

4. Materiales y Métodos

4.1. Cultivo de *Arabidopsis thaliana*

4.1.1. Origen del mutante albino

El mutante albino empleado en este trabajo se aisló en un experimento de mutagénesis con metanosulfonato de etilo (EMS), que fue realizado en el año 2008 por Héctor Candela, durante su estancia postdoctoral en el Plant Gene Expression Center. El EMS es un agente alquilante responsable de mutaciones puntuales, en concreto transiciones del tipo G/C→A/T (Greene *et al.*, 2003). La estirpe mutagenizada fue el mutante recesivo *bp-9* (*brevipedicellus-9*; Venglat *et al.*, 2002), cuyo fondo genético es parcialmente el de la estirpe Landsberg *erecta* (*Ler*). Por su letalidad, el mutante albino caracterizado en este trabajo se identificó en una familia M₃, obtenida a partir de la autofecundación de plantas M₂ que no manifestaban el fenotipo.

4.1.2. Selección de mutantes en las familias F₂

La población cartográfica se obtuvo cruzando plantas de fenotipo silvestre (heterocigóticas para la mutación) con la estirpe silvestre Columbia-0 (Col-0). La progenie F₁ de estos cruzamientos, compuesta en un 50% por plantas heterocigóticas para la mutación, se autofecundó para identificar familias F₂ que segregasen plántulas albinas. Se seleccionaron 87 plantas albinas y 170 plantas de fenotipo silvestre, con las que se establecieron sendos grupos de plantas (*pools*) de los que se aisló colectivamente el ADN genómico (ADNg) para su secuenciación posterior. Las plantas se congelaron a -80°C hasta el momento de purificar su ADN genómico.

4.1.3. Medio de cultivo sólido para *Arabidopsis thaliana*

Las plantas de *Arabidopsis thaliana* se cultivaron en medio de Murashige y Skoog (MS, Murashige y Skoog, 1962), cuya composición se describe a continuación: NH₄NO₃ 20,6 mM; H₃BO₃ 0,1 mM; CaCl₂ 3 mM; CoCl₂·6H₂O 0,1 μM; CuSO₄·5H₂O 0,1 μM; Na₂EDTA 0,11 mM; MgSO₄ 1,5 mM; MnSO₄·H₂O 0,1 mM; NaMoO₄·2H₂O 1 μM; KI 5 μM; KNO₃ 18,8 mM; KH₂PO₄ 1,24 mM; y ZnSO₄·7H₂O 30 μM.

Para preparar 1 L de medio MS, se añaden 900 mL de agua destilada a un matraz de 2 L. Posteriormente, se disuelven 4,3 g/L de sales de Murashige y Skoog (Duchefa), 10 g de sacarosa (concentración final de 1% m/v) y 0,5 g/L de MES (ácido 2-[N-morfolino]-etano sulfónico) con ayuda de un agitador magnético. El pH de la disolución se ajusta a 5,7 con KOH 5 M. Finalmente, tras enrasar el volumen a 1 L, se añaden 6,5 g/L de Plant Agar (Duchefa) y se esteriliza en autoclave a 121°C durante 20 min. Tras la esterilización del medio de cultivo, alicuotamos el medio en cajas de Petri de 15 cm de diámetro, una vez

vermiculita y turba en proporción 2:2:1 (v/v/v) que había sido esterilizada previamente en autoclave. Las macetas se subirrigaron con una disolución nutritiva comercial (abono Universal líquido Carrefour), manteniendo un nivel de 2 ó 3 cm de la misma en las bandejas. La disolución contiene: nitrógeno (N) total: 6% (incluye N amoniacal: 2,8%; N ureico: 2,2%; y N nítrico: 1,0%); pentóxido de fósforo (P_2O_5): 6%; óxido de potasio (K_2O): 6%; micronutrientes: cobre (Cu): 0,002%; hierro (Fe): 0,025%; manganeso (Mn): 0,014%; molibdeno (Mo): 0,001%; y zinc (Zn) quelado con EDTA: 0,004%. Las plantas se irrigaron semanalmente con agua y con una dilución al 0,5% (v/v) de la disolución nutritiva descrita cada dos semanas.

Las plantas cultivadas en cajas de Petri se transfirieron a macetas, con ayuda de pinzas metálicas esterilizadas con alcohol, a los 26 días tras la siembra. Los cultivos en maceta se mantuvieron a $23\pm 1^\circ C$ en una cámara climática bajo iluminación continua de 5.000 lx suministrada por tubos fluorescentes F28 T5/D y Leuci T5-28W. El riego se interrumpió cuando cesó la producción de flores, con el fin de que se secan los frutos y recolectar las semillas.

4.1.6. Recolección de semillas

El secado de las plantas tuvo lugar en la cámara climática. Una vez que las plantas completaron su ciclo de vida y se secaron, se recolectaron las semillas. Los restos vegetales se eliminaron con ayuda de un colador y las semillas se almacenaron en tubos *ependorf* a $4^\circ C$.

4.2. Manipulación de ácidos nucleicos

4.2.1. Purificación de ADN genómico para ultrasecuenciación

Realizamos dos purificaciones de ADN_g utilizando un kit para plantas (*GeneJET Plant Genomic DNA Purification Mini Kit*) siguiendo el protocolo del fabricante (Thermo Scientific), que se resume a continuación. El tejido (120 mg del mutante albino y 220 mg de las plantas de fenotipo silvestre) se trituró en morteros de porcelana estériles distintos y en presencia de nitrógeno líquido. El material vegetal se transfirió a tubos *ependorf*, a los que añadimos 350 μL de tampón de lisis A. Tras agitar durante 10-15 s con un vórtex, añadimos 50 μL de tampón de lisis B y 20 μL de RNasa A. Los tubos se incubaron, con agitación periódica, en un baño a $65^\circ C$ durante 10 min. A continuación, añadimos 130 μL de disolución de precipitación a cada tubo, que agitamos por inversión e incubamos durante 5 min en hielo. A continuación, centrifugamos ambos tubos durante 5 min a 13.300 rpm (16.300g). Transferimos los sobrenadantes a tubos limpios, a los que añadimos 400 μL de disolución de unión (*binding solution*) y 400 μL de etanol al 96%. Transferimos parte del volumen (600-700 μL) a las columnas suministradas con el kit y centrifugamos los tubos a 8000 rpm (5.900g) durante 1 min. La operación se repitió con el volumen restante,

descartando cada vez el sobrenadante tras su paso por la columna. Las columnas se lavaron dos veces con 500 μ L de tampón de lavado 1. El tampón de lavado se eliminó centrifugando durante 1 min a 10.000 rpm (9.200g). Por último, eluimos el ADN_g retenido en la columna a tubos *ependorf* estériles añadiendo 100 μ L de tampón de elución al centro de la columna, incubando a temperatura ambiente durante 5 min y centrifugando a 10.000 rpm durante 5 min.

4.2.2. Purificación de ADN_g para genotipado de plantas individuales

Los mutantes albinos se recolectaron de las cajas de Petri a los 26 días tras la siembra y se congelaron a -80°C en tubos *ependorf*. También se congelaron plantas control de fenotipo silvestre. Se purificó el ADN_g de 35 plantas albinas y 3 plantas silvestres de la F₂ con el propósito de estudiar la cosegregación del albinismo con la mutación candidata. Además, purificamos ADN_g a partir de 20 hojas de otras tantas plantas F₂ fenotípicamente silvestres, con el propósito de identificar plantas portadoras de la mutación candidata con las que purificar ARN total. El protocolo seguido para esta purificación de ADN_g se resume a continuación. Añadimos 250 μ L de tampón de extracción (Tris 100 mM pH 8; EDTA 50 mM a pH 8; NaCl 500 mM) a cada tubo *ependorf* y trituramos el tejido con palillos de plástico hasta conseguir una disolución homogénea. Añadimos otros 250 μ L de tampón de extracción, hasta alcanzar un volumen final de 500 μ L, y agitamos los tubos por inversión. Añadimos 35 μ L de SDS 20% (m/v) a cada tubo e incubamos durante 5 min a 65°C en un baño, agitando periódicamente los tubos por inversión. Transcurrido este tiempo, añadimos 130 μ L de acetato de potasio 5 M a cada tubo, e incubamos los tubos en hielo durante 8 min. Posteriormente, centrifugamos los tubos a 13.000 rpm (15.600g) durante 10 min. Transferimos el sobrenadante a tubos nuevos, a los que añadimos 640 μ L de isopropanol y 60 μ L de acetato de sodio 3 M. Tras agitar los tubos, se incubaron en hielo durante 20 min y se centrifugaron a 13000 rpm durante 10 min. El precipitado resultante se lavó con 300 μ L de etanol 70%. Este etanol se eliminó tras centrifugar una última vez a 13.000 rpm durante 5 min. Una vez seco, el precipitado se resuspendió en un volumen de 25 a 50 μ L de agua destilada estéril.

