

Máster en Mejora Genética Vegetal

Trabajo Fin de Máster

**Identificación de regiones genómicas implicadas en el control del tamaño y peso
de la semilla de judía mediante GWAS**



*Identification of genomic regions involved in the control of the size and weight in common
bean seed using GWAS*

Brezo Mateos Suárez



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA

Director: Prof. Dr. Jaime Prohens

Co-director: Dr. Juan José Ferreira

ib
SERIDA

Resumen

El peso y la morfología de la semilla son caracteres de gran relevancia en *Phaseolus vulgaris* L. debido a que determinan la clase comercial y el rendimiento de los cultivares. En este estudio se llevó a cabo un análisis GWAS con 3099 SNP (distancia > 500pb, valores perdidos < 5%, MAF > 0.01) sobre un panel de diversidad de 308 genotipos (SERIDA-Panel) previamente genotipados mediante GBS. Se detectaron 231 asociaciones significativas que implicaban 146 SNP. Estos SNP fueron agrupados en 25 regiones etiquetadas por al menos 2 SNP y situadas a una distancia física superior a 1 Mb. 9 regiones detectadas se solapaban con QTL asociados con caracteres de semilla en estudios previos. 11 de estas regiones contenían 16 genes candidatos propuestos por su homología con genes de la especie modelo *Arabidopsis thaliana* que participan en el control del fenotipo de la semilla. Los resultados de este trabajo permitieron validar algunos QTL e identificar nuevas regiones relacionadas con el control del fenotipo de la semilla.

Summary

Seed weight and morphology are very important traits in *Phaseolus vulgaris* L. breeding programs because they determine the market class and yield of each line. A diversity panel of 308 bean genotypes (SERIDA-Panel) previously genotyped by GBS was grown in two different campaigns, and a GWAS using 3099 SNP (distance > 500pb, < 5% lost values, MAF > 0.01) was conducted to investigate the architecture of the genetic control of seven seed traits. 231 significant associations involving 146 SNP were detected. These SNP were grouped in 25 regions tagged by two SNP filtering by physical distance (< 1Mb). 9 detected regions were overlapping QTL previously associated with seed traits. 11 regions contained 16 candidate genes proposed because of their homology with *Arabidopsis thaliana* genes involved in the control of the seed phenotype. The finding of this work allowed validating some QTL and identifying new regions involved in the control of seed phenotype.

Abreviaturas

GWAS, Estudio de asociación del genoma completo

Mb, Mega pares de bases

LG, grupo de ligamiento

MAF, mínima frecuencia alélica

SNP, polimorfismo para un nucleótido

AGRADECIMIENTOS

Al grupo de Mejora Genética Vegetal del SERIDA-Villaviciosa:

A Juan José Ferreira, por cambiar de idea en Mayo de 2014 y abrirme las puertas del laboratorio de Genética Vegetal del SERIDA, por la pasión por el trabajo que muestra día a día y que siempre ha tratado de transmitirme (quiero pensar que con éxito) y por las mil y una correcciones.

A Ana Campa por estar siempre dispuesta a escuchar, por su dedicación al trabajo y por ser un gran ejemplo de mujer en la ciencia.

A Ester Murube, por ser una gran compañera (y aún mejor revisora) y un fondo inagotable de bibliografía.

A Marcos Bueno, con quien he compartido el día a día en el laboratorio, por soportar las pequeñas tragedias diarias del proceso de elaboración de un TFM sin dejar nunca de bromear.

A Jaime Prohens por tutorar a distancia, ejercer de puente y resolver todas nuestras dudas y al resto de profesores del máster de mejora por todos los buenos recuerdos y los conocimientos que nos han dado estos dos años.

A mi familia por toda la paciencia que han demostrado y el apoyo incondicional que han sido estos dos últimos años.

Este trabajo ha sido desarrollado en el Servicio Regional de Investigación y Desarrollo Agroalimentario (SERIDA) del Principado de Asturias y en parte financiado por el proyecto AGL2017-87050R del Ministerio de Economía y Competitividad, Gobierno de España.

Imágenes de portada: semillas de los genotipos que integran el panel de diversidad de judía SERIDA-Panel y campos de cultivo del centro. Archivo del SERIDA-Villaviciosa.

ÍNDICE

INTRODUCCIÓN.....	6
La judía común	6
Diversidad genética local	7
Control genético del fenotipo de semilla.....	8
Mapeo por asociación.....	9
Objetivo	11
MATERIAL Y MÉTODOS	12
Material vegetal	12
Fenotipado	12
Genotipado	12
Análisis de asociaciones	13
Correspondencia con QTL y genes candidatos descritos	14
RESULTADOS	15
Variación morfológica y correlaciones entre caracteres.....	15
Principales regiones genómicas implicadas	18
Correspondencia con QTL y genes candidatos descritos	18
DISCUSIÓN	23
CONCLUSIONES.....	27
REFERENCIAS.....	28
Bibliografía	28
ANEXOS	34

ÍNDICE DE TABLAS

Tabla 1. Distribución de los 3099 SNP por cromosoma.....	13
Tabla 2. Principales regiones genómicas implicadas en las asociaciones SNP-carácter más significativas. Se consideran regiones aquellas etiquetadas por al menos 2SNP y separadas por más de 1Mb.....	20
Tabla 3. Co-localización de QTL descritos (González et al., 2016 ¹ ; Cichy et al., 2009 ² ; Pérez Vega et al., 2010 ³ ; Blair et al., 2006 ⁴ ; Sandhu et al., 2018 ⁵ ; McClean et al., 2018 ⁶) para las dimensiones o peso de la semilla y las regiones detectadas en este trabajo.....	21
Tabla 4. Co-localización de genes candidatos para el peso de la semilla propuesto por Schmutz et al (2014) y las regiones detectadas en este trabajo.*: Genes candidatos también detectados en un panel de genotipos mesoamericanos (Schmutz et al., 2014).....	22
Tabla 1 (ANEXOS). SNP significativamente asociados con el carácter área de la semilla ($p < 0,01$) en el análisis MLM.....	35
Tabla 2 (ANEXOS). SNP significativamente asociados con el carácter perímetro de la semilla ($p < 0,01$) en el análisis MLM.....	36
Tabla 3 (ANEXOS). SNP significativamente asociados con el carácter longitud de la semilla ($p < 0,01$) en el análisis MLM.....	37
Tabla 4 (ANEXOS). SNP significativamente asociados con el carácter anchura de la semilla ($p < 0,01$) en el análisis MLM.....	38
Tabla 5 (ANEXOS). SNP significativamente asociados con la relación longitud/anchura ($p < 0,01$) en el análisis MLM.....	40
Tabla 6 (ANEXOS). SNP significativamente asociados con la circularidad ($p < 0,01$) en el análisis MLM.....	42
Tabla 7 (ANEXOS). SNP significativamente asociados con el peso de 25 semillas ($p < 0,01$) en el análisis MLM.....	43

ÍNDICE DE FIGURAS

Figura 1. Representación de la distribución de los 3099 SNP seleccionados a los largo de los 11 cromosomas de judía.....	13
Figura 2. Correlaciones observadas entre los caracteres morfológicos de semilla medidos. Se indica la significación (*, $p < 0,05$; **, $0,05 > p > 0,01$; *** $p < 0,01$) y el valor de r en la parte superior. La diagonal muestra la distribución de los caracteres y la mitad inferior representa la dispersión de los valores.....	15
Figura 3. A) Gráfico Manhattan mostrando 3099 SNP distribuidos en 11 cromosomas de judía indicando en color rojo los SNP asociados con área de la semilla para un umbral de $p < 0,01$. B) Gráfico Manhattan mostrando 3099 SNP distribuidos en 11 cromosomas de judía indicando en color rojo los SNP asociados con Peso de la semilla para un umbral de $p < 0,01$	16
Figura 4. Gráfico de cajas obtenido con el paquete ggplot2 para el peso de la semilla y para los haplotipos de tres SNP significativamente asociado a este carácter. Mediante el test T-student, en los tres SNP diferencias significativas ($p < 0.01$) entre haplotipos. La línea horizontal representa la mediana.....	18

INTRODUCCIÓN

La judía común

La judía común (*Phaseolus vulgaris* L.) es una de las leguminosas más importantes destinadas al consumo humano y constituye una importante fuente de proteínas, calorías y micronutrientes (Broughton et al., 2003). Es una especie diploide con un genoma relativamente pequeño ($2n=22$; 521,1Mb Schmutz et al., 2014).

P. vulgaris es una planta herbácea de ciclo anual con cuyas hojas son compuestas, alternas y trifoliadas. Sus flores son papilionáceas típicas con una quilla que envuelve los 11 estambres junto con el pistilo, morfología que favorece una reproducción predominantemente autógama, si bien en la especie se han descrito una bajas tasas de alogamia (Ferreira et al., 2000). El tallo es delgado y presenta un número variable de nudos y entrenudos, y su longitud difiere según la variedad o genotipo. Dentro de esta especie se han descrito cuatro tipos de hábito de crecimiento diferentes según el tallo termine en inflorescencia floral, el número y longitud de los entrenudos y la capacidad o incapacidad para trepar de la planta (Debouck e Hidalgo 1985).

Esta especie es altamente polimórfica en el fenotipo de planta, vaina o semilla, lo que ha permitido distintas formas de aprovechamiento dependiendo de la variedad: en forma de vaina verde en el momento en el que la semilla comienza a desarrollarse (fréjoles, judía de verdeo, *snap* o *green beans*) o en forma de grano tostado o cocido tras un proceso de rehidratación (judía grano o drybean).

A nivel mundial la producción de judía seca es mayoritaria (26'8 millones de toneladas en 29'4 millones de hectáreas frente a 23 millones de toneladas en 1'6 millones de hectáreas de la judía verde) y se centra en el continente asiático, donde se produce el 45% del total, destacando el papel de Myanmar (5'1 millones de toneladas) y la India (3'8 millones de toneladas). El resto de producción a nivel mundial se distribuye principalmente entre América y África, con un 26'1 y un 24'2% respectivamente, destacando la producción de Brasil (2'6 millones de toneladas) y Estados Unidos (1'6 millones de toneladas) (FAOstat, 2016). Por su parte el cultivo de judía de verdeo se localiza mayoritariamente en Asia, donde se produce el 91,9% principalmente en China (18'7 millones de toneladas) (FAOstat, 2016). En España la superficie de cultivo de esta especie se ha incrementado paulatinamente en la última década llegando a 9400 hectáreas, con una producción total de 17800 toneladas (MAPAMA, 2016).

P. vulgaris fue domesticada por las culturas precolombinas de América Central y del Sur (Gepts et al., 1986; Gepts, 1998) en dos eventos paralelos de domesticación que tuvieron lugar sobre dos pools genéticos silvestres que habían divergido previamente: el mesoamericano (asociado al centro homónimo y distribuido entre el norte de México y Colombia) y el andino (distribuido desde el sur de Perú al noroeste de Argentina) (Kwak&Gepts, 2009). Estos procesos de domesticación han dado lugar a la existencia de dos grupos de germoplasma dentro del material domesticado bien diferenciados y asociados respectivamente al centro Mesoamericano y al Andino (Bitocchi et al., 2012) así como en los parientes silvestres de la especie, que actualmente se distribuyen entre el norte de México y el noroeste de Argentina.

Con la colonización europea de América la especie se introdujo en la Península Ibérica, desde donde pasó a Europa y se dispersó por todo el mundo. Como consecuencia de los procesos de domesticación y posterior dispersión y adaptación a nuevos ambientes fue surgiendo un gran número de variedades que difieren en caracteres morfológicos y fisiológicos, incluyendo caracteres de interés agronómico tales como el tamaño o la calidad de la semilla. Esta variación ha sido ampliamente explotada tanto en programas de mejora como en estudios de diversidad (Pérez-Vega et al., 2009).

Diversidad genética local

La Península Ibérica reúne una importante diversidad de *P. vulgaris* que fue puesta de manifiesto en el trabajo de prospección y caracterización de Puerta Romero (1961) en el que ya se describen algunas de las variedades tradicionales actuales como los tipos faba granja, verdina, riñón, canellini, canela, ganxet, gernikesa, etc. Esta amplia diversidad probablemente guarda relación con el papel de la Península Ibérica como vía de dispersión del cultivo desde América a otros territorios y por ello algunos autores consideran la Península Ibérica como un centro de diversidad secundario de la especie (Santalla et al., 2002; Rodiño et al., 2003).

