UNIVERSIDAD NACIONAL AUTONOMA DE MEXICO

Maestría y Doctorado en Ciencias Bioquímicas

Caracterización molecular y funcional de mutantes de *A. thaliana* afectadas en la biogénesis del cloroplasto.

TESIS

QUE PARA OPTAR POR EL GRADO DE:

Doctor en Ciencias

PRESENTA:

M. en C. Luis Alberto de Luna Valdez

Tutor:

Dr. Ángel Arturo Guevara García (IBt, UNAM)

Comité tutor:

Dra. Rosario Vera Estrella (IBt, UNAM)

Dra. Ana Paulina Barba de la Rosa (IPICyT, CONACyT)

Cuernavaca, Morelos. Agosto, 2020

La ciencia es, en mi opinión, la actividad más hermosa creada por el ser humano. Es una lástima que en la actualidad solo conozcamos a su prima menos agraciada: "La investigación".

Para Luz de Luna y Fernanda Gómez

Agradecimientos

A las instituciones.

Le agradezco al CONACyT y la UNAM por el financiamiento proporcionado a este trabajo a través de los proyectos DGAPA-UNAM IN210917, IN207214, CONACyT FC-2015/96 y CB-2015/251848, 220534 y de la beca con numero 240088.

A los académicos.

Agradezco a los doctores Gladys Cassab López, José Luis Reyes Taboada, Joseph Dubrovsky, Mario Serrano Ortega y Omar Pantoja Ayala por sus comentarios acerca de este documento y del trabajo.

Agradezco también a los doctores Ángel Arturo Guevara García, Rosario Vera Estrella y Ana Paulina Barba de la Rosa por haber formado parte del comité tutoral encargado de la evaluación del desarrollo de este trabajo a lo largo de toda su duración.

A las personas.

Finalmente, me gustaría agradecerles a todas las personas que tuvieron un impacto grande en mi desarrollo personal y profesional durante el periodo en el que desarrolle mi trabajo de investigación.

- Dr. Arturo Guevara, por mostrarme que los investigadores son personas con defectos y virtudes de los que se puede aprender más que solo ciencia.
- Dra. Patricia León, por enseñarme que a veces para demostrar una idea, hay que trabajar en su contra.

- Dra. Rosario Vera, por mostrarme que las ideas son muy poco cuando no están acompañadas de trabajo duro y desdén por el fracaso.
- Dra. Elizabeth Cordoba. Ni siquiera voy a tratar de enumerar la cantidad de cosas que aprendí de Elizabeth, porque son demasiadas. Elizabeth es una persona crítica y comprensiva, de ideas flexibles y valores inquebrantables; Elizabeth es tal vez una de las personas que más influencia ha tenido en mi desarrollo personal y profesional hasta ahora. Dispuesta en todo momento a discutir mis ideas más recientes o a platicar sobre las especies de plantas más raras en nuestras respectivas colecciones, encontré en Elizabeth algo mucho más significativo que un compañero de laboratorio o, inclusive, más significativo que un asesor. No me queda la menor duda de que ni mi trabajo ni yo mismo seríamos lo que hoy somos sin la influencia y el apoyo de Elizabeth. Gracias Elizabeth.
- A mi tutor. Aunque en papel solo hay una persona, estoy convencido plenamente de que mi verdadero tutor es una entidad abstracta compuesta por diferentes aspectos de las personas que he nombrado hasta ahora: Arturo Guevara, Rosario Vera, Patricia León y Elizabeth Cordoba. Gracias porque, en conjunto, fueron el tutor que necesitaba.
- A Luz de Luna y Fernanda Gómez, porque lo son todo para mi. Son, sin lugar a duda, las personas más importantes en mi vida. Gracias por todas las veces que lograron hacerme olvidar en menos de un minuto el décimo experimento fallido de la semana. No tienen idea de lo completo que soy con ustedes. Gracias.

Tabla de contenido.

Abreviaturas.

accD	Acetíl-CoA carboxílasa
DNA	Acido Desoxirribonucleico
RNA	Ácido Ribonucleico
ADP	Adenosina difosfato
ATP	Adenosina trifosfato
N-Terminal	Región amino terminal de una proteína
CAH1	Anhidrasa Carbónica 1
BN-PAGE	Blue Native Poliacrylamide Gel Electrophoresis
CTE	Cadena de transporte de electrones
CLB/clb	Chloroplast biogenesis
CGE1	Chloroplastic GrpE 1
CGE2	Chloroplastic GrpE 2
cla1	Cloroplastos alterados
LHC	Complejo captador de luz
BiFC	Complementación Bimolecular de la Fluorescencia
ROS	Especies reactivas del oxígeno
PAP	Fosfoadenosina 5´-fosfato
PSI	Fotosistema I
PSII	Fotosistema II
cYFP	Fragmento C-terminal del reportero YFP
nYFP	Fragmento N-terminal del reportero YFP
PhANGs	Genes Nucleares Asociados a la fotosíntesis
HA	Hemaglutinina de la Influenza Humana
LB	Medio Lisogenico
MEcPP	Metil-D-eritrol 2,4-ciclodifosfato
MEP	Metil-Eritrol-Fosfato
MgProtoIX	Mg-Protoporfirina IX
NADPH	Nicotinamida adenina dinucleótido fosfato
pEG	pEarly Gate 103
HRP	Peroxidasa de rábano
YFP	Proteína Amarilla Fluorescente
CFP	Proteína Cian Fluorescente
HSP	Proteína de Heat-shock
PBP1	Proteína de Unión a PYK10
RPL21	Proteína ribosomal cloroplástica L21
GFP	Proteína Verde Fluorescente
pCE	pSPYCE
pNE	pSPYNE
ER	Retículo Endoplásmico
NEP	RNA Polimerasa codificada en el Núcleo
PEP	RNA Polimerasa codificada en el Plástido

rrn16S	rRNA cloroplástico 16S	Wt	Tipo Silvestre
RuBisCO	Ribulosa Bisfosfato Carboxilasa/Oxigenasa	FRET	Transferencia de Energía por Resonancia de Förster
CDS	Secuencia Codificante	Toc	Translocón de la membrana Externa del Cloroplasto
psbA	Subunidad D1 del centro de reacción del fotosistema II	Tic	Translocón de la membrana Interna del Cloroplasto
RBCS	Subunidad pequeña de la enzima RuBisCO	β-cyc	β-ciclocitral
SVR3	Supressor of Variegation 3		

Resumen.

La vida en la biósfera depende en gran medida de la capacidad de los organismos fotoautótrofos para generar biomasa y oxígeno a partir de agua y del carbono inorgánico disponible en la atmósfera, esto mediante un proceso bioquímico conocido como fotosíntesis. En los eucariontes, la fotosíntesis se lleva a cabo dentro de unos organelos especializados conocidos como cloroplastos, los cuales, a lo largo de la evolución, se han integrado a la fisiología y al desarrollo vegetal. Debido a la importancia de estos organelos, este trabajo se centra en profundizar en el entendimiento de los mecanismos moleculares que subyacen a la biogénesis de los cloroplastos en *Arabidopsis thaliana*. A través de un análisis proteómico comparativo, se identificaron 3 proteínas (AtRAB GTPasa E1B (ATRABE1B), PYK10-binding protein 1 (PBP1) y Chloroplast GrpE 1 (CGE1)) que la mutación de sus genes correspondientes genera defectos en la pigmentación, lo cual es un fenotipo distintivo de las plantas con defectos en el desarrollo del cloroplasto. El análisis molecular de estas mutantes puso de manifiesto que están afectadas en la acumulación de transcritos que son usados como marcadores de la biogénesis y el funcionamiento de los cloroplastos. Por otro lado, utilizando ensayos de complementación genética en bacterias, demostramos que CGE1 y su homólogo CGE2 son ortólogos funcionales de las cochaperonas GrpE de *E. coli*, indicando que la función de ambas proteínas en los cloroplastos de plantas superiores está relacionada a la función que tienen junto a las chaperonas Heat-shock Protein 70 (HSP70). Mediante técnicas de microscopia de fluorescencia encontramos que CGE1 y CGE2 se localizan en el estroma y la envoltura del cloroplasto, y que interaccionan en distintas configuraciones con la HSP70 cloroplástica cpHsc70-1. Además, encontramos que solamente CGE1 y cpHsc70-1 están coreguladas en respuesta a estrés por calor, sugiriendo que ambas trabajan en conjunto para regular los mismos procesos en el cloroplasto. Finalmente, un análisis bioquímico de líneas mutantes en CGE1 y CPHSC70-1, reveló que ambas tienen defectos en la acumulación de los oligómeros del Light-Harvesting Complex II, los cuales están involucrados en la fotosíntesis. En conclusión, demostramos la participación de la proteína CGE1 en el funcionamiento de los cloroplastos, como parte del sistema de chaperonas y cochaperonas que opera dentro de los cloroplastos de las plantas superiores.