4.2.3. Amplificaciones mediante PCR

En este trabajo hemos utilizado los cebadores de la Tabla 1, que fueron sintetizados por un proveedor comercial (Sigma-Aldrich). Las reacciones de PCR se prepararon en tubos *ependorf* de 0,2 mL de pared fina, en un volumen final de 20 μ L. La composición de las reacciones fue la siguiente: 6,3 μ L de agua desionizada, 2 μ L de *DreamTaq Buffer* 10x, 1,6 μ L de una mezcla de los 4 dNTP a una concentración de 10 mM (con cada uno de ellos a 2,5 mM), 4 μ L de los cebadores *forward* y *reverse* correspondientes a 2,5 μ M cada uno, 0,1

μL de la enzima *DreamTaq* a $5 \text{ U}/\mu\text{L}$ y $2 \mu\text{L}$ de molde apropiado. Durante la preparación de cada reacción, todos los reactivos se mantuvieron en hielo hasta el momento de su uso.

Tabla 1.- Nombre y secuencia de los cebadores empleados

Cebador	Secuencia nucleotídica (5'→3')
At2g04030-F	CGTCGTGGAACACAAATCAC
At2g04030-R	AGGCTTACCCATAGCGGTTT
At2g04030-cds-F	ggggacaagttgtacaaaaaagcaggctTAATGGCTCCTGCTTTGAGTA
At2g04030-cds-R1	ggggaccactttgtacaagaaagctgggtCAATCTTGCCAAGGATCACTC

Se indica en letras minúsculas la secuencia de los sitios *attB1* y *attB2*, que permitirán la clonación del producto de PCR mediante la tecnología Gateway.

Las reacciones de PCR se incubaron en un termociclador *T100 Thermal Cycler* (Bio-Rad). La temperatura y la duración de cada una de las etapas del programa de amplificación fueron las que se muestran a continuación. La etapa inicial se programó a 95°C durante 2 min. Tras esta, se realizaron 35 ciclos en las siguientes condiciones: (1) una etapa de desnaturalización a 95°C 30 s; (2) una etapa de hibridación a una temperatura comprendida entre 55°C y 64°C , dependiendo de la temperatura de fusión (T_m) de los cebadores presentes en la mezcla de reacción, durante 30 s; y (3) una etapa de síntesis a 72°C durante 45 s, tiempo que se determinó teniendo en cuenta que la enzima *DreamTaq* sintetiza ADN a razón de 1 kb cada 30 s. Finalmente, se añadió una etapa final de extensión a 72°C durante 5 min.

4.2.4. Electroforesis en gel de agarosa

Las muestras de ADN se visualizaron en geles de agarosa al 1% (m/v) en tampón de electroforesis TAE 1x (0,5 g de agarosa en 50 mL de TAE 1x), que se preparó a partir de una disolución madre a concentración 50x (Tris-HCl 2 M; ácido acético 5,71%; EDTA 50 mM; pH 8,0). La tinción de los geles se realizó con $1,25 \mu\text{L}$ de SafeView, que se añadieron a la disolución de agarosa antes de que se solidificase. Además, en la primera y última calle de cada gel se cargaron $4 \mu\text{L}$ de marcador de peso molecular (1 kb DNA Ladder de GeneCraft a una concentración de $0,1 \mu\text{g}/\mu\text{L}$). Las muestras de ADN purificado se prepararon con $1 \mu\text{L}$ de ADN, $8 \mu\text{L}$ de agua desionizada y $1 \mu\text{L}$ de tampón de carga 10x (DNA Loading Buffer 10x de 5prime). Por último, la electroforesis se realizó con una fuente de alimentación a 90 V (LabNet) durante 30 min.

4.2.5. Purificación de productos de PCR

Para llevar a cabo este paso se utilizó el *GeneJET PCR Purification Kit*, siguiendo las instrucciones del fabricante (Thermo Scientific). Los productos purificados se visualizaron en geles de agarosa al 1% en tampón de electroforesis TAE 1x.

4.3. Análisis bioinformático

Para la secuenciación del genoma, recurrimos a los servicios de la empresa STAB VIDA (Caparica, Portugal), que dispone de un secuenciador masivamente paralelo Illumina HiSeq 2500. Los fragmentos se secuenciaron por ambos extremos con lecturas de 101 nucleótidos (*paired-end reads*). Para evaluar la calidad de las lecturas obtenidas y determinar la presencia en las mismas de secuencias derivadas de los adaptadores añadidos durante la preparación de las muestras, utilizamos el programa FastQC (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>). Hemos utilizado en programa PEAR (Paired-End reAd mergeR; <http://sco.h-its.org/exelixis/web/software/pear/>) para detectar solapamientos entre las dos lecturas derivadas de cada fragmento secuenciado. En caso de solapamiento, PEAR ensambla ambas lecturas, produciendo secuencias de mayor longitud que las lecturas originales (Zhang *et al.*, 2014). Para eliminar los adaptadores detectados, utilizamos el programa Trimmomatic (versión 0.32; Bolger *et al.*, 2014), al que se le proporcionó un archivo con la secuencia de los adaptadores mediante la opción ILLUMINACLIP. También se utilizaron las opciones LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 y MINLEN:36.

El alineamiento de las lecturas obtenidas al genoma de referencia se llevó a cabo con el programa Bowtie2 (versión 2.1.0), que emplea un algoritmo de alineamiento muy eficiente basado en la transformación de Burrows-Wheeler (Langmead *et al.*, 2009; Langmead y Salzberg, 2012). Como genoma de referencia, a la que alineamos las lecturas, utilizamos la versión más reciente del genoma de *Arabidopsis thaliana* (versión TAIR10), cuya secuencia descargamos en formato FastA (Lamesch *et al.*, 2012). El índice del genoma de referencia se preparó con el programa bowtie2-build, especificando su secuencia con la opción -f. El alineamiento de las lecturas emparejadas se llevó a cabo mediante el programa bowtie2, con las siguientes opciones: -x (especifica el índice del genoma al que deben alinearse las lecturas, preparado anteriormente con bowtie2-build), --qc-filter (descarta las lecturas de baja calidad), -p (especifica el número de procesadores disponibles; en nuestro caso es igual a 8), -X (especifica el tamaño de máximo de los fragmentos que, en nuestro caso, fue fijado arbitrariamente en 650 nucleótidos), --no-discordant (descarta las parejas de lecturas cuyos alineamientos no son concordantes), --no-mixed (descarta las lecturas cuya pareja no ha sido eliminada) y --no-unal (descarta aquellas lecturas para las que no se detectan alineamientos).

Los archivos SAM resultantes, que contienen la descripción de los alineamientos de las lecturas, fueron convertidos a formato BAM mediante el comando `view` de `samtools` (Li *et al.*, 2009). Los archivos BAM correspondientes se ordenaron según la posición de las lecturas alineadas en el genoma de referencia mediante el comando `sort` de `samtools`. Los archivos así ordenados y correspondientes a la misma muestra, se combinaron en un único archivo mediante el comando `merge` de `samtools`. El archivo resultante se indizó con el comando `index` de `samtools`. Utilizamos el comando `AddOrReplaceReadGroups` del programa `piccard-tools` (versión 2.5.0) para asignar las lecturas a grupos. El archivo BAM resultante fue procesado con el programa GATK (Genome Analysis Toolkit; versión 3.6; McKenna *et al.*, 2010), utilizando las opciones `-T RealignerTargetCreator` y `-T IndelRealigner`. Estas dos opciones permiten seleccionar una lista de sitios del genoma adyacentes a inserciones y deleciones que posiblemente requieren que los alineamientos de las lecturas sean corregidos, y llevar a cabo las correcciones necesarias. Utilizamos, por último, el programa GATK con la opción `-T UnifiedGenotyper`, para identificar las bases que ocupan cada una de las posiciones del genoma (operación denominada, en inglés, *base calling*). El resultado se guardó en un archivo en formato VCF (Danecek *et al.*, 2011), que fue procesado manualmente mediante pequeñas rutinas escritas en lenguajes Perl y AWK, ejecutadas en la línea de comandos de Linux, y Microsoft Excel.