Un reflejo de esta diversidad local son las seis producciones locales de judía grano que actualmente se encuentran amparadas por marcas de calidad diferenciada (MAPAMA, 2016): IGP Faba Asturiana, IGP Faba de Lourenza, IGP judía del Barco de Ávila, IGP Alubia de la Bañeza, DOP Mongeta del Ganxet y DOP Fesols de Santa Pau. Asimismo existen estatus de protección menores para cultivos más locales como la alubia tolosana y la alubia pinta alavesa

en el País Vasco, la alubia verdina en Asturias, el corico montañés en Cantabria y la alubia de Anguiano en La Rioja.

En gran medida la diversidad local de la especie está reunida y preservada en la colección del Centro de Recursos Fitogenéticos-Madrid. Este centro conserva la colección base nacional (colección a largo plazo) de esta especie y una colección activa que actualmente mantiene 3383 accesiones, en su mayor parte fruto de prospecciones en el territorio nacional. A partir de esta colección se generó una colección nuclear de 211 entradas empleando los datos de pasaporte y de semilla con el fin de representar la diversidad reunida en la colección (De la Rosa et al 2000; Pérez-Vega et al 2009). En el año 2010 se llevó a cabo una actualización eliminando entradas redundantes que habían sido detectadas en las caracterizaciones posteriores e incorporando nuevas entradas provenientes de áreas geográficas poco representadas, de modo que actualmente la colección nuclear española consta de 202 entradas. (INIA: <http://wwwx.inia.es/inventarionacional/>).

Control genético del fenotipo de semilla

El fenotipo de semilla se puede descomponer en diferentes caracteres para su caracterización: colores, distribución del color, forma, dimensiones o peso de la semilla. Las características de la semilla son de especial interés en esta especie dado que determinan el tipo comercial y constituyen uno de los objetivos principales de su cultivo. Además el peso de la semilla constituye un componente del rendimiento junto con el nº de vainas/planta y nº de semilla/vaina (Kamfwa et al 2015).

Dada la importancia de la morfología y las dimensiones de la semilla se han desarrollado numerosos estudios con el fin de esclarecer el control genético de estos caracteres. Park et al. (2000) propuso QTL asociados con el control del peso de la semilla en Pv03, Pv04, Pv06 y Pv08, con la longitud en Pv08 y Pv02 y con la anchura en los grupos de ligamiento Pv04, Pv06 y Pv11 en una población recombinante derivada del cruce de las líneas PC-50 x XAN-159 (Park et al., 2000). Blair et al. (2006) detectó QTL relacionados con el rendimiento de la semilla en Pv02, Pv03, Pv04 y Pv09 en una población derivada de un retrocruzamiento de las líneas ICA Cerinza (parental recurrente) x G224404 (donante silvestre) (Blair et al., 2006). Por su parte Cichy et al.,(2009) propone QTL para el control del peso de la semilla en Pv01, asociado al entorno del gen *fin* en una población RIL derivada del cruce de las líneas G19833 y AND696 (Cichy et al., 2009). En otro trabajo Pérez-Vega et al. (2010) localizó

14 QTL significativamente asociados a las dimensiones y peso de la semilla en una población RIL derivada del cruzamiento de las líneas Xana x Cornell (Pérez-Vega et al., 2010). Finalmente González et al (2017) describen 12 QTL para estos mismos caracteres analizando las poblaciones MA (PHA0419 /Beluga) y 8 en la población AM (Beluga/PHA0399) (González et al., 2017).

Además de los QTL se han descrito genes mayores que se relacionan con el control genético del peso y otros aspectos morfológicos de la semilla como el gen *P/p*, que determina la presencia de color y está asociado con el peso de la semilla (Sax, 1923) de forma que se considera el primer marcador asociado a un carácter cuantitativo. Otro ejemplo sería el gen *Asp/asp*, que condiciona el brillo de la cubierta y está asociado con el peso de la semilla y otros caracteres como la respuesta a la hidratación, la intensidad de color o la producción (Sandhu et al., 2018).

Mapeo por asociación

Tradicionalmente el estudio de la herencia de los caracteres de semillas se abordaba a través del análisis genético directo (*forward genetic*) mediante el estudio de segregación en poblaciones biparentales. Los avances en técnicas de secuenciación de ADN que permiten un genotipado masivo (p.ej. *genotyping by sequencing*; Elshire et al., 2011) así como la disponibilidad del genoma de la especie secuenciado y anotado (Schmutz et al., 2014) permiten el desarrollo de otro tipo de estrategia: el mapeo de asociación.

El mapeo de asociación, también conocido como mapeo de desequilibrio de ligamiento (*linkage disequilibrium mapping*) ha surgido como una herramienta para resolver la base genética de la expresión de caracteres aprovechando la variabilidad histórica fruto de sucesivos eventos de recombinación evolutivos a nivel de población (Nordborg & Tavare, 2002; Risch & Merikangas, 1996) lo que aumenta la potencia del análisis frente al uso de poblaciones biparentales entre los que se dan pocos eventos de recombinación (Burghardt et al., 2017). El mapeo de asociación ofrece como ventajas respecto a los mapas de ligamiento el incremento de la resolución, una menor inversión de tiempo y un mayor número de alelos (Yu & Buckler, 1996) y puede realizarse sobre genes candidatos (*candidate-gene association mapping*) o sobre la variación del genoma completo (*genome-wide association mapping*, "GWAS" o *genome-scan*) (Risch & Merikangas, 1996). Este último permite un estudio de varios caracteres

mediante el análisis de su asociación con cientos de miles de marcadores moleculares distribuidos por todo el genoma (Zhu et al., 2008).

La mayor ventaja que presenta este último es que mientras que en poblaciones biparentales se investiga la variación entre los parentales, en un GWAS se explora un mayor rango de variación fruto de numerosos eventos de recombinación y, por tanto, ofrece la posibilidad de estudiar la arquitectura del control genético de un carácter (Korte and Farlow, 2013; Burghardt et al., 2017). Esta estrategia de análisis está comenzando a ser utilizada en el análisis del control genético del fenotipo de semilla, particularmente en su peso.

En el estudio del control genético del peso de la semilla de judía cabría destacar el estudio de Schmutz et al. (2014) en el cual se proponen 139 genes candidatos para el control del peso de la semilla mediante su homología con genes de *Arabidopsis thaliana* descritos por Van Daele et al. (2012), validando algunos de ellos en un análisis GWAS realizado sobre un panel de genotipos mesoamericanos (Schmutz et al., 2014; Van Daele et al., 2012). Posteriormente en el trabajo de Moghaddam et al. (2016) se investiga la asociación con el peso de la semilla en un panel de genotipos del acervo mesoamericano sin localizar asociaciones relevantes, aunque en una subpoblación de este panel se detectan asociaciones muy significativas con SNP localizados en los cromosomas 8 y 10; se proponen dos genes candidatos para el control del peso de la semilla: Phvul.010G017600 y Phvul.008G013300 (Moghaddam et al., 2016). En 2018 McClean et al. confirman la asociación entre el gen P y el peso de la semilla y proponen la implicación del gen candidato Phvul.007G200300 (McClean et al., 2018).

Objetivo

El SERIDA dispone de un Panel de diversidad (SERIDA-Panel) constituido por 308 líneas de judía (Campa et al., 2018) que ha sido establecido tanto para el análisis de la diversidad genética local como para estudios de asociación que complementen los trabajos de mapeo en poblaciones biparentales. Este panel ha sido genotipado mediante GBS (*genotyping by sequencing*) con 9972 SNP.

El objetivo de este trabajo fue identificar las regiones genómicas implicadas en el control de 7 caracteres relacionados con las dimensiones y peso de la semilla mediante un análisis de asociación (GWAS) realizado sobre un panel de 308 líneas de judía previamente genotipadas mediante 'genotyping by sequencing'(GBS).

MATERIAL Y MÉTODOS

Material vegetal

Se utilizó el Panel SERIDA de diversidad de judía establecido a partir del material reunido en la colección de semillas del SERIDA Villaviciosa, Asturias. Este panel está compuesto por 308 líneas e incluye: 213 líneas derivadas de acciones locales, la mayoría incluidas en la colección nuclear española de judías (De la Rosa et al., 2000; Pérez-Vega et al., 2009); 60 líneas derivadas de cultivares élite de judía de verdeo; 35 líneas derivadas de genotipos relevantes dentro de la especie como los genotipos secuenciados y anotados G198833 y BAT93 (Schmutz et al., 2014).

Fenotipado

La caracterización se desarrolló en las campañas 2016 y 2017 en invernadero (ciclo de cultivo marzo-julio) con parcelas de 8-10 plantas/línea de 0.75 m distribuidas al azar por el invernadero. Se tomó una imagen de 25 semillas de cada línea con un escáner de mesa y con el programa *SmartGrain* (Tanabata et al. 2012) se midieron los siguientes parámetros: **área** (mm²), **perímetro** (mm), **longitud** (mm), **anchura** (mm), **relación longitud/anchura** (L/A) y **circularidad**. También se valoró el peso de 25 semillas (g) en cuatro muestras tomadas al azar de cada línea empleando una balanza de precisión. En el momento del fenotipado las semillas contenían un 11% de humedad.

Genotipado

El panel fue genotipado mediante GBS (Elshire et al. 2011) por la empresa at BGI-Tech (Copenhagen, Denmark) usando el enzima de restricción *ApeKI*. Los tags (secuencias de 64 bases con SNP) fueron alineados con el genoma de la especie disponible en las bases de datos NCBI (BioProject: PRJNA41439). En total se generaron un total de 9972 SNP distribuidos por los 11 cromosomas de judía, que fueron filtrados considerando la distancia física (>500pb), porcentaje de valores perdidos (<5%) y frecuencia alélica mínima (MAF>0,01), resultando en 3099 (**Tabla1, Figura 1**).

Tabla 1. Distribución de los 3099 SNP por cromosoma

Cromosoma	Pv01	Pv02	Pv03	Pv04	Pv05	Pv06	Pv07	Pv08	Pv09	Pv10	Pv11
N. SNP	291	399	359	199	222	238	349	317	307	179	239

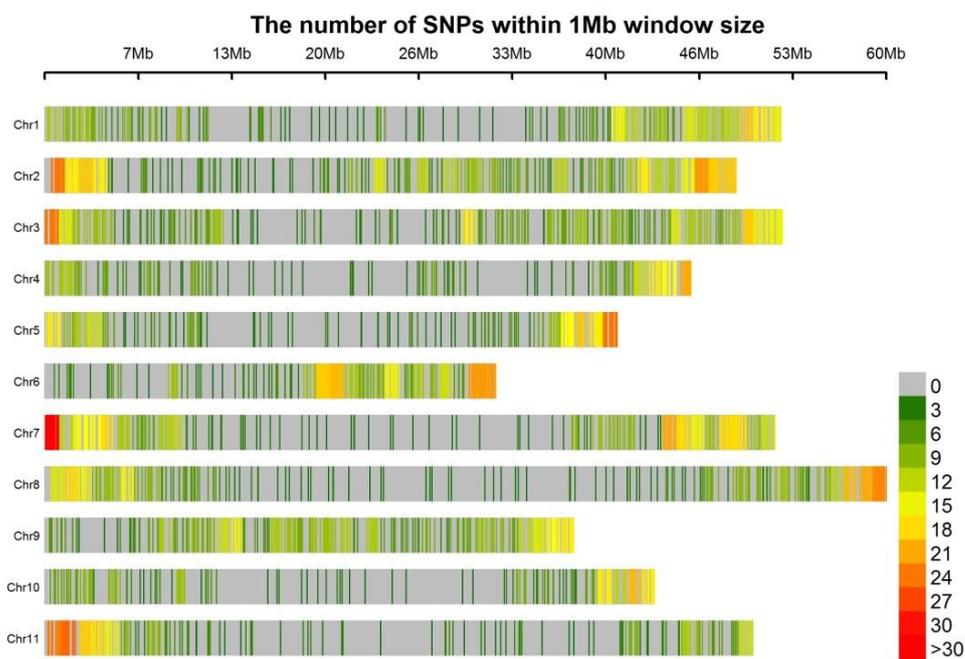


Figura 1. Representación de la distribución de los 3099 SNP seleccionados a lo largo de los 11 cromosomas de judía.

Análisis de asociaciones

Con los datos reunidos en las dos campañas se estimó la media para cada carácter en cada línea del Panel y se calculó la correlación de Pearson entre los caracteres empleando los paquetes 'Corplot' y 'Performance Analytics' del software R (R CoreTeam, 2016).

Para la identificación de asociaciones significativas entre SNP y caracteres se utilizó el método lineal mixto (MLM) implementado en TASSEL V5.2 (Bradbury et al., 2007) según la siguiente ecuación $Y = X\alpha + P\beta + K\mu + e$, en la que Y es el fenotipo estimado, X es el genotipo, P es el efecto de la estructura de la población obtenido a partir de un análisis de componentes principales, K es la matriz de relaciones ('kinship matrix') y e son los efectos residuales. Como umbral de significación para estos múltiples tests de asociación realizados se utilizó el q -value,

una modificación de la tasa de falsos descubrimientos (Storey & Tibshirani 2003) que fue obtenido con la ayuda del paquete Q value.

Los resultados de las asociaciones se representaron como Manhattan plots y QQ plots mediante el paquete de R QQman (Turner, 2018).

Correspondencia con QTL y genes candidatos descritos

Para el estudio de la correspondencia entre los resultados obtenidos en este análisis y QTL previamente publicados se localizó la posición física de sus marcadores asociados o flanqueantes en el genoma de judía, en aquellos casos en los que se disponía de la secuencia (Phaseolus Genes, 2018; <http://phaseolusgenes.bioinformatics.ucdavis.edu/>). La posición física se obtuvo mediante el alineamiento (BLAST) de la secuencia amplificada o cebadores con el genoma de judía; versión 2. (Disponible en Phytozome; www.phytozome.net).

En el caso de la lista de genes candidatos propuestos por Schmutz et al. (2014) se llevó a cabo un proceso similar: se localizó en el genoma (Phytozome, versión 2) la secuencia candidata para obtener su posición física y su función anotada.

RESULTADOS

Variación morfológica y correlaciones entre caracteres

Se observó una amplia variación para todos los caracteres medidos; por ejemplo la variable longitud de la semilla osciló entre 22 y 13 mm mientras que el peso de 25 semillas varió entre 28,98 g de una línea de tipo fabada (oblonga, de tamaño muy grande) y 4,58 g de una línea del tipo 'navy' (redonda y muy pequeña).

Se detectaron correlaciones significativas ($p < 0,05$) entre las 7 variables medidas salvo entre el peso de 25 semillas y la relación longitud/anchura (Figura 2). La correlación entre área y perímetro de la semilla resultó ser la de mayor valor con $r > 0,90$.

El análisis MLM con los 3099 SNP y 7 caracteres de semillas detectó un total de 231 asociaciones significativas SNP-carácter que implican a 146 SNP diferentes distribuidos por los 11 cromosomas. Las **Tablas 1 a 7** en el anexo de este trabajo recogen las características de estas asociaciones para cada carácter.

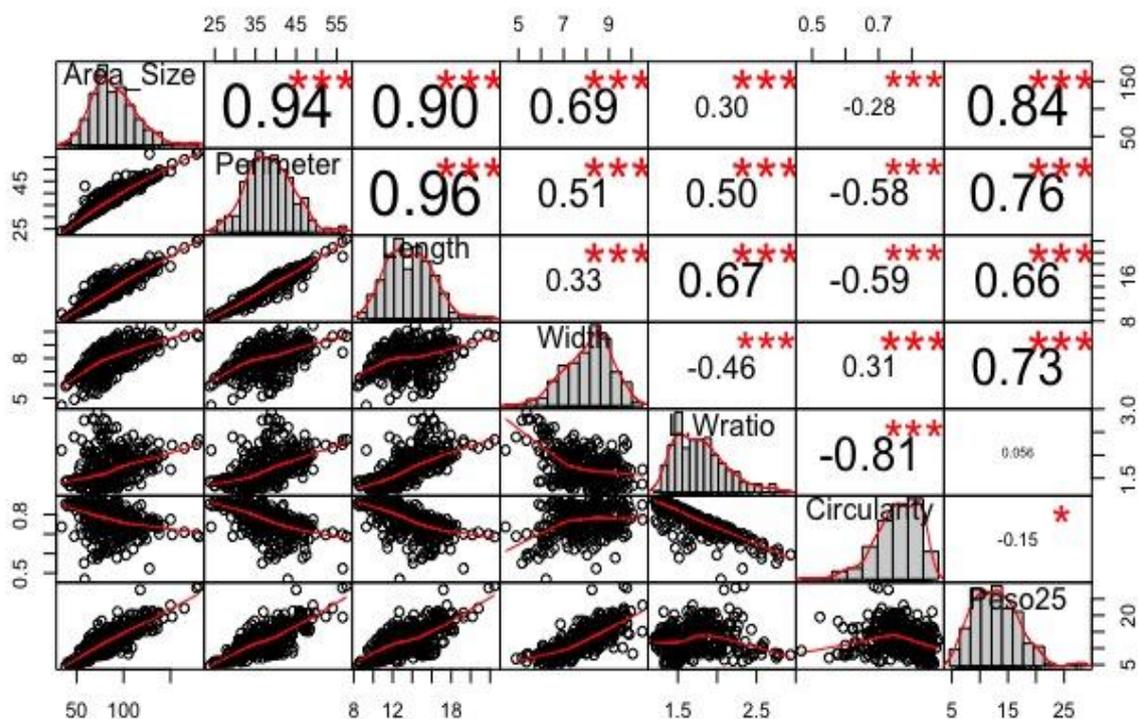


Figura 2. Correlaciones observadas entre los caracteres morfológicos de semilla medidos. Se indica la significación (*, $p < 0,05$; **, $0,05 > p > 0,01$; *** $p < 0,01$) y el valor de r en la parte superior. La diagonal muestra la distribución de los caracteres y la mitad inferior representa la dispersión de los valores.

Representando los resultados en forma de gráficas Manhattan resulta destacable la presencia de SNP significativamente asociados a los caracteres área y peso de 25 semillas situados en los telómeros del cromosoma 1, 3, 7 y 10 (Figura 3). También se detectan SNP significativos aislados que pueden ser artefactos dado que los SNP del entorno no muestran asociaciones significativas.

Para el carácter **área de la semilla** se detectaron un total de 29 SNP significativamente asociados que distribuyen en 10 cromosomas (Tabla 1 anexos; Figura 3A). El mayor número de SNP se detectó en el cromosoma 10 entre las posiciones 39,4 y 42,1 Mb. Lo mayores valores de significación [$-\log_{10}(p) = 4,51$] se detectaron para el SNP s8_55946412, localizado en el cromosoma 8.

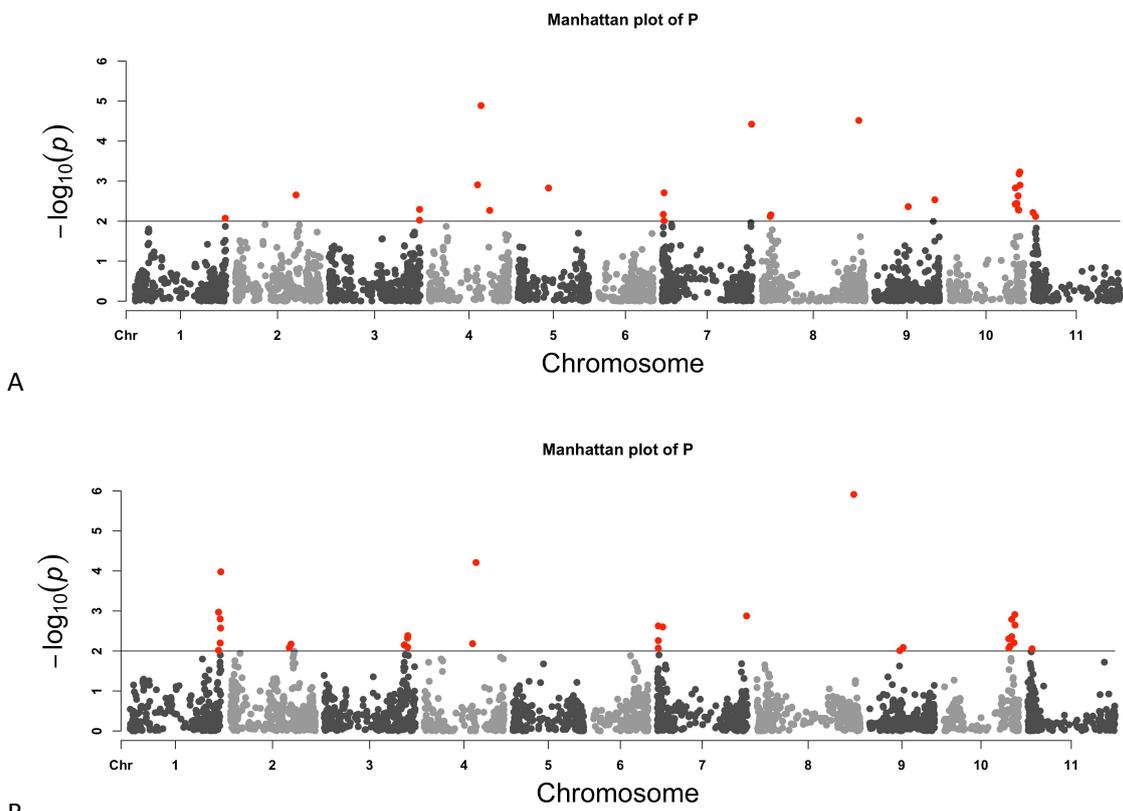


Figura 3. A) Gráfico Manhattan mostrando 3099 SNP distribuidos en 11 cromosomas de judía indicando en color rojo los SNP asociados con área de la semilla para un umbral de $p < 0,01$. **B)** Gráfico Manhattan mostrando 3099 SNP distribuidos en 11 cromosomas de judía indicando en color rojo los SNP asociados con Peso de la semilla para un umbral de $p < 0,01$.

Para el carácter **perímetro de la semilla** se detectaron un total de 21 SNP significativamente asociados que distribuyen en 8 cromosomas (Tabla 2 anexos). El mayor número de SNP se detectó en el cromosoma 10 entre la posición 41,2 y 42,1 Mb. Lo mayores

valores de significación $[-\lg_{10}(p) = 5.05]$ se detectaron para el SNP s8_55946412, localizado en el cromosoma 8.

En relación con el carácter **longitud de la semilla** se detectaron un total de 17 SNP significativamente asociados que distribuyen en 7 cromosomas (**Tabla 3** anexos). Los cromosomas 4 y 9 disponían del mayor número de SNP distribuidos entre las posiciones 16.2-45.9 y 20.0-35.4 respectivamente. Siete SNP mostraron unos valores de significación mayores de 3, siendo el más significativo s8_55946412 $[-\lg_{10}(p) = 5.616]$.

Para el carácter **anchura de la semilla** se detectaron un total de 46 SNP significativamente asociados y localizados en 10 cromosomas (**Tabla 4** anexos). Muchos de estos SNP se localizaron en los cromosomas 7 (11 SNP) y 10 (13 SNP). Siete SNP mostraron unos valores de significación mayores de 3, destacando s10_38396717 el más significativo $[-\lg_{10}(p) = 4.188]$.

Para el **índice longitud/anchura** se detectaron un total de 43 SNP significativamente asociados y localizados en 9 cromosomas (**Tabla 5** anexos). Muchos de estos SNP se localizaron en los cromosomas 2 (9 SNP) y 7 (16 SNP). Seis SNP mostraron unos valores de significación mayores de 3, siendo el más significativo s4_43792143 $[-\lg_{10}(p) = 4.122]$

En cuanto al **índice de circularidad** se detectaron un total de 30 SNP significativamente asociados (**Tabla 6** anexos). Estos SNP se localizaron en 9 cromosomas aunque mayoritariamente en los cromosomas 2 (8) y 7 (12). Los valores de significación oscilaban entre 2 y 3, destacando s4_43792143 $[-\lg_{10}(p) = 3.492]$.

Finalmente, para el carácter **peso de la semilla** se detectaron un total de 33 SNP significativamente asociados (**Tabla 7 anexos; Figura 3B**). Estos SNP se localizaron en 9 cromosomas, destacando Pv10 por ser el que disponía del mayor número de SNP (9). Tres SNP mostraron unos valores de significación mayores de 3, destacando s8_55946412 $[-\lg_{10}(p) = 5.914]$.