Abstract.

Life in the biosphere depends on the ability of photoautotrophic organisms to generate biomass from water and the inorganic carbon present in the atmosphere, by a biochemical process called photosynthesis. In eukaryotes, photosynthesis takes place inside specialized organelles known as chloroplasts, which throughout the evolution have been integrated into the physiology and development of plants. Given the pivotal role of chloroplasts, this study focuses on widening the current understanding of the molecular mechanisms that underlie chloroplast biogenesis in *Arabidopsis thaliana*. Through a comparative proteomic analysis, we identified three proteins (AtRAB GTPasa E1B (ATRABE1B), PYK10-binding protein 1 (PBP1) y Chloroplast GrpE 1 (CGE1)) whose mutation cause pigmentation deficiencies, which is a distinctive phenotype of plants with chloroplast development defects. Molecular analysis of these mutants unveiled that they are affected in the accumulation of several transcripts known to be markers of chloroplast biogenesis and functionality. Furthermore, using genetic complementation assays in bacteria, we demonstrated that CGE1 and the paralogue CGE2 are functional orthologs of the GrpE cochaperone from *E. coli*; indicating that the function of the two CGEs is linked to that of Heat-shock Protein 70 (HSP70) chaperones. Using a fluorescence microscopy approach, we found that both CGE1 and CGE2 are localized in the chloroplast's stroma and envelope, and they can establish direct physical interactions with the chloroplastic HSP70 protein cpHsc70-1. We also found that only CGE1 and cpHsc70-1 are coregulated in response to heat stress, suggesting that these two proteins work together to regulate the same processes in the chloroplasts. Finally, a biochemical analysis revealed that plants having mutations in CGE1 and CPHSC70-1 genes have defects in the accumulation of LHCII oligomers, which are involved in photosynthesis. In conclusion, we demonstrated the participation of CGE1 in chloroplast functionality as a part of the chaperone and cochaperone system operating in the chloroplasts of land plants.

1. Introducción.

La vida en la biósfera, en especial la de los organismos heterótrofos, depende en gran medida de la actividad de unos organelos de las células vegetales y de las algas conocidos como cloroplastos. Dichos organelos son el lugar en donde reside toda la maquinaria enzimática necesaria para llevar a cabo la fotosíntesis; el proceso bioquímico mediante el que varios componentes inorgánicos como el agua, la luz y el dióxido de carbono, son convertidos en los compuestos orgánicos que son utilizados como fuente de energía y biomasa por los seres vivos. Debido a la gran importancia de los cloroplastos en la biología, la elucidación de los mecanismos celulares y moleculares que subyacen su biogénesis y funcionamiento resulta indispensable para comprender de manera íntegra los procesos fundamentales en los que se basa la vida.

1.1. El origen y la diversidad de los plástidos.

Los cloroplastos son organelos semiautónomos que se originaron hace aproximadamente 900 millones de años, mediante procesos de endosimbiosis entre cianobacterias de vida libre y eucariontes primitivos; los cuales integraron a estas bacterias a su desarrollo y metabolismo [Shih y Matzke, 2013]. A lo largo de la evolución, los cloroplastos se han diversificado formando un grupo de organelos relacionados conocidos como plástidos, cuyo origen real aún es tema de debate puesto que los mecanismos celulares y moleculares, así como las condiciones metabólicas que facilitaron la endosimbiosis aún son desconocidas [Keeling, 2010]. Por ejemplo, existen evidencias que indican que los eucariontes primitivos que dieron origen a los linajes vegetales (Archaeplastida) tuvieron infecciones con algunos patógenos intracelulares del

género Chlamydia, puesto que se han encontrado genes de origen Chlamydial codificados en los genomas de los organismos descendientes de los Archaeplastida; esto llevó a la generación de una serie de hipótesis que sugieren que la presencia del patógeno Chlamydial propició las condiciones metabólicas en las que tener y mantener endosimbiontes autotróficos fue beneficioso para los eucariontes en cuestión [Huang *et al*, 2007]. A pesar de que las evidencias filogenéticas indican que existe una correlación entre el establecimiento del cloroplasto y la infección con Chlamydia, el estudio de los genes propuestos como facilitadores de este proceso ha dado resultados contradictorios [Baum, 2013; Ball *et al*, 2015].

Actualmente, existen tres tipos de plástidos: primarios, secundarios y terciarios. Los plástidos primarios son organelos rodeados por dos membranas, éstos representan un sólo evento de endosimbiosis entre una bacteria fotosintética y una célula eucarionte no fotosintética. Estos plástidos se encuentran en cuatro linajes de eucariontes: las clorófitas y las plantas (cloroplastos), las rodófitas (rodoplasto) y las glaucófitas (muroplasto) [Keeling, 2010] (Figura 1). Por otro lado, los plástidos secundarios y terciarios son organelos rodeados por más de dos membranas, estos representan procesos endosimbióticos entre eucariontes fotosintéticos y eucariontes no fotosintéticos. Estos plástidos han sufrido modificaciones como la disminución del número original de membranas que los rodean y la pérdida del núcleo del eucarionte fotosintético simbionte, este tipo de plástidos se encuentran comúnmente en los euglénidos y los dinoflagelados [Keeling, 2010] (Figura 1).

Los plástidos modernos presentan varias características que son vestigios de sus ancestros bacterianos. Por ejemplo:

- Como las bacterias, los plástidos tienen genomas de DNA organizados en cromosomas circulares [Kolodner *et al*, 1975].
- Los ribosomas plastídicos son del tipo 70S, al igual que los bacterianos; además de que los mRNAs cloroplásticos tienen secuencias similares a las regiones Shine-Dalgarno que facilitan la traducción de los mRNAs bacterianos [Hirose *et al*, 2004].
- Los cloroplastos se dividen mediante un mecanismo similar a la fisión binaria de los procariontes, utilizando estructuras proteínicas en forma de anillos que constriñen la región ecuatorial durante la división [Yoshida *et al*, 2012].

A pesar de la diversidad de plástidos, en este trabajo usaremos el término “plástido” para referirnos a los, igualmente diversos, plástidos primarios presentes en las plantas verdes, que son el objeto de este estudio (Figura 2).

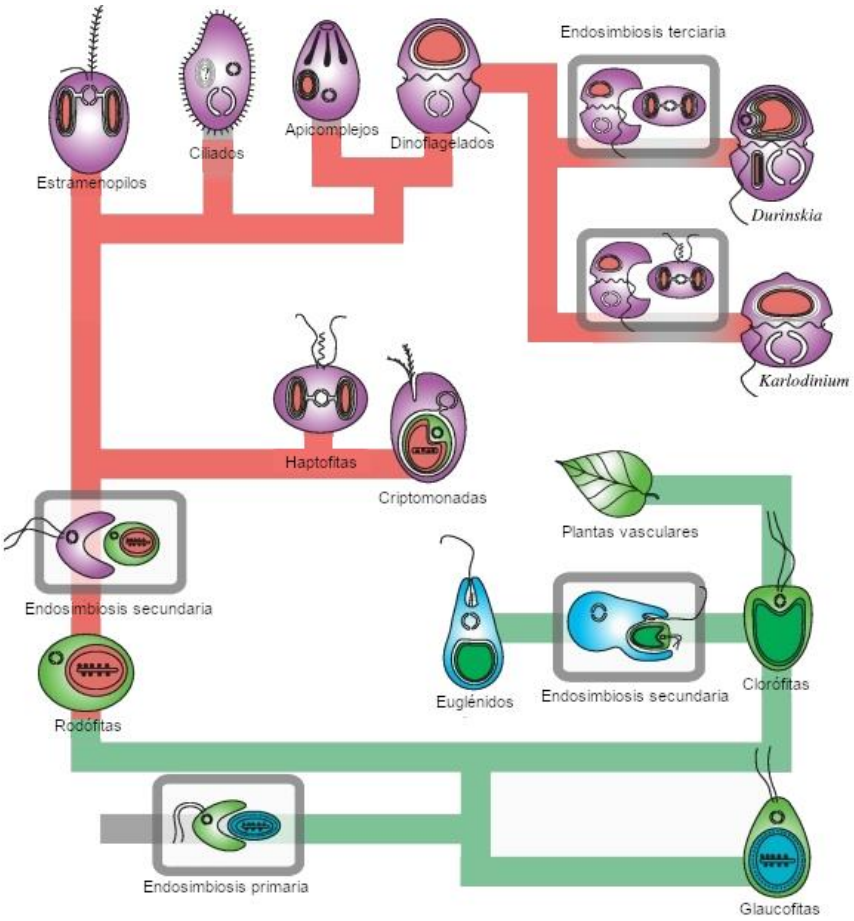


Figura 1. Tipos de endosimbiosis. Se presentan en marcos grises los tipos principales de endosimbiosis y los linajes modernos que presentan a los plástidos correspondientes. Imagen adaptada de Keeling, 2010.