Se representaron para cada cromosoma todas las posiciones y frecuencias relativas al mutante, la media móvil de los datos anteriores, la media móvil de las posiciones y frecuencias del silvestre y el valor conocido como *boost*, que permite definir una región en la que se encuentre con una gran probabilidad la mutación a estudio. La fórmula para calcular el valor de *boost* es: $B_v = 1/|1 - \max(\theta_{tar}, 1 - \theta_{tar}) / \max(\theta_{obs}, 1 - \theta_{obs})|$ (Sun y Schneeberger, 2015).

Algunos apartados de la sección de Materiales y Métodos se han modificado a partir de trabajos anteriores realizados en el laboratorio de Héctor Candela (como el de E.M. Rodríguez Alcocer, 2014, citado en la Bibliografía).

5. Resultados

5.1. Fenotipo y modo de herencia de un mutante albino

Para determinar el modo de herencia del mutante a estudio, cruzamos plantas fenotípicamente silvestres portadoras de la mutación albina en heterocigosis con plantas de la estirpe silvestre Columbia-0 (Col-0). No encontramos plantas albinas en la progenie F_1 de estos cruzamientos, como cabe esperar si la mutación se hereda como un carácter recesivo. Con el fin de obtener la población F_2 necesaria para la realización de un experimento de cartografía mediante secuenciación, permitimos que las plantas de la progenie F_1 se autofecundaran. Las poblaciones F_2 resultantes fueron sembradas en cajas de Petri que contenían medio MS suplementado con sacarosa (véase el apartado 4.1.3 de la página 14). El fenotipo albino reapareció en 15 de los 85 individuos de la Familia 2, un resultado que es compatible con la razón fenotípica 3:1 (silvestre:mutante), característica del modo de herencia monogénico recesivo (Tabla 2). Sin embargo, el fenotipo mutante apareció en un número de individuos significativamente mayor que el esperado en la Familia 1, quizá por la presencia de mutaciones adicionales que distorsionan la segregación fenotípica en dicha familia. En conjunto, el fenotipo albino reapareció en 87 de los 257 individuos estudiados, aproximadamente un tercio de las plantas de la generación F_2 .

Tabla 2.- Segregación del fenotipo mutante albino en la progenie F_2 de cruzamientos entre plantas portadoras de la mutación a estudio y la estirpe silvestre Col-0

Familia F_2	Clases fenotípicas ^a		Hipótesis	Segregación	
	Silvestre	Albino		X^2	P
1	100	72	3:1	25,19 ^b	5,2×10 ⁻⁷
2	70	15	3:1	2,07 ^b	0,15

^aNúmero de individuos asignados a cada clase fenotípica. ^bLos valores de X^2 han sido calculados aplicando la corrección de Yates por haber sólo 2 clases fenotípicas. Al nivel de significación $\alpha=0,05$, el valor crítico del estadístico de la prueba X^2 es 3,84 (con 1 grado de libertad).

Por otro lado, sembramos la población F_2 en medio MS carente de sacarosa con el fin de estudiar posibles diferencias en el desarrollo de las plantas mutantes. En estas condiciones, las plántulas presentaron un menor desarrollo, particularmente por el reducido tamaño de sus cotiledones, lo que dificultó aún más la tarea de localizarlas en las cajas de Petri. En la Figura 1 se muestran los fenotipos de la estirpe silvestre y del mutante albino en presencia de sacarosa.



Figura 1.- Fenotipo de un mutante albino de *Arabidopsis thaliana* cultivado en medio suplementado con sacarosa. (A) Mutante albino. (B) Estirpe silvestre Landsberg *erecta*. Las barras de escala representan 5 mm. En ambos casos, las fotografías se tomaron a los 26 días tras la siembra.

5.2. Obtención de ADN para su secuenciación

Las 87 plantas fenotípicamente mutantes identificadas en la población cartográfica F₂ (72 de la Familia 1 y 15 de la Familia 2) se recolectaron en tubos *ependorf*, que se mantuvieron a -80°C hasta la purificación del ADN genómico (ADNg). El ADNg se purificó tras combinar los tejidos de las plantas de ambas familias, siguiendo el protocolo descrito en el apartado 4.2.1 de la página 16. Procedimos de la misma manera con las plantas fenotípicamente silvestres. Las dos muestras (de plantas mutantes y silvestres, respectivamente) se enviaron congeladas en hielo seco a la empresa STAB VIDA (Caparica, Portugal) para su secuenciación con un secuenciador Illumina HiSeq 2500.

La calidad y cantidad del ADN de las muestras fue evaluada por medio de una electroforesis en gel de agarosa (Figura 2) y mediante un fluorímetro Qubit. La concentración de ADN fue de 15,4 ng/μL en la muestra obtenida a partir de las plántulas albinas y 13,2 ng/μL en la muestra obtenida a partir de plántulas de fenotipo silvestre. La Figura 2 muestra una única banda, correspondiente a ADN de alto peso molecular, en las calles correspondientes a ambas muestras.

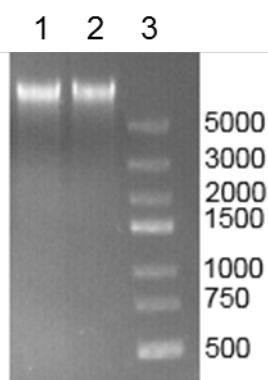


Figura 2.- Electroforesis en gel de agarosa de las muestras de ADN genómico. Aspecto de las muestras enviadas tras someterlas a electroforesis en gel de agarosa al 1% con un voltaje de 120 V durante 20 min. Las calles del gel fueron cargadas con: (1) 1 μ L de la muestra de ADN genómico obtenida a partir de las plántulas albinas, (2) 1 μ L de la muestra de ADN genómico obtenida a partir de las plántulas de fenotipo silvestre, y (3) marcador de peso molecular. Se indica el tamaño, en pares de bases, de las bandas del marcador de peso molecular.

5.3. Procesamiento de las secuencias obtenidas

El número de fragmentos secuenciados para las distintas muestras se presenta en la Tabla 3, que también resume los resultados del procesamiento de las lecturas (*reads*) con los programas PEAR y Trimmomatic. La muestra de ADN obtenida a partir de los mutantes albinos (muestra G-2) se secuenció en dos tandas (A y B). Disponemos de la secuencia nucleotídica de 80.980.764 lecturas de la muestra G-2, resultado de secuenciar cada uno de los 40.490.382 fragmentos de ADN (20.507.732 de la tanda A y 19.982.650 de la tanda B) por ambos extremos (*paired-end reads*; lecturas emparejadas). La muestra de ADN obtenida a partir de los individuos silvestres (muestra G-3) también se secuenció en dos tandas (A y B). Para la muestra G-3, disponemos de la secuencia nucleotídica de 111.204.158 lecturas, resultado de secuenciar un total de 55.602.079 fragmentos de ADN (28.604.938 de la tanda A y 26.997.141 de la tanda B) por ambos extremos. La gran mayoría de las lecturas obtenidas posee una longitud de 101 nucleótidos.

Tabla 3.- Número de fragmentos secuenciados y resultado del procesamiento de las secuencias obtenidas con PEAR

Categoría	Muestra			
	G-2 (A)	G-2 (B)	G-3 (A)	G-3 (B)
Fragmentos	20.507.732	19.982.650	28.604.938	26.997.141
Fragmentos con lecturas solapantes	7.967.500 (38,851%)	7.343.716 (36,750%)	14.161.049 (49,506%)	12.689.511 (47,003%)
Fragmentos con lecturas no solapantes	12.539.763 (61,140%)	12.637.993 (63,245%)	14.427.004 (50,435%)	14.290.941 (52,935%)
Fragmentos descartados	469 (0,002%)	941 (0,005%)	16.885 (0,059%)	16.689 (0,062%)

Hemos utilizado el programa PEAR para combinar la secuencia de las lecturas *forward* y *reverse* derivadas de un mismo fragmento, una operación que resulta posible si el tamaño del fragmento secuenciado es inferior a la suma de las longitudes de las lecturas individuales. Cuando PEAR detecta el solapamiento de dos lecturas, combina sus secuencias y las reemplaza por una única lectura de mayor longitud. Tras el procesamiento de la muestra G-2 con PEAR, se fusionaron las lecturas derivadas de 15.311.216 fragmentos y se descartaron, por problemas de calidad de los datos, las lecturas de 1.410 fragmentos (Tabla 3). Tras el procesamiento de la muestra G-3, se fusionaron las lecturas de 26.850.560 fragmentos y desecharon las lecturas de 33.574 fragmentos. En las etapas posteriores del análisis bioinformático, las lecturas fusionadas por PEAR fueron procesadas como lecturas individuales (*single-end reads*) y las restantes fueron procesadas como lecturas emparejadas (*paired-end reads*).

La Tabla 4 resume las frecuencias de cada nucleótido en las diversas muestras. Como indican los datos de esta tabla, el contenido en GC alcanza el 37,24% en los fragmentos secuenciados.

Tabla 4.- Frecuencias nucleotídicas determinadas empíricamente a partir de la secuencia de las lecturas

Muestra	Frecuencia			
	A	C	G	T
G-2 (A)	0,312083	0,186569	0,187602	0,31374
G-2 (B)	0,313074	0,185596	0,18674	0,314591
G-3 (A)	0,312901	0,185862	0,186835	0,314402
G-3 (B)	0,314115	0,184835	0,185959	0,315091
Promedio ^a	0,31310347	0,18566903	0,18673282	0,31449488

^aLos valores se han calculado como la media ponderada de las frecuencias de cada nucleótido teniendo en cuenta el número de fragmentos secuenciados en cada muestra.

Tras el procesamiento con PEAR, procesamos las secuencias resultantes con Trimmomatic, un programa que permite detectar y eliminar los adaptadores añadidos a cada fragmento durante la preparación de las muestras previa a la secuenciación, que ocasionalmente persisten en las secuencias obtenidas. El programa también recorta (*trim*) las bases de menor calidad situadas en los extremos de las lecturas, una operación que aprovecha las puntuaciones Phred contenidas en los archivos en formato FastQ (Cock *et al.*, 2010). La Tabla 5 recoge los resultados del procesamiento con Trimmomatic de las lecturas emparejadas (*paired-end*). En línea con la mayor calidad obtenida típicamente en la

secuenciación de las lecturas *forward*, se observa que el procesamiento con Trimmomatic descarta, en todos los casos, un mayor número de lecturas *reverse*.

Tabla 5.- Resultados del procesamiento de lecturas emparejadas con Trimmomatic

Fragmentos	Muestra			
	G-2 (A)	G-2 (B)	G-3 (A)	G-3 (B)
Procesados ^a	12.539.763	12.637.993	14.427.004	14.290.941
Persisten ambas lecturas	12.124.743 (96,690%)	12.194.406 (96,490%)	13.934.168 (96,580%)	13.770.307 (96,360%)
Persiste sólo la lectura <i>forward</i>	373.589 (2,989%)	364.478 (2,880%)	432.361 (3,000%)	429.857 (3,010%)
Persiste sólo la lectura <i>reverse</i>	25.578 (0,200%)	53.551 (0,420%)	35.689 (0,250%)	57.578 (0,400%)
Se descartan ambas lecturas	15.853 (0,130%)	25.558 (0,200%)	24.786 (0,170%)	33.199 (0,230%)

^aNótese que el número de fragmentos procesados coincide con el número de “Fragmentos con lecturas no solapantes” recogido en la Tabla 3.

La Tabla 6 recoge los resultados del procesamiento con Trimmomatic de las lecturas individuales (*single-end*).

Tabla 6.- Resultados del procesamiento de lecturas individuales con Trimmomatic

Muestra	Muestra			
	G-2 (A)	G-2 (B)	G-3 (A)	G-3 (B)
Procesadas ^a	7.967.500	7.343.716	14.161.049	12.689.511
Persisten	7.951.199 (99,800%)	7.318.810 (99,660%)	14.136.575 (99,830%)	12.658.132 (99,750%)
Descartadas	16.301 (0,200%)	24.906 (0,340%)	24.474 (0,170%)	31.379 (0,250%)

^aNótese que el número de lecturas individuales procesadas coincide con el número de “Fragmentos con lecturas solapantes” recogido en la Tabla 3.

5.4. Alineamiento de las lecturas al genoma de referencia

Tras el procesamiento con PEAR y Trimmomatic, las lecturas fueron alineadas a la versión más reciente disponible de la secuencia del genoma de *Arabidopsis thaliana* (TAIR10) utilizando el programa Bowtie2. Este programa utiliza un algoritmo basado en la

transformación de Burrows-Wheeler para alinear las lecturas a la secuencia del genoma de referencia. En el caso de las lecturas emparejadas, Bowtie2 determina si los alineamientos de las dos lecturas de un mismo fragmento son concordantes o discordantes, en función de la distancia y orientación relativa de las mismas. Los resultados del alineamiento de las lecturas al genoma de referencia se presentan en la Tabla 7.

Tabla 7.- Resultados del alineamiento de las lecturas al genoma con Bowtie2

Clasificación	Muestra			
	G-2 (A)	G-2 (B)	G-3 (A)	G-3 (B)
Pares de lecturas ^a	12.124.743	12.194.406	13.934.168	13.770.307
Sin alineamientos	549.250 (4,53%)	583.744 (4,79%)	516.994 (3,71%)	557.908 (4,05%)
Con un alineamiento concordante	8.400.209 (69,28%)	8.440.021 (69,21%)	10.223.172 (73,37%)	10.092.067 (73,29%)
Con más de un alineamiento concordante	3.175.284 (26,19%)	3.170.641 (26,00%)	3.194.002 (22,92%)	3.120.332 (22,66%)
Porcentaje de parejas alineadas	95,47%	95,21%	96,29%	95,95%
Lecturas individuales ^b	8.350.366	7.736.839	14.604.625	13.145.567
Sin alineamientos	491.628 (5,89%)	440.331 (5,69%)	577.975 (3,96%)	502.151 (3,82%)
Con un alineamiento	4.853.275 (58,12%)	4.448.243 (57,50%)	9.147.817 (62,64%)	8.161.184 (62,08%)
Con más de un alineamiento	3.005.463 (35,99%)	2.847.815 (36,81%)	4.878.833 (33,41%)	4.482.232 (34,10%)
Porcentaje de lecturas alineadas	94,11%	94,31%	96,04%	96,18%

^a Nótese que el número de pares de lecturas coincide con el número de fragmentos para los que “Persisten ambas lecturas” tras el procesamiento con Trimmomatic, recogido en la Tabla 5. ^b El número de lecturas individuales de esta Tabla es la suma del número de lecturas individuales que “Persisten” (Tabla 6) y el número de fragmentos para los que “Persiste sólo la lectura *forward*” o “Persiste sólo la lectura *reverse*” tras el procesamiento con Trimmomatic (Tabla 5).

Mediante el examen de los alineamientos de las parejas de lecturas a la secuencia del genoma de referencia es posible evaluar empíricamente *a posteriori* cómo se distribuye el tamaño de los fragmentos secuenciados por ambos extremos. Dicha distribución se representa en la Figura 3.

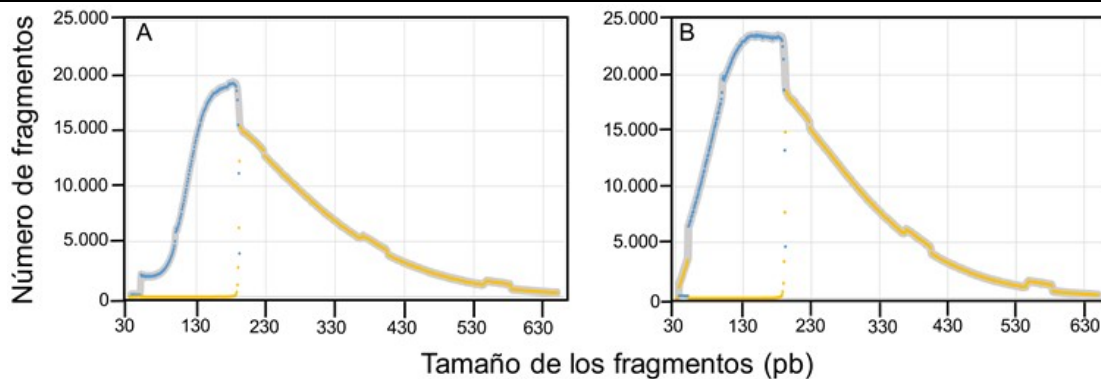


Figura 3.- Distribución por tamaño de los fragmentos secuenciados. (A) Distribución de tamaños de los fragmentos secuenciados para el mutante albino. (B) Distribución de tamaños de los fragmentos secuenciados para la estirpe silvestre. En azul se indican fragmentos resultantes de la fusión de lecturas forward y reverse por PEAR, para los que el tamaño del fragmento es igual al tamaño de la secuencia resultante de la fusión. En amarillo se muestran los tamaños deducidos a partir de las parejas de secuencias alineadas por Bowtie2. En gris se indica la suma de ambos valores.

5.5. Determinación de la posición del gen mutado usando cartografía mediante secuenciación

Una vez completado el alineamiento de las lecturas al genoma de referencia, los archivos resultantes (con formato BAM), fueron procesados con los programas GATK y Samtools. La Tabla 8 describe la asignación de las secuencias alineadas a los diferentes cromosomas. La cobertura ha sido estimada a la baja asumiendo que todas las secuencias alineadas poseen la misma longitud (101 nucleótidos). Como puede observarse, las secuencias derivadas de los genomas cloroplástico y mitocondrial son particularmente abundantes, un aspecto que debe ser considerado en el diseño del experimento.

En esta etapa se examinan las lecturas alineadas para determinar los nucleótidos que ocupan cada posición del genoma, que se almacenan en archivos con formato VCF. Con el fin de identificar posiciones del genoma que pudieran ser utilizadas como marcadores moleculares en la cartografía mediante secuenciación del gen a estudio, examinamos primero los resultados obtenidos para la muestra G-3 (compuesta por un grupo de individuos de la población cartográfica F_2 fenotípicamente silvestres). Seleccionamos, para su uso como marcadores moleculares, todas aquellas posiciones del genoma que reuniesen los siguientes criterios: (a) las lecturas presentasen dos alelos distintos en dicha posición; (b) la frecuencia relativa del alelo de referencia en dicha posición estuviese comprendida entre 0,3

y 0,7; (c) la suma de las frecuencias absolutas para los alelos considerados fuese igual al número de lecturas que ocupan dicha posición, y (d) que la cobertura de dicha posición estuviese comprendida entre 20 y 120, dos valores que definen un intervalo relativamente amplio en torno a la cobertura media estimada para los cromosomas 1 a 5 (Tabla 8). El criterio (c) excluye todas aquellas posiciones en las que aparecen más de dos alelos distintos, lo que es indicativo de secuencias alineadas en la posición incorrecta (como podría suceder en el caso de genes duplicados, por ejemplo). El criterio (d) excluye de manera efectiva todas aquellas secuencias correspondientes a la fracción repetitiva del genoma, que deben ser excluidas por las razones ya expuestas para el criterio (c). Estos criterios fueron reunidos por 222.090 posiciones del genoma, según el análisis de los datos obtenidos a partir de la muestra G-3. La distribución por cromosoma de dichas posiciones fue: 43.899 (cromosoma 1), 32.755 (cromosoma 2), 53.876 (cromosoma 3), 509 (cromosoma 4), y 91.051 (cromosoma 5). Como se aprecia, ninguna posición de los genomas mitocondrial y cloroplástico superó los criterios descritos.

Tabla 8.- Cobertura y número de secuencias alineadas a cada cromosoma

Cromosoma	Tamaño	G-2		G-3	
		Secuencias alineadas	Cobertura (aprox.)	Secuencias alineadas	Cobertura (aprox.)
1	30.427.671	9.878.943	32,79	15.636.972	51,90
2	19.698.289	8.932.796	45,80	13.197.527	67,67
3	23.459.830	10.148.806	43,69	15.865.718	68,31
4	18.585.056	7.269.431	39,50	11.518.160	62,59
5	26.975.502	9.380.373	35,12	14.815.323	55,47
Mitocondria	366.924	1.990.539	547,92	1.297.401	353,59
Cloroplasto	154.478	13.926.668	9.105,46	7.598.111	4.967,76

Para identificar la posición del gen a estudio, procedimos a estudiar las frecuencias alélicas en la muestra G-2 (compuesta por un grupo de 87 plantas fenotípicamente albinas). Se tabuló la frecuencia del alelo de referencia (Col-0) en la muestra G-2 en las 222.090 posiciones seleccionadas a partir de los datos de muestra G-3. Los datos obtenidos para cada cromosoma se han representado gráficamente en la Figura 4. Como puede apreciarse, la representación gráfica de las frecuencias alélicas en la muestra G-2 permitió identificar una única región del cromosoma 2, que es candidata a contener el gen cuyas mutaciones causan albinismo en las plántulas de *Arabidopsis thaliana*. Seleccionamos una región

amplia, de aproximadamente 400 kilobases, comprendida entre las posiciones 1.100.313 y 1.494.673 del cromosoma 2, para la búsqueda de la mutación causante del fenotipo albino.

El análisis descrito en los párrafos anteriores fue repetido variando los criterios descritos para la selección de marcadores moleculares a partir de la muestra G-3. En particular, se repitió el análisis modificando el criterio (b) para incluir posiciones con un rango mayor de frecuencias alélicas (entre 0,15 y 0,85). Estas nuevas condiciones, con un número de posiciones seleccionadas sensiblemente mayor (262.422), condujeron al mismo resultado.

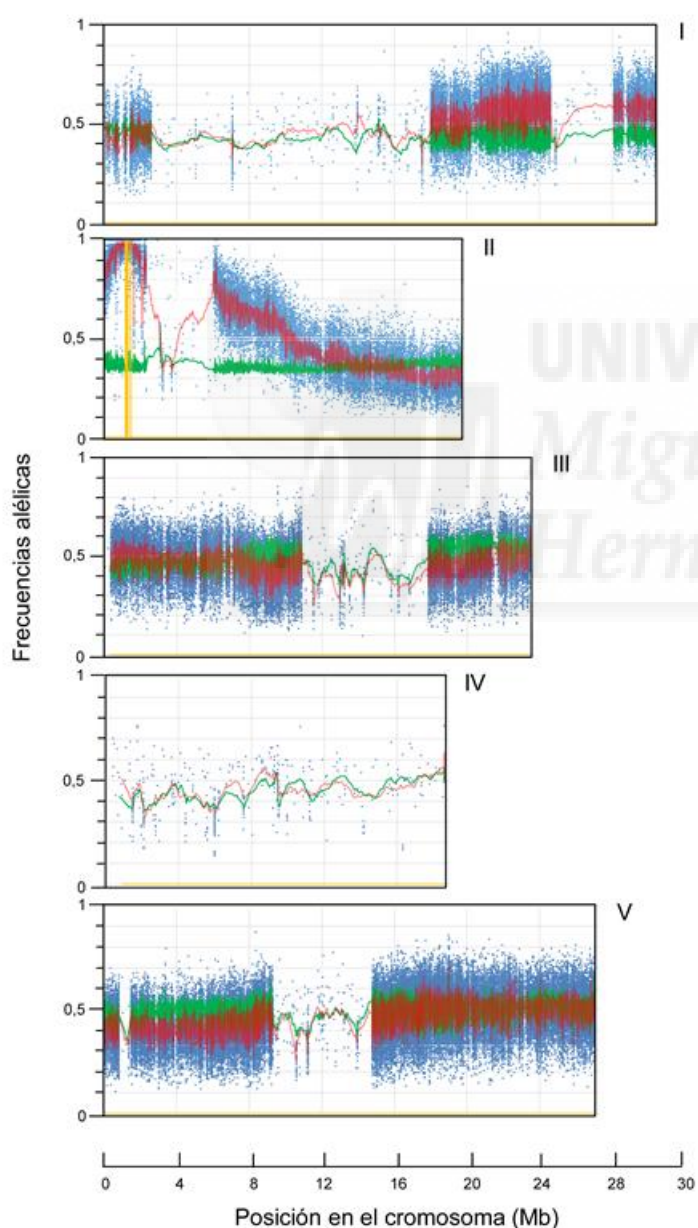


Figura 4.- Representación de las frecuencias alélicas en las muestras G-2 y G-3. Los puntos azules representan las frecuencias alélicas en la muestra G-2 para los marcadores considerados individualmente. La existencia de bloques con distinta abundancia de puntos azules pone de manifiesto que el genoma del mutante albino seleccionado es una mezcla de dos genomas distintos, uno de ellos muy parecido al genoma de la estirpe Col-0. Para cada cromosoma, la línea roja representa la media móvil de las frecuencias alélicas determinadas para la muestra G-2. La línea verde representa la media móvil de las frecuencias alélicas en la muestra G-3 (control). Por razones de claridad, no se han representado las frecuencias alélicas de los marcadores individuales en la muestra G-3. La línea amarilla representa el valor de *boost* correspondiente a los valores de la media móvil de la muestra G-2. El valor de *boost* es una transformación matemática de los datos utilizada por algunos autores para destacar la posición con frecuencia alélica máxima (véase el apartado 4.3).

Tabla 9.- Genes candidatos seleccionados en este trabajo

Gen	Cromosoma (posición)	Mutación (nucleótidos)	Mutación (aminoácidos)	Anotación
At2g03760	2 (1.150.012)	C → T	D → D	<i>ARABIDOPSIS THALIANA SULFOTRANSFERASE1</i>
	2 (1.150.015)	G → A	S → S	
	2 (1.150.030)	C → T	C → C	
At2g04030	2 (1.283.497)	G → A	No codificante	<i>HEAT SHOCK PROTEIN 90.5</i>
At2g04050	2 (1.337.410)	G → A	S → S	<i>DTXL1</i> (MATE efflux family protein)
At2g04080	2 (1.357.595)	C → T	V → V	<i>DTXL3</i> (MATE efflux family protein)
At2g04170	2 (1.418.980)	C → T	P → P	TRAF-like family protein
At2g04300	2 (1.494.653)	C → T	L → L	Leucine-rich repeat protein kinase
	2 (1.494.673)	G → A	T → T	

5.6. Identificación de genes candidatos en el intervalo seleccionado

La principal ventaja de los métodos de cartografía mediante secuenciación es que no sólo permiten identificar rápidamente la localización cromosómica del gen de interés, sino que también pueden conducir a la identificación de la lesión causante del fenotipo mutante. Para intentar identificar la mutación responsable del fenotipo albino, preparamos una lista con todas las posiciones del genoma de la muestra G-2 ocupadas por un nucleótido distinto del presente en el genoma de Col-0. Dicha lista fue cribada utilizando varios criterios. (1) En primer lugar, se excluyeron todos los cambios identificados previamente en otro mutante (afectado en el desarrollo de flores y frutos) aislado en el mismo experimento de mutagénesis (Eva Rodríguez-Alcocer y Héctor Candela, sin publicar). Se cruzaron las listas de cambios detectados en ambos mutantes y se descartaron todos los presentes en ambas listas, ya que, muy probablemente, dichos cambios deben corresponder a polimorfismos entre la estirpe Col-0 y el parental mutagenizado. (2) En segundo lugar, se aplicó un segundo filtrado, en el que se retuvieron únicamente las transiciones G/C→A/T, ya que éste es el tipo de mutaciones inducido por el mutágeno empleado (EMS). Este segundo filtro redujo la lista de mutaciones candidatas a 127. (3) En tercer lugar, descartamos todas

La densidad de loci polimórficos en los distintos cromosomas varió entre 1 marcador cada 300 pb (cromosoma 5) y 1 marcador cada 36 kb (cromosoma 4), dependiendo de la diferente contribución de regiones ricas y pobres en polimorfismos a cada cromosoma. Como también se observa en la Figura 4, la transición entre regiones ricas y pobres en polimorfismos de un mismo cromosoma se produce abruptamente en todos los casos.

Si bien ya existen numerosas aplicaciones bioinformáticas que permiten analizar los resultados de experimentos de cartografía mediante secuenciación (Hartwig *et al.*, 2012; Ossowski *et al.*, 2008; Schneeberger y Weigel, 2011), las herramientas disponibles no poseen la flexibilidad necesaria para afrontar situaciones como la descrita en el párrafo anterior. Por ello, hemos preferido llevar a cabo el análisis de los datos manualmente, lo que nos ha permitido seleccionar los parámetros más adecuados en cada etapa del mismo. Para analizar los datos provenientes de una población F_2 , las versiones más recientes del programa SHOREmap (Schneeberger *et al.*, 2009) requieren que se resecuencie el genoma de una muestra correspondiente a plantas fenotípicamente mutantes, así como los genomas de los parentales no mutagenizados. En nuestro análisis, hemos secuenciado únicamente dos muestras. Hemos utilizado la información obtenida a partir del grupo (*pool*) de plantas silvestres para seleccionar un conjunto muy fiable de marcadores polimórficos bialélicos en la población cartográfica. Ello es posible gracias a que la frecuencia esperada para los alelos derivados del parental mutagenizado varía entre 1/2 (en el caso de un locus bialélico no ligado a la mutación causante del fenotipo) y 1/3 (en el caso de un locus bialélico ligado a la mutación causante del fenotipo). Con una cobertura suficiente, estas frecuencias garantizan la detección de la inmensa mayoría de los loci bialélicos existentes para su uso posterior como marcadores a analizar en la muestra derivada del mutante. En nuestro experimento, la cobertura media para los cromosomas del genoma nuclear osciló entre 52x y 68x, más que suficiente para permitir la detección de ambos alelos para un locus polimórfico.

El sesgo observado en las frecuencias alélicas determinadas en esas mismas posiciones utilizando los datos que hemos obtenido al secuenciar los mutantes es compatible con la presencia de una única mutación recesiva en el cromosoma 2. Nuestro estudio de las diferencias existentes entre la secuencia de los individuos silvestres y los mutantes nos ha permitido seleccionar una mutación que, presumiblemente, perturba el procesamiento de los intrones (*splicing*) en el gen At2g04030. La selección de esta mutación no ha estado exenta de dificultades, y ha requerido excluir secuencialmente polimorfismos existentes entre los parentales (que hemos detectado por comparación con los resultados del trabajo de Fin de Grado de Eva Rodríguez Alcocer), mutaciones de tipos diferentes a los causados por el metanosulfonato de etilo (EMS; que típicamente induce transiciones G/C→A/T; Greene *et al.*, 2003), y mutaciones sinónimas o localizadas en regiones no

codificantes. La mutación identificada es una transición G→A que afecta al último nucleótido del séptimo intrón del gen At2g04030, que codifica una proteína de la familia Hsp90 de proteínas de choque térmico, denominado Hsp90.5 (Feng *et al.*, 2014; Inoue *et al.*, 2013; Oh *et al.*, 2014). El nucleótido afectado se corresponde con el de la posición -1 del sitio aceptor del *splicing*, que se encuentra absolutamente conservado, como se muestra en el LOGO de la Figura 6.

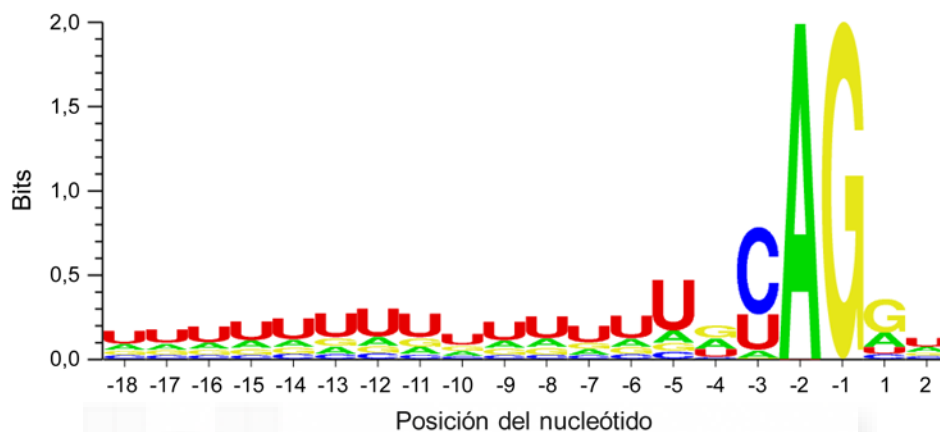
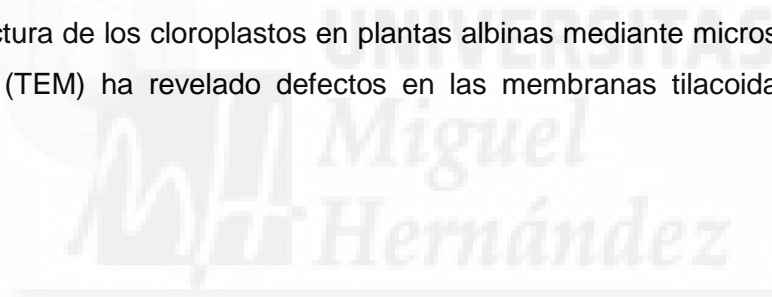


Figura 6.- Representación mediante un LOGO de la conservación de la secuencia de los sitios aceptores del *splicing* de *Arabidopsis thaliana*. Las posiciones indicadas con números negativos pertenecen al intrón precedente. Las posiciones indicadas con números positivos corresponden al exón posterior. El valor máximo que admite una secuencia de nucleótidos es 2 bits y se alcanza cuando una posición es ocupada siempre por el mismo nucleótido. La gráfica ha sido elaborada por nosotros a partir de datos de 1000 intrones diferentes utilizando el programa WebLogo (Crooks *et al.*, 2004).

Las proteínas de choque térmico de 90 kDa (Hsp90) funcionan como chaperonas moleculares, bien ayudando al plegamiento *de novo* de otras proteínas, o bien ayudando a prevenir su desnaturalización. Estas proteínas constan de tres dominios conservados: (1) un dominio de unión a ATP en el extremo N-terminal; (2) un dominio central encargado de la unión de proteínas cliente (*client*); y (3) un dominio de dimerización en el extremo C-terminal (Oh *et al.*, 2014). El genoma de *Arabidopsis thaliana* contiene 7 genes que codifican otras tantas proteínas Hsp90: AtHsp90.1, AtHsp90.2, AtHsp90.3 y AtHsp90.4 constituyen la subfamilia citosólica, AtHsp90.5 (Cao *et al.*, 2003) se localiza en el cloroplasto, AtHsp90.6 en la mitocondria y AtHsp90.7 en el retículo endoplásmico, respectivamente (Feng *et al.*, 2014). El gen At2g04030 de *Arabidopsis thaliana* codifica la proteína cloroplástica de choque térmico Hsp90.5, que pertenece a una subfamilia de chaperonas encargadas de permitir el paso de proteínas a través de las membranas del cloroplasto (Inoue *et al.*, 2013).

En los estudios realizados por Feng *et al.* (2014), se observó que las plantas de *Arabidopsis thaliana* que presentaban alguna alteración en el gen AtHsp90.5 poseían un

fenotipo albino, como el observado en nuestro mutante. Existen otros mutantes afectados en este mismo gen, incluyendo uno que causa resistencia a clorato (*cr88*; Feng *et al.*, 2014; Inoue *et al.*, 2013). La disponibilidad de estos alelos mutantes nos permitirá realizar ensayos de complementación para confirmar que hemos identificado correctamente el gen mediante. También se han descrito alelos letales embrionarios del gen *AtHsp90.5*. Por este motivo, el gen también ha sido descrito como *EMBRYO DEFECTIVE 1956* (*EMB1956*; Tzafirir *et al.*, 2004). Un problema asociado a las mutaciones letales embrionarias es que impiden la caracterización de sus efectos durante las fases del ciclo de vida posteriores a la embriogénesis (Candela *et al.*, 2011). Nuestro alelo mutante alcanza el estadio de plántula cuando se cultiva en ausencia de sacarosa, y llega a producir algunas hojas sólo si se cultiva en presencia de sacarosa, lo que sugiere que es un alelo hipomorfo (de pérdida de función parcial). El fenotipo postembrionario de las mutaciones en el gen *AtHsp90.5*, sin embargo, ya ha sido caracterizado gracias a la disponibilidad de otros alelos hipomorfos, cuyo desarrollo progresa hasta etapas posteriores a la germinación (Feng *et al.*, 2014), y también gracias a la disponibilidad de plantas transgénicas en la que el gen endógeno se silencia por cosupresión (Feng *et al.*, 2014; Oh *et al.*, 2014). Entre otros defectos, el estudio de la ultraestructura de los cloroplastos en plantas albinas mediante microscopía electrónica de transmisión (TEM) ha revelado defectos en las membranas tilacoidales (Feng *et al.*, 2014).



7. Conclusiones y proyección futura

Como parte de la investigación que se realiza en el laboratorio de Héctor Candela, hemos iniciado la caracterización de un gen de la planta modelo *Arabidopsis thaliana*, cuyas mutaciones de pérdida de función causan un fenotipo albino y detienen el desarrollo en el estadio de plántula. Hemos determinado que el fenotipo albino de este mutante se transmite con un modo de herencia monogénico recesivo. Para determinar la base molecular de este fenotipo mutante, hemos seleccionado plantas mutantes y silvestres para identificar el gen siguiendo una estrategia de cartografía mediante secuenciación (*mapping-by-sequencing*).

Hemos purificado ADN genómico a partir de sendos grupos de plantas mutantes y silvestres. Tras comprobar la calidad del ADN obtenido, fue enviado para su secuenciación por medio de un equipo de secuenciación por síntesis Illumina HiSeq 2500. Hemos realizado el análisis bioinformático de las secuencias obtenidas. El número total de fragmentos secuenciados fue de 40.490.382 para las plantas albinas, y de 55.602.079 para las plantas silvestres. Dicho análisis bioinformático ha comprendido las siguientes etapas: (1) evaluación de la calidad y número de las secuencias obtenidas, (2) ensamblaje de las lecturas *forward* y *reverse* solapantes, (3) eliminación de adaptadores y secuencias de baja calidad, (4) alineamiento de las lecturas de cada muestra al genoma de referencia, (5) identificación de las posiciones del genoma que corresponden a loci polimórficos, y (6) determinación de las frecuencias alélicas para los alelos de dichos loci. La comparación de las frecuencias alélicas en las muestras derivadas del mutante y del silvestre, nos ha permitido determinar que la mutación responsable del fenotipo albino reside en el cromosoma 2. El examen de las mutaciones contenidas en una región candidata de aproximadamente 400 kb del cromosoma 2, nos ha permitido identificar 9 transiciones G/C→A/T en 6 genes distintos de esta región. Las 8 mutaciones localizadas en regiones codificantes son sinónimas y no alteran la secuencia de las proteínas correspondientes. La mutación restante sustituye un residuo de guanina muy conservado en el séptimo intrón del gen At2g04030. Este gen codifica un miembro de la familia Hsp90 de proteínas de choque térmico.

En cuanto a los experimentos a realizar de manera inmediata, ya hemos diseñado cebadores para genotipar la mutación candidata en plantas individuales. Los resultados obtenidos hasta la fecha, con un número limitado de plantas, indican que la mutación cosegrega absolutamente con el fenotipo albino. También hemos diseñado cebadores y hemos conseguido amplificar la región codificante completa del gen, si bien esos resultados no han sido plasmados en este trabajo. Pretendemos generar construcciones para complementar el fenotipo mutante, mediante la expresión de una copia silvestre del gen en plantas transgénicas.

8. Bibliografía

- Andersson, I. y Backlund, A. (2008). Structure and function of RuBisCO. *Plant Physiol. Biochem.* **46**: 275-291.
- Bolger, A.M., Lohse, M. y Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**: 2114-2120.
- Bortiri, E., Jackson, D. y Hake, S. (2006). Advances in maize genomics: the emergence of positional cloning. *Curr. Opin. Plant Biol.* **9**: 164-171.
- Bryant, N., Lloyd, J., Sweeney, C., Myouga, F. y Meinke, D. (2011). Identification of nuclear genes encoding chloroplast-localized proteins required for embryo development in *Arabidopsis*. *Plant Physiol.* **155**: 1678-1689.
- Candela, H., Casanova-Sáez, R. y Micol, J.L. (2015). Getting started in mapping-by-sequencing. *J. Integr. Plant Biol.* **57**: 606-612.
- Candela, H., Pérez-Pérez, J.M. y Micol, J.L. (2011). Uncovering the post-embryonic functions of gametophytic- and embryonic-lethal genes. *Trends Plant Sci.* **16**: 336-345.
- Cao, D., Froehlich, J.E., Zhang, H. y Cheng, C.L. (2003). The chlorate-resistant and photomorphogenesis-defective mutant *cr88* encodes a chloroplast-targeted HSP90. *Plant J.* **33**: 107-118.
- Carles, C.C., Choffnes-Inada, D., Reville, K., Lertpiriyapong, K. y Fletcher, J.C. (2005). *ULTRAPETALA1* encodes a SAND domain putative transcriptional regulator that controls shoot and floral meristem activity in *Arabidopsis*. *Development* **132**: 897-911.
- Cock, P.J., Fields, C.J., Goto, N., Heuer, M.L. y Rice, P.M. (2010). The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res.* **38**: 1767-1771.
- Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E. (2004). WebLogo: A sequence logo generator. *Genome Res.* **14**: 1188-1190.
- Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., McVean, G., Durbin, R. y 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. *Bioinformatics* **27**: 2156-2158.
- Fellerer, C., Schweiger, R., Schöngruber, K., Soll, J. y Schwenkert, S. (2011). Cytosolic HSP90 cochaperones HOP and FKBP interact with freshly synthesized chloroplast preproteins of *Arabidopsis*. *Mol. Plant* **4**: 1133-1145.

- Feng, J., Fan, P., Jiang, P., Lv, S., Chen, X. y Li, Y. (2014). Chloroplast-targeted Hsp90 plays essential roles in plastid development and embryogenesis in *Arabidopsis* possibly linking with VIPP1. *Physiol. Plant.* **150**: 292-307.
- Fletcher, J.C. (2001). The *ULTRAPETALA* gene controls shot and floral meristem size in *Arabidopsis*. *Development* **128**: 1323-1333.
- Gallavotti, A. y Whipple, C.J. (2015). Positional cloning in maize (*Zea mays* subsp. *mays*, Poaceae). *Appl. Plant Sci.* **3**: 1400092.
- Greene, E.A., Codomo, C.A., Taylor, N.E., Henikoff, J.G., Till, B.J., Reynolds, S.H., Enns, L.C., Burtner, C., Johnson, J.E., Odden, A.R., Comai, L. y Henikoff, S. (2003). Spectrum of chemically induced mutations from a large-scale reverse-genetic screen in *Arabidopsis*. *Genetics* **164**: 731-740.
- Hartwig, B., James, G.V., Konrad, K., Schneeberger, K. y Turck, F. (2012). Fast isogenic mapping-by-sequencing of ethyl methanesulfonate-induced mutant bulks. *Plant Physiol.* **160**: 591-600.
- Inoue, H., Li, M. y Schnell, D.J. (2013). An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. *Proc. Natl. Acad. Sci. USA* **110**: 3173-3178.
- Kobayashi, Y., Takusagawa, M., Harada, N., Fukao, Y., Yamaoka, S., Kohchi, T., Hori, K., Ohta, H., Shikanai, T. y Nishimura, Y. (2016). Eukaryotic components remodeled chloroplast nucleoid organization during the green plant evolution. *Genome Biol. Evol.* **8**: 1-16.
- Lamesch, P., Berardini, T.Z., Li, D., Swarbreck, D., Wilks, C., Sasidharan, R., Muller, R., Dreher, K., Alexander, D.L., Garcia-Hernandez, M., Karthikeyan, A.S., Lee, C.H., Nelson, W.D., Ploetz, L., Singh, S., Wensel, A. y Huala, E. (2012). The *Arabidopsis* Information Resource (TAIR): improved gene annotation and new tools. *Nucleic Acids Res.* **40 (Database issue)**: D1202-10.
- Langmead, B., Trapnell, C, Pop, M, y Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**: R25.
- Langmead, B. y Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**: 357-359.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. y 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**: 2078-2079.
- Li, H. y Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **15**: 1754-1760.
- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernysky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M. y DePristo, M.A. (2010). The

- Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**: 1297-1303.
- Metzker, M.L. (2010). Sequencing technologies - the next generation. *Nat. Rev. Genet.* **11**: 31-46.
- Michelmore, R.W., Paran, I. y Kesseli, R.V. (1991). Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Genetics* **88**: 9828-9832.
- Murashige, T. y Skoog, F. (1962). A revised medium for rapid growth and bio-assays with tobacco tissue cultures. *Physiol. Plant* **15**: 473-497.
- Myouga, F., Akiyama, K., Motohashi, R., Kuromori, T., Ito, T., Iizumi, H., Ryusui, R., Sakurai, T. y Shinozaki, K. (2010). The Chloroplast Function Database: a large-scale collection of *Arabidopsis Ds/Spm-* or T-DNA-tagged homozygous lines for nuclear-encoded chloroplast proteins, and their systematic phenotype analysis. *Plant J.* **61**: 529-542.
- Oh, S.E., Yeung, C., Babaei-Rad, R. y Zhao, R. (2014). Cosuppression of the chloroplast localized molecular chaperone HSP90.5 impairs plant development and chloroplast biogenesis in *Arabidopsis*. *BMC Res. Notes* **7**: 643.
- Ossowski, S., Schneeberger, K., Clark, R.M., Lans, C., Warthmann, N. y Weigel, D. (2008). Sequencing of natural strains of *Arabidopsis thaliana* with short reads. *Genome Res.* **18**: 2024-2033.
- Pareek, C.S., Smoczynski, R. y Tretyn, A. (2011). Sequencing technologies and genome sequencing. *J. Appl. Genet.* **52**: 413-435.
- Pérez-Pérez, J.M., Candela, H., Robles, P., Quesada, V., Ponce, M.R. y Micol, J.L. (2009). Lessons from a search for leaf mutants in *Arabidopsis thaliana*. *Int. J. Dev. Biol.* **53**: 1623-1634.
- Powikrowska, M., Oetke, S., Jensen, P.E. y Krupinska, K. (2014). Dynamic composition, shaping and organization of plastid nucleoids. *Front. Plant Sci.* **5**: 424.
- Roberts, R.J., Carneiro, M.O. y Schatz, M.C. (2013). The advantages of SMRT sequencing. *Genome Biol.* **14**: 405.
- Rodríguez Alcocer, Eva María (2014). Cartografía mediante secuenciación de un mutante que perturba el desarrollo del fruto en *Arabidopsis thaliana*. Trabajo de Fin de Grado. Facultad de Ciencias Experimentales. Universidad Miguel Hernández de Elche.
- Sanger, F. y Coulson, A.R. (1975). A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. *J. Mol. Biol.* **94**: 441-448.

- Schneeberger, K. y Weigel, D. (2011). Fast-forward genetics enabled by new sequencing technologies. *Trends Plant Sci.* **16**: 282-288.
- Schneeberger, K., Ossowski, S., Lanz, C., Juul, T., Petersen, A.H., Nielsen, K.L., Jørgensen, J.E., Weigel, D. y Andersen, S.U. (2009). Shoremap: simultaneous mapping and mutation identification by deep sequencing. *Nat. Methods* **6**: 550-551.
- Shin, S.C., Ahn, D.H., Kim, S.J., Lee, H., Oh, T.J., Lee, J.E. y Park, H. (2013). Advantages of Single-Molecule Real-Time sequencing in high-GC content genomes. *PLoS ONE* **8**: e68824.
- Sun, H. y Schneeberger, K. (2015). SHOREmap v3.0: fast and accurate identification of causal mutations from forward genetic screens. *Methods Mol. Biol.* **1284**: 381-395.
- Thudi, M., Li, Y., Jackson, S.A., May, G.D. y Varshney, R.K. (2012). Current state-of-art of sequencing technologies for plant genomics research. *Brief. Functional Genomics* **11**: 3-11.
- Tomizioli, M., Lazar, C., Brugière, S., Burger, T., Salvi, D., Gatto, L., Moyet, L., Breckels, L.M., Hesse, A.M., Lilley, K.S., Seigneurin-Berny, D., Finazzi, G., Rolland, N. y Ferro, M. (2014). Deciphering thylakoid sub-compartments using a mass spectrometry-based approach. *Mol. Cell. Proteomics* **13**: 2147-2167.
- Tu, J., Ge, Q., Wang, S., Wang, L., Sun, B., Yang, Q., Bai, Y. y Lu, Z. (2012). Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis. *BMC Genomics* **13**: 43.
- Tucker, T., Marra, M. y Friedman, J.M. (2009). Massively parallel sequencing: the next big thing in genetic medicine. *Am. J. Hum. Genet.* **85**: 142-154.
- Tzafrir, I., Pena-Muralla, R., Dickerman, A., Berg, M., Rogers, R., Hutchens, S., Sweeney, T.C., McElver, J., Aux, G., Patton, D. y Meinke, D. (2004). Identification of genes required for embryo development in *Arabidopsis*. *Plant Physiol.* **135**: 1206-1220.
- Venglat, S.P., Dumonceaux, T., Rozwadowski, K., Parnell, L., Babic, V., Keller, W., Martienssen, R., Selvaraj, G. y Datla, R. (2002). The homeobox gene *BREVIPEDICELLUS* is a key regulator of inflorescence architecture in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* **99**: 4730-4735.
- Yanhu, L., Lu, W. y Li, Y. (2015). The principle and application of the single-molecule real-time sequencing technology. *Yi Chuan* **37**: 259-268.
- Zhang, J., Chiodini, R., Badr, A. y Zhang, G. (2011). The impact of next-generation sequencing on genomics. *J. Genet. Genomics* **38**: 95-109.
- Zhang, J., Kobert, K., Flouri, T., y Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. *Bioinformatics* **30**: 614-620.

Zuryn, S., Le Gras, S., Jamet, K. y Jarriault, S. (2010). A strategy for direct mapping and identification of mutations by whole-genome sequencing. *Genetics* **186**: 427-430.