Dado que este es el carácter de semilla más estudiado por su relevancia, se exploró la existencia de diferencias significativas (T-student) entre haplotipos para los SNP con mayores valores en la significación en la asociación, localizados en los telómeros de los cromosomas 1, 3, 7 y 10. Se encontraron diferencias significativas en todos los casos excepto para dos SNP localizados en el telómero del cromosoma 3 (s3_47879520, s3_47996483). La **Figura 4** muestra mediante un diagrama de cajas la variación encontrada para el Peso de semilla entre

haplotipos de tres SNP (s1_51995080, s7_663223, s10_39183928) localizados en los telómeros de los cromosomas 1, 7 y 11 (véase **Figura 2a** también).

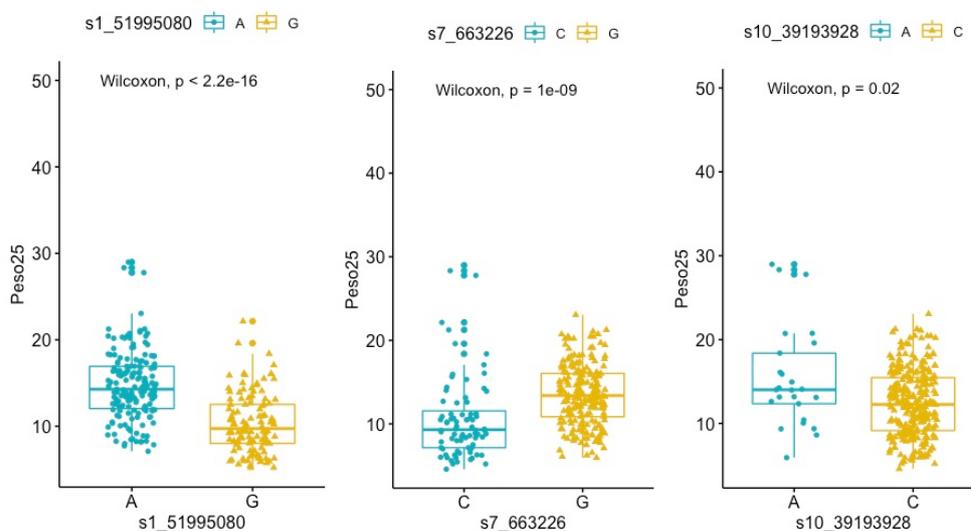


Figura 4. Gráfico de cajas obtenido con el paquete ggplot2 para el peso de la semilla y para los haplotipos de tres SNP significativamente asociado a este carácter. Mediante el test T-student, en los tres SNP diferencias significativas ($p < 0.01$) entre haplotipos. La línea horizontal representa la mediana.

Principales regiones genómicas implicadas

Considerando la posición física de los SNP se definieron regiones genómicas etiquetadas por el menos dos SNP y separadas entre ellas por al menos 1 Mb. Bajo estos criterio se detectó un total de 25 regiones (**Tabla 2**) estaban etiquetadas por un promedio de 4,8 SNP oscilando entre 2 y los 17 que abarca la región Pv10_41. El tamaño de estas regiones varió entre 0,01 y 3,5 Mb y se localizaron en todos los cromosomas excepto el 6, destacando los cromosomas 2 y 7 con 5 y 6 regiones respectivamente. Además resulta llamativa la disposición en tándem y marcadamente telomérica de las regiones Pv02_40, Pv02_45 y Pv02_47 en el cromosoma Pv02 y las regiones Pv07_28, Pv07_31 Pv07_37 y Pv07_39 en el cromosoma Pv07.

Asimismo se observa una diferenciación de las regiones en cuanto a los caracteres con los que están asociadas significativamente; ejemplo de ello sería el carácter Peso de 25 semillas se localizó en 8 regiones mientras que el Área de la semilla aparece asociada a 11 regiones (ver **Tabla 2**).

Correspondencia con QTL y genes candidatos descritos

Se estudió la correspondencia entre las regiones candidatas detectadas y la posición física de los marcadores de QTL previamente asociados a tamaño o peso de la semilla (**Tabla 3**)

Tabla 2. Principales regiones genómicas implicadas en las asociaciones carácter-SNP significativas. Se consideran regiones aquellas etiquetadas por al menos 2 SNP y separadas por más de 1 Mb

Región	Chr	Posición	n°SNP	Caracteres						
Pv01_	1	44.7_44.8	3	Ratio L/A						
Pv01_	1	50.1_51.4	8	Peso25	Anchura	Área				
Pv02_	2	0.66_0.67	2	Ratio L/A						
Pv02_	2	34.8_38.0	9	Área	Longitud	Peso25	Anchura			
Pv02_	2	40.5_42.2	5	Anchura	Circularidad	Ratio L/A				
Pv02_	2	45.90_45.98	3	Ratio L/A	Circularidad					
Pv02_	2	47.75_47.98	2	Perímetro	Longitud	Ratio L/A	Circularidad			
Pv03_	3	47.09_47.19	2	Peso25						
Pv03_	3	48.9_50.1	4	Peso25	Anchura					
Pv03_	3	53.0_53.03	2	Área	Perímetro	Longitud	Anchura			
Pv04_	4	34.14_35.27	2	Área	Perímetro	Longitud	Peso25			
Pv04_	4	44.0_45.9	3	Anchura	Circularidad	Ratio L/A	Perímetro	Longitud		
Pv05_	5	15.5_17.7	3	Área	Perímetro	Longitud	Ratio L/A	Circularidad		
Pv07_	7	0.64_12.8	6	Peso25	Área	Perímetro	Anchura			
Pv07_	7	4.1_5.6	11	Perímetro	Anchura					
Pv07_	7	28.4_28.42	2	Ratio L/A						
Pv07_	7	31.9_32.6	7	Ratio L/A						
Pv07_	7	36.8_37.8	6	Circularidad	Longitud	Ratio L/A				
Pv07_	7	39.1_39.5	3	Área	Longitud	Anchura	Perímetro	Peso25		
Pv08_	8	5.07_6.2	5	Área	Longitud	Anchura	Perímetro			
Pv09_	9	13.06_13.68	2	Ratio L/A	Circularidad					
Pv09_	9	20.04_20.04	7	Área	Perímetro	Longitud				
Pv09_	9	37.7_38.0	2	Ratio L/A	Anchura					
Pv10_	10	38.6_42.1	17	Peso25	Área	Anchura	Perímetro	Longitud	Circularidad	
Pv11_	11	1.15_2.71	5	Anchura	Ratio L/A	Área	Longitud	Peso25		

. Los resultados indican que la posición de nueve regiones identificadas en este trabajo (Pv01_44, Pv01_51, Pv02_36, Pv02_40, Pv07_01, Pv07_04, Pv07-39, Pv10_41 y Pv11_01) se superpone con la posición física de QTL previamente mapeados en diferentes poblaciones biparentales. Resulta interesante destacar que algunas regiones corresponden con QTL identificados en trabajos independientes: Pv01_44, Pv02_36 y Pv11_01.

Región	Chr	Posición	QTL	Marcadores	Ref	Posición de los marcadores
Pv01_44	1	44.7_44.8	SL-1.2MA	BMD045 - FIN	1	Chr01:44257277..44258304/ -
			SeedW_LP2000	Fin (gene TFL1y)	2	Chr01:44856374..44857816
Pv01_51	1	50.1_51.4	SW-1AM	IAC023 - BMB356	1	Chr01:50001762..50002487 / Chr01:50687551..50690651
			SL-1AM	IAC023 - BMB356	1	Chr01:50001762..50002487 / Chr01:50687551..50690651
			SWI-2MA	IAC013 - BM221	1	Chr01:51023984..51024203 /
Pv02_06	2	0.66_0.67				
Pv02_36	2	34.8_38.0	SL2	SW13	3	Chr02:36923562..36923987
			SW2.1	BM143	4	
			100-GW	BM143	5	Chr02:33623552..33623570/ Chr02:33623400..33623420
Pv02_40	2	40.5_42.2	SL-2AM	BMC280 - PVESTBR008	1	Chr02:42426509..42426637 / (Chr02:42426504..42426525
Pv02_45	2	45.90_45.98				
Pv02_47	2	47.75_47.98				
Pv03_47	3	47.09_47.19				
Pv03_49	3	48.9_50.1				
Pv03_53	3	53.0_53.03				
Pv04_34	4	34.14_35.27				
Pv04_44	4	44.0_45.9				
Pv05_15	5	15.5_17.7				
Pv07_01	7	0.64_1.28	SW_MO16	gen Asp	6	1,17 - 1,59 Mb Pv07 en mapa XC
Pv07_04	7	4.1_5.6	SW7.1	Phs	4	5.1- 5.4Mb mapa Pv07 XC
Pv07_28	7	28.4_28.42				
Pv07_31	7	31.9_32.6				
Pv07_37	7	36.8_37.8				
Pv07_39	7	39.1_39.5	WI7	gen P	6	
Pv08_05	8	5.07_6.2				
Pv09_13	9	13.06_13.68				
Pv09_20	9	20.04_20.04				
Pv09_37	9	37.7_38.0				
Pv10_41	10	38.6_42.1	SL10	PVBR135	3	Chr10:41126781..41126800 / Chr10:41126644..41126665
Pv11_01	11	1.15_2.71	SW11a	BM022	4	Chr11:2368475..2369539
			SWI-11.1AM	BM240 - PVESTBR057	1	(Chr11:3034455..3034476 / Chr11:3033977..3033998)

Tabla 3. Co-localización de QTL descritos (González et al., 2016¹; Cichy et al., 2009²; Pérez Vega et al., 2010³; Blair et al., 2006⁴; Sandhu et al., 2018⁵; McClean et al., 2018⁶) para las dimensiones o peso de la semilla y las regiones detectadas en este trabajo.

Paralelamente se llevó a cabo un estudio de la correspondencia entre las regiones acotadas y los 139 genes candidatos para el peso de la semilla propuestos por Schmutz et al. (2014) a partir de un estudio de homología con genes implicados en el control de este carácter en *Arabidopsis thaliana* (Van Daele et al., 2012; Schmutz et al., 2014). Se observó que 11 de las regiones propuestas incluyen un total de 16 genes candidatos; cuatro los cuales habían sido detectados en un estudio GWAS similar desarrollado sobre un panel de genotipos mesoamericanos (Schmutz et al., 2014).

Tabla 4. Co-localización de genes candidatos para el peso de la semilla propuesto por Schmutz et al (2014) y las regiones detectadas en este trabajo.*: Genes candidatos también detectados en un panel de genotipos mesoamericanos (Schmutz et al., 2014)

Nombre	Chr	Posición	Gen candidato	Localización del gen (pb)	
Pv01_44	1	44.7_44.8			
Pv01_51	1	50.1_51.4	Phvul.001G261500	50901610	50903533
Pv02_0.6	2	0.66_0.67			
Pv02_36	2	34.8_38.0	Phvul.002G191500*	35550216	35553534
Pv02_40	2	40.5_42.2	Phvul.002G246800	41852748	41859007
Pv02_45	2	45.90_45.98	Phvul.002G282200	45145416	45150459
			Phvul.002G285000	45392829	45395648
Pv02_47	2	47.75_47.98			
Pv03_47	3	47.09_47.19			
Pv03_49	3	48.9_50.1	Phvul.003G253100*	49243161	49247295
			Phvul.003G264600	50336718	50343534
Pv03_53	3	53.0_53.03			
Pv04_34	4	34.14_35.27			
Pv04_44	4	44.0_45.9			
Pv05_15	5	15.5_17.7			
Pv07_01	7	0.64_1,28			
Pv07_04	7	4.1_5.6	Phvul.007G064800	5789085	5791121
Pv07_28	7	28.4_28.42	Phvul.007G167900	28395643	28399785
Pv07_31	7	31.9_32.6			
Pv07_37	7	36.8_37.8			
Pv07_39	7	39.1_39.5	Phvul.007G269400	39056533	39058538
			Phvul.007G269500	39071636	39073133
Pv08_05	8	5.07_6.2			
Pv09_13	9	13.06_13.68	Phvul.009G078800	13187509	13190472
			Phvul.009G081800	13517847	13521865
Pv09_20	9	20.04_20.04			
Pv09_37	9	37.7_38.0			
Pv10_41	10	38.6_42.1	Phvul.010G117100	39543750	39545984
Pv11_01	11	1.15_2.71	Phvul.011G013500*	1022670	1025112
			Phvul.011G014000*	1087210	1089801

La comparación entre los resultados de las **Tablas 3 y 4** permite identificar siete regiones que contienen QTL y genes candidatos descritos: Pv01_51, Pv02_36, Pv02_40, Pv07_04, Pv07_39, Pv10_41, Pv11_01.