Como ya se mencionó anteriormente, durante la evolución de los linajes vegetales, los cloroplastos se integraron al metabolismo y al desarrollo de las células eucariontes. En el caso de los organismos multicelulares, esta integración dio como resultado la generación de distintos organelos especializados en llevar a cabo procesos específicos, según las funciones o requerimientos particulares de las células que los portan (Figura 2). En las plantas vasculares, los cloroplastos pertenecen a un grupo diverso de organelos interconvertibles que tienen características morfológicas y funcionales distintivas [Whatley, 1978]. Por ejemplo, los cromoplastos presentes en las flores y los frutos están especializados en la síntesis y acumulación de pigmentos, mientras que la

producción y almacenamiento de lípidos, proteínas y almidón ocurren en los oleinoplastos, proteinoplastos y amiloplastos, respectivamente [Neuhaus y Emes, 2000] (Figura 2). Además del almacenamiento de macromoléculas, recientemente se han descrito tipos particulares de cloroplastos en los cuales (a pesar de contener pigmentos fotosintéticos) no se realiza la fotosíntesis de manera primordial, sino que están especializados en ser centros de integración de señales celulares [Beltrán *et al*, 2018]. Todos estos organelos se desarrollan a partir de precursores conocidos como proplástidos, los cuales están presentes en las células embrionarias y meristemáticas de las plantas, y en respuesta a señales intracelulares y ambientales son capaces de modificar su desarrollo para generar cualquier tipo de plástido necesario [Jarvis y López-Juez, 2013] (Figura 2).

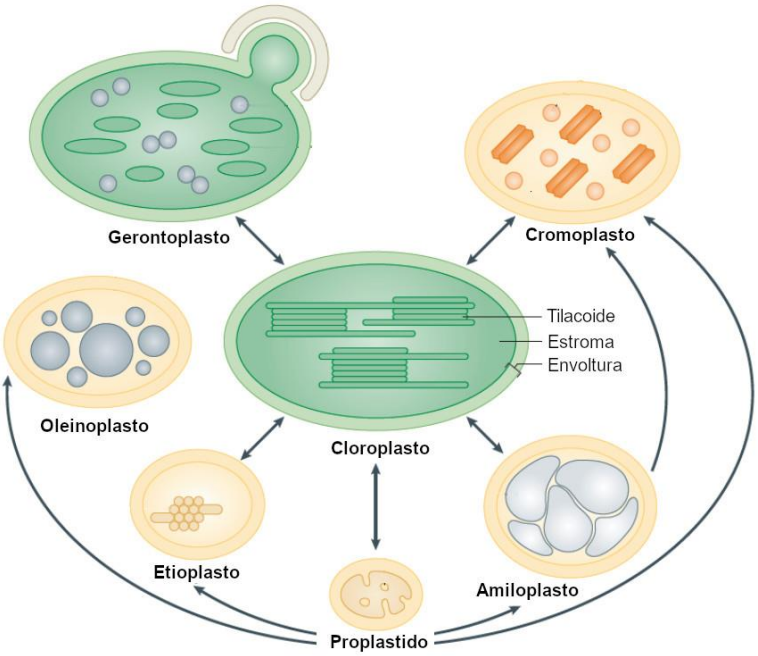


Figura 2. Tipos de plástidos. Se muestra un diagrama de los plástidos más comunes en las células vegetales y las interconversiones que se pueden dar entre ellos. Además, se presenta la estructura básica (envoltura, estroma y tilacoides) de los cloroplastos. Imagen adaptada de Jarvis y Lopéz-Juez, 2013.

1.2. Estructura y función de los cloroplastos.

A pesar de que la importancia de todos los plástidos es bien reconocida, el cloroplasto es por mucho el plástido más estudiado hasta la fecha; siendo sujeto de innumerables estudios que han puesto de manifiesto su participación en una amplia gama de procesos fisiológicos y metabólicos. Por ejemplo, se sabe que los cloroplastos son la fuente de hormonas indispensables como el ácido abscísico, el ácido jasmónico, el ácido salicílico y las giberelinas; además de que están encargados de la síntesis de galactolípidos y, a través de la fotosíntesis (su función más característica), sintetizan todos los esqueletos de carbono necesarios para el crecimiento de la planta [Chen *et al*, 2018].

Los cloroplastos son organelos celulares que tienen una estructura bien definida que consta de: una envoltura formada por dos unidades de membrana (i.e. dos bicapas de lípidos) que rodean a un compartimento interno llamado estroma, y la membrana del tilacoide, la cual es una unidad de membrana que forma una red de cúmulos (grana), que a su vez encierran a otro compartimento fluido conocido como el lumen del tilacoide (Figura 2). Cada una de estas partes, o subcompartimentos, lleva a cabo diferentes funciones que son esenciales para el funcionamiento del cloroplasto. Por ejemplo, la envoltura del cloroplasto contiene a una gran parte de la maquinaria proteínica necesaria para mediar el transporte de proteínas del citoplasma al cloroplasto [Waters y Langdale, 2009], además de que es el sitio en el que suceden algunas reacciones esenciales para la replicación del DNA cloroplástico y la división de los cloroplastos [Powikrowska *et al*, 2014; Oldenburg y Bendich, 2015, Chen *et al*, 2018]. Durante la fase luminosa de la fotosíntesis, la separación entre el estroma y el lumen del tilacoide que genera la membrana de éste, es esencial para la producción de moléculas energéticas, debido a

que las enzimas que producen el ATP y el NADPH+ tienen orientaciones especificas en esta membrana que les permiten llevar a cabo su función. Además, los complejos macromoleculares que participan en la fase luminosa, como los fotosistemas PSI y PSII, están anclados a la membrana del tilacoide [Gao *et al*, 2018]. Por último, en el estroma se encuentran muchas enzimas con funciones muy diversas, como el ciclo de Calvin-Benson para la fijación de carbono y la vía MEP para la síntesis de isoprenoides [Perello *et al*, 2016; Sharkey, 2019].

En vista del papel fundamental que tienen los cloroplastos en el desarrollo vegetal, muchos esfuerzos han sido realizados por un sinnúmero de grupos de investigación para determinar los procesos más esenciales involucrados en su desarrollo.

1.3. La biogénesis del cloroplasto.

En términos generales, el desarrollo o biogénesis de los cloroplastos es un proceso que implica una serie de cambios estructurales y funcionales en los proplástidos, los cuales culminan en la generación de cloroplastos fotosintéticamente activos. A pesar de que la definición es muy directa, en realidad se trata de un proceso complejo en el que intervienen varios mecanismos específicos que se regulan unos a otros, y cuyos componentes moleculares aún son, en su mayoría, desconocidos. Algunos de los mecanismos más importantes que median la biogénesis de los cloroplastos son: La señalización anterógrada, la señalización retrograda, el desarrollo de la membrana del tilacoide y el establecimiento de la maquinaria fotosintética [Waters y Langdale, 2009]. A continuación, se presenta una descripción general de cada uno de estos mecanismos.

1.3.1. La señalización anterógrada.

De manera general, se conocen como anterógradas a las señales derivadas del genoma nuclear que afectan la expresión de genes, el desarrollo o el funcionamiento de los cloroplastos. Debido a su origen endosimbiótico, los cloroplastos conservan un genoma de DNA que, aunque reducido, es completamente funcional [Martin *et al*, 1998]. Los genomas cloroplásticos son de aproximadamente 145 Kb y codifican para un promedio de 128 genes, entre los que se encuentran solamente 84 proteínas [Sato *et al*, 1999]. En este sentido, es importante mencionar que los cloroplastos tienen proteomas compuestos por hasta 3, 500 proteínas diferentes [van Wijk y Baginsky, 2011; Lande *et al*, 2017], las cuales están, en su mayoría, codificadas en el genoma nuclear; estos genes son transcritos en el núcleo y traducidos en el citoplasma, mientras que las proteínas resultantes son importadas al cloroplasto a través de mecanismos postraduccionales que las mueven a través de las membranas de la envoltura del cloroplasto. Puesto que la mayor parte de las proteínas esenciales para todos los procesos que se llevan a cabo dentro del cloroplasto son importadas por medio de mecanismos postraduccionales, resulta claro que tanto el desarrollo como el funcionamiento de los cloroplastos dependen por completo de estos mecanismos.

En la actualidad se conocen varios mecanismos implicados en el importe postraduccional de proteínas al cloroplasto, si bien el mecanismo general de la operación de estos mecanismos ya está determinado, muchos detalles aún son desconocidos. A continuación, se describen brevemente las características más importantes de los mecanismos de importe de proteínas al cloroplasto.

1.3.1.1. Vía canónica de importe de proteínas al cloroplasto: el sistema Toc/Tic.

La vía canónica de importe de proteínas al cloroplasto involucra el transporte de las proteínas blanco a través de la envoltura del organelo, utilizando un complejo de translocadores denominados translocón de la membrana externa (Toc) y translocón de la membrana interna (Tic) del cloroplasto [Leister, 2003; Sjuts *et al*, 2017].

Como su nombre lo indica, el complejo Toc se encarga de regular el paso de las preproteínas a través de la membrana externa de la envoltura del cloroplasto. Uno de los componentes principales de este complejo es Toc75, una proteína con dominios transmembranales que forma el poro por el cual las preproteínas atraviesan la membrana [Schnell *et al*, 1994; Hinnah *et al*, 2002]. El gen *Toc75* está conservado en todos los linajes vegetales y su mutación en estado homocigoto causa letalidad embrionaria, dejando de manifiesto la gran relevancia de la función de esta proteína [Baldwin *et al*, 2005]. Además, se sabe que algunos componentes del Toc son capaces de dar especificidad al transporte de proteínas. Por ejemplo, la subunidad Toc159, implicada en el reconocimiento de las preproteínas, está codificada por cuatro genes en arabidopsis *AtToc159, AtToc132, AtToc120* y *AtToc90* [Sjuts *et al*, 2017]. La mutante de arabidopsis en uno de estos genes (*AtToc159*) es albina y no sobrevive para generar hojas verdaderas [Bauer *et al*, 2000]. Esto indica que las proteínas Toc159 no son redundantes, pues no pueden compensar la mutación en *AtToc159* [Kubis *et al*, 2004]. De manera similar, la sobre-expresión de *AtToc159* no es capaz de rescatar el fenotipo de coloración verde pálida mostrada por las mutaciones en *AtToc132* y *AtToc120* [Kubis *et al*, 2004]. En conjunto, estas observaciones indican que las distintas isoformas de Toc159 son capaces de reconocer a distintas poblaciones de proteínas blanco.

Por otro lado, en la membrana interna del cloroplasto se encuentran los componentes del complejo Tic, los cuales están encargados de facilitar el importe de las preproteínas que emergen de la membrana externa a través del Toc [Schnell y Blobel, 1993]. La evidencia más reciente sugiere que existen dos tipos de complejos Tic, el complejo que contiene a la proteína Tic110 y el que contiene a Tic20; ambas proteínas tienen la capacidad de formar poros en la membrana, por lo que se supone que son el componente principal de sus respectivos complejos de importe [Heins *et al*, 2002; Balsera *et al*, 2009; Kovács-Bogdán *et al*, 2011]. Tanto Tic110 como Tic20 han sido purificadas en complejos proteínicos que incluyen a otras proteínas Tic, sin embargo, nunca han sido encontradas formando parte de los mismos complejos macromoleculares [Inaba *et* al, 2003; Kikuchi *et al*, 2009; Kikuchi *et al*, 2013]. Además, las mutantes hipomórficas en arabidopsis de los genes *AtTic110* y *AtTic20* tienen defectos en la fotomorfogénesis, mientras que las mutantes nulas son letales [Chen *et al*, 2002; Kovacheva *et al*, 2005]. Estos datos indican que tanto AtTic110 como AtTic20 son componentes no redundantes y esenciales en el importe de proteínas al cloroplasto, lo cual apoya la hipótesis de que el importe de proteínas al cloroplasto está mediado por dos complejos Tic independientes. Sin embargo, aún se desconocen los detalles que diferencian a estos dos complejos a nivel funcional.

Además de los complejos Toc y Tic, en el estroma de los cloroplastos existe un sistema diverso de chaperoninas, chaperonas y cochaperonas que están encargadas de mediar el importe, el plegamiento y el procesamiento de los péptidos importados. Por ejemplo, las chaperonas Hsp90C, Hsp93/ClpC y pcHsc70 son consideradas el motor que impulsa el importe de proteínas, puesto que se unen a las preproteínas en tránsito y promueven

el importe mediante la hidrólisis de ATP [Sjuts *et al*, 2017]. Cabe mencionar que la inhibición de la actividad ATPasa de Hsp90C resulta en defectos en el importe de proteínas, sin embargo, bajo estas condiciones la interacción de las preproteínas con los componentes del Toc y del Tic no se ve afectada, indicando que la hidrolisis de ATP en el estroma es un paso esencial e independiente en el importe de proteínas al cloroplasto [Inoue *et al*, 2013]. Además, las mutantes de arabidopsis afectadas en los genes que codifican para estas chaperonas tienen defectos importantes en el importe de proteínas y el desarrollo de los cloroplastos. Por ejemplo, las mutantes sencillas en los genes *AtcpHsc70-1* y *AtcpHsc70-2* tienen cloroplastos con una capacidad reducida para importar proteínas al cloroplasto, generando plantas con fenotipos de cotiledones y hojas variegadas a causa de defectos severos en el desarrollo de los cloroplastos; mientras que las dobles mutantes en estos dos genes son letales [Su y Li, 2008; Su y Li, 2010]. En otros sistemas biológicos como bacterias y levaduras, se sabe que las chaperonas de la familia Hsp70 funcionan junto con cochaperonas, las cuales son esenciales para que las chaperonas lleven a cabo su función de manera eficiente [Liberek *et* al, 1991; Fink, 1999; Mayer, 2010]. En particular, las cochaperonas de la familia DnaJ están encargadas de reconocer a las proteínas blanco de las chaperonas Hsp70, mientras que las cochaperonas de la familia GrpE son intercambiadores de nucleótidos que regulan el intercambio de ADP por ATP en el sitio activo de las proteínas Hsp70 [Schröder *et al*, 1993; Suh *et* al, 1999; Brehmer *et* al, 2004; Sugimoto *et al*, 2008; Mayer, 2010;]. En los linajes vegetales existen homólogos tanto de las proteínas DnaJ como de las GrpEs. En los cloroplastos de las plantas superiores, las DnaJs forman una familia de 19 miembros con funciones diversas, como el ensamblaje de los fotosistemas y la oligomerización de

proteínas clave para la formación de las membranas de los tilacoides [Liu *et al*, 2005; Chen *et* al, 2010; Chiu *et al*, 2013]. Por otro lado, la función de las GrpEs en plantas superiores aun no es muy clara; como se discutirá más adelante, los únicos trabajos que exploran el papel de las GrpEs en los cloroplastos están enfocados en algas y plantas basales, y no existen estudios que aborden su función en el desarrollo y funcionamiento de los cloroplastos de plantas superiores. Cabe mencionar, que la caracterización de estas proteínas en arabidopsis es una de las partes importantes de este trabajo y los resultados serán discutidos a detalle más adelante.

De igual importancia en el importe de proteínas a los plástidos, son los péptidos de tránsito. Estos elementos presentes en las preproteínas, son secuencias N-terminales que contienen los motivos necesarios para interaccionar con el sistema de translocación, y median la entrada al cloroplasto. A pesar de los esfuerzos que se han dirigido a la identificación de secuencias consenso que permitan definir a los péptidos de tránsito, poco avance se ha logrado debido a que los péptidos de tránsito identificados experimentalmente comparten muy poca identidad. Hoy en día se reconocen algunas características importantes que comparten los péptidos de tránsito a cloroplasto, por ejemplo, un alto contenido de aminoácidos hidroxilados, carencia de aminoácidos ácidos, carencia de estructuras secundarias y la presencia de residuos de prolina en algunas posiciones a lo largo del péptido [von Heijne et al, 1989; von Heijne y Nishikawa, 1991; Lee *et al*, 2018]. A pesar de estas limitaciones, se han desarrollado algoritmos como ChloroP 1.1 (http://www.cbs.dtu.dk/services/ChloroP/), que permiten una buena predicción de la existencia de péptidos de tránsito en las regiones N-terminales de las proteínas. Por otro lado, algunos estudios han identificado a 7 grupos de motivos de

secuencia que son recurrentes en los péptidos de transito de las proteínas cloroplásticas. En estos experimentos se encontró que estos motivos son indispensables para el importe de las proteínas que los portan, y se determinó que hasta el 51% de las proteínas analizadas tienen los motivos característicos de al menos uno de los 7 grupos [Lee *et al*, 2008]. Sin embargo, el hecho de que sólo una fracción de los péptidos de tránsito analizados contuviera a alguno de estos grupos de motivos, indica que existen mecanismos adicionales de importe de proteínas al cloroplasto que usan este tipo de péptidos de transito con motivos desconocidos.

1.3.1.2. Vías no canónicas de importe de proteínas al cloroplasto.

En la actualidad se sabe que no todas las proteínas cloroplásticas interaccionan con el sistema Toc/Tic del modo canónico descrito en la sección anterior, sino que utilizan alguna de las llamadas vías no canónicas para llegar al cloroplasto. Las vías no canónicas se han denominado así debido a que los ejemplos disponibles de proteínas importadas a través de estas vías son muy limitados; de hecho, se estima que solamente el 11.4% de las proteínas cloroplásticas utilizan alguna de estas vías para entrar al cloroplasto [Armbruster *et al*, 2009]. Por ejemplo, las proteínas de la membrana externa de la envoltura no tienen péptidos de transito canónicos; sin embargo, interaccionan con las proteínas citoplásmicas AKR2 y HSP17.8, las cuales las dirigen hacia la membrana externa, donde, a pesar de no tener péptidos de tránsito, interaccionan con Toc75 para mediar su inserción en la membrana sin tener que interaccionar con los componentes del Tic [Hofmann y Theg, 2005; Lee *et al*, 2013]. Además, algunas proteínas de la membrana externa tienen péptidos de tránsito similares a los que llevan a otras proteínas al estroma del cloroplasto, por lo que inician su tránsito a través del complejo Toc, pero

una vez dentro del poro del translocón, otras señales en su estructura primaria hacen que se desasocien del Toc y se integren a la membrana. Un ejemplo de este tipo de transporte es TOC75 [Inoue y Keegstra, 2003]. Un mecanismo similar se ha descrito para el importe de proteínas a la membrana interna de la envoltura del cloroplasto, pero en este mecanismo, las proteínas atraviesan todo el complejo Toc y se integran a la membrana interna mientras atraviesan el Tic [Viana *et al*, 2010].

Por otro lado, se conoce un mecanismo que utiliza los componentes de la vía secretoria para movilizar proteínas sintetizadas en el citoplasma hasta los cloroplastos. La hipótesis general en torno a este mecanismo indica que las proteínas son sintetizadas en el retículo endoplásmico (ER) y transportadas al aparato de Golgi (Golgi), donde son glicosiladas y luego transportadas en vesículas hasta el cloroplasto. En la actualidad, los ejemplos de proteínas transportadas por esta vía son escasos: solo se sabe que la proteína LHCPII de Euglenas [Sulli *et al*, 1999], tres amilasas de arroz [Chen *et al*, 2004; Asatsuma *et al*, 2005] y una anhidrasa carbónica de arabidopsis [Villarejo *et al*, 2005] utilizan la vía secretoria para llegar a los cloroplastos. Además, la información disponible sobre los detalles celulares y moleculares que subyacen y regulan esta vía es muy limitada. Por ejemplo, se sabe que la región N-terminal de la proteína LHCPII de Euglenas dirige la inserción de esta proteína en la membrana del ER de manera cotraduccional, y luego es transportada hacia Golgi y al cloroplasto en el lado citoplásmico de las vesículas secretorias [Sulli *et al*, 1999]. De manera similar, se sabe que la region N-terminal de las proteínas αAmy3 y αAmy8 es capaz de dirigir a otras proteínas al cloroplasto a través de la vía secretoria [Chen *et al*, 2004]. Estos datos indican que es posible que los motivos de proteína que median las interacciones de las preproteínas con los componentes de la

vía secretoria se encuentren en los péptidos de tránsito de dichas preproteínas. Hasta la fecha, no se han logrado identificar componentes moleculares que regulen de manera directa el tránsito de proteínas al cloroplasto a través de esta vía, ni en algas ni en plantas superiores. Sin embargo, algunas de las evidencias recabadas en nuestro trabajo, las cuales serán discutidas más adelante, sientan una base sólida para la identificación de un posible regulador de este proceso en arabidopsis.

Finalmente, además de la señalización anterógrada relacionada al importe de proteínas, se sabe que existen factores transcripcionales que regulan la expresión de genes cloroplásticos codificados en el núcleo. Por ejemplo, los factores transcripcionales GLK y GNC regulan la expresión de genes involucrados en la síntesis de clorofila y de proteínas de los fotosistemas, los cuales promueven el desarrollo de los cloroplastos [Zubo *et al*, 2018]. De este modo, es evidente que el núcleo ejerce una regulación directa sobre los mecanismos de desarrollo y funcionamiento de los cloroplastos a nivel transcripcional y postraduccional (importe de proteínas).

1.3.2. La señalización retrograda.

A pesar del estricto control que ejerce el núcleo sobre los cloroplastos, estos últimos también son capaces de modular la expresión de genes nucleares. En términos generales, la señalización retrógrada se trata de señales derivadas del plástido que afectan la transcripción de genes codificados en el núcleo, los cuales a su vez impactan diversos procesos celulares e inclusive pueden afectar el estado funcional o de desarrollo de los cloroplastos. Por ejemplo, cuando plántulas son tratadas con compuestos que son inhibidores directos del metabolismo cloroplástico, los niveles de expresión de genes

cloroplásticos codificados en el núcleo disminuyen, esto implica la existencia de mecanismos que detectan el estado del cloroplasto y emiten señales capaces de impactar la expresión de genes nucleares [Waters y Langdale, 2009].

La señalización retrograda se divide en dos categorías conceptuales que corresponden a los procesos que dichas señales controlan. A continuación, se describen brevemente ambas categorías.

1.3.2.1. Control Biogénico.

El control biogénico incluye a los mecanismos de señalización retrograda que operan durante el desarrollo de los cloroplastos, es decir durante las etapas de transición entre proplástido y cloroplasto [Jarvis y López-Juez, 2013]. Las señales biogénicas están encargadas de regular los niveles de expresión de los genes cloroplásticos que codifican a los componentes del aparato fotosintético. En la actualidad, se han descrito dos tipos principales de señales biogénicas: las derivadas de la síntesis de tetrapirroles y las generadas por la expresión de genes en el cloroplasto [Hernández-Verdeja y Strand, 2018]. Se sabe que la síntesis de tetrapirroles en el cloroplasto es una fuente de señales retrogradas, puesto que mutantes afectadas en esta vía presentan niveles alterados de expresión de genes nucleares asociados a la fotosíntesis (PhANGs). Por ejemplo, en plantas de tipo silvestre la inhibición química del desarrollo del cloroplasto genera una disminución en la transcripción de *LHCB1*, un PhANG que es un componente importante en los complejos captadores de luz (LHC) del fotosistema II (PSII). Por otro lado, en presencia del inhibidor, la disminución en la transcripción de *LHCB1* no sucede en plantas mutantes afectadas en la síntesis de tretrapirroles (gun2-gun6) [Mochizuki *et al*,

2001; Woodson *et al*, 2011]. Estas evidencias indican que la vía de síntesis de tetrapirroles, es una fuente de señales que regulan negativamente la expresión de los PhANGs cuando hay defectos en el desarrollo del cloroplasto. De manera similar, la inhibición química y las mutaciones en los componentes de la transcripción y la traducción plastídica, afectan la expresión de los PhANGs, indicando que los mecanismos de expresión de genes cloroplásticos también son una fuente de señales biogénicas [Sullivan y Gray, 1999; Woodson *et al*, 2013; Chan *et al*, 2016]. A pesar de que se conoce la existencia de estas dos fuentes de señales biogénicas, aún no se conoce la identidad exacta de todas las moléculas señalizadoras ni de los componentes de las cascadas de transducción involucradas. En la actualidad se conocen solo dos moléculas que funcionan como señales biogénicas, la Mg-Protoporfirina-IX (MgProtoIX) y el hemo [Hernández-Verdeja y Strand, 2018]. Ambas moléculas son derivadas de la síntesis de tetrapirroles pero regulan de manera distinta la expresión de los PhANGs: MgProtoIX funciona como un regulador negativo, mientras que el hemo es un regulador positivo [Woodson *et al*, 2011; Barajas-López *et al*, 2013]. En conjunto, se cree que ambas moléculas están encargadas de regular muy finamente la expresión de los PhANGs durante el desarrollo del cloroplasto.

1.3.2.2. Control operacional.

El control operacional se refiere a la señalización que se establece entre los cloroplastos maduros y funcionales y el genoma nuclear [Jarvis y López-Juez, 2013]. En general, se acepta que las señales operacionales están encargadas de regular la actividad cloroplástica en respuesta a condiciones de estrés que afectan el funcionamiento de los cloroplastos. Se conocen dos tipos principales de señales operacionales: las derivadas de la cadena de transporte de electrones (CTE) y especies reactivas del oxígeno (ROS) y las derivadas de vías metabólicas [Hernández-Verdeja y Strand, 2018]. Se sabe que los desbalances en la CTE debidos a cambios ambientales, como fluctuaciones en la intensidad de la luz, pueden producir varios tipos de ROS como $^1O_2$, O_2^- y H_2O_2, los cuales son capaces de afectar la expresión de genes nucleares [de Souza *et al*, 2017]. Por ejemplo, el $^1O_2$ producido en condiciones de intensidad luminosa alta es capaz de inducir la trascripción de genes que codifican para proteínas antioxidantes, y también inicia una respuesta de muerte celular [Page *et al*, 2017]. Se sabe que ambas respuestas están mediadas por las proteínas cloroplásticas EXECUTER 1 y 2, y que la respuesta de muerte es independiente de la toxicidad del $^1O_2$, puesto que las mutantes para estas proteínas no muestran ninguna de estas respuestas [Lee *et al*, 2007]. De manera similar, el H_2O_2 promueve la síntesis de ácido abscísico, que a su vez induce la transcripción de una peroxidasa cloroplástica codificada en el núcleo que se encarga de eliminar el H_2O_2 [Galvez-Valdivieso *et al*, 2009].

Por otro lado, se han identificado algunos metabolitos que funcionan como señales operacionales que se producen a partir de algunas vías metabólicas en el cloroplasto, como el fosfoadenosina 5´-fosfato (PAP), β-ciclocitral (β-cyc) y metil-D-eritrol 2,4-

ciclodifosfato (MEcPP). Al igual que las señales de ROS, se cree que estos metabolitos están encargados de mediar respuestas a estrés [Hernández-Verdeja y Strand, 2018]. Por ejemplo, el PAP se acumula en plantas expuestas a sequía, y es capaz de mediar la activación de la transcripción de genes nucleares relacionados con las respuestas ante estrés abiótico [Estavillo *et al*, 2011]. De manera similar, el estrés foto oxidativo genera la acumulación de β-cyc a partir de β-caroteno en el cloroplasto, el cual aumenta la síntesis de la hormona vegetal ácido salicílico para impactar la transcripción de genes que eliminan las ROS generadas por el daño foto oxidativo [Lv *et al*, 2015]. Finalmente, la producción de MEcPP se ha relacionado a señales abióticas como heridas, aumento en la intensidad luminosa y estrés oxidativo; además, se sabe que el MEcPP no solo modifica la expresión de genes de respuesta a estrés en plantas y bacterias, sino que es capaz de alterar las interacciones entre el DNA y las histonas en Chlamydia [Ostrovsky *et al*, 1998; Xiao *et al*, 2012; Chan *et al*, 2016]. Esta evidencia sugiere que este metabolito podría ser no solo relevante en modelos vegetales, sino que es parte de un mecanismo ancestral de respuesta a estrés abiótico que está conservado desde bacterias. Cabe mencionar que, hasta la fecha, no hay evidencia clara que indique cuales son los transductores de señales involucrados en transducir las señales de estos metabolitos a través del citoplasma hasta el núcleo para ejercer su efecto.

En conjunto, toda la información presentada en las secciones 1.3.1 y 1.3.2 deja de manifiesto que los cloroplastos no son organelos aislados o simples fábricas de compuestos, sino que existe una vasta y compleja red de comunicación que mantiene sus actividades coordinadas con las del resto de la célula vegetal, reforzando la noción actual que propone que los cloroplastos son centros de integración de señales

intracelulares y extracelulares, que son clave para regular las respuestas celulares y fisiológicas ante el estrés.

1.3.3. La biogénesis de los tilacoides.

Uno de los eventos más notables durante la biogénesis de los cloroplastos, es el proceso mediante el cual se genera y organiza la membrana de los tilacoides. Los tilacoides son una red organizada de cúmulos (grana) y lamelas que se extienden dentro del estroma, y que forman un compartimento llamado lumen del tilacoide. Tanto la membrana de los tilacoides como el lumen son esenciales para establecer la compartimentalización necesaria para que la fotosíntesis se lleve a cabo de manera eficiente. En la membrana de los tilacoides se encuentran embebidos los fotosistemas I y II (PSI y PSII), que son los complejos macromoleculares encargados de realizar las reacciones luminosas de la fotosíntesis [Koochak *et al*, 2018]. A nivel celular, la red de lamelas y grana que forman los tilacoides es la característica distintiva de los cloroplastos maduros, y es justamente la presencia y el estado de esta red una de las características utilizadas por los investigadores para determinar el grado de madurez de un cloroplasto. Los proplástidos, que son los plástidos indiferenciados precursores de los cloroplastos, contienen una red simple de membranas internas llamadas protilacoides, los cuales son el punto de partida para la generación de los tilacoides [Waters y Langdale, 2009].

Hay tres puntos clave en la formación de los tilacoides:

- El importe de lípidos. Para crecer, la membrana del tilacoide necesita de lípidos, estos son sintetizados en unos subcompartimentos conocidos como plastoglóbulos, que están adheridos a la membrana del protilacoide [Kelly y

Dörmann, 2004; Kobayashi *et al*, 2007; Bréhélin y Kessler, 2008]. Además, los lípidos también se transportan hacia los protilacoides en vesículas que emergen de la membrana interna de la envoltura del cloroplasto. Algunos componentes moleculares de este sistema de tráfico vesicular se han comenzado a identificar en los últimos años, por ejemplo, las proteínas VIPP1 y THF1, cuyas mutantes presentan fenotipos de desarrollo deficiente o ausencia de los tilacoides, respectivamente [Kroll *et al*, 2001; Wang *et al*, 2004].

- El importe de proteínas. Las proteínas que están en la membrana y el lumen de los tilacoides son responsables de las funciones de estos. Se conocen cuatro mecanismos que están encargados de llevar proteínas desde el estroma hasta los tilacoides, la evidencia experimental indica que estos mecanismos no son redundantes, es decir, que cada uno transporta a una subpoblación especifica de proteínas del tilacoide [Jarvis y López-Juez, 2013]. Adicionalmente, algunas evidencias sugieren que algunos de los componentes de la membrana del tilacoide, como proteínas de los LHCs, enzimas para la síntesis de pigmentos y moléculas de clorofila, se mueven en las vesículas que viajan desde la membrana interna del cloroplasto hasta los protilacoides [Morré *et al*, 1991; Shimoni *et al*, 2005].

- Arquitectura de la red. Como se mencionó, las grana y las lamelas que forman los tilacoides maduros en los cloroplastos, son una de las características que definen a estos últimos. En la actualidad, se han descrito algunos factores moleculares que son importantes en este proceso. Por ejemplo, la GTPasa FZL, que se localiza en la membrana interna del cloroplasto y en las membranas de los tilacoides.

Aunque las plantas mutantes *fzl* no son deficientes en la generación de tilacoides, los cloroplastos contienen proporciones inusuales de grana y lamelas, además de que acumulan vesículas en el estroma, lo que indica que FZL es importante en el mantenimiento y remodelación de la arquitectura de la red de tilacoides [Gao *et al*, 2006].

1.3.4. La fotosíntesis.

Como se mencionó, la fotosíntesis es la característica definitiva de los cloroplastos, se trata de un proceso tan vital para el desarrollo vegetal y de toda la vida en la biósfera, que es y ha sido ampliamente estudiado durante décadas. No es la intención de esta sección hacer una descripción detallada de todos los procesos bioquímicos y moleculares que constituyen y regulan a la fotosíntesis, sin embargo, se abordarán algunos puntos importantes que sentarán la base de la discusión de los resultados descritos más adelante.

La fotosíntesis es un sistema complejo que consiste en varios procesos interdependientes que están separados espacialmente gracias a la compartimentalización que existe dentro de los cloroplastos. Por un lado, las reacciones oscuras o de fijación de carbono se llevan a cabo en el estroma, puesto que ahí es donde se localizan las enzimas necesarias para este proceso: como la RuBisCO (Ribulosa Bisfosfato Carboxilasa/Oxigenasa). Por otra parte, la luz es capturada en los LHCs del PSI y del PSII, los cuales transmiten la energía capturada hasta el centro de reacción, donde se llevan a cabo las reacciones fotoquímicas que promueven la formación de un

potencial electroquímico a través de la membrana del tilacoide, la generación de poder reductor y, en el caso del PSII, la generación de O_2 [Liu et al, 2019].

Los fotosistemas y sus LHCs son complejos macromoleculares formados por una gran cantidad de subunidades individuales que están codificadas tanto en el genoma nuclear como en el cloroplástico. Por ejemplo, el PSII es un complejo homodimérico, y cada uno de los monómeros está formado a su vez por 20 subunidades independientes y una gran cantidad de moléculas de clorofila [Loll *et al*, 2005]. De manera similar, existen dos tipos de LHCIIs, el LHCII principal que es un trímero de proteínas independientes de la familia *Lhc*, mientras que los LHCIIs secundarios son monómeros de otras proteínas de la misma familia [Wei *et al*, 2016; Gao *et al*, 2018]. Cuando los dímeros del PSII interaccionan con los LHCs principales o secundarios, se forma un supercomplejo. Cuando los supercomplejos interaccionan de manera lateral en la membrana del tilacoide, se forman megacomplejos que además pueden formar arreglos semicristalinos cuando interaccionan con los megacomplejos embebidos en la membrana de los tilacoides adyacentes [Kouřil et al, 2012; Theis y Schroda, 2016]. Debido a que la energía luminosa absorbida por los pigmentos de los LHCs se mueve hasta los centros de reacción por resonancia de Förster (FRET), resulta evidente que la estequiometría de los fotosistemas, así como su organización en complejos, supercomplejos, megacomplejos y arreglos semicristalinos es esencial para su función.

Debido a su alta complejidad, se sabe que hay sistemas de chaperonas moleculares encargados de ensamblar los fotosistemas y mantener su integridad. Por ejemplo, se sabe que la chaperona cpHsc70-1 está involucrada en el ensamblaje del fotosistema II durante la biogénesis del cloroplasto y que también participa en la reparación del centro

de reacción cuando este sufre daño foto oxidativo [Schroda *et al*, 1999; Yokthongwattana *et al*, 2001]. De manera similar, se ha propuesto que algunas co-chaperonas de la familia DnaJ participan en la homeostasis de los LHCs en plantas [Chen *et al*, 2010].

A pesar de todo el conocimiento acumulado en torno a los mecanismos biológicos que modulan el desarrollo de los cloroplastos, aún existen muchas preguntas sin responder como: ¿cuál es la naturaleza de los transductores y de las señales tempranas que inician el proceso de desarrollo de los proplástidos?, ¿qué otros mecanismos celulares de importe de proteínas al cloroplasto existen?, ¿son los plástidos inmaduros capaces de modular el desarrollo vegetal?, etc.

2. Antecedentes.

En este contexto, nuestro grupo de investigación está interesado en esclarecer los mecanismos moleculares que regulan el desarrollo del cloroplasto. Con ese objetivo en mente, en el año 2004 se compiló una colección de mutantes afectadas en la pigmentación denominada *clb* (chloroplast biogenesis), la cual incluye plantas de color verde pálido, amarillas y albinas [Gutiérrez-Nava *et al*, 2004]. Estas plantas muestran diferentes niveles de afectación en el desarrollo del cloroplasto, y a través de su caracterización molecular y funcional, se ha logrado la identificación de genes implicados en este importante proceso de diferenciación [Guevara-García *et al*, 2005; Chateigner-Boutin *et al*, 2008].

Con la intención de profundizar en el conocimiento de los factores involucrados en la biogénesis de los cloroplastos, se abordó una estrategia de proteómica comparativa utilizando algunas líneas de la colección *clb* [de Luna-Valdez, 2012]. En resumen, se determinaron y compararon los perfiles de expresión de proteínas en geles de acrilamida de dos dimensiones, de plantas tipo silvestre de 8 días de desarrollo y de las líneas mutantes *cla1* (cloroplastos alterados), *clb5* y *clb19* de 16 días de crecimiento. A partir de los resultados de este análisis, se seleccionaron 26 proteínas posiblemente involucradas en la biogénesis de los cloroplastos, entre las cuales se encontraron chaperonas moleculares, proteínas de unión a RNA, enzimas involucradas en la eliminación de ROS, proteínas involucradas en el importe de proteínas al cloroplasto y proteasas cloroplásticas [de Luna-Valdez, 2012]. Posteriormente, se adquirieron y analizaron los fenotipos de líneas mutantes para las 26 proteínas identificadas en busca de evidencias de defectos en el desarrollo de los cloroplastos, en particular, deficiencias

en la pigmentación. Estos estudios fenotípicos resultaron en la identificación de tres líneas mutantes que exhiben claramente fenotipos de defectos en pigmentación y desarrollo (Figura 3) [de Luna-Valdez, 2012], en cuya caracterización se enfoca la propuesta principal de este proyecto.

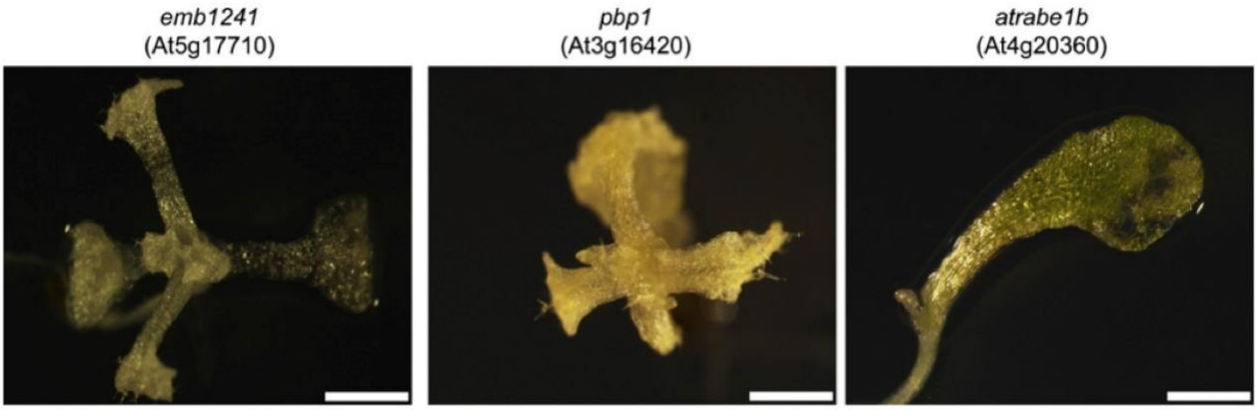


Figura 3. Fenotipo de las mutantes *atrabe1b*, *emb1241* y *pbp1*. Fenotipo de líneas mutantes para genes que codifican para las proteínas posiblemente involucradas en la biogénesis del cloroplasto. Se muestran plántulas de 16 días de desarrollo de las tres líneas mutantes *atrabe1b*, *emb1241* y *pbp1*. Escala: 10 µm.

Además, se realizaron algunos análisis *in silico* en busca de evidencias que apoyaran la participación de las tres proteínas afectadas en las líneas mutantes referidas. Particularmente, se identificaron interactomas parciales para cada una de las tres proteínas candidatas y su posible función fue inferida con base en su anotación en las bases de datos TAIR (http://www.arabidopsis.org) y UniProt (http://www.uniprot.org), las cuales recopilan evidencias bioinformáticas y experimentales. A continuación, se ofrece una breve descripción de estas proteínas, las cuales son el objeto de estudio de este trabajo.

Proteína EMB1241.

La proteína EMB1241 contiene un dominio de aproximadamente 180 aminoácidos conocido como GrpE (http://smart.embl.de/smart/show_motifs.pl?GENOMIC=1&DO_PFAM=DO_PFAM&INCLUDE_SIGNALP=INCLUDE_SIGNALP&ID=3702.AT5G17710.2-P), que en *E.coli* ha sido caracterizado como una proteína co-chaperona con función de intercambiador de nucleótidos para la chaperona HSP70, la cual está involucrada en las respuestas a estrés [Mayer *et al*, 2010]. Las proteínas pertenecientes a la familia GrpE están distribuidas a lo largo de toda la filogenia, en plantas existen 485 secuencias con homología a la proteína GrpE de *E. coli*, las cuales se encuentran distribuidas en 85 especies vegetales (http://pfam.sanger.ac.uk/family/PF01025). El genoma de *Arabidopsis thaliana* contiene cuatro genes que codifican para proteínas con el dominio GrpE, dos de las cuales están predichas para localizarse en las mitocondrias y dos en los cloroplastos.

Por otro lado, utilizando la herramienta de búsqueda de interacciones proteína-proteína STRING (http://string-db.org/), se detectaron 10 proteínas posiblemente interactoras de EMB1241 [de Luna-Valdez, 2012], entre las cuales destacan las proteínas identificadas como At4g39960, At2g22360, que están anotadas como proteínas co-chaperonas DnaJ, las cuales trabajan en conjunto con las proteínas GrpE en algunos procesos celulares en *E.coli* [Mayer *et al*, 2010]. Además, resulta especialmente interesante que entre los interactores de EMB1241, se detectaron las proteínas cpHsc70-1 y cpHsc70-2, que son conocidas chaperonas moleculares cloroplásticas que están involucradas en el plegamiento y en la translocación al plástido de proteínas indispensables para la función de este, siendo parte esencial del sistema Toc/Tic. Estas observaciones sugieren que la

proteína EMB1241 podría formar parte del sistema canónico de importe de proteínas, facilitando la función de las chaperonas Hsp70 gracias a su función de intercambiador de nucleótidos. Adicionalmente, esta hipótesis podría explicar el fenotipo albino de las plantas mutantes para la proteína EMB1241, puesto que las plantas afectadas en la expresión de las proteínas cpHsc70-1 y cpHsc70-2 muestran fenotipos de deficiencia en la pigmentación [Su y Li, 2008; Su y Li, 2010].

Cabe resaltar que no existe información sobre la función de las GrpEs cloroplásticas (incluida EMB1241) de las plantas superiores, por lo que resulta interesante indagar en el papel que podría llevar a cabo en los cloroplastos de estos organismos.

Proteína PBP1.

La proteína PYK10-binding protein 1 (PBP1) es una proteína de 298 aminoácidos que consiste en dos dominios repetidos de lectinas de unión a manosa del tipo Jacalina [Barre *et al*, 2001]. En arabidopsis, PBP1 interacciona directamente con la enzima β-glucosidasa PYK10, la cual es el constituyente principal de los cuerpos ER que se forman en el retículo endoplásmico, cataliza la hidrólisis de compuestos con enlaces O-glucosídicos y está involucrada en el establecimiento de las relaciones mutualistas entre arabidopsis y el hongo endofítico *Piriformospora indica* [Nagano *et al*, 2005; Sherameti *et al*, 2018]. Existen datos que indican que PBP1 es capaz de mejorar la actividad hidrolítica de PYK10 y que podría actuar como chaperona molecular durante el plegamiento de esta enzima en respuesta a estrés biótico [Nagano *et al*, 2005].

Los datos encontrados acerca de la función de la proteína PBP1 no nos permiten generar hipótesis directas sobre la posible relación existente entre esta proteína y el desarrollo de los cloroplastos. Sin embargo, dado el fenotipo albino de la mutante *pbp1* (Figura 3) y tomando en consideración que PBP1 es una proteína residente del ER, podemos hipotetizar que PBP1 regula algún proceso en el ER cuya afectación impacta negativamente el desarrollo de los cloroplastos. Como se mencionó en la sección 1.3.1, existe un mecanismo celular poco descrito que utiliza componentes de la vía secretoría, como el ER, para llevar proteínas citoplásmicas al cloroplasto; en este contexto, y dada su descripción como posible chaperona [Nagano *et al*, 2005], cabe la posibilidad de que PBP1 sea uno de los componentes moleculares que regulan esta vía de importe de proteínas al cloroplasto.

Proteína ATRABE1B.

La Rab GTPasa E1B de arabidopsis (ATRABE1B) es un factor de elongación tipo Tu que actúa en la traducción de transcritos cloroplásticos. En las bacterias, estos factores median la unión de los aminoacil-tRNAs al sitio A de los ribosomas, en una reacción que depende de la hidrólisis de GTP [Pape *et al*, 1998]. En el genoma de arabidopsis existen otros siete factores de elongación tipo Tu, la mutación de unos de ellos: SVR3 (Supressor of Variegation 3), es capaz de suprimir el fenotipo de variegación provocado por la mutación de VAR2, una proteína involucrada en la división de los cloroplastos en arabidopsis [Liu *et al*, 2010]. Además, se sabe que la función de SVR3 es indispensable para el desarrollo de los cloroplastos cuando las plantas se exponen a estrés por frío [Liu *et al*, 2010]. Estos datos nos permiten hipotetizar que el papel de la proteína ATRABE1B en el desarrollo de los cloroplastos pudiera no estar vinculado con su función como factor de elongación, sino que, al igual que la proteína SVR3, podría estar involucrada en la división de los cloroplastos o inclusive la biogénesis de las membranas internas.

A pesar de las hipótesis formuladas sobre la posible participación de ATRABE1B en la división del plástido, la determinación *in silico* de proteínas interactoras, utilizando la base de datos STRING (http://string-db.org/) [de Luna-Valdez, 2012], no muestra evidencias que apoyen esta hipótesis, sino que sugieren que ATRABE1B funciona solamente como parte de los ribosomas cloroplásticos. Todas las proteínas identificadas como interactores de ATRABE1B forman parte de las distintas subunidades de los ribosomas 70S, de hecho, un análisis de enriquecimiento generado con la herramienta "GO-Enrichment" disponible en STRING, reveló que los términos ontológicos más enriquecidos entre esas proteínas son "síntesis de proteínas" (P value= 5.24×10^{-7}), como

proceso biológico, y "componente estructural de ribosomas" (P value= $2.15x10^{-8}$), tomando en cuenta su función [de Luna-Valdez, 2012]. Como se mencionó anteriormente, la maquinaria de expresión genética de los cloroplastos (incluyendo la traducción) modula el desarrollo de estos organelos a través de la producción de señales retrogradas biogénicas [Hernández-Verdeja y Strand, 2018]. Se ha reportado que la inhibición química, así como mutaciones en algunos componentes moleculares que median la traducción cloroplástica generan plantas con defectos en la pigmentación y en la morfología foliar [Sullivan y Gray, 1999; Woodson *et al*, 2013; Parker *et al*, 2014; Zoschke and Bock, 2018], esto es consistente con el fenotipo mostrado por las mutantes *atrabe1b* (Figura 3), reforzando la noción de que ATRABE1B está involucrada en la traducción de los transcritos cloroplásticos.

Actualmente no existe información exacta enfocada en la dilucidación de la relación que existe entre ATRABE1B y el desarrollo de los cloroplastos. Sin embargo, el fenotipo mostrado por las mutantes en esta proteína (Figura 3), es indicativo de defectos severos en el desarrollo de los cloroplastos y de la plántula, por lo que el estudio a profundidad de esta mutante resulta de gran interés para nuestro grupo de investigación.

3. Justificación.

A pesar de los diversos estudios que se han realizado en torno a la dilucidación de los mecanismos moleculares que subyacen al desarrollo de los cloroplastos, este proceso es aún poco comprendido. De tal manera, creemos que la caracterización detallada de los genes y proteínas afectados en líneas mutantes para las proteínas EMB1241, PBP1 y ATRABE1B, puede conducir a la identificación de componentes y procesos moleculares novedosos implicados en el desarrollo de este organelo.

4. Hipótesis.

Debido a los fenotipos de deficiencia en la pigmentación mostrados por las líneas mutantes *emb1241*, *pbp1* y *atrabe1b*, las proteínas EMB1241, PBP1 y ATRABE1B deben de estar involucradas en procesos indispensables para el desarrollo de los cloroplastos de *A. thaliana*.

5. Objetivos.

- **Objetivo general.**

  Analizar la posible participación de las proteínas EMB1241, PBP1 y ATRABE1B en el proceso de desarrollo de los cloroplastos de *A. thaliana*.

- **Objetivos particulares.**

  1. Caracterizar a nivel fenotípico las líneas mutantes *emb1241*, *pbp1* y *atrabe1b*.
  2. Caracterizar a nivel molecular las líneas mutantes *emb1241*, *pbp1* y *atrabe1b*.
  3. Caracterizar a nivel funcional a la proteína PBP1.
  4. Caracterizar a nivel funcional a la proteína EMB1241.

6. Materiales y Métodos.

6.1. Material vegetal y condiciones de crecimiento.

Para los ensayos de expresión transitoria de proteínas, se cultivaron plantas de *Nicotiana benthamiana* durante 6 semanas en substrato Metro-Mix 300 (Sun Gro Horticulture, USA) suplementado con un fertilizante de liberación prolongada (Osmocote Smart-release, The Scotts Miracle-Gