>C. chinense</i>	6,64	3,59	108,96	62,88	37,40	1,15
DE <i>C. chinense</i>	5,38	0,99	23,54	13,90	9,95	0,17
CV <i>C. chinense</i>	0,81	0,28	0,22	0,22	0,27	0,15
PI355808	-	-	149,50	46,25	64,50	0,95
PI585254	-	-	125,30	64,43	48,50	1,04
PI585256	-	-	105,33	49,00	48,00	0,83
PI585257	-	-	148,00	65,67	53,00	0,85
PI593920	0,95	2,78	122,25	49,00	51,00	0,91
PI593924	-	-	141,25	51,50	56,00	0,95
PI595907	1,23	3,78	145,25	55,50	57,50	1,05
BGV013699	-	-	114,00	56,00	57,00	1,18
CGN22792	-	-	138,00	44,25	74,50	1,00
PI439488	-	-	110,25	82,25	47,25	0,97
PI560941	-	-	117,75	57,50	64,75	1,11
PI573337	1,13	3,64	108,00	130,00	39,00	1,08
Media <i>C. frutescens</i>	1,10	3,40	127,07	62,61	55,08	0,99
DE <i>C. frutescens</i>	0,14	0,54	16,50	23,66	9,55	0,11
CV <i>C. frutescens</i>	0,13	0,16	0,13	0,38	0,17	0,11

Pes: peso del fruto (g). LonPe: longitud del pedicelo del fruto (cm). AIP: altura de la planta (cm). AnP: ancho de la plata (cm). LT: longitud del tallo (cm). DT: diámetro del tallo (cm). PorH2O: porcentaje de agua del fruto. PesSe: peso seco por fruto (g). DE: desviación estándar. CV: coeficiente de variación.

Tabla A.6. Media de los datos de fenotipado de cada accesión para los parámetros cuantitativos de planta y fruto. (continuación).

Accesión	Pes	LonPe	AIP	AnP	LT	DT
BGV005852	4,76	5,37	152,00	103,50	48,50	1,19
BGV006064	16,75	7,24	157,75	45,05	30,00	1,07
BGV011957	14,31	6,03	126,25	82,25	35,25	0,88
PI257133	9,00	4,08	180,50	81,00	47,50	1,42
PI257135	7,43	6,03	127,50	76,50	31,00	1,08
PI355813	7,25	5,05	143,00	67,00	39,50	0,72
PI585239	10,33	4,58	146,00	116,00	26,00	1,21
PI585244	6,36	6,19	155,33	92,67	49,33	1,24
PI585249	7,49	5,14	142,25	62,50	52,25	1,06
PI593932	6,19	5,40	123,33	61,67	39,67	1,10
PI595905	7,04	5,07	124,00	96,00	38,00	1,27
BGV013703	3,82	3,99	121,00	55,00	36,50	0,72
CAP1366	10,55	9,54	159,67	46,67	81,33	1,22
CGN17174	5,51	5,50	127,00	82,50	19,75	0,92
CGN21479	3,48	4,19	114,33	84,50	44,00	0,95
PI215739	5,16	4,04	91,67	66,00	27,33	0,80
PI290982	5,96	5,51	158,50	64,75	65,00	1,38
BGV008233	2,44	4,33	134,00	130,00	67,00	1,11
BGV013986	0,29	3,64	170,50	59,25	42,50	1,06
CGN17042	0,25	3,23	113,00	49,75	52,50	1,09
CGN23205	0,24	2,38	91,50	67,00	30,50	0,67
PI590505	1,78	3,16	100,63	81,67	33,67	1,20
PI596061	1,86	5,75	120,00	101,50	30,50	0,88
Media <i>C. baccatum</i>	6,01	5,02	133,90	77,07	42,07	1,05
DE <i>C. baccatum</i>	4,27	1,50	23,99	22,15	14,64	0,21
CV <i>C. baccatum</i>	0,71	0,30	0,18	0,29	0,35	0,20
BGV013300	11,01	3,69	71,50	70,50	15,50	1,03
PI355394	17,56	4,05	141,67	75,67	63,00	0,89
PI585262	5,31	3,82	78,67	63,00	18,67	1,23
PI585269	11,21	3,27	65,00	99,00	17,00	0,80
PI585271	5,69	3,52	75,50	67,50	27,00	0,99
PI585273	19,10	5,03	126,00	141,00	22,00	1,88
PI585275	11,53	3,86	120,75	65,67	40,00	0,80
BGV013702	-	-	48,00	46,67	12,67	0,85
CAP217	19,96	3,75	83,75	60,13	34,30	0,88
CAP357	9,41	6,13	48,50	49,50	25,00	0,76
CAP1486	15,53	3,96	78,00	76,00	41,03	1,11
CGN22108	6,98	4,30	95,00	95,67	25,00	0,80
CGN23768	26,61	4,46	65,50	57,50	21,00	1,43
Media <i>C. pubescens</i>	13,32	4,15	84,45	74,45	27,86	1,03
DE <i>C. pubescens</i>	6,52	0,77	29,03	25,24	13,79	0,32
CV <i>C. pubescens</i>	0,49	0,19	0,34	0,34	0,49	0,31

Pes: peso del fruto (g). LonPe: longitud del pedicelo del fruto (cm). AIP: altura de la planta (cm). AnP: ancho de la plata (cm). LT: longitud del tallo (cm). DT: diámetro del tallo (cm). PorH2O: porcentaje de agua del fruto. PesSe: peso seco por fruto (g). DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer.

Accesión	P	A	WMH	MW	HMW	MH	CH	AR	AG	AB	AL	ALV	AAV	ABV	HA	CHR	LD
BGHZ0082	94,99	607,29	25,33	26,63	27,50	28,74	30,50	182,96	83,43	39,91	104,85	48,32	33,03	45,44	53,99	55,87	1,43
BGHZ0097	275,80	3043,51	30,70	36,53	104,69	113,47	112,92	160,87	69,18	34,24	91,77	42,17	31,05	40,69	52,67	51,27	5,69
BGHZ0102	314,82	5240,51	61,22	71,69	88,98	102,16	109,65	153,65	65,50	33,65	87,29	39,37	31,20	37,50	51,07	48,19	3,78
BGHZ0106	373,27	2505,12	17,68	29,67	139,43	163,75	162,49	178,77	91,97	49,79	107,49	49,31	28,11	41,85	56,31	50,56	4,60
BGHZ0113	329,85	4848,39	49,16	57,23	104,71	116,94	121,77	142,12	69,28	41,67	86,48	38,42	22,70	33,62	56,39	40,83	3,95
BGHZ0179	215,46	2657,42	53,94	57,19	50,88	60,66	64,90	157,80	69,81	35,35	90,81	41,28	29,85	38,79	52,40	49,55	3,29
BGHZ0181	209,15	1951,55	32,55	36,52	73,58	78,97	79,29	156,37	77,60	43,21	95,77	43,45	28,17	37,76	53,76	47,23	8,85
BGHZ0206	386,06	7004,46	62,36	74,04	113,07	131,28	139,52	151,97	96,83	65,66	102,41	46,11	16,42	29,44	61,04	33,78	4,01
BGHZ0213	403,07	6113,25	47,06	56,40	143,18	154,99	155,57	173,35	83,78	38,44	102,43	46,22	28,75	44,36	57,08	53,93	3,18
BGHZ0224	223,65	3230,71	67,26	68,63	54,40	58,89	64,12	167,21	86,68	54,59	104,32	46,36	26,57	35,84	53,83	44,72	1,39
BGHZ0280	58,06	233,36	16,53	17,12	16,40	17,01	19,81	192,78	102,14	54,28	116,25	53,76	28,57	44,73	57,63	53,49	1,60
BGHZ0282	379,49	4317,11	30,33	45,45	144,15	155,08	153,64	156,33	79,90	43,75	94,13	43,04	25,49	36,77	54,53	44,82	6,59
BGHZ0284	313,27	1407,54	11,18	25,68	118,65	137,28	136,76	170,87	79,02	36,99	97,77	45,44	30,78	42,85	54,61	53,21	3,38
BGHZ0337	280,26	4566,05	59,41	66,21	82,69	92,28	97,30	146,64	78,01	43,77	89,56	41,23	22,33	35,57	57,35	42,63	4,84
BGHZ0338	339,88	4643,51	43,24	54,17	107,22	117,35	121,53	145,27	58,03	29,00	79,56	36,11	28,98	36,52	50,12	46,11	6,87
BGHZ0413	324,46	4547,48	42,74	55,36	106,15	119,65	122,25	158,03	82,01	39,55	98,22	43,83	27,12	37,08	55,62	45,24	3,31
BGHZ0418	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BGHZ0447	238,18	1925,09	20,96	30,69	87,80	97,72	97,97	170,50	76,81	39,74	99,10	45,13	32,16	42,39	52,61	53,55	2,86
BGHZ0463	154,37	1505,99	39,69	44,31	43,63	45,83	47,64	155,90	99,23	54,96	99,40	46,67	15,55	35,97	66,59	39,21	1,83
BGHZ0775	360,01	6570,16	80,81	86,95	87,52	107,29	122,54	138,61	71,87	38,42	83,26	38,20	21,88	33,15	56,79	39,83	5,15
BGHZ0782	121,24	903,82	26,59	29,52	38,72	41,23	42,46	211,26	149,10	29,68	113,38	66,30	10,69	64,69	80,33	65,62	2,33
BGHZ0958	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BGHZ0969	95,11	615,20	25,90	27,00	28,06	29,11	31,24	193,02	96,45	51,37	114,31	52,98	30,62	45,04	55,80	54,54	1,57
BGHZ0971	304,68	4281,01	45,94	53,75	93,42	106,00	110,05	155,79	79,15	38,48	95,16	43,41	27,63	36,99	55,59	48,35	3,07
BGHZ1109	332,23	5381,14	74,20	79,11	69,96	97,25	110,49	174,55	97,49	54,24	107,64	49,93	23,67	40,31	59,68	46,92	2,62
BGHZ1114	78,78	277,57	10,20	11,62	30,84	32,94	33,77	190,61	101,38	50,98	112,99	53,26	27,65	45,91	58,71	53,39	1,26
BGHZ1174	226,92	3149,89	65,63	67,72	59,26	65,82	64,54	175,37	93,25	43,94	105,14	49,71	26,35	43,18	57,68	51,00	1,19
BGHZ1375	326,65	2544,21	20,57	26,76	132,91	144,65	144,12	171,76	81,75	43,89	101,41	46,15	30,26	40,60	53,49	50,90	5,11
BGHZ1521	528,74	4615,45	23,28	39,46	197,04	224,84	229,27	175,25	75,13	35,16	97,94	44,76	33,24	43,02	52,25	54,51	5,41
BGHZ1753	271,10	1749,33	15,48	27,12	105,92	117,13	118,70	172,88	81,35	43,23	101,67	45,49	28,71	42,16	54,58	51,06	4,55
BGHZ2154	283,04	2152,54	19,24	28,06	106,46	120,09	118,42	168,99	84,35	41,79	99,53	46,23	27,94	41,53	56,10	50,15	3,37
BGHZ2546	266,76	4474,41	68,95	73,73	72,37	79,70	82,74	167,56	91,06	51,96	103,29	47,28	24,26	38,07	58,36	45,40	1,45
BGHZ2561	243,03	3479,73	74,47	77,89	45,65	55,22	59,79	158,18	64,12	32,58	89,66	41,95	32,28	39,45	50,63	51,03	1,57
BGHZ3337	441,71	5586,08	31,20	47,20	167,69	184,00	183,88	160,52	67,65	37,51	93,07	41,39	32,49	38,44	49,02	50,38	3,44
BGHZ5912	266,89	4226,95	74,86	77,57	61,55	68,85	75,08	164,04	78,45	41,46	96,66	44,16	29,22	39,44	54,33	48,63	1,61
BR113-05	255,63	3092,18	40,16	46,17	81,81	94,04	94,56	170,33	87,46	42,27	101,53	46,87	27,21	41,48	56,31	50,80	3,96
C-431	161,41	1706,01	53,18	54,51	35,27	39,83	42,69	178,77	86,40	46,26	105,83	48,35	30,10	42,45	54,62	52,10	2,02
CAP1661	367,00	6189,48	50,90	58,90	131,25	139,70	139,71	142,07	69,88	38,44	85,81	38,20	24,36	33,37	54,19	41,38	4,43
CGN17234	416,71	5071,06	29,99	49,59	158,26	171,58	171,47	157,69	80,51	43,62	94,70	43,43	25,48	37,47	56,20	45,63	2,48
CGN17235	185,59	895,22	12,81	16,58	72,35	82,11	81,87	188,00	85,61	42,61	108,49	49,36	35,02	45,48	52,51	57,11	2,37
CGN20507	353,47	5614,20	48,77	65,45	115,62	127,68	129,99	151,97	70,14	36,36	90,53	39,62	29,39	37,25	52,52	48,37	3,41

P: perímetro. A: área. WMH: anchura en la mitad de la altura. MW: anchura máxima. HMW: anchura en la mitad de la anchura. MH: altura máxima. CH: altura en la curva. AR: rojo promedio. AG: verde promedio. AB: azul promedio. AL: luminosidad media. ALV: media del valor L. AAV: media del valor A. ABV: media del valor a. HA: media del valor h. HA: tono promedio. CHR: cromatismo promedio. LD: grado de lobulato. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	PA	PT	FSIEI	FSIEII	CFSI	PFB	DFB	FST	ELL	CIR	REC	SH	PIA	OVO	VA	HAV	WWP
BGHZ0082	0,55	0,25	1,08	1,09	1,19	0,88	0,56	1,57	0,06	0,07	0,52	0,01	0,02	0,31	0,06	0,18	0,33
BGHZ0097	0,56	0,25	3,12	3,44	3,74	1,13	0,48	2,34	0,09	0,35	0,46	0,02	0,07	0,50	0,24	1,29	0,17
BGHZ0102	0,64	0,19	1,42	1,45	1,74	1,11	0,50	2,23	0,11	0,17	0,47	0,03	0,09	0,48	0,27	1,23	0,21
BGHZ0106	0,55	0,25	5,61	8,44	9,58	1,09	0,60	1,93	0,15	0,45	0,34	0,04	0,00	0,45	0,51	1,98	0,22
BGHZ0113	0,62	0,21	2,05	2,14	2,48	0,89	0,63	1,48	0,09	0,24	0,49	0,05	0,13	0,30	0,42	1,06	0,31
BGHZ0179	0,85	0,15	1,06	0,95	1,18	0,95	0,59	1,66	0,10	0,11	0,53	0,04	0,19	0,37	0,24	0,47	0,28
BGHZ0181	0,60	0,20	2,19	2,31	2,50	0,92	0,63	1,49	0,08	0,26	0,49	0,02	0,01	0,34	0,23	0,57	0,28
BGHZ0206	0,68	0,22	1,78	1,82	2,24	1,16	0,58	2,02	0,11	0,22	0,48	0,05	0,36	0,54	0,34	1,58	0,17
BGHZ0213	0,53	0,30	2,77	3,23	3,36	1,08	0,49	2,16	0,09	0,33	0,49	0,01	0,02	0,49	0,41	1,61	0,17
BGHZ0224	0,64	0,20	0,87	0,84	0,96	0,86	0,55	1,49	0,06	0,07	0,54	0,01	0,01	0,25	0,17	0,31	0,39
BGHZ0280	0,55	0,26	1,00	1,00	1,18	0,81	0,55	1,47	0,05	0,05	0,52	0,01	0,02	0,24	0,04	0,08	0,38
BGHZ0282	0,57	0,23	3,41	4,46	4,78	1,09	0,54	2,03	0,12	0,38	0,42	0,01	0,06	0,45	0,50	1,62	0,20
BGHZ0284	0,55	0,25	5,72	10,89	12,26	1,33	0,54	2,56	0,19	0,46	0,26	0,05	0,01	0,63	0,48	2,10	0,12
BGHZ0337	0,60	0,24	1,39	1,31	1,60	0,98	0,64	1,52	0,08	0,13	0,53	0,05	0,08	0,34	0,27	0,74	0,30
BGHZ0338	0,63	0,20	2,20	2,63	2,96	1,12	0,68	1,58	0,10	0,28	0,53	0,01	0,07	0,46	0,33	1,13	0,15
BGHZ0413	0,67	0,18	2,25	2,53	2,99	1,21	0,55	2,38	0,10	0,29	0,44	0,04	0,18	0,55	0,28	1,57	0,16
BGHZ0418	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BGHZ0447	0,56	0,24	3,27	4,18	4,73	1,26	0,54	2,36	0,11	0,39	0,37	0,01	0,05	0,56	0,31	1,55	0,14
BGHZ0463	0,67	0,18	1,04	1,09	1,19	0,89	0,53	1,74	0,07	0,09	0,47	0,01	0,02	0,34	0,13	0,38	0,31
BGHZ0775	0,75	0,17	1,24	1,09	1,49	0,81	0,50	1,80	0,11	0,14	0,47	0,03	0,03	0,37	0,61	1,08	0,31
BGHZ0782	0,55	0,25	1,36	1,39	1,52	0,91	0,44	2,03	0,08	0,13	0,46	0,01	0,02	0,37	0,10	0,39	0,29
BGHZ0958	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
BGHZ0969	0,55	0,25	1,06	1,08	1,19	0,83	0,51	1,62	0,06	0,06	0,51	0,01	0,02	0,28	0,06	0,17	0,35
BGHZ0971	0,59	0,23	1,97	2,06	2,52	1,05	0,56	2,00	0,09	0,23	0,47	0,03	0,02	0,41	0,38	1,12	0,25
BGHZ1109	0,68	0,17	1,30	1,06	1,54	0,53	0,59	0,88	0,12	0,19	0,52	0,11	0,28	0,34	0,67	0,85	0,30
BGHZ1114	0,55	0,25	2,83	2,99	3,22	0,90	0,47	1,95	0,08	0,32	0,41	0,00	0,01	0,36	0,07	0,32	0,31
BGHZ1174	0,72	0,15	0,97	0,91	1,04	0,88	0,50	1,81	0,07	0,07	0,53	0,04	0,09	0,31	0,16	0,43	0,34
BGHZ1375	0,55	0,24	5,49	6,40	6,97	1,03	0,46	2,32	0,10	0,44	0,39	0,00	0,01	0,46	0,28	1,77	0,26
BGHZ1521	0,58	0,22	5,77	8,70	9,90	1,06	0,58	1,67	0,14	0,46	0,32	0,06	0,01	0,46	0,60	3,67	0,21
BGHZ1753	0,55	0,25	4,42	6,73	7,57	1,39	0,56	2,58	0,14	0,44	0,33	0,00	0,00	0,59	0,39	1,51	0,12
BGHZ2154	0,55	0,25	4,31	5,50	6,30	1,32	0,51	2,51	0,13	0,42	0,37	0,01	0,04	0,62	0,28	1,81	0,13
BGHZ2546	0,59	0,23	1,05	1,01	1,17	0,95	0,47	1,93	0,08	0,09	0,50	0,02	0,04	0,36	0,24	0,70	0,29
BGHZ2561	0,67	0,18	0,71	0,61	0,77	0,84	0,68	1,24	0,07	0,16	0,58	0,02	0,04	0,18	0,25	0,25	0,43
BGHZ3337	0,56	0,25	4,02	5,46	5,99	1,16	0,55	2,26	0,13	0,42	0,38	0,01	0,05	0,50	0,66	2,79	0,17
BGHZ5912	0,64	0,20	0,92	0,84	1,04	0,95	0,55	1,67	0,09	0,10	0,54	0,02	0,02	0,34	0,19	0,61	0,30
BR113-05	0,65	0,19	2,09	2,18	2,40	1,13	0,50	2,32	0,09	0,25	0,47	0,03	0,12	0,44	0,21	1,08	0,24
C-431	0,56	0,25	0,74	0,67	0,81	0,83	0,59	1,43	0,07	0,13	0,54	0,02	0,04	0,22	0,14	0,23	0,39
CAP1661	0,77	0,20	2,41	2,60	2,84	1,13	0,52	2,19	0,09	0,28	0,49	0,04	0,19	0,52	0,23	1,55	0,18
CGN17234	0,52	0,27	3,41	5,22	5,73	1,32	0,63	2,23	0,14	0,41	0,35	0,01	0,03	0,54	0,60	2,86	0,16
CGN17235	0,55	0,26	5,04	5,75	6,31	0,94	0,49	2,03	0,10	0,43	0,42	0,00	0,01	0,40	0,17	0,89	0,26
CCN20507	0,56	0,24	2,00	2,28	2,58	1,20	0,53	2,28	0,10	0,25	0,46	0,03	0,12	0,49	0,42	1,48	0,17

PA: área del pericarpio. PT: grosor del pericarpio. FSIEI: índice externo de la forma del fruto I. FSIEII: índice externo de la forma del fruto II. CFSI: índice de la forma curva del fruto. PFB: bloqueo proximal del fruto. DFB: bloqueo distal del fruto. FST: forma triangular del fruto. ELL: elipse. CIR: circular. REC: rectangular. SH: altura en los hombros. PIA: área de la hendidura proximal. OVO: ovoide. VA: V. Asimetría. HAV: H. Asimetría Ov. WWP: posición de la anchura máxima. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	P	A	WMH	MW	HMW	MH	CH	AR	AG	AB	AL	ALV	AAV	ABV	HA	CHR	LD
CGN23293	370,70	6508,54	69,03	79,61	95,76	109,47	120,22	138,02	74,89	40,62	84,04	38,84	20,31	32,69	58,90	38,56	6,43
OI27-05	400,70	7334,01	57,45	73,77	125,62	146,19	151,80	162,43	77,69	41,09	94,55	43,23	27,78	38,47	53,14	49,26	3,87
PA-145	213,42	2127,79	32,43	37,60	66,47	75,12	77,62	156,07	68,70	36,03	90,31	41,05	30,69	38,00	51,04	48,89	6,04
PI309-05	361,22	5993,11	78,97	83,00	63,43	95,02	114,11	138,39	69,41	42,38	87,47	37,75	23,27	33,16	55,06	40,58	4,14
CAP1371	243,31	3241,92	43,57	52,46	73,52	81,05	84,79	171,56	92,19	52,92	105,61	48,24	27,68	38,78	55,50	48,79	4,55
CAP1539	81,23	310,17	11,61	12,40	31,41	33,52	34,48	187,93	102,34	56,37	114,95	52,96	26,44	43,31	58,28	50,80	1,32
CGN16805	205,71	2535,52	53,56	57,72	52,88	60,61	62,77	181,19	100,77	55,11	111,19	51,67	24,74	41,76	59,67	48,64	1,68
CGN16897	297,04	4907,51	77,71	82,34	63,57	80,19	86,06	165,19	96,07	60,35	104,56	47,30	20,46	33,62	59,24	39,44	1,67
CGN16899	141,08	1253,26	41,89	43,10	33,41	37,75	39,87	180,07	90,82	45,95	108,09	49,09	31,55	43,40	54,94	53,69	2,26
CGN21475	125,99	1014,62	42,49	43,24	26,92	30,21	31,84	188,01	92,46	50,40	112,12	51,02	31,71	43,80	54,18	53,66	1,99
CGN22161	311,54	4673,38	78,52	81,65	58,75	79,50	86,82	167,45	93,66	49,84	103,48	47,95	23,73	38,33	58,42	45,38	3,20
C-87	247,78	3075,83	42,94	49,17	74,63	87,95	95,82	182,12	89,89	39,58	104,30	49,52	29,75	46,47	57,70	55,52	3,52
C-94	139,29	1132,38	28,70	31,07	44,62	46,31	47,69	182,92	123,17	80,14	123,79	56,75	15,87	35,78	66,08	39,14	3,77
C-204	62,40	258,97	15,93	16,84	18,78	19,78	21,80	196,45	120,68	70,95	125,83	58,43	21,63	42,48	63,84	47,94	1,63
CGN21460	60,68	147,42	6,95	8,40	23,80	25,22	26,55	171,78	80,29	32,30	95,79	45,83	30,91	44,89	56,13	54,20	-
CGN21526	104,04	485,80	13,11	14,77	41,41	42,84	43,55	185,00	84,21	36,48	104,19	48,72	34,37	47,25	54,52	58,21	-
CGN21527	82,65	236,24	7,68	10,22	32,45	34,98	36,20	177,07	82,66	33,59	99,13	47,02	31,71	46,38	55,65	56,23	-
CGN22102	147,00	1055,18	21,12	29,73	46,44	51,11	53,31	182,28	99,49	49,51	109,56	51,94	26,09	43,61	59,55	51,40	5,17
CGN22783	40,29	85,75	6,18	6,90	15,65	16,21	17,44	175,73	90,78	41,56	102,25	48,36	27,70	43,80	57,92	51,86	5,68
CGN23254	230,69	2505,40	36,51	41,67	73,36	80,51	83,52	147,68	58,65	31,36	84,06	37,45	31,94	36,50	48,78	48,35	4,83
CGN24355	202,75	1054,97	13,35	23,09	74,76	83,57	84,69	175,10	89,43	45,70	105,58	48,25	28,02	42,40	56,32	50,86	2,58
BGV005857	251,74	2480,70	28,76	38,70	87,73	98,12	103,09	168,02	71,69	35,04	96,34	43,36	33,36	42,28	50,68	54,03	6,12
BGV005981	226,40	2118,05	27,33	32,56	82,22	87,79	89,61	175,00	82,20	39,68	101,50	46,48	31,41	42,81	53,73	53,20	5,72
BGV006008	236,90	2711,27	33,26	40,28	84,25	89,60	91,20	176,15	92,52	43,03	103,13	48,76	26,77	44,13	58,89	51,77	6,23
BGV006055	274,36	2601,32	27,05	37,11	100,17	110,50	113,73	175,47	80,19	36,35	98,88	45,83	31,73	44,00	54,21	54,39	4,43
PI241670	237,99	977,88	11,20	27,31	64,75	100,60	103,01	186,69	98,66	48,60	110,71	51,97	27,74	45,43	58,55	53,09	-
PI585238	183,84	642,51	8,75	25,20	49,19	74,08	77,73	182,62	87,55	44,53	106,83	49,03	31,42	44,18	54,57	54,60	-
PI585246	199,75	1648,01	23,39	30,24	75,71	78,86	79,55	185,26	93,57	45,43	108,52	50,54	31,21	45,26	56,92	54,74	2,80
BGV013687	249,51	2640,32	36,86	42,76	76,45	85,67	89,10	167,67	90,59	43,06	98,56	46,76	24,52	40,76	59,14	47,60	5,90
BGV013691	234,05	2645,01	33,60	40,95	82,91	89,90	90,00	173,52	81,09	38,64	99,80	46,11	31,61	43,36	53,33	53,52	3,28
CAP524	45,86	106,28	7,03	7,62	17,79	18,08	19,55	184,08	96,86	33,96	104,33	50,78	27,49	49,85	61,14	57,85	-
CGN20808	40,61	82,60	6,39	7,32	14,94	16,23	17,68	169,95	105,90	43,43	100,95	50,80	17,90	44,88	68,14	48,85	-
CGN22155	158,69	1121,68	19,49	22,54	59,42	63,70	65,17	184,44	92,36	47,38	108,24	49,90	30,04	44,25	55,83	53,53	2,09
PI213915	160,13	839,76	14,22	21,64	56,15	64,84	66,99	178,78	86,42	41,52	103,59	47,74	31,46	43,84	54,48	54,06	-
PI439212	150,72	643,84	10,79	18,92	56,37	63,29	65,97	187,83	84,09	36,69	105,88	49,39	34,98	47,27	53,97	58,66	3,88
PI543188	192,54	1820,98	27,90	32,84	64,82	70,47	74,06	170,87	70,77	36,04	97,24	44,13	33,98	42,48	51,37	54,48	4,62
CGN17172	430,56	6711,92	44,47	59,28	156,10	170,29	171,50	156,38	74,42	33,53	91,00	40,81	29,78	39,59	55,00	50,33	5,56
CGN21486	308,75	6217,61	59,30	68,81	97,62	103,14	105,32	146,59	82,18	46,69	90,91	41,82	20,66	33,57	58,50	39,59	6,46
CGN23239	262,12	2173,20	22,94	35,75	88,36	102,71	105,61	158,34	68,74	35,98	91,18	41,56	30,67	39,04	51,73	49,48	7,91
CAP68	181,07	2140,76	59,07	59,79	41,16	45,11	47,51	185,94	99,64	53,78	112,76	52,32	26,77	43,43	58,34	50,74	1,99
CAP242	268,89	3753,49	48,99	54,47	81,69	91,93	97,47	161,06	70,61	34,20	89,28	42,05	28,49	40,85	53,40	49,16	6,72

P: perímetro. A: área. WMH: anchura en la mitad de la altura. MW: anchura máxima. HMW: altura en la curva. MH: altura en la curva. AR: rojo promedio. AG: verde promedio. AB: azul promedio. AL: luminosidad media. ALV: media del valor L. AAV: media del valor a. ABV: media del valor b. HA: tono promedio. CHR: croma promedio. LD: grado de lóbulos. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	PA	PT	FSIEI	FSIEII	CFSI	PFB	FST	ELL	CIR	REC	SH	PIA	OVO	VA	HAV	WWP
CGN23293	0,72	0,13	1,37	1,36	1,72	0,99	1,63	0,11	0,15	0,59	0,01	0,03	0,35	0,34	0,80	0,30
OI27405	0,56	0,25	2,03	2,20	2,65	1,08	1,84	0,10	0,26	0,43	0,05	0,21	0,33	0,67	1,76	0,24
PA-145	0,65	0,18	1,93	2,05	2,34	1,09	1,65	0,09	0,24	0,52	0,03	0,15	0,42	0,21	0,67	0,25
PB09-05	0,81	0,17	1,16	0,86	1,41	0,97	2,00	0,13	0,14	0,60	0,03	0,04	0,23	0,43	0,51	0,41
CAP1371	0,58	0,22	1,54	1,71	1,93	1,09	1,76	0,10	0,18	0,49	0,01	0,03	0,44	0,33	0,79	0,22
CAP1539	0,54	0,26	2,69	2,62	2,95	0,74	1,55	0,06	0,30	0,40	0,01	0,03	0,25	0,09	0,27	0,39
CGNI6805	0,56	0,25	1,05	1,03	1,23	0,98	2,27	0,11	0,12	0,47	0,06	0,24	0,41	0,14	0,67	0,27
CGNI6897	0,69	0,21	0,98	0,83	1,09	0,94	1,75	0,10	0,12	0,55	0,08	0,27	0,34	0,27	0,74	0,31
CGNI6899	0,56	0,25	0,87	0,82	0,99	0,80	1,39	0,06	0,10	0,50	0,04	0,11	0,21	0,12	0,20	0,40
CGN21475	0,60	0,22	0,70	0,64	0,75	0,73	1,26	0,05	0,14	0,51	0,04	0,12	0,17	0,14	0,13	0,42
CGN22161	0,74	0,14	0,98	0,75	1,20	0,93	1,46	0,11	0,14	0,58	0,07	0,21	0,24	0,42	0,49	0,37
C-87	0,56	0,24	1,83	1,75	2,25	0,96	3,81	0,09	0,21	0,41	0,10	0,03	0,44	0,45	1,02	0,23
C-94	0,55	0,25	1,49	1,55	1,67	1,01	2,56	0,07	0,13	0,56	0,03	0,13	0,41	0,06	0,40	0,29
C-204	0,62	0,21	1,17	1,18	1,34	0,86	1,64	0,06	0,08	0,50	0,02	0,07	0,30	0,03	0,13	0,33
CGN21460	-	-	3,06	3,41	3,78	0,86	1,63	0,08	0,35	0,40	0,00	0,00	0,32	0,06	0,26	0,32
CGN21526	-	-	2,92	3,20	3,39	1,02	1,93	0,08	0,33	0,48	0,00	0,01	0,42	0,06	0,40	0,23
CGN21527	-	-	3,38	4,04	4,77	1,03	1,62	0,09	0,39	0,39	0,00	0,00	0,41	0,12	0,47	0,24
CGN22102	0,63	0,20	1,73	2,22	2,58	1,28	2,07	0,11	0,24	0,44	0,02	0,04	0,51	0,19	0,63	0,18
CGN22783	0,62	0,20	2,37	2,54	2,73	0,82	1,35	0,07	0,28	0,45	0,01	0,01	0,24	0,04	0,12	0,36
CGN23254	0,68	0,17	1,94	2,04	2,28	1,04	1,65	0,10	0,24	0,49	0,02	0,05	0,40	0,27	0,84	0,23
CGN24355	0,55	0,25	3,82	5,63	5,99	1,19	1,94	0,14	0,42	0,36	0,00	0,02	0,53	0,38	1,13	0,17
BGV005857	0,55	0,24	2,54	2,99	3,77	0,96	1,66	0,11	0,33	0,45	0,01	0,01	0,26	0,40	0,93	0,38
BGV005981	0,60	0,21	2,71	3,13	3,41	1,13	1,79	0,09	0,33	0,49	0,01	0,01	0,46	0,24	0,79	0,16
BGV006008	0,56	0,23	2,24	2,59	2,84	1,06	1,82	0,09	0,28	0,48	0,02	0,04	0,39	0,23	0,86	0,27
BGV006055	0,55	0,24	3,05	3,73	4,26	0,92	1,57	0,11	0,37	0,41	0,02	0,00	0,25	0,48	1,24	0,34
PI241670	-	-	4,11	5,99	9,22	0,95	1,49	0,18	0,45	0,22	0,06	0,01	0,30	0,67	2,05	0,43
PI585238	-	-	3,49	5,99	8,77	1,38	2,17	0,20	0,44	0,21	0,01	0,01	0,61	0,60	1,35	0,16
PI585246	0,63	0,20	2,65	3,17	3,38	1,03	1,52	0,10	0,33	0,45	0,00	0,00	0,39	0,28	0,67	0,25
BGV013687	0,80	0,10	1,98	2,13	2,38	1,00	1,77	0,09	0,24	0,51	0,03	0,07	0,32	0,25	0,72	0,28
BGV013691	0,56	0,25	2,10	2,38	2,52	1,09	2,87	0,09	0,26	0,43	0,02	0,06	0,50	0,18	1,12	0,19
CAP524	-	-	2,38	2,56	2,74	0,72	1,07	0,06	0,28	0,47	0,02	0,01	0,17	0,04	0,07	0,44
CGN20808	-	-	2,19	2,37	2,72	0,77	1,41	0,07	0,28	0,38	0,00	0,00	0,26	0,07	0,16	0,35
CGN22155	0,56	0,25	2,80	3,06	3,28	1,03	1,90	0,08	0,32	0,45	0,01	0,02	0,41	0,14	0,61	0,26
PI213915	-	-	3,20	3,94	4,37	1,15	2,39	0,11	0,39	0,34	0,00	0,00	0,50	0,28	0,94	0,27
PI439212	0,55	0,25	3,51	5,31	6,24	1,22	1,89	0,14	0,41	0,34	0,00	0,00	0,48	0,29	1,11	0,22
PI543188	0,65	0,24	2,09	2,50	2,70	0,88	1,45	0,08	0,27	0,50	0,01	0,01	0,31	0,19	0,54	0,28
CGNI1712	0,63	0,20	3,00	3,30	3,62	1,10	2,35	0,11	0,34	0,44	0,02	0,07	0,44	0,57	1,64	0,23
CGN21486	0,60	0,22	1,64	1,70	1,72	0,98	1,37	0,07	0,18	0,55	0,02	0,07	0,32	0,33	0,68	0,31
CGN23239	0,63	0,18	2,93	3,89	4,70	1,31	2,71	0,12	0,37	0,36	0,03	0,14	0,47	0,45	1,61	0,15
CAP68	0,56	0,25	0,76	0,70	0,81	0,78	1,35	0,05	0,12	0,53	0,01	0,03	0,20	0,18	0,17	0,44
CAP242	0,65	0,18	1,71	1,72	2,04	0,78	1,14	0,08	0,20	0,46	0,06	0,03	0,23	0,44	0,76	0,37

PA: área del pericarpio. PT: grosor del pericarpio. FSIEI: índice externo de la forma del fruto I. FSIEII: índice externo de la forma del fruto II. CFSI: índice de la forma curva del fruto. PFB: bloqueo proximal del fruto. DFB: bloqueo distal del fruto. FST: forma triangular del fruto. ELL: elipse. CIR: circular. REC: rectangular. SH: altura en los hombros. PIA: área de la hendidura proximal. OVO: ovoide. VA: V. Asimetría. HAV: H. Asimetría Ov. WWP: posición de la anchura máxima. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	P	A	WMH	MW	HMW	MH	CH	AR	AG	AB	AL	ALV	AAV	ABV	HA	CHR	LD
CAP913	311,72	5388,14	71,06	75,27	79,50	91,40	101,39	152,00	121,57	23,04	83,73	51,74	0,20	51,32	89,80	51,36	2,47
CGN16800	111,56	823,30	36,46	37,09	25,57	27,64	29,27	200,09	105,10	57,77	120,88	56,19	29,80	45,60	57,19	54,09	1,58
CGN16881	257,19	3984,00	70,72	74,92	60,34	70,79	75,01	187,48	156,31	53,07	113,20	65,12	-1,61	54,85	91,69	54,89	2,04
CGN16835	409,91	3294,65	19,72	37,79	150,66	172,61	173,10	175,79	72,73	47,29	100,25	45,34	34,89	43,32	50,62	55,65	3,70
CGN17193	391,34	4713,87	31,92	44,91	139,95	162,83	163,12	178,16	84,25	32,33	103,03	48,04	30,25	44,39	55,28	53,79	4,54
CGN23235	300,30	5335,04	65,19	71,08	83,99	93,09	101,97	148,03	65,13	29,81	83,57	38,54	28,21	37,80	52,98	47,26	3,25
CGN21581	256,67	2618,30	32,60	42,28	79,69	97,48	96,67	151,63	60,96	32,44	85,74	38,38	31,99	37,51	49,71	49,28	3,41
CGN16916	238,57	3198,24	52,69	56,15	65,26	76,28	81,76	160,94	83,41	43,19	96,02	44,60	25,27	38,95	57,05	46,57	4,78
CGN16917	269,43	3985,31	55,71	61,07	77,21	87,52	90,87	150,38	78,82	40,04	87,79	41,86	23,19	37,07	58,19	43,89	5,25
CGN21480	234,51	3020,45	46,87	51,34	74,75	81,03	82,59	159,73	75,15	34,34	91,30	42,76	28,40	41,10	55,33	50,46	2,67
CGN22825	301,18	2494,72	22,48	33,96	113,36	128,21	129,47	157,73	75,23	39,81	92,92	42,65	27,71	38,58	53,93	47,57	-
CGN16833	236,90	2711,27	33,26	40,28	84,25	89,60	91,20	176,15	92,52	43,03	103,13	48,76	26,77	44,13	58,89	51,77	6,23
CGN17010	232,21	3113,54	52,15	54,93	64,37	71,43	77,59	165,00	84,15	40,87	97,49	45,24	25,98	40,30	57,17	48,02	3,37
CGN17190	272,18	3905,65	56,87	62,33	66,61	81,68	91,33	150,41	68,24	36,56	87,89	39,73	28,14	36,48	52,22	46,14	2,68
CGN17191	365,99	4014,02	28,09	45,08	131,53	151,11	149,75	170,74	81,69	40,31	99,27	45,88	29,80	42,11	54,71	51,71	4,92
CGN22154	261,58	3528,69	55,03	61,75	67,42	79,55	89,45	152,21	74,06	37,09	89,03	41,23	26,15	38,32	54,51	46,13	4,16
CGN24358	416,16	3000,35	17,59	35,42	154,05	177,57	177,44	181,53	81,36	41,54	105,28	47,64	34,31	43,63	51,88	55,29	4,27
CAP1049	380,60	4776,67	32,90	46,57	144,27	159,22	157,89	163,00	76,74	37,86	94,48	43,34	29,62	40,55	53,59	50,43	4,79
CAP1051	331,55	2655,23	21,75	33,66	119,38	139,28	140,86	160,78	73,99	38,94	94,36	43,05	29,64	39,03	52,65	48,70	6,59
CAP491	394,96	2371,16	14,07	32,16	153,27	168,24	166,68	176,78	100,35	55,30	109,72	51,29	21,77	40,95	60,83	46,75	4,04
CAP1254	128,96	484,23	9,57	13,15	50,99	56,49	58,13	209,75	190,31	107,17	149,14	77,02	-7,18	46,20	98,88	46,76	2,47
CGN16808	265,50	3312,71	40,63	48,05	90,64	101,22	99,68	168,33	80,64	39,35	96,12	45,40	31,21	41,25	53,52	50,88	3,84
CGN17214	231,82	2940,65	78,80	79,62	36,35	49,14	53,17	172,71	86,51	49,33	104,47	46,43	27,88	39,84	55,03	48,67	3,77
CGN16807	190,85	2179,97	62,79	62,93	35,51	43,62	47,53	196,37	104,12	55,27	119,42	55,08	28,07	45,83	58,33	53,82	2,13
CGN16817	218,23	2549,37	77,64	78,04	31,70	42,24	47,11	173,78	89,70	53,74	106,89	48,66	25,59	39,21	56,08	47,40	2,82
CGN17147	191,09	2354,88	55,63	57,98	48,23	52,88	55,33	179,33	89,52	43,91	105,03	49,08	29,09	44,44	56,64	52,93	1,55
CGN17227	252,73	3434,40	47,35	55,90	77,77	87,41	89,32	168,21	83,54	35,30	95,75	45,72	27,70	43,79	57,60	51,96	2,48
CGN22127	297,78	2625,85	23,21	34,59	108,96	125,65	127,16	171,62	80,13	40,50	97,62	45,93	30,66	42,17	53,90	52,31	3,66
CAP607	155,98	1213,81	22,27	25,10	59,76	61,94	62,64	155,10	62,64	32,91	89,30	40,39	32,53	38,86	49,97	50,56	4,36
CAP609	164,88	1358,74	25,26	31,03	54,11	58,61	59,31	163,89	75,08	35,46	93,76	43,57	29,80	41,52	54,30	51,14	-
CAP614	106,88	540,18	14,68	17,21	40,34	42,43	43,27	169,49	85,35	46,98	101,51	46,93	27,40	40,63	55,95	49,07	3,05
CAP977	139,86	992,86	22,45	26,50	44,53	49,21	50,88	169,03	83,51	45,76	101,04	46,19	28,25	39,96	54,68	49,03	3,74
CAP979	365,22	4312,93	34,23	39,19	143,18	155,08	154,64	155,21	64,57	30,30	87,27	39,94	32,49	40,03	50,92	51,24	6,70
CAP1538	199,17	1942,90	31,28	35,00	65,02	70,79	71,77	163,07	79,08	43,12	96,94	44,37	27,80	39,80	54,64	47,79	4,94
Media C. annuum	246,04	2860,66	37,70	44,59	78,04	88,30	91,12	169,39	85,93	43,30	100,21	46,59	27,15	41,25	56,78	49,91	3,77
DE C. annuum	102,00	1831,64	20,78	20,46	38,96	43,63	43,44	14,94	18,39	10,64	10,36	5,94	6,57	4,65	7,34	4,87	1,67
CV C. annuum	0,41	0,64	0,55	0,46	0,50	0,49	0,48	0,09	0,21	0,25	0,10	0,13	0,24	0,11	0,13	0,10	0,44
PM647	40,72	85,90	6,58	7,12	15,73	16,38	17,79	149,98	88,70	52,47	95,27	43,88	19,19	33,57	60,09	38,56	4,94
PM670	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Media C. annuum var. glabriusculum	40,72	85,90	6,58	7,12	15,73	16,38	17,79	149,98	88,70	52,47	95,27	43,88	19,19	33,57	60,09	38,56	4,94
DE C. annuum var. glabriusculum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CV C. annuum var. glabriusculum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

P: perímetro. A: área. WMH: anchura en la mitad de la altura. MW: anchura máxima. HMW: altura en la mitad de la anchura. MH: altura máxima. CH: altura en la curva. AR: rojo promedio. AG: verde promedio. AB: azul promedio. AL: luminosidad media. ALV: media del valor L. AAV: media del valor a. ABV: media del valor b. HA: tono promedio. CHR: croma promedio. LD: grado de lóbulo. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	PA	PT	FSIEI	FSIEII	CFSI	PFB	DFB	FST	ELL	CIR	REC	SH	PIA	OVO	VA	HAV	WWP
CAP913	0,67	0,18	1,21	1,11	1,41	0,95	0,56	1,85	0,09	0,11	0,57	0,03	0,15	0,36	0,30	0,72	0,28
CGN16800	0,56	0,25	0,75	0,71	0,82	0,79	0,61	1,31	0,05	0,11	0,54	0,01	0,02	0,20	0,08	0,11	0,41
CGN16881	0,62	0,21	0,95	0,86	1,06	0,93	0,51	1,86	0,09	0,11	0,51	0,05	0,16	0,33	0,25	0,63	0,30
CGN16835	0,55	0,25	4,59	7,54	8,59	1,41	0,61	2,33	0,17	0,45	0,34	0,02	0,03	0,51	0,61	2,40	0,09
CGN17193	0,56	0,25	3,61	4,48	5,07	1,29	0,51	2,49	0,11	0,40	0,38	0,01	0,08	0,57	0,48	2,44	0,15
CGN23235	0,62	0,21	1,29	1,27	1,51	0,97	0,66	1,48	0,08	0,12	0,54	0,01	0,02	0,35	0,29	0,74	0,30
CGN21581	0,66	0,18	2,34	2,50	3,05	1,22	0,42	3,05	0,11	0,29	0,40	0,04	0,39	0,60	0,37	1,39	0,12
CGN16916	0,66	0,18	1,37	1,31	1,57	0,96	0,51	1,78	0,09	0,14	0,50	0,05	0,07	0,33	0,27	0,74	0,33
CGN16917	0,66	0,19	1,44	1,34	1,59	0,99	0,53	1,88	0,09	0,14	0,52	0,04	0,06	0,32	0,35	0,77	0,31
CGN21480	0,67	0,18	1,59	1,61	1,78	1,01	0,48	2,03	0,08	0,17	0,48	0,03	0,09	0,40	0,23	0,78	0,26
CGN22825	-	-	3,77	5,15	5,78	1,18	0,52	2,30	0,12	0,41	0,34	0,01	0,01	0,49	0,46	1,84	0,18
CGN16833	0,56	0,23	1,70	1,64	1,89	0,92	0,58	1,66	0,08	0,18	0,53	0,04	0,10	0,31	0,26	0,64	0,33
CGN17010	0,67	0,17	1,35	1,27	1,56	0,95	0,58	1,60	0,08	0,14	0,50	0,06	0,14	0,31	0,27	0,54	0,33
CGN17190	0,64	0,20	1,32	1,19	1,57	1,00	0,54	1,97	0,10	0,13	0,55	0,03	0,13	0,37	0,40	0,63	0,28
CGN17191	0,56	0,25	3,34	4,57	5,41	1,46	0,43	3,29	0,12	0,40	0,34	0,02	0,09	0,62	0,48	2,61	0,12
CGN22154	0,70	0,16	1,31	1,26	1,61	0,98	0,63	1,55	0,09	0,14	0,50	0,04	0,18	0,29	0,40	0,67	0,28
CGN24358	0,57	0,25	5,07	8,89	10,12	1,38	0,52	2,60	0,18	0,45	0,28	0,06	0,01	0,60	0,62	3,03	0,12
CAP1049	0,56	0,25	3,47	4,36	4,86	1,32	0,47	2,70	0,11	0,39	0,41	0,02	0,06	0,59	0,36	2,35	0,13
CAP1051	0,55	0,24	4,13	5,51	6,24	1,29	0,46	2,91	0,14	0,43	0,35	0,04	0,01	0,61	0,36	2,04	0,12
CAP491	0,55	0,25	5,47	10,43	11,61	1,40	0,53	2,66	0,18	0,46	0,28	0,07	0,01	0,63	0,62	2,09	0,11
CAP1254	0,55	0,25	4,21	5,41	6,01	1,03	0,66	1,54	0,10	0,42	0,40	0,03	0,00	0,42	0,15	0,80	0,23
CGN16808	0,56	0,25	2,11	2,27	2,57	1,08	0,43	2,63	0,09	0,25	0,44	0,03	0,18	0,49	0,29	1,23	0,21
CGN17214	0,62	0,20	0,60	0,47	0,67	0,56	0,57	1,10	0,08	0,22	0,54	0,04	0,02	0,12	0,57	0,19	0,48
CGN16807	0,56	0,25	0,69	0,59	0,76	0,84	0,66	1,28	0,07	0,16	0,58	0,01	0,01	0,20	0,19	0,25	0,42
CGN16817	0,62	0,20	0,55	0,41	0,61	0,63	0,62	1,20	0,09	0,25	0,55	0,02	0,01	0,06	0,42	0,10	0,50
CGN17147	0,56	0,25	0,92	0,89	1,02	0,89	0,47	1,93	0,07	0,08	0,51	0,02	0,07	0,32	0,11	0,41	0,34
CGN17227	0,56	0,25	1,57	1,66	1,90	1,09	0,42	2,71	0,10	0,18	0,43	0,04	0,19	0,48	0,24	1,08	0,20
CGN22127	0,55	0,25	3,73	4,76	5,72	1,19	0,53	2,25	0,12	0,40	0,36	0,00	0,00	0,47	0,39	1,90	0,24
CAP607	0,55	0,25	2,50	2,71	2,84	1,01	0,61	1,65	0,08	0,29	0,49	0,02	0,05	0,37	0,12	0,52	0,28
CAP609	-	-	1,88	2,18	2,38	1,20	0,65	1,90	0,10	0,24	0,52	0,03	0,13	0,51	0,14	0,68	0,16
CAP614	0,55	0,25	2,47	2,78	2,87	1,09	0,52	2,18	0,09	0,30	0,45	0,01	0,03	0,48	0,08	0,49	0,21
CAP977	0,66	0,18	1,83	1,94	2,20	1,07	0,58	1,93	0,09	0,22	0,50	0,03	0,13	0,46	0,14	0,50	0,20
CAP979	0,62	0,21	4,01	4,30	4,67	0,99	0,49	2,11	0,09	0,39	0,47	0,01	0,02	0,39	0,32	1,59	0,25
CAP1538	0,56	0,24	1,94	2,05	2,26	1,06	0,67	1,61	0,09	0,23	0,53	0,03	0,15	0,43	0,21	0,65	0,21
Media C. annuum	0,61	0,22	2,32	2,84	3,26	1,02	0,56	1,92	0,10	0,26	0,46	0,03	0,07	0,39	0,30	0,99	0,27
DE C. annuum	0,07	0,04	1,30	2,17	2,49	0,18	0,08	0,49	0,03	0,12	0,08	0,02	0,08	0,12	0,17	0,73	0,09
CV C. annuum	0,11	0,16	0,56	0,77	0,76	0,18	0,14	0,26	0,30	0,47	0,18	0,80	1,14	0,31	0,57	0,73	0,35
PM647	0,62	0,20	2,32	2,35	2,68	0,68	0,56	1,24	0,06	0,28	0,40	0,01	0,01	0,13	0,05	0,12	0,46
PM670	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Media C. annuum var. glabriusculum	0,62	0,20	2,32	2,35	2,68	0,68	0,56	1,24	0,06	0,28	0,40	0,01	0,01	0,13	0,05	0,12	0,46
DE C. annuum var. glabriusculum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CV C. annuum var. glabriusculum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

PA: área del pericarpio. PT: grosor del pericarpio. FSIEI: índice externo de la forma del fruto I. FSIEII: índice externo de la forma del fruto II. CFSI: índice de la forma curva del fruto. PFB: bloqueo proximal del fruto. DFB: bloqueo distal del fruto. FST: forma triangular del fruto. ELL: elipsoide. CIR: circular. REC: rectangular. SH: altura en los hombros. PIA: área de la hendidura proximal. OVO: ovoide. VA: V. Asimetría. HAV: H. Asimetría Ov. WWP: posición de la anchura máxima. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	P	A	WMH	MW	HMW	MH	CH	AR	AG	AB	AL	ALV	AAV	ABV	HA	CHR	LD
CGN23259	122,45	756,91	19,25	21,21	46,33	47,99	49,47	201,75	130,16	60,56	124,06	60,89	18,97	49,96	69,47	53,31	3,12
CGN17040	40,83	101,78	9,15	9,69	13,82	14,86	15,65	195,67	126,45	18,78	100,91	58,80	15,90	61,88	75,63	64,00	-
PI593922	128,22	641,49	13,31	17,75	48,31	52,18	54,05	172,18	77,25	39,42	99,55	45,15	32,71	42,01	52,21	54,76	5,36
PI585252	201,49	1804,50	30,25	33,37	76,02	79,63	81,99	175,00	75,16	38,66	100,41	44,91	35,78	42,05	50,19	54,84	3,32
PI593929	123,37	902,50	24,79	28,50	42,87	44,34	46,32	179,99	74,42	32,52	98,34	45,25	35,81	46,09	52,26	58,94	3,76
PI585278	143,37	1250,14	32,15	34,51	46,91	48,00	50,57	182,15	90,42	51,22	109,77	49,45	31,11	41,63	53,24	51,93	3,48
PI585253	124,61	916,70	25,10	26,30	44,62	46,07	46,72	180,34	78,82	37,38	102,26	46,55	33,94	43,98	52,60	55,90	2,23
BGV005890	136,61	1095,84	28,65	31,54	44,77	46,80	49,44	177,62	78,49	34,27	99,68	46,09	33,92	45,23	52,67	56,16	6,40
PI224427	95,85	434,04	14,66	18,04	31,88	33,05	36,10	186,75	101,80	38,13	107,16	53,28	24,02	50,08	64,38	55,43	6,26
PI360725	137,94	517,70	9,77	15,50	52,44	58,00	59,55	178,86	75,03	38,12	102,41	45,94	36,02	43,20	50,18	56,28	4,84
PI593933	69,98	319,23	16,40	17,09	24,12	24,52	26,18	199,29	136,34	23,02	103,82	61,12	12,55	62,76	78,75	64,12	2,97
BGV013705	144,01	1040,73	25,64	29,25	46,66	52,51	55,13	180,70	89,31	48,34	108,14	48,96	30,81	41,83	54,12	51,54	4,28
CAP472	40,60	106,90	9,15	9,45	14,63	14,90	16,02	194,86	173,12	69,09	124,23	70,75	-7,29	55,62	97,66	56,11	2,52
CAP868	189,21	1218,03	21,76	27,91	63,73	72,77	74,70	190,38	154,19	33,21	105,22	64,97	-0,75	61,82	90,71	61,87	4,30
CAP1364	73,83	205,46	8,03	10,21	29,69	30,55	31,86	217,27	139,37	17,66	110,56	64,47	18,14	67,74	75,01	70,19	-
CGN17220	100,43	472,20	14,09	16,30	37,46	39,20	40,46	173,63	77,65	41,29	101,06	45,53	33,67	41,61	51,00	53,24	3,17
CGN19211	81,91	395,80	16,97	17,95	29,50	30,43	31,76	190,00	139,46	73,21	123,87	61,54	9,55	44,19	77,88	45,25	2,17
CGN23565	126,03	676,22	16,74	18,64	48,12	52,04	53,18	185,69	147,30	22,92	100,48	63,12	1,92	63,97	88,37	64,09	-
PI260478	121,76	661,84	16,39	19,66	45,69	48,51	50,29	190,92	85,57	40,62	107,99	49,38	34,93	47,32	53,03	58,27	4,58
PI315015	201,85	1577,12	23,80	28,50	75,42	79,36	82,28	144,30	87,65	47,91	89,50	42,15	16,53	33,22	64,03	37,42	3,77
PI315020	100,04	569,90	20,24	22,00	33,69	35,66	37,40	181,36	94,16	49,42	108,60	50,24	28,11	43,06	57,12	51,67	4,76
BGV014749	85,59	478,66	22,61	22,87	24,87	26,19	28,39	201,06	97,37	40,97	113,90	54,50	33,84	51,26	56,35	61,59	1,47
CGN17045	90,89	519,02	23,82	25,87	25,59	27,15	30,29	188,73	82,57	42,45	108,88	48,77	36,25	45,65	51,54	58,36	4,64
PI543184	163,69	1307,10	27,82	31,23	54,92	56,92	60,00	182,86	86,93	38,72	104,25	48,35	32,55	45,80	54,67	56,68	5,56
PI543203	105,11	333,10	8,07	11,87	41,27	43,59	45,28	195,35	182,27	120,46	148,72	73,78	-8,08	36,37	102,37	37,28	3,49
Media C. chinense	117,99	732,12	19,14	21,81	41,73	44,21	46,12	185,87	107,25	43,93	108,15	53,76	22,84	48,33	65,02	55,57	3,93
DE C. chinense	42,86	449,69	7,28	7,59	15,97	17,36	17,62	13,58	33,49	20,91	11,85	8,98	14,34	9,05	16,27	7,53	1,31
CV C. chinense	0,36	0,61	0,38	0,35	0,38	0,39	0,38	0,07	0,31	0,48	0,11	0,17	0,63	0,19	0,25	0,14	0,33
PI555808	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI585254	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI585256	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI585257	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI593920	76,73	299,52	12,62	13,34	29,04	29,63	31,51	169,67	75,73	41,79	99,47	44,87	31,92	40,40	53,20	51,92	7,23
PI593924	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI595907	57,81	199,71	11,33	12,27	20,32	21,02	22,82	174,08	69,73	28,24	95,15	44,06	36,36	45,79	51,83	58,90	2,76
BGV013699	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CGN22792	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI439488	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI560941	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI573337	102,12	299,17	7,73	12,99	35,78	41,29	43,73	189,21	176,89	113,76	142,13	71,70	-8,30	36,33	102,92	37,27	4,80
Media C. frutescens	78,89	266,14	10,56	12,86	28,38	30,65	32,69	177,65	107,45	61,26	112,25	53,54	19,99	40,84	69,32	49,36	4,93
DE C. frutescens	22,23	57,52	2,54	0,55	7,75	10,17	10,51	10,25	60,21	45,97	25,96	15,73	24,60	4,75	29,11	11,04	2,24
CV C. frutescens	0,28	0,22	0,24	0,04	0,27	0,33	0,32	0,06	0,56	0,75	0,23	0,29	1,23	0,12	0,42	0,22	0,45

P: perímetro. A: área. WMH: anchura en la mitad de la altura. MW: anchura máxima. HMW: altura en la mitad de la anchura. MH: altura máxima. CH: altura en la curva. AR: rojo promedio. AG: verde promedio. AB: azul promedio. AL: luminosidad media. ALV: media del valor L. AAV: media del valor a. ABV: media del valor b. HA: tono promedio. CHR: croma promedio. LD: grado de lóbulo. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	PA	PT	FSIEI	FSIEII	CFSI	PFB	DFB	FST	ELL	CIR	REC	SH	PIA	OVO	VA	HAV	WWP
CGN23259	0,55	0,25	2,33	2,56	2,75	0,75	0,73	1,10	0,07	0,27	0,48	0,00	0,00	0,06	0,13	0,11	0,58
CGN17040	-	-	1,43	1,47	1,60	0,71	0,51	1,43	0,05	0,13	0,46	0,01	0,00	0,22	0,03	0,06	0,40
PI593922	0,59	0,22	2,97	3,66	3,93	1,11	0,56	1,99	0,10	0,36	0,45	0,00	0,00	0,38	0,16	0,55	0,25
PI585252	0,55	0,25	2,36	2,61	2,87	0,69	0,47	1,44	0,08	0,29	0,40	0,00	0,00	0,20	0,22	0,50	0,45
PI593929	0,56	0,25	1,59	1,71	1,81	0,74	0,53	1,57	0,06	0,18	0,41	0,01	0,00	0,22	0,13	0,27	0,43
PI585278	0,63	0,19	1,35	1,39	1,56	0,88	0,50	1,81	0,06	0,13	0,49	0,01	0,01	0,25	0,13	0,34	0,41
PI585253	0,58	0,23	1,70	1,73	1,87	0,73	0,52	1,41	0,06	0,18	0,46	0,01	0,01	0,21	0,10	0,25	0,43
BGV005890	0,44	0,29	1,49	1,61	1,77	0,81	0,56	1,35	0,07	0,16	0,47	0,01	0,01	0,25	0,16	0,30	0,40
PI224427	0,55	0,24	1,76	2,19	2,47	0,90	0,72	1,20	0,08	0,24	0,46	0,02	0,01	0,21	0,14	0,25	0,34
PI360725	0,55	0,25	3,68	5,38	6,13	1,20	0,55	2,05	0,13	0,41	0,35	0,03	0,00	0,50	0,17	0,90	0,19
PI593933	0,48	0,30	1,41	1,47	1,52	0,77	0,48	1,66	0,05	0,12	0,47	0,01	0,01	0,26	0,04	0,16	0,39
BGV013705	0,55	0,24	1,80	1,91	2,12	0,84	0,30	2,85	0,07	0,22	0,37	0,01	0,02	0,33	0,19	0,54	0,37
CAP472	0,55	0,25	1,57	1,60	1,75	0,61	0,66	0,93	0,04	0,15	0,47	0,01	0,01	0,07	0,03	0,03	0,45
CAP868	0,63	0,20	2,63	3,02	3,50	0,93	0,39	2,50	0,11	0,34	0,35	0,01	0,02	0,30	0,30	0,93	0,41
CAP1364	-	-	3,01	3,43	3,76	0,89	0,56	1,57	0,10	0,36	0,38	0,00	0,00	0,25	0,10	0,38	0,38
CGN17220	0,65	0,18	2,42	2,71	2,87	0,89	0,58	1,50	0,08	0,30	0,44	0,01	0,01	0,32	0,09	0,33	0,32
CGN19211	0,58	0,23	1,71	1,72	1,83	0,71	0,39	2,05	0,05	0,17	0,44	0,01	0,01	0,24	0,05	0,18	0,40
CGN23565	-	-	2,88	3,02	3,36	0,75	0,47	1,63	0,08	0,32	0,38	0,00	0,00	0,25	0,14	0,43	0,38
PI260478	0,55	0,24	2,47	2,74	2,95	0,91	0,53	1,93	0,10	0,31	0,40	0,00	0,00	0,34	0,13	0,53	0,34
PI315015	0,64	0,19	2,89	3,23	3,49	0,85	0,54	1,59	0,08	0,34	0,43	0,00	0,00	0,25	0,27	0,61	0,38
PI315020	0,55	0,25	1,64	1,69	1,88	0,77	0,49	1,61	0,07	0,18	0,42	0,00	0,00	0,30	0,08	0,27	0,33
BGV014749	0,55	0,25	1,13	1,09	1,23	0,78	0,47	1,68	0,04	0,05	0,51	0,01	0,01	0,21	0,04	0,12	0,43
CGN17045	0,56	0,24	1,06	1,07	1,30	0,72	0,54	1,35	0,06	0,09	0,43	0,02	0,01	0,24	0,13	0,19	0,38
PI543184	0,52	0,26	1,81	2,01	2,23	0,91	0,58	1,58	0,09	0,23	0,47	0,01	0,01	0,30	0,17	0,44	0,37
PI543203	0,55	0,25	3,65	5,16	5,72	0,95	0,89	1,10	0,12	0,41	0,45	0,01	0,00	0,19	0,15	0,31	0,37
Media C. chinense	0,56	0,24	2,11	2,41	2,65	0,83	0,54	1,64	0,08	0,24	0,43	0,01	0,01	0,25	0,13	0,36	0,38
DE C. chinense	0,05	0,03	0,75	1,13	1,27	0,13	0,12	0,43	0,02	0,10	0,04	0,01	0,01	0,09	0,07	0,23	0,07
CV C. chinense	0,09	0,12	0,36	0,47	0,48	0,16	0,22	0,26	0,30	0,43	0,10	0,95	1,02	0,35	0,53	0,64	0,18
PI355808	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI585254	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI585256	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI585257	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI593920	0,55	0,23	2,21	2,29	2,50	0,72	0,58	1,25	0,07	0,26	0,47	0,00	0,00	0,21	0,05	0,15	0,41
PI593924	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI595907	0,54	0,25	1,71	1,79	1,99	0,78	0,62	1,29	0,06	0,19	0,44	0,01	0,01	0,20	0,07	0,14	0,38
BGV013699	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CGN22792	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI439488	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI560941	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
PI573337	0,55	0,25	3,22	4,66	5,81	0,91	0,87	0,98	0,11	0,40	0,32	0,05	0,01	0,08	0,21	0,28	0,51
Media C. frutescens	0,55	0,24	2,38	2,91	3,43	0,81	0,69	1,17	0,08	0,29	0,41	0,02	0,00	0,16	0,11	0,19	0,44
DE C. frutescens	0,00	0,01	0,77	1,53	2,07	0,10	0,16	0,17	0,03	0,11	0,08	0,02	0,00	0,07	0,09	0,08	0,07
CV C. frutescens	0,01	0,06	0,32	0,53	0,60	0,12	0,23	0,14	0,35	0,37	0,19	1,20	0,91	0,42	0,76	0,42	0,16

P.A.: área del pericarpio. PT: grosor del pericarpio. FSIEI: índice externo de la forma del fruto I. FSIEII: índice externo de la forma del fruto II. CFSI: índice de la forma curva del fruto. PFB: bloqueo proximal del fruto. DFB: bloqueo distal del fruto. FST: forma triangular del fruto. ELL: elipse. CIR: circular. REC: rectangular. SH: altura en los hombros. PIA: área de la hendidura proximal. OVO: ovoide. VA: V. Asimetría. HAV: H. Asimetría Ov. WWP: posición de la anchura máxima. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	P	A	WMH	MW	HMW	MH	CH	AR	AG	AB	AL	ALV	AAV	ABV	HA	CHR	LD
BGV005852	103,14	605,97	18,87	19,73	38,08	39,84	41,14	175,67	94,21	48,62	104,96	49,16	26,20	41,62	58,19	49,17	2,94
BGV006064	240,73	1985,49	24,59	29,06	93,09	101,74	102,69	185,88	80,95	40,61	105,70	47,86	35,28	45,35	51,25	56,90	4,04
BGV011957	216,95	1606,97	20,96	27,55	83,92	88,85	90,94	184,23	79,37	42,23	106,49	47,83	36,35	43,95	49,86	56,80	4,28
PI257133	120,69	911,94	30,62	31,86	34,50	37,08	40,15	189,62	94,04	47,97	112,64	51,47	32,27	44,77	54,25	55,27	2,63
PI257135	201,59	1033,18	14,53	20,31	77,43	88,35	88,99	195,03	81,02	39,22	108,69	49,02	36,94	47,88	51,47	60,57	3,15
PI355813	168,73	1094,26	18,41	24,47	60,69	67,22	68,37	184,78	96,72	46,85	111,30	51,85	28,66	44,62	56,67	51,58	3,48
PI585239	159,80	1056,80	19,67	22,12	61,93	66,23	67,48	189,01	90,91	44,56	108,92	50,10	32,56	45,76	54,69	56,25	3,14
PI585244	99,12	621,61	22,65	23,52	33,37	34,41	35,87	199,34	143,47	35,14	111,04	63,38	9,34	60,22	81,18	61,00	1,72
PI585249	156,26	1018,41	18,25	21,47	60,18	64,75	65,57	182,13	94,08	47,33	107,95	50,24	28,33	43,50	56,48	52,49	3,94
PI593932	167,48	996,59	16,91	19,42	67,39	70,98	70,96	186,39	90,63	41,70	107,31	50,12	31,46	46,45	55,94	56,26	2,53
PI595905	190,65	1219,35	17,53	23,58	74,85	80,39	82,31	167,73	78,12	43,41	99,92	44,28	31,33	38,99	51,52	50,47	3,91
BGV013703	133,72	617,79	12,49	15,62	53,44	57,68	57,67	179,28	87,98	41,46	104,43	48,56	30,11	44,24	55,84	53,66	2,11
CAP1366	170,86	1186,25	19,40	22,20	69,28	72,68	72,10	206,78	131,42	35,10	113,38	61,70	18,48	59,54	72,90	62,50	2,58
CGN1714	153,82	886,30	16,51	19,98	59,50	64,71	66,04	175,99	82,62	42,79	103,64	46,73	31,29	42,65	53,66	52,70	4,22
CGN21479	114,22	565,01	14,52	16,09	44,70	47,48	48,57	200,69	131,47	37,25	111,97	60,74	16,88	57,76	73,68	60,15	1,99
PI215739	144,47	809,80	15,96	18,36	57,24	61,60	61,61	176,40	74,15	37,64	101,71	45,65	34,74	43,10	51,14	55,43	4,51
PI290982	154,77	1099,99	21,22	23,99	57,26	60,04	61,45	194,29	121,42	29,84	105,47	57,62	18,83	57,17	72,02	60,42	2,90
BGV008233	133,06	505,87	10,26	14,25	52,07	58,25	58,87	204,99	133,40	15,31	103,67	63,92	10,81	67,00	80,86	67,93	3,81
BGV013986	33,89	67,75	6,19	6,57	12,94	13,20	14,58	173,52	83,54	42,42	101,60	46,57	30,41	41,99	54,18	51,95	-
CGN17042	43,34	100,54	7,25	7,87	15,85	16,42	18,32	153,52	82,74	36,71	89,51	43,28	22,52	40,29	60,78	45,87	-
CGN23205	30,62	63,49	6,70	7,32	10,61	10,87	13,02	166,55	111,51	71,67	112,10	51,92	14,62	33,42	66,54	36,54	3,63
PI590505	97,36	336,96	9,91	13,05	36,37	40,57	41,89	207,68	141,89	21,40	107,80	63,94	12,97	65,47	78,77	66,88	2,35
PI596061	81,02	297,21	10,73	12,01	31,90	33,19	34,18	183,58	94,84	50,45	110,12	50,66	28,77	43,13	56,34	51,91	2,63
Media C. baccatum	135,49	812,50	16,27	19,15	51,59	55,50	56,64	185,35	100,46	40,86	106,54	52,03	26,05	47,78	60,79	55,34	3,17
DE C. baccatum	55,43	479,62	6,04	6,77	22,21	24,40	24,12	13,67	23,08	10,71	5,32	6,48	8,59	8,84	10,39	6,80	0,82
CV C. baccatum	0,41	0,59	0,37	0,35	0,43	0,44	0,43	0,07	0,23	0,26	0,05	0,12	0,33	0,19	0,17	0,12	0,26
BGV013300	111,94	760,70	26,27	27,76	32,42	35,49	37,41	159,58	50,67	24,65	86,79	38,15	39,66	40,76	45,91	56,65	4,41
PI355394	133,37	1111,13	30,19	32,18	41,41	44,14	45,89	164,10	56,02	25,11	90,25	40,03	39,40	41,78	46,99	57,79	4,41
PI585262	109,57	565,66	16,38	18,87	40,74	42,88	44,51	173,90	66,28	31,14	96,49	43,17	38,20	43,56	48,41	57,99	3,44
PI585269	106,67	727,26	26,15	28,44	30,42	32,38	35,61	159,22	64,93	28,39	88,72	40,17	33,55	40,48	50,16	52,74	1,50
PI585271	98,19	599,67	22,31	23,55	31,94	34,83	36,27	144,11	49,16	26,21	81,00	35,82	33,87	36,58	46,53	48,77	3,67
PI585273	161,28	1432,56	28,07	30,18	59,06	61,06	62,02	161,62	51,73	23,51	86,91	38,63	39,31	41,81	46,73	57,40	3,35
PI585275	157,11	1083,42	20,80	24,50	58,64	62,63	64,33	159,64	56,79	27,03	89,21	39,61	38,82	41,40	47,38	56,28	4,50
BGV013702	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CAP217	143,98	1251,32	32,42	34,49	44,68	47,39	49,91	153,35	57,26	29,43	86,84	38,23	35,35	38,45	46,92	51,77	4,52
CAP357	115,43	806,00	28,40	29,72	32,70	35,60	38,95	177,63	131,87	74,48	118,68	57,97	8,50	40,52	77,40	41,49	3,74
CAP1486	119,42	890,09	30,92	31,87	32,76	35,27	38,34	167,21	59,50	30,26	92,82	40,79	38,47	41,16	46,89	56,36	4,09
CGN22108	116,25	722,18	21,30	22,67	41,79	43,29	43,56	157,92	58,18	28,11	87,82	38,61	35,89	39,49	47,74	53,40	3,39
CGN23768	160,45	1576,31	38,81	40,29	45,72	50,64	54,13	168,04	129,56	20,27	88,65	55,89	2,37	57,61	87,71	57,68	5,12
Media C. pubescens	127,80	960,53	26,83	28,71	41,02	43,80	45,91	162,19	69,33	30,72	91,18	42,26	31,95	41,97	53,23	54,03	3,84
DE C. pubescens	22,57	328,98	6,06	5,81	9,90	10,15	9,85	8,97	29,13	14,12	9,42	7,09	12,63	5,24	13,92	4,90	0,92
CV C. pubescens	0,18	0,34	0,23	0,20	0,24	0,23	0,21	0,06	0,42	0,46	0,10	0,17	0,40	0,12	0,26	0,09	0,24

P: perímetro. A: área. WMH: anchura en la mitad de la altura. MW: anchura máxima. HMW: altura en la mitad de la anchura. MH: altura máxima. CH: altura en la curva. AR: rojo promedio. AG: verde promedio. AB: azul promedio. AL: luminosidad media. ALV: media del valor L. AAV: media del valor a. ABV: media del valor b. HA: tono promedio. CHR: croma promedio. LD: grado de lóbulos. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Tabla A.7. Media de los datos de fenotipado de los frutos de cada accesión para los parámetros cuantitativos obtenidos con el programa Tomato Analyzer. (continuación).

Accesión	PA	PT	FSIEI	FSIEII	CFSI	PFB	DFB	FST	ELL	CIR	REC	SH	PIA	OVO	VA	HAV	WWP
BGV005852	0,55	0,25	2,04	2,05	2,18	0,69	0,57	1,24	0,05	0,23	0,46	0,01	0,01	0,15	0,10	0,17	0,45
BGV006064	0,61	0,21	3,53	3,85	4,29	0,81	0,40	2,13	0,08	0,37	0,39	0,02	0,01	0,26	0,26	0,94	0,40
BGV011957	0,55	0,25	3,25	4,09	4,45	0,90	0,74	1,26	0,10	0,37	0,44	0,01	0,00	0,22	0,27	0,83	0,41
PI257133	0,66	0,18	1,16	1,12	1,29	0,80	0,53	1,56	0,05	0,08	0,49	0,02	0,02	0,25	0,10	0,21	0,38
PI257135	0,55	0,25	4,44	5,38	6,24	0,83	0,49	1,74	0,11	0,42	0,31	0,04	0,00	0,23	0,28	1,11	0,41
PI355813	0,66	0,18	2,80	3,38	3,73	1,06	0,52	2,22	0,10	0,35	0,40	0,01	0,04	0,38	0,24	0,77	0,29
PI585239	0,55	0,25	2,98	3,17	3,46	0,77	0,58	1,37	0,08	0,33	0,44	0,01	0,01	0,14	0,20	0,37	0,49
PI585244	0,60	0,22	1,46	1,48	1,60	0,67	0,55	1,30	0,04	0,13	0,45	0,01	0,03	0,14	0,09	0,12	0,46
PI585249	0,55	0,25	3,04	3,34	3,67	0,92	0,47	2,03	0,08	0,34	0,43	0,00	0,01	0,30	0,17	0,65	0,36
PI593932	0,55	0,25	3,66	3,99	4,24	0,87	0,40	2,26	0,08	0,37	0,43	0,00	0,01	0,30	0,16	0,65	0,36
PI595905	0,55	0,25	3,45	4,29	4,87	0,97	0,54	1,85	0,10	0,39	0,39	0,02	0,01	0,32	0,30	1,00	0,30
BGV013703	0,55	0,25	3,80	4,34	4,73	0,93	0,52	1,83	0,09	0,38	0,41	0,01	0,00	0,33	0,15	0,60	0,30
CAP1366	0,48	0,30	3,22	3,62	3,97	0,92	0,58	1,62	0,08	0,36	0,43	0,00	0,02	0,26	0,21	0,77	0,38
CGN17174	0,63	0,20	3,16	3,66	3,98	0,81	0,54	1,60	0,08	0,35	0,40	0,01	0,00	0,20	0,18	0,58	0,43
CGN21479	0,55	0,25	2,99	3,08	3,36	0,75	0,54	1,40	0,07	0,33	0,45	0,01	0,01	0,17	0,12	0,29	0,44
PI215739	0,55	0,25	3,42	3,64	3,96	0,91	0,42	2,26	0,09	0,36	0,43	0,01	0,00	0,35	0,14	0,59	0,31
PI290982	0,65	0,20	2,54	2,77	2,99	0,91	0,66	1,43	0,08	0,30	0,49	0,01	0,04	0,21	0,16	0,35	0,43
BGV008233	0,55	0,25	4,18	5,11	5,86	0,84	0,58	1,51	0,09	0,42	0,35	0,01	0,00	0,16	0,19	0,77	0,48
BGV013986	-	-	2,02	2,11	2,32	0,70	0,62	1,15	0,06	0,23	0,45	0,01	0,01	0,12	0,04	0,05	0,47
CGN17042	-	-	2,07	2,20	2,54	0,77	0,62	1,26	0,06	0,25	0,43	0,01	0,01	0,20	0,06	0,13	0,38
CGN23205	0,55	0,25	1,49	1,59	1,78	0,68	0,68	1,00	0,05	0,17	0,42	0,01	0,00	0,01	0,05	0,01	0,57
PI590505	0,54	0,26	3,19	3,69	4,16	0,91	0,52	1,82	0,10	0,37	0,38	0,00	0,00	0,32	0,15	0,55	0,33
PI596061	0,55	0,25	2,85	3,05	3,27	0,86	0,59	1,51	0,07	0,32	0,47	0,01	0,01	0,23	0,08	0,27	0,40
Media C. baccatum	0,57	0,24	2,90	3,26	3,61	0,84	0,55	1,62	0,08	0,31	0,42	0,01	0,01	0,23	0,16	0,51	0,40
DE C. baccatum	0,05	0,03	0,86	1,11	1,27	0,10	0,08	0,37	0,02	0,09	0,04	0,01	0,01	0,09	0,08	0,32	0,07
CV C. baccatum	0,08	0,12	0,30	0,34	0,35	0,12	0,15	0,23	0,24	0,29	0,10	0,75	1,01	0,38	0,47	0,63	0,17
BGV013300	0,55	0,25	1,28	1,24	1,43	0,84	0,52	1,66	0,07	0,11	0,49	0,03	0,08	0,29	0,11	0,24	0,35
PI355394	0,55	0,25	1,38	1,38	1,52	0,86	0,58	1,50	0,06	0,12	0,50	0,03	0,07	0,28	0,12	0,28	0,34
PI585262	0,55	0,25	2,29	2,53	2,79	0,93	0,52	1,84	0,08	0,28	0,41	0,02	0,01	0,37	0,14	0,46	0,25
PI585269	0,55	0,25	1,18	1,19	1,34	0,94	0,61	1,55	0,07	0,10	0,53	0,03	0,10	0,34	0,09	0,24	0,28
PI585271	0,48	0,30	1,49	1,51	1,66	0,79	0,43	1,89	0,06	0,14	0,43	0,02	0,03	0,29	0,09	0,26	0,36
PI585273	0,55	0,25	2,03	2,12	2,24	0,79	0,62	1,26	0,06	0,23	0,48	0,02	0,03	0,25	0,14	0,34	0,34
PI585275	0,48	0,30	2,57	2,91	3,23	0,96	0,53	1,87	0,09	0,32	0,42	0,01	0,02	0,34	0,17	0,67	0,32
BGV013702	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
CAP217	0,56	0,25	1,38	1,39	1,54	0,73	0,60	1,23	0,06	0,12	0,49	0,03	0,04	0,20	0,12	0,22	0,40
CAP357	0,55	0,25	1,17	1,09	1,35	0,65	0,68	0,97	0,05	0,08	0,52	0,05	0,02	0,07	0,09	0,06	0,50
CAP1486	0,56	0,25	1,12	1,07	1,25	0,78	0,61	1,30	0,06	0,08	0,53	0,04	0,09	0,22	0,11	0,15	0,38
CGN22108	0,55	0,25	1,89	1,94	2,06	0,73	0,60	1,22	0,06	0,22	0,43	0,01	0,05	0,18	0,13	0,26	0,43
CGN23768	0,56	0,24	1,24	1,18	1,38	0,79	0,52	1,56	0,06	0,09	0,49	0,04	0,09	0,24	0,13	0,31	0,39
Media C. pubescens	0,54	0,26	1,58	1,63	1,82	0,82	0,57	1,49	0,07	0,16	0,48	0,03	0,05	0,26	0,12	0,29	0,36
DE C. pubescens	0,03	0,02	0,49	0,61	0,64	0,09	0,07	0,29	0,01	0,08	0,04	0,01	0,03	0,08	0,03	0,16	0,07
CV C. pubescens	0,05	0,08	0,31	0,37	0,35	0,12	0,12	0,20	0,15	0,52	0,09	0,41	0,60	0,33	0,22	0,53	0,18

PA: área del pericarpio. PT: grosor del pericarpio. FSIEI: índice externo de la forma del fruto I. FSIEII: índice externo de la forma del fruto II. CFSI: índice de la forma curva del fruto. PFB: bloqueo proximal del fruto. DFB: bloqueo distal del fruto. FST: forma triangular del fruto. ELL: elipsoide. CIR: circular. REC: rectangular. SH: altura en los hombros. PIA: área de la hendidura proximal. OVO: ovoide. VA: V. Asimetría. HAV: H. Asimetría Ov. WWP: posición de la anchura máxima. Todas las medidas de longitudes están tomadas en milímetros. DE: desviación estándar. CV: coeficiente de variación.

Anexos de figuras

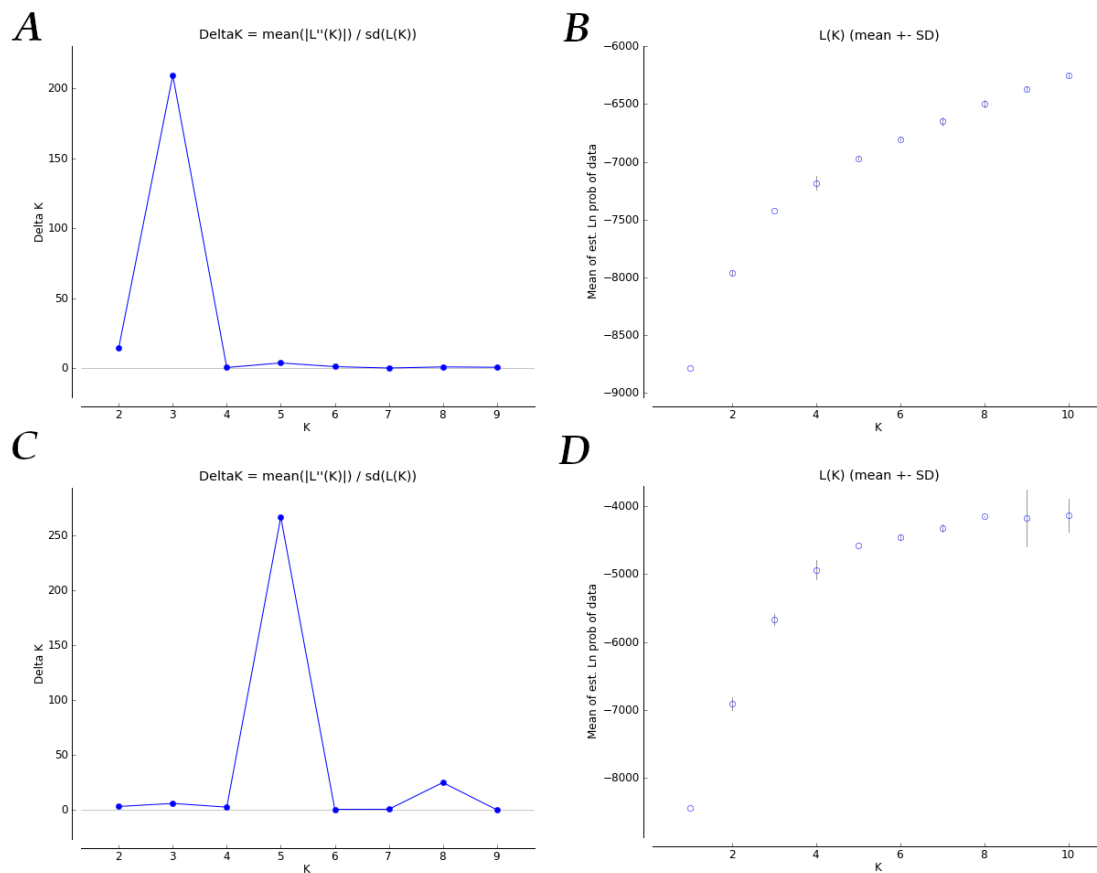


Figura A.1. Métodos de evaluación de la estructura genética inferida con el análisis STRUCTURE obtenidos con STRUCTURE HARVESTER para los grupos *C. annuum* y no-*C. annuum* y empleando 28 microsatélites. A. Estadístico ΔK de Evanno en función de K , tras diez réplicas para las 148 accesiones de *C. annuum*. B. Media del logaritmo de la probabilidad de los datos $\text{Pr}(X/K)$, para las 148 accesiones de *C. annuum*. C. Estadístico ΔK de Evanno en función de K , tras diez réplicas para las 82 accesiones que no pertenecen a *C. annuum*. D. Media del logaritmo de la probabilidad de los datos $\text{Pr}(X/K)$ para las 82 accesiones que no pertenecen a *C. annuum*.

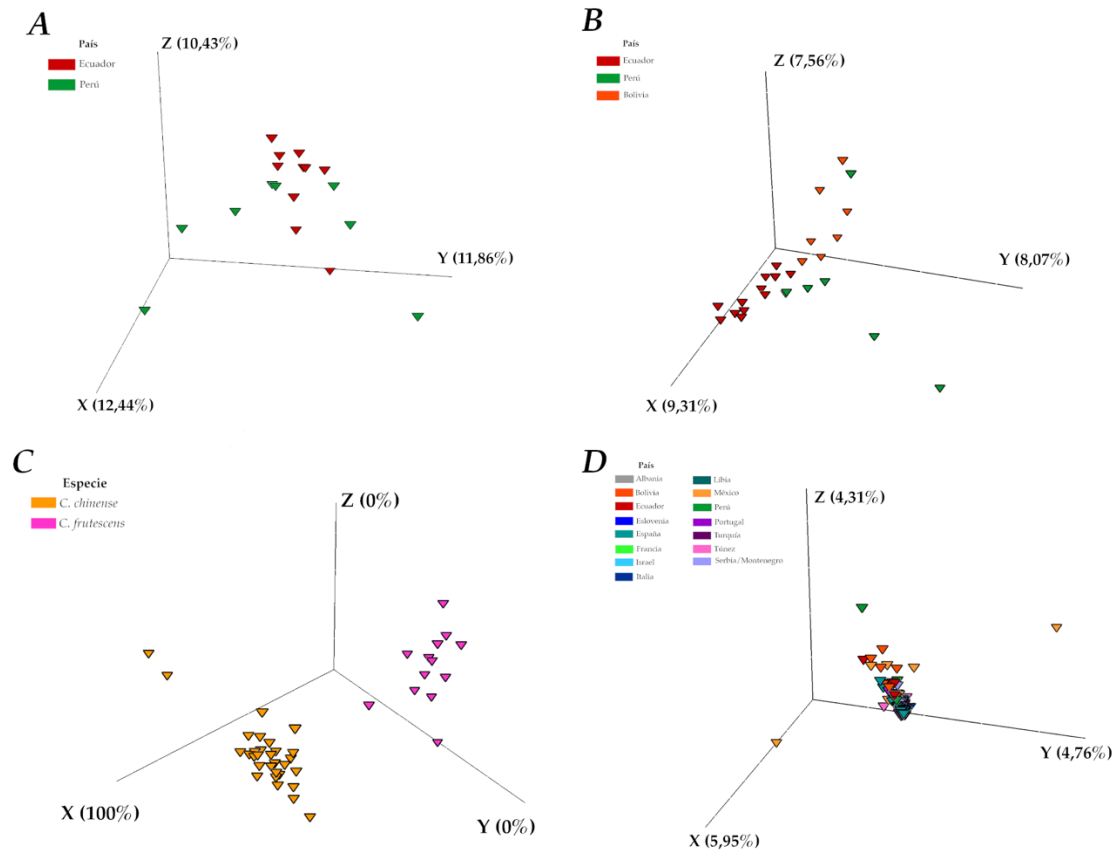


Figura A.2. Análisis factoriales de correspondencia (AFCs) calculados para cada uno de los subgrupos encontrados en el AFC global (Figura 1.5). AFC para las 16 accesiones de *C. pubescens*. B. AFC para las 24 accesiones de *C. baccatum*. C. AFC para las 39 accesiones de *C. chinense* (28) + *C. frutescens* (11). D. AFC para las 147 accesiones de *C. annuum*.

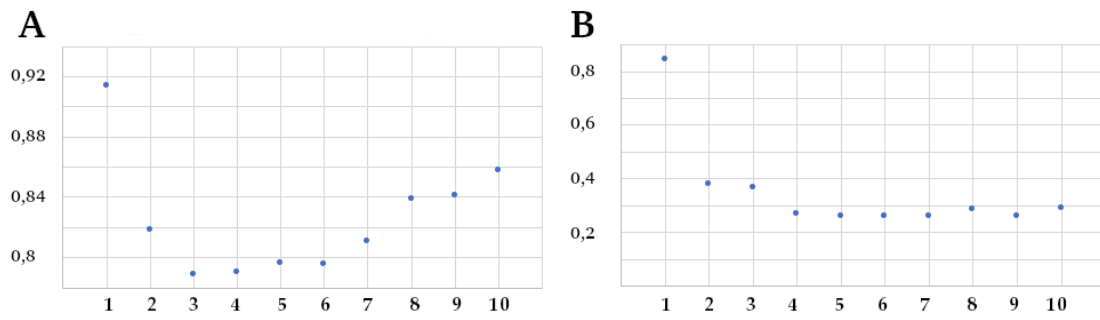


Figura A.3. Resultados de validación cruzada (*cross-validation*) empleada para inferir el número de clústeres en el programa ADMIXTURE. A. Validación cruzada para el conjunto de *C. annuum* que muestra una estructura más probable en $K=3$ clústeres. B. Validación cruzada para el conjunto del complejo *C. annuum* que muestra una estructura más probable en $K=7$ clústeres.

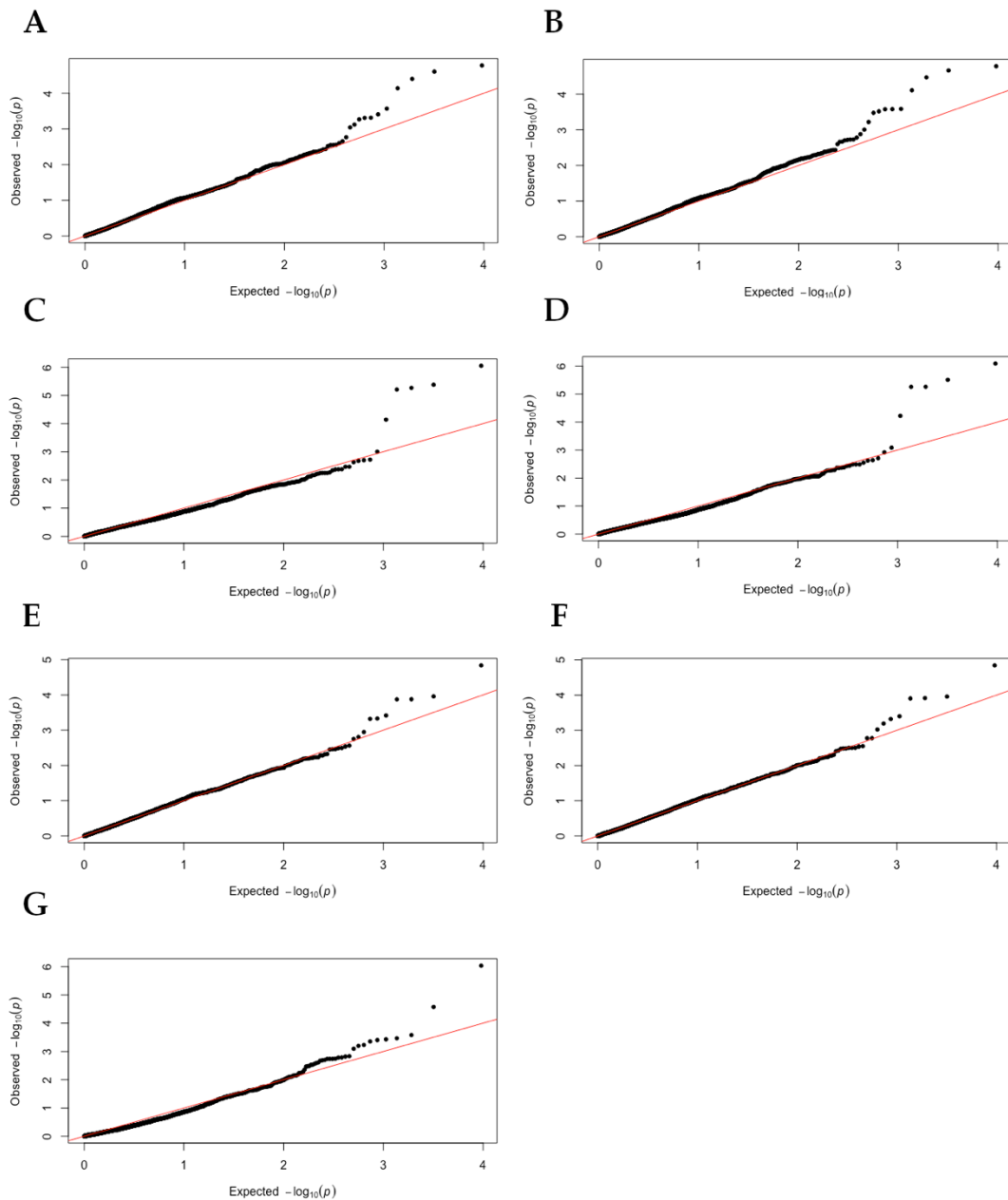


Figura A.4. Gráficos *quantile-quantile* (Q-Q) obtenidos para los parámetros con asociaciones significativas en *C. annuum*. A. Gráfico Q-Q para el parámetro longitud del tallo (LT) con el modelo MLM+Q+K. B. Gráfico Q-Q para el parámetro longitud del tallo (LT) con el modelo MLM+K. C. Gráfico Q-Q para el parámetro verde promedio (AG) con el modelo MLM+Q+K. D. Gráfico Q-Q para el parámetro verde promedio (AG) con el modelo MLM+K. E. Gráfico Q-Q para el parámetro media del valor L (ALV) con el modelo MLM+Q+K. F. Gráfico Q-Q para el parámetro media del valor L (ALV) con el modelo MLM+K. G. Gráfico Q-Q para el parámetro área de la hendidura proximal (PIA) con el modelo GLM+Q. En todos los casos, la línea roja indica el $-\log_{10}(p\text{-valor})$ esperado y los puntos negros el $-\log_{10}(p\text{-valor})$ observado para cada SNP.