DISCUSIÓN

En este trabajo se llevó a cabo un estudio de asociación utilizando una matriz de 3099 SNP obtenidos mediante GBS y siete caracteres morfológicos de la semilla sobre el panel de diversidad del SERIDA. Los resultados revelaron que el Panel SERIDA presenta una amplia diversidad para el fenotipo de semilla y que los siete caracteres medidos (área, perímetro, longitud, anchura, ratio longitud/anchura, circularidad y peso de 25 semillas) están correlacionados significativamente. Estas correlaciones ya habían sido descritas en trabajos previos (Pérez-Vega et al., 2010) y sugieren la implicación de genes comunes en el control de estos caracteres (muchas de las regiones identificadas están implicadas en el control genético de más de un carácter) aunque también se detectaron regiones específicas de carácter.

El control genético de las dimensiones de la semilla ha sido objeto de estudios previos, aunque el carácter más comúnmente estudiado es el peso debido a su relación directa con el rendimiento (Kafwa et al., 2015). Tradicionalmente estos trabajos empleaban poblaciones biparentales como poblaciones de líneas recombinantes (Blair et al., 2006; Pérez-Vega et al., 2010). El desarrollo de nuevas técnicas de genotipado como el GBS (genotipado por secuenciación) (Elshire, 2011) junto con la disponibilidad del genoma de referencia de *P. vulgaris* secuenciado y anotado (Schmutz et al., 2014) ha permitido el desarrollo de estudios de asociación de genoma completo (GWAS) para el estudio de la arquitectura genómica de los caracteres en paneles de individuos no relacionados. La potencia del análisis GWAS para identificar genes se basa en cinco puntos fundamentales: la complejidad del carácter, su heredabilidad, el número de accesiones y las relaciones entre ellas y la densidad de variantes genómicas. Este análisis daría el óptimo resultado cuando no hay interferencia entre el parentesco y la variación fenotípica, los responsables de la variación fenotípica son pocos loci de efecto mayor y la mayor parte de la variación fenotípica es el resultado de la genética y no del ambiente (Burghardt et al., 2017).

El análisis MLM reveló 231 asociaciones significativas SNP-carácter que implican a 146 SNP localizados en los 11 cromosomas y que pueden ser agrupados en un total de 25 regiones separadas por una distancia mínima de 1Mb y etiquetadas por al menos 2 SNP. Habitualmente los estudios de asociación solo consideran los genes más próximos al SNP más significativo (Moghaddam et al., 2016), sin embargo este criterio puede conducir a errores ya que se ha visto que ocasionalmente los SNP pueden afectar a genes distantes (Visel et al., 2009). Asimismo estos trabajos generalmente utilizan una alta densidad de SNP aplicando criterios

poco restrictivos en el filtrado. Por todo ello en este trabajo se ha optado por aplicar un filtrado restrictivo (distancia física >500pb, porcentaje de valores perdidos <5%) y definir regiones candidatas a partir de la posición de los SNP significativamente asociados a los caracteres analizados. Este enfoque facilita también el análisis de la correspondencia entre las regiones definidas en este trabajo y los QTL descritos.

Bajo estos criterios se observa que las posiciones de 9 de las 25 regiones detectadas en este trabajo corresponden con QTL previamente descritos a partir del análisis de poblaciones biparentales. Sin embargo, no fue posible estudiar esta correspondencia en todos los casos ya que la secuencia de algunos marcadores subyacentes a QTL de semilla no se encontraba disponible o no se localizó (ej. Park et al 2000).

En la región Pv01_44, localizada en la posición 44.7-44.8 y relacionada con el ratio longitud/anchura se localizaron los QTL SL-1.2MA y SeedW_LP2000, propuestos por González et al. (2016) y Cichy et al. (2009) para el control de la longitud y el peso de la semilla, respectivamente. En la región Pv01_51, también ubicada en el cromosoma 1 y asociada con el peso de 25 semillas, la anchura y el área, se encontraron los QTL SW-1AM, SL-1AM y SW1-2MA propuestos por González et al. (2016) para el control del peso de la semilla (SW-1AM y SW1-2MA) y su longitud (SL-1AM).

En la región Pv02_36, localizada en la posición 34.8-38.0 del cromosoma 2 y relacionada con el control genético del área, la longitud, la anchura y el peso de 25 semillas, se ubicaron tres QTL, SL2, SW2.1, 100-GW, propuestos por Pérez-Vega et al., (2010). La región Pv02_40, en el cromosoma 2, se encuentra situada en la posición 40.5-42.2 y asociada con el control de la anchura, la circularidad y el ratio longitud/anchura. En ella se encuentra el QTL SL-2AM, propuesto por González et al., (2016).

En la región Pv10_41, ubicada en la posición 38.6-42.1 del cromosoma 10 y asociada con el peso de 25 semillas, el área, la anchura, el perímetro, la longitud y la circularidad, se localizó el QTL SL10, propuesto por Pérez-Vega et al., (2010).

En la región Pv11_01, localizada en el cromosoma 11 en posición 1.15-2.71 y relacionada con el control genético de la anchura, el ratio longitud anchura, el área, la longitud y el peso de 25 semillas, se localizaron los QTL SW11a y SWI-11.1AM (Blair et al., 2011; González et al., 2016) relacionados con el control genético del peso de la semilla.

En algunos casos, se pudo establecer una correspondencia mediante la comparación de mapas genéticos a pesar de que la secuencia de los marcadores no se encontrara

disponible; este es el caso de gen *Asp/asp*, que controla el brillo de la semilla (Lamprecht, 1940) y que fue mapeado entre dos SNP localizados en la posición 1.17-1.59 Mb del cromosoma Pv07 en la población de líneas recombinantes Xana/Cornell (Campa et al., 2010). Este gen ha sido significativamente asociado a diferentes caracteres de semillas entre los que se encuentra el peso (Sandhu et al., 2018) por lo que fue posible establecer una correspondencia con la región Pv07_01, detectada en este trabajo. También es el caso del gen *fin*, un gen que controla el hábito de crecimiento determinado frente a indeterminado y que parece estar controlado por **PvTFL1y**, un gen homólogo del gen **Terminal Flower 1 (TFL1)** descrito en *Arabidopsis thaliana* (Kwak et al., 2012). La posición del gen PvTFL1y (Chr01:44856374..44857816) permitió localizar los QTL SL-1.2MA y SW_LP2000 y relacionarlos con la región Pv01_44, detectada en el telómero del cromosoma Pv01.

De forma similar se utilizó el gen *P* para establecer una correspondencia entre el QTL WI7 detectado en la población Xana/Cornell (Pérez-Vega et 2010) y la región Pv07_39 detectada en este trabajo. El gen *P*, que determina el color de la semilla, fue físicamente posicionado en la región 39.31 – 42.52 y en su entorno se sitúa un QTL que afecta a las dimensiones y peso de la semilla (McClellan et al 2018; Sax 1923). Esta posición ha permitido establecer una correspondencia entre el QTL WI7 y la región Pv07_39.

En algunos casos los QTL asociados con el tamaño o peso de la semilla previamente descritos no se superponen exactamente con las regiones propuestas pero se encuentran en sus proximidades. Este es el caso de los QTL sw2.2, sw3.1, sw6.1 sw8.1 sw8.2 sw9.1 y sw10.1 encontrados por Blair et al. (2006) en la población Bc2F2:5 ICA Cerinza x G24404; no se superponen con ninguna de las regiones descritas en este trabajo, si bien en algunos casos se encuentran muy próximos a éstas. Ejemplo de esta situación sería el QTL sw10.1 asociado al marcador BM221 localizado en la posición 44.21 Mb (Chr10:44214961..44215240) próxima a la región Pv10_41. Esta situación puede ser debida a la baja saturación de los mapas, que dificulta un posicionamiento preciso de los QTL. Del mismo modo una baja densidad de SNP en determinadas regiones genómicas puede condicionar la detección de estos QTL en el panel. Otros factores como la contribución del QTL a la expresión del carácter o la composición del panel también influyen en la detección de asociaciones significativas (Burghardt et al., 2017, Korte & Farlow, 2013).

Paralelamente a la publicación del genoma de *P. vulgaris* se propusieron 139 genes candidatos para el peso de la semilla a partir de su homología con genes implicados en el control de este carácter en *A. thaliana* (Schmutz et al., 2014; Van Daele et al 2012). Muchos de

estos genes también están implicado en el proceso de domesticación, dado que uno de los rasgos que más ha variado entre las poblaciones silvestres y variedades domesticadas ha sido el fenotipo de la semilla (Schmutz et al., 2014). Los resultados revelaron que 11 regiones detectadas en este trabajo contienen 16 de estos genes, cuatro de los cuales (**Phvul.002G191500**, **Phvul.003G253100**, **Phvul.011G013500** y **Phvul.011G014000**) también habían sido detectados por Schmutz et al. (2014) en un análisis GWAS utilizando un panel de genotipos perteneciente al acervo genético Mesoamericano.

El análisis *in silico* de estos 16 genes candidatos reveló que codificaban elementos reguladores, factores de activación de citoquininas y giberelinas, una región con repeticiones ricas en leucina y proteínas de unión a histonas, elementos que participan en la regulación de la división celular y el crecimiento y en la morfología final de la semilla, y que mostraban homología con los descritos por Van Daele et al. (2012) en *Arabidopsis thaliana*.

Los resultados obtenidos constituyen una etapa más en la identificación y validación de los genes o secuencias de ADN que contribuyen a la expresión del fenotipo de semilla y aportan nuevos datos para la validación de QTL implicados en el control del fenotipo de semilla al detectar las mismas regiones genómicas implicadas en diferentes análisis independientes. En este sentido, el desarrollo de isolíneas específicas de QTL (líneas que varíen sólo para región genómica que contenga un QTL) podría ser un paso más en la verificación de la implicación de estas regiones.

Asimismo estos resultados podrían ayudar en la mejora genética del rendimiento o de las características morfológicas de la semilla de judía destinada para el consumo de grano mediante la selección asistida y el diseño de marcadores específicos para los QTL más relevantes para cada carácter.

CONCLUSIONES

- El análisis de asociación desarrollado en el panel SERIDA de 308 líneas genotipadas con 3099 SNP ha permitido detectar asociaciones significativas con los caracteres de la semilla valorados.
- Los SNP significativamente asociados se agruparon en 25 regiones genómicas localizadas en 10 de los 11 cromosomas de *P. vulgaris*. Nueve de estas regiones se superponen con la posición de QTL previamente descritos en el análisis de poblaciones biparentales.
- Este estudio confirma la utilidad del Panel SERIDA para estudios GWAS ya que 9 de las 25 regiones identificadas se solapan con la posición de QTL previamente descritos por otros autores, y 11 de estas regiones contienen 16 genes candidatos homólogos de genes implicados en el control genético del peso de la semilla en la especie modelo *Arabidopsis thaliana*.
- Teniendo en cuenta los QTL descritos en la bibliografía, las 25 regiones detectadas en este análisis y la presencia de genes candidatos, se puede considerar que las principales regiones implicadas en el control de las dimensiones y peso de la semillas son Pv01_51, Pv02_36, Pv07_04, Pv07_39, Pv10_41, Pv11_01.
- Los resultados de este trabajo proporcionan nuevas regiones implicadas en el control del fenotipo de semilla que deberían ser validadas en nuevos estudios.

REFERENCIAS

Bibliografía

- Bitocchi, E. , Bellucci, E. , Giardini, A. , Rau, D. , Rodriguez, M. , Biagetti, E. , Santilocchi, R. , SpagnolettiZeuli, P. , Gioia, T. , Logozzo, G. , Attene, G. , Nanni, L. and Papa, R. (2013), Molecular analysis of the parallel domestication of the common bean (*Phaseolus vulgaris*) in Mesoamerica and the Andes. *New Phytol*, 197: 300-313. doi:[10.1111/j.1469-8137.2012.04377.x](https://doi.org/10.1111/j.1469-8137.2012.04377.x).
- Blair, M.W., Astudillo C., Rengifo J., Beebe S.E., Graham R. (2011) *QTL analyses for seed iron and zinc concentrations in an intra-gene pool population of Andean common beans (Phaseolus vulgaris L.)* in *Theoretical Applied Genetics* 122(3):511-21.
- Blair, M.W., Iriarte, G. & Beebe, S. (2006), *Microsatellite marker diversity in common bean (Phaseolus vulgaris L.)* in *Theoretical Applied Genetics* 112: 1149. <https://doi.org/10.1007/s00122-006-0217-2>.
- Bradbury, P., Zhang Z., Kroon D., Casstevens T., Ramdos Y., Buckler E. (2007) *TASSEL: software for association mapping of complex traits in diverse samples* in *Bioinformatics*, 23(19):2633-35.
- Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P., Vanderleyden, J., (2003) *Beans (Phaseolus spp.) – model food legume* in *Plant and Soil*, 252: 55.
- Burghardt, L.T., Young, N.D., Tiffin, P. (2017) *A guide to genome-wide association mapping in plants* in *Current Protocols in Plant Biology*, 2:22-38.
- Campa, A., Murube, E., Ferreira, J.J., (2018) *Genetic diversity gathered in a common bean panel established for genome-wide association study* in *Annual Reports in Bean Improvement Cooperative*, 61, en prensa.
- Campa, A., Pérez-Vega, E., Pascual, A., Ferreira, J. J., (2010) *Genetic analysis and molecular mapping of quantitative trait loci in common bean against Pythium ultimum* in *Phytopathology*, 100(12):1315-20. Doi:10.1094/PHYTO-06-10-0161
- Cichy, K.A., Caldas, G. V., Snapp, S. S., Blair, M. W, (2009) *QTL Analysis of seed iron, zinc, and phosphorus levels in an andean bean population* in *Crop Science*, 49:1742.1750.

- De la Rosa, L., Lázaro, A., Varela, F., (2000) *Racionalización de la colección española de Phaseolus vulgaris L.* p. 55–62. in II Seminario de judía de la Península Ibérica. Actas de Asociación Española de Leguminosas, Villaviciosa (Asturias), Spain.
- Debouck, D.G., Hidalgo, R., (1985) *Morfología de la planta de frijol común* . In: López Genes, M., Fernández, O., Fernando, O.; Schoonhoven, A. van (eds.). Frijol: Investigación y producción. Programa de las Naciones Unidas (PNUD); Centro Internacional de Agricultura Tropical (CIAT), Cali, CO. p. 7-41.
- Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S., Mitchell S.E. *A robust, simple Genotyping-by-Sequencing (GBS) approach for high diversity species* (2011) in Plos One, 6(5):10.
- Ferreira, J. J., Álvarez, E., Fueyo, M. A., Roca, A., Giraldez, R., (2000). *Determination of the outcrossing rate of Phaseolus vulgaris L. Using seed protein markers* in Euphytica, 113:259-263.
- Gepts, P., (1998) *Origin and evolution of common bean: past events and recent trends* in Horticultural Science, 33:1124–1130.
- Gepts, P., Osborn, T.C., Rashka, K., Bliss, F.A., (1986) *Phaseolin-protein Variability in Wild Forms and Landraces of the Common Bean(Phaseolus vulgaris): Evidence for Multiple Centers of Domestication* in Economic Botany, 40: 451.
- González A. M., Yuste-Lisbona, F.J., Fernández-Lozano, A., Lozano, R., Santalla M. (2017) *Genetic Mapping and QTL Analysis in Common Bean*. In: Pérez de la Vega, M., Santalla, M., Marsolais, F.m (eds) The Common Bean Genome. Compendium of PlantGenomes. Springer, Cham.
- González, A. M., Yuste-Lisbona, F. J., Saburido, S., Bretones, S., De Ron, A. M., Lozano, R., & Santalla, M. (2016). *Major Contribution of Flowering Time and Vegetative Growth to Plant Production in Common Bean As Deduced from a Comparative Genetic Mapping* in Frontiers in Plant Science, 7, 1940. <http://doi.org/10.3389/fpls.2016.01940>.
- Kamfwa, K., Cichy, K.A., Kelly, J.D. (2015) *Genome-wide association analysis of symbiotic nitrogen fixation in common bean* in Theoretical and Applied Genetics, 128: 1999.
- Karl Sax, H., McPhee, C., (1923) *Color factors in bean hybrids* in Journal of Heredity 14(5):205–208, <https://doi.org/10.1093/oxfordjournals.jhered.a102316>.

- Korte, A., Farlow, A., (2013) *The advantages and limitations of trait analysis with GWAS: a review* in *Plant Methods*, 9:29.
- Kwak, M., Gepts, P., (2009) *Structure of genetic diversity in the two major gene pools of common bean (*Phaseolus vulgaris* L., Fabaceae)* in *Theoretical and Applied Genetics*, 118: 979.
- Kwak, M., Toro, O., Debouck, D. G., Gepts, P., (2012) *Multiple origins of the determinate growth habit in domesticated common bean (*Phaseolus vulgaris*)* in *Annals of Botany* 110(8): 1573–1580.
- Lamprecht, H (1940) *Zur Genetik von *Phaseolus vulgaris*. XVII-XVIII. Zwei neue Gene für Abzeichen auf der Testa, Punc und Mip, sowie über die Wirkung von V und Inh.* In *Hereditas* 26:292-304.
- McClellan, P., Bett, K. E., Stonehouse, R., Lee, R., Pflieger, S., Moghaddam, S. M., Geffroy, V., Miklas, P., Mamidi, S., (2018) *White seed color in common bean (*Phaseolus vulgaris*) results from convergent evolution in the P (pigment) gene* in *New Phytologist* 219(3):1112-1123. doi: 10.1111/nph.15259.
- Moghaddam S. M., Mamidi, S., Osorno, J.M., Lee, R., Brick, M., Kelly, J., Miklas, P., Urrea, C., Song, Q., Cregan, P., Grimwood, J., Schmutz, J., McClellan, P.E., (2016) *Genome-wide association study identifies candidate loci underlying agronomic traits in a Middle American diversity panel of common bean* in *Plant Genome* 9(3) doi:10.3835/plantgenome2016.02.0012.
- Nordborg, M., Tavaré, S., (2002) *Linkage disequilibrium: what history has told us* in *Trends in Genetics*, 18(2): 83-90.
- Park, S. O., Coyne, D. P., Jung, G., Skroch, P. W., Arnaud-Santana, E., Steadman, J.R., Ariyaranthe, H. M., Nienhuis, J., (2000) *Mapping of QTL for seed size and shape traits in common bean* in *Journal of the American Society Horticultural Science* 125:466–475.
- Pérez-Vega, E., Campa, A., De la Rosa, L., Giraldez, R., Ferreira, J. J., (2009) *Genetic Diversity in a Core Collection Established from the Main Bean Genebank in Spain* in *Crop Science*, 49:1377-1386. doi:10.2135/cropsci2008.07.0409.
- Pérez-Vega, E., Pañeda, A., Rodríguez-Suárez, C., Campa, A., Giraldez, R., (2010) *Mapping of QTL for morpho-agronomic and seed quality traits in a RIL population of common*

- bean (Phaseolus vulgaris L.)* in *Theoretical and Applied Genetics* 120: 1367. <https://doi.org/10.1007/s00122-010-1261-5>.
- Puerta-Romero, J., (1961) *Variedades de judía cultivadas en España* in Ministerio de Agricultura, Pesca y Alimentación. Madrid, España. 789 p.
- R Core Team 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org/>.
- Risch, N., Merikangas, K., 1996. The future of genetic studies of complex human diseases. *Science*. 273(5281):1516–1517.
- Rodiño, A., Santalla, M., De Ron, A., Singh S.P., (2003) *A core collection of common bean from the Iberian peninsula* in *Euphytica* 131: 165. <https://doi.org/10.1023/A:1023973309788>.
- Sandhu, K. S., You, F. M., Conner, R. L., Balasubramanian, P. M, Hou, A., (2018) *Genetic analysis and QTL mapping of the seed hardness trait in a black common bean (Phaseolus vulgaris) recombinant inbred line (RIL) population* in *Molecular Breeding* 38: 34. <https://doi.org/10.1007/s11032-018-0789-y>.
- Santalla, M., Rodiño, A., De Ron, A., (2002) *Allozyme evidence supporting southwestern Europe as a secondary center of genetic diversity for the common bean* in *Theoretical and Applied Genetics* 104: 934. <https://doi.org/10.1007/s00122-001-0844-6>.
- Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S.B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. A., Chovatia, M., Gepts, P., Goodstein, D. M., Gonzales, M., Hellsten, U., Hyten, D. L., Jia, G., Kelly, J. D., Kudrna, D., Lee, R., Richard, M. M., Miklas, P. N., Osorno, J. M., Rodrigues, J., Thareau, V., Urrea, C. A., Wang, M., Yu, Y., Zhang, M., Wing, R. A., Cregan, P. B., Rokhsar, D. S., Jackson, S. A., (2014) *A reference genome for common bean and genome-wide analysis of dual domestications* in *Nature Genetics*, 46(7):707-713. DOI: 10.1038/ng.3008.
- Storey, J. D., Tibshirani, R., (2003) *Statistical significance for genome wide studies*. *PNAS*, 100(16):9440-9445.

- Tanabata, T., Shibaya, T., Hori, K., Ebana, K., Yano, M., (2012) *SmartGrain: High-throughput phenotyping software for measuring seed shape through image analysis* in *Plant Physiology* 160(4): 1871–1880.
- Turner, S., (2018) *Qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots* in *Journal of Open Source Software*, 3(25):731 <https://doi.org/10.21105/joss.00731>.
- Van Daele, I., Gonzalez, N., Vercauteren, I., de Smet, L., Inze, D., Roldán-Ruiz, I., Vuylsteke, M., (2012) *A comparative study of seed yield parameters in Arabidopsis thaliana mutants and transgenics* in *Plant Biotechnology Journal*, 10:488–500.
- Visel, A., Rubin, E. M., Pennacchio, L. A. (2009) *Genomic views of distant-actin enhancers* in *Nature*, 46:199–205.
- Yu, J., Buckler, E.S. (2006) *Genetic association mapping and genome organization of maize* in *Current Opinion in Biotechnology* 17:155–160.
- Zhu, C., Gore, M., Buckler, E. S., Yu, J., (2008) *Status and Prospects of Association Mapping in Plants* in *Plant Genome* 1:5-20. doi:10.3835/plantgenome2008.02.0089.

Webs

CRF-Madrid

<http://wwwx.inia.es/inventarionacional/>

última consulta: 16/07/2018

CRF-INIA

<http://www.crf.inia.es/judias/>

última consulta 16/07/2018

Ministerio de agricultura, pesca y alimentación

<https://www.mapama.gob.es/es/>

última consulta: 16/07/2018

Marcas de calidad (MAPAMA)

<https://www.mapama.gob.es/es/alimentacion/temas/calidad-agroalimentaria/calidad-diferenciada/>

última consulta: 16/07/2018

Phaseolus genes

<http://phaseolusgenes.bioinformatics.ucdavis.edu/>

última consulta: 16/07/2018

Phytozome

<https://phytozome.jgi.doe.gov/pz/portal.html#>

última consulta: 16/07/2018



Tabla 1. SNP significativamente asociados con el **carácter área** de la semilla ($p < 0,01$) en el análisis MLM. Se indican: coordenadas genómicas (Cr, cromosoma y posición física en pb); p-valor de la asociación representado como $-\log_{10}(p)$; frecuencia alélica mínima (MAF); y porcentaje de variabilidad explicada por cada SNP (R^2).

SNP	Chr	Posición tag (V2)	Pol	$-\log_{10}(p)$	MAF	R^2
s1_51779298	1	Chr01:51062799..51062862 (+)	A/T	2,072	0,468	0,023
s2_35091874	2	Chr02:35833226..35833289 (-)	C/G	2,654	0,245	0,032
s3_51843345	3	Chr03:53005170..53005235 (-)	T/G	2,025	0,123	0,023
s3_51868687	3	Chr03:53030484..53030547 (+)	T/A	2,293	0,245	0,027
s4_27768193	4	Chr04:16188553..16188616 (-)	G/T	2,907	0,292	0,036
s4_29757350	4	Chr04:34137292..34137355 (-)	A/G	4,888	0,19	0,065
s4_34708537	4	Chr04:35269988..35270051 (-)	G/A	2,268	0,486	0,026
s5_17539222	5	Chr05:15588232..15588295 (+)	G/A	2,827	0,403	0,036
s7_633265	7	Chr07:650516..650579 (+)	G/A	2,169	0,383	0,025
s7_1066580	7	Chr07:1084387..1084450 (+)	A/G	2,008	0,247	0,023
s7_1090742	7	Chr07:1108527..1108590 (+)	C/T	2,708	0,253	0,032
s7_51283516	7	Chr07:39574655..39574718 (+)	C/T	4,423	0,255	0,058
s8_5080354	8	Chr08:5074042..5074105 (+)	T/A	2,117	0,392	0,025
s8_5497128	8	Chr08:5492175..5492238 (+)	C/A	2,156	0,331	0,024
s8_55946412	8	Chr08:59380789..59380852 (-)	C/T	4,517	0,193	0,059
s9_19503651	9	Chr09:20040175..20040238 (+)	G/A	2,362	0,159	0,027
s9_34774046	9	Chr09:35368558..35368621 (-)	T/C	2,533	0,25	0,03
s10_38333216	10	Chr10:39430517..39430580 (-)	G/A	2,424	0,343	0,03
s10_38396717	10	Chr10:39526879..39526942 (-)	T/A	2,83	0,329	0,035
s10_39185557	10	Chr10:40284905..40284968 (+)	C/A	2,431	0,088	0,028
s10_39193928	10	Chr10:40293254..40293316 (+)	C/A	2,449	0,086	0,029
s10_39978402	10	Chr10:41088440..41088503 (-)	A/T	2,633	0,324	0,032
s10_40091019	10	Chr10:41198897..41198960 (+)	G/A	2,287	0,306	0,027
s10_40329757	10	Chr10:41487036..41487099 (-)	T/C	2,275	0,199	0,028
s10_40536728	10	Chr10:41561615..41561678 (+)	G/A	3,18	0,153	0,04
s10_40979207	10	Chr10:42000131..42000193 (+)	T/C	3,229	0,15	0,04
s10_41090517	10	Chr10:42111336..42111399 (+)	A/G	2,9	0,15	0,036
s11_260055	11	Chr11:254364..254427 (+)	A/C	2,215	0,339	0,026
s11_1689653	11	Chr11:1686318..1686381 (+)	T/C	2,116	0,272	0,024

Tabla 2. SNP significativamente asociados con el **carácter perímetro** de la semilla ($p < 0,01$) en el análisis MLM. Se indican: coordenadas genómicas (Cr, cromosoma y posición física en pb); p-valor de la asociación representado como $-\log_{10}(p)$; frecuencia alélica mínima (MAF); y porcentaje de variabilidad explicada por cada SNP (R^2).

SNP	Chr	Posición tag v2	Pol	$-\log_{10}(p)$	MAF	R^2
s2_47173033	2	Chr02:47746388..47746451 (+)	G/A	2,719	0,328	0,033
s3_43234504	3	Chr03:44378203..44378266 (-)	G/A	2,001	0,047	0,023
s3_51843345	3	Chr03:53005170..53005235 (-)	T/G	2,371	0,123	0,028
s4_29757350	4	Chr04:34137292..34137355 (-)	G/A	3,246	0,19	0,041
s4_43792143	4	Chr04:45971741..45971804 (+)	T/G	2,87	0,261	0,036
s5_1024594	5	Chr05:1034638..1034701 (+)	T/A	2,06	0,387	0,024
s5_17539222	5	Chr05:15588232..15588295 (+)	G/A	3,147	0,403	0,04
s7_1090742	7	Chr07:1108527..1108590 (+)	C/T	2,033	0,253	0,023
s7_10531655	7	Chr07:10731839..10731902 (+)	T/A	2,187	0,065	0,025
s7_50879045	7	Chr07:39166060..39166123 (-)	T/C	2,054	0,464	0,023
s7_50909179	7	Chr07:39194374..39194437 (+)	A/G	2,796	0,353	0,034
s7_51283516	7	Chr07:39574655..39574718 (+)	C/T	4,88	0,255	0,065
s8_6194586	8	Chr08:6249447..6249510 (+)	C/T	2,197	0,085	0,025
s8_55946412	8	Chr08:59380789..59380852 (-)	C/T	5,05	0,193	0,068
s9_33898583	9	Chr09:20040175..20040238 (+)	G/C	2,389	0,207	0,028
s10_39978402	10	Chr10:41088440..41088503 (-)	A/T	2,818	0,324	0,035
s10_40091019	10	Chr10:41198897..41198960 (+)	G/A	2,442	0,306	0,029
s10_40329757	10	Chr10:41487036..41487099 (-)	T/C	2,302	0,199	0,028
s10_40536728	10	Chr10:41561615..41561678 (+)	G/A	2,714	0,153	0,034
s10_40979207	10	Chr10:42000131..42000193 (+)	T/C	2,523	0,15	0,03
s10_41090517	10	Chr10:42111336..42111399 (+)	A/G	2,267	0,15	0,027

Tabla 3. SNP significativamente asociados con el **carácter longitud** de la semilla ($p < 0,01$) en el análisis MLM. Se indican: coordenadas genómicas (Cr, cromosoma y posición física en pb de la secuencias tag que incluye el SNP); p-valor de la asociación representado como $-\log_{10}(p)$; frecuencia alélica mínima (MAF); y porcentaje de variabilidad explicada por cada SNP (R^2).

SNP	Chr	Posición tag	Pol	$-\log_{10}(p)$	MAF	R^2
s2_35091874	2	Chr02:35833226..35833289 (-)	C/G	2,439	0,245	0,029
s2_47173033	2	Chr02:47746388..47746451 (+)	G/A	3,593	0,328	0,046
s3_51843345	3	Chr03:53005170..53005235 (-)	T/G	2,387	0,123	0,028
s3_51868687	3	Chr03:53030484..53030547 (+)	T/A	2,153	0,245	0,025
s4_27768193	4	Chr04:16188553..16188616 (-)	G/T	2,015	0,292	0,023
s4_29757350	4	Chr04:34137292..34137355 (-)	A/G	4,135	0,19	0,055
s4_34708537	4	Chr04:35269988..35270051 (-)	G/A	2,387	0,486	0,028
s4_43792143	4	Chr04:45971741..45971804 (+)	C/T	3,012	0,261	0,038
s5_17539222	5	Chr05:15588232..15588295 (+)	G/A	4,006	0,403	0,054
s7_48741309	7	Chr07:37085151..37085214 (+)	C/G	2,424	0,203	0,028
s7_51283516	7	Chr07:39574655..39574718 (+)	C/T	3,289	0,255	0,041
s8_5497128	8	Chr08:5492175..5492238 (+)	C/A	2,015	0,331	0,023
s8_55946412	8	Chr08:59380789..59380852 (-)	C/T	5,616	0,193	0,077
s9_19503651	9	Chr09:20040175..20040238 (+)	G/A	2,233	0,159	0,026
s9_33898583	9	Chr09:20040175..20040238 (+)	G/C	2,585	0,207	0,031
s9_34720152	9	Chr09:20040175..20040238 (+)	A/T	2,481	0,261	0,031
s9_34774046	9	Chr09:35368558..35368621 (-)	T/C	3,442	0,25	0,044

Tabla 4. SNP significativamente asociados con el **carácter anchura** de la semilla ($p < 0,01$) en el análisis MLM. Se indican: coordenadas genómicas (Cr, cromosoma y posición física en pb); p-valor de la asociación representado como $-\log_{10}(p)$; frecuencia alélica mínima (MAF); y porcentaje de variabilidad explicada por cada SNP (R^2).

SNP	Chr	Pos V2	Pol	$-\log_{10}(p)$	MAF	R^2
s1_51531291	1	Chr01:50814659..50814722 (-)	G/C	2,201	0,09	0,024
s1_51835134	1	Chr01:51118611..51118659 (-)	T/C	2,003	0,457	0,022
s2_2913819	2	Chr02:2868495..2868558 (-)	C/G	2,255	0,48	0,025
s2_17371758	2	Chr02:17750828..17750891 (+)	T/C	2,585	0,285	0,029
s2_35663954	2	Chr02:36295463..36295526 (+)	A/G	2,139	0,319	0,023
s2_35786641	2	Chr02:36418809..36418872 (+)	C/T	2,558	0,322	0,029
s2_36155857	2	Chr02:36805586..36805648 (-)	A/G	2,854	0,325	0,034
s2_36261610	2	Chr02:36911227..36911290 (+)	G/A	2,793	0,312	0,033
s2_36395852	2	Chr02:37050582..37050645 (-)	T/C	2,084	0,303	0,023
s2_36649418	2	Chr02:37309908..37309971 (+)	C/T	2,699	0,352	0,031
s2_37004268	2	Chr02:38086557..38086620 (+)	A/T	2,896	0,359	0,034
s2_37439123	2	Chr02:40505314..40505377 (-)	A/G	2,67	0,373	0,03
s3_47996483	3	Chr03:49131292..49131355 (+)	G/C	2,504	0,155	0,028
s3_48956412	3	Chr03:50110557..50110620 (+)	T/C	2,05	0,078	0,022
s3_51868687	3	Chr03:53030484..53030547 (+)	T/A	2,688	0,245	0,032
s4_41984665	4	Chr04:44087494..44087557 (+)	A/T	2,02	0,251	0,022
s6_31660835	6	Chr06:30910885..30910948 (+)	A/G	2,268	0,059	0,025
s7_1263354	7	Chr07:1281414..1281477 (+)	C/G	2,708	0,068	0,031
s7_3509864	7	Chr07:3568737..3568800 (+)	G/A	2,002	0,164	0,021
s7_4257560	7	Chr07:4278109..4278172 (-)	G/C	2,254	0,331	0,025
s7_4524725	7	no sale	A/T	2,402	0,315	0,027
s7_4755652	7	Chr07:4823959..4824022 (+)	T/C	2,228	0,318	0,024
s7_5242200	7	Chr07:5317110..5317173 (+)	T/C	2,279	0,317	0,025
s7_5368821	7	Chr07:5442363..5442426 (-)	T/G	2,444	0,304	0,027
s7_5442911	7	Chr07:5516442..5516505 (-)	G/T	2,363	0,293	0,026
s7_5525199	7	Chr07:5598835..5598898 (-)	T/G	2,359	0,298	0,026
s7_50879045	7	Chr07:39166060..39166123 (-)	T/C	2,516	0,464	0,029
s7_51283516	7	Chr07:39574655..39574718 (+)	C/T	4,151	0,255	0,051
s8_5497128	8	Chr08:5492175..5492238 (+)	C/A	2,04	0,331	0,022
s8_6289935	8	Chr08:6249447..6249510 (+)	T/C	2,276	0,309	0,025
s8_8250759	8	Chr08:8337675..8337738 (+)	T/C	2,002	0,287	0,022
s9_37301796	9	Chr09:38072208..38072271 (-)	G/T	3,211	0,155	0,038
s10_22705425	10	Chr10:15997304..15997367 (-)	C/T	2,193	0,062	0,024
s10_38333216	10	Chr10:39430517..39430580 (-)	G/A	3,358	0,343	0,042
s10_38396717	10	Chr10:39526879..39526942 (-)	T/A	4,188	0,329	0,057
s10_38783810	10	Chr10:39915924..39915987 (-)	A/G	2,701	0,307	0,032
s10_38913819	10	Chr10:40046100..40046163 (+)	A/G	3,382	0,311	0,04
s10_39180865	10	Chr10:40280166..40280229 (+)	T/C	3,345	0,288	0,041
s10_39185557	10	Chr10:40284905..40284968 (+)	C/A	2,051	0,088	0,022
s10_39232286	10	Chr10:40331767..40331830 (+)	A/T	2,398	0,261	0,027

s10_39978402	10	Chr10:41088440..41088503 (-)	A/T	2,192	0,324	0,024
s10_40329757	10	Chr10:41487036..41487099 (-)	T/C	2,088	0,199	0,024
s10_40536728	10	Chr10:41561615..41561678 (+)	G/A	2,793	0,153	0,033
s10_40979207	10	Chr10:42000131..42000193 (+)	T/C	3,238	0,15	0,038
s10_41090517	10	Chr10:42111336..42111399 (+)	A/G	2,478	0,151	0,028
s11_1526553	11	Chr11:1518854..1518917 (-)	G/A	2,057	0,225	0,023

Tabla 5. SNP significativamente asociados con la relación longitud/anchura ($p < 0,01$) en el análisis MLM. Se indican: coordenadas genómicas (Cr, cromosoma y posición física en pb); p-valor de la asociación representado como $-\log_{10}(p)$; frecuencia alélica mínima (MAF); y porcentaje de variabilidad explicada por cada SNP (R^2).

SNP	Chr	Posición V2	Pol	$-\log_{10}(p)$	MAF	R^2
s1_45417533	1	Chr01:44712245..44712314 (-)	C/T	2,223	0,238	0,025
s1_45457600	1	Chr01:44752177..44752240 (-)	G/A	2,873	0,253	0,033
s1_45594842	1	Chr01:44888486..44888549 (-)	T/C	2,328	0,148	0,027
s2_709003	2	Chr02:667806..667869 (+)	A/G	2,382	0,36	0,027
s2_712930	2	Chr02:671678..671741 (+)	A/T	2,168	0,061	0,024
s2_1627195	2	Chr02:1584068..1584131 (+)	C/T	2,513	0,133	0,029
s2_40711052	2	Chr02:41238182..41238245 (-)	A/G	2,46	0,369	0,028
s2_45346661	2	Chr02:45904466..45904529 (-)	C/T	2,417	0,395	0,026
s2_45422622	2	Chr02:45977526..45977589 (-)	C/T	2,466	0,381	0,027
s2_45423669	2	Chr02:45978594..45978657 (+)	C/G	2,541	0,392	0,029
s2_47173033	2	Chr02:47746388..47746451 (+)	G/A	5,128	0,328	0,065
s2_47385469	2	Chr02:47984479..47984542 (+)	C/T	2,087	0,48	0,023
s3_4320940	3	Chr03:4466099..4466162 (-)	T/C	2,291	0,235	0,026
s4_42314191	4	Chr04:44419464..44419527 (+)	G/T	3,589	0,408	0,042
s4_43792143	4	Chr04:45971741..45971804 (+)	C/T	4,122	0,261	0,053
s5_17217569	5	Chr05:17759050..17759113 (+)	G/C	2,876	0,269	0,035
s5_17539222	5	Chr05:15588232..15588295 (+)	G/A	2,434	0,403	0,029
s5_19893620	5	Chr05:17759050..17759113 (+)	T/C	2,311	0,284	0,027
s5_20044966	5	Chr05:21626650..21626713 (+)	G/C	2,489	0,274	0,029
s7_4204587	7	Chr07:4270231..4270294 (+)	T/A	2,036	0,386	0,022
s7_4212507	7	Chr07:4278109..4278172 (-)	C/T	2,058	0,376	0,022
s7_40034923	7	Chr07:28402571..28402634 (+)	T/G	2,276	0,182	0,024
s7_40055678	7	Chr07:28423370..28423433 (+)	C/G	2,253	0,194	0,025
s7_41527486	7	Chr07:29892297..29892360 (-)	G/C	2,035	0,342	0,023
s7_43451986	7	Chr07:31812319..31812382 (-)	A/C	2,219	0,293	0,025
s7_43609346	7	Chr07:31959748..31959811 (+)	T/C	2,213	0,285	0,025
s7_43812838	7	Chr07:32158916..32158979 (-)	A/C	2,59	0,296	0,03
s7_43886926	7	Chr07:32233387..32233450 (-)	C/T	2,921	0,264	0,034
s7_44025276	7	Chr07:32371858..32371921 (+)	A/G	2,558	0,301	0,031
s7_44162685	7	Chr07:32509487..32509550 (-)	T/G	2,959	0,287	0,035
s7_44259174	7	Chr07:32611999..32612062 (+)	T/A	2,917	0,261	0,035
s7_44265146	7	Chr07:32617965..32618028 (-)	T/C	3,43	0,3	0,041
s7_44305649	7	Chr07:32658445..32658508 (-)	T/C	2,801	0,259	0,034
s7_48409250	7	Chr07:36785264..36785327 (+)	C/T	2,235	0,112	0,025
s7_48741309	7	Chr07:37085151..37085214 (+)	C/G	3,133	0,203	0,037
s8_55946412	8	Chr08:59380789..59380852 (-)	C/T	3,927	0,193	0,047
s9_12670503	9	Chr09:13065661..13065724 (-)	T/C	2,69	0,159	0,032
s9_32337929	9	no mach	T/C	2,069	0,149	0,022
s9_34720152	9	Chr09:20040175..20040238 (+)	A/T	2,511	0,261	0,029
s9_35615822	9	Chr09:20040175..20040238 (+)	G/T	2,038	0,125	0,022
s9_36466985	9	Chr09:37216961..37217024 (+)	T/A	3,169	0,176	0,038

s11_1620036	11	Chr11:1616739..1616802 (+)	T/C	2,333	0,395	0,026
s11_7045229	11	Chr11:1882828..1882891 (+)	G/T	2,697	0,061	0,031

Tabla 6. SNP significativamente asociados con **la circularidad** ($p < 0,01$) en el análisis MLM. Se indican: coordenadas genómicas (Cr, cromosoma y posición física en pb); p-valor de la asociación representado como $-\log_{10}(p)$; frecuencia alélica mínima (MAF); y porcentaje de variabilidad explicada por cada SNP (R^2).

SNP	Chr	Pos V2	Pol	$-\log_{10}(p)$	MAF	R^2
s1_50223721	1	Chr01:49512407..49512470 (-)	G/A	2,496	0,054	0,029
s2_29058024	2	Chr02:30323163..30323226 (+)	G/C	2,504	0,079	0,031
s2_39810109	2	Chr02:40675810..40675873 (-)	C/G	2,008	0,362	0,022
s2_40123221	2	Chr02:40675810..40675873 (-)	T/G	2,427	0,44	0,03
s2_40191996	2	Chr02:40745549..40745612 (-)	T/G	2,403	0,433	0,028
s2_45346661	2	Chr02:45904466..45904529 (-)	C/T	2,171	0,395	0,025
s2_45422622	2	Chr02:45977526..45977589 (-)	C/T	2,13	0,381	0,024
s2_45423669	2	Chr02:45978594..45978657 (+)	C/G	2,125	0,392	0,025
s2_47173033	2	Chr02:47746388..47746451 (+)	G/A	2,565	0,328	0,03
s3_1177029	3	Chr03:1302446..1302509 (+)	A/C	2,175	0,051	0,025
s4_41984665	4	Chr04:44087494..44087557 (+)	A/T	2,229	0,251	0,026
s4_43792143	4	Chr04:45971741..45971804 (+)	C/T	3,492	0,262	0,046
s5_4876693	5	Chr05:4913417..4913480 (+)	C/T	2,218	0,03	0,026
s5_19893620	5	Chr05:17759050..17759113 (+)	T/C	2,32	0,284	0,028
s5_20044966	5	Chr05:21626650..21626713 (+)	G/C	2,879	0,274	0,038
s7_4036378	7	Chr07:4101390..4101453 (+)	T/C	2,498	0,331	0,03
s7_4117127	7	Chr07:4182128..4182191 (+)	T/A	2,194	0,333	0,025
s7_4204587	7	Chr07:4270231..4270294 (+)	T/A	2,83	0,386	0,035
s7_4212507	7	Chr07:4278109..4278172 (-)	C/T	2,796	0,376	0,034
s7_4257560	7	Chr07:4278109..4278172 (-)	G/C	2,089	0,331	0,024
s7_4524725	7	no mach	A/T	2,079	0,315	0,023
s7_4755652	7	Chr07:4823959..4824022 (+)	T/C	2,06	0,318	0,024
s7_5242200	7	Chr07:5317110..5317173 (+)	T/C	2,128	0,317	0,026
s7_48019091	7	Chr07:36385020..36385083 (-)	A/G	2,783	0,061	0,033
s7_48530923	7	Chr07:36905941..36906004 (-)	C/A	2,533	0,272	0,031
s7_48653874	7	Chr07:36997082..36997145 (-)	C/T	2,416	0,279	0,029
s7_48659893	7	Chr07:37003102..37003165 (+)	C/A	2,449	0,266	0,029
s8_55946412	8	Chr08:59380789..59380852 (-)	C/T	2,388	0,193	0,028
s9_13245138	9	Chr09:13688966..13689029 (+)	T/C	2,167	0,28	0,025
s10_40632854	10	Chr10:41656717..41656780 (-)	T/C	2,316	0,159	0,027

Tabla 7. SNP significativamente asociados con el **peso de 25 semillas** ($p < 0,01$) en el análisis MLM. Se indican: coordenadas genómicas (Cr, cromosoma y posición física en pb); p-valor de la asociación representado como $-\log_{10}(p)$; frecuencia alélica mínima (MAF); y porcentaje de variabilidad explicada por cada SNP (R^2).

SNP	Chr	Posición tag	Pol	$-\log_{10}(p)$	MAF	R^2
s1_50822498	1	Chr01:50105511..50105574 (-)	G/T	2,015	0,219	0,023
s1_50842559	1	Chr01:50125657..50125720 (+)	C/T	2,971	0,197	0,038
s1_51779298	1	Chr01:51062799..51062862 (-)	A/T	2,801	0,468	0,034
s1_51790864	1	Chr01:51074394..51074457 (+)	T/A	2,2	0,196	0,025
s1_51995080	1	Chr01:51281708..51281771 (+)	A/G	2,572	0,415	0,032
s1_52137885	1	Chr01:51430006..51430069 (-)	G/A	3,98	0,168	0,053
s2_34110322	2	Chr02:34838627..34838690 (-)	G/C	2,084	0,148	0,024
s2_35091874	2	Chr02:35833226..35833289 (-)	C/G	2,172	0,245	0,025
s3_45966243	3	Chr03:47089727..47089919 (-)	A/C	2,146	0,298	0,025
s3_46066595	3	Chr03:47193437..47193500 (+)	G/A	2,155	0,29	0,025
s3_47838531	3	Chr03:48972415..48972478 (-)	T/C	2,322	0,276	0,027
s3_47879520	3	Chr03:49013365..49013428 (-)	T/C	2,096	0,277	0,024
s3_47996483	3	Chr03:49131292..49131355 (+)	G/C	2,383	0,155	0,028
s4_27768193	4	Chr04:16188553..16188616 (-)	G/T	2,183	0,292	0,026
s4_29757350	4	Chr04:34137292..34137355 (-)	A/G	4,212	0,19	0,055
s7_627376	7	Chr07:644609..644672 (-)	G/A	2,071	0,395	0,026
s7_633265	7	Chr07:650516..650579 (+)	G/A	2,625	0,39	0,032
s7_663226	7	Chr07:680494..680557 (+)	G/C	2,26	0,278	0,027
s7_3218398	7	no mach	A/G	2,6	0,155	0,031
s7_51283516	7	Chr07:39574655..39574718 (+)	C/T	2,876	0,255	0,035
s8_55946412	8	Chr08:59380789..59380852 (-)	C/T	5,914	0,193	0,082
s9_17589973	9	Chr09:18115884..18115947 (+)	T/C	2,01	0,14	0,023
s9_19503651	9	Chr09:20040175..20040238 (+)	G/A	2,089	0,159	0,024
s10_37444122	10	no mach	A/C	2,305	0,444	0,03
s10_37576060	10	Chr10:38678940..38679003 (+)	T/A	2,076	0,439	0,025
s10_38333216	10	Chr10:39430517..39430580 (-)	G/A	2,123	0,343	0,026
s10_38396717	10	Chr10:39526879..39526942 (-)	T/A	2,315	0,329	0,027
s10_39185557	10	Chr10:40284905..40284968 (+)	C/A	2,36	0,088	0,028
s10_39193928	10	Chr10:40293254..40293316 (+)	C/A	2,79	0,086	0,035
s10_40536728	10	Chr10:41561615..41561678 (+)	G/A	2,206	0,153	0,025
s10_40979207	10	Chr10:42000131..42000193 (+)	T/C	2,91	0,15	0,035
s10_41090517	10	Chr10:42111336..42111399 (+)	A/G	2,646	0,151	0,033
s11_2532897	11	Chr11:2714686..2714747 (+)	A/G	2,055	0,031	0,023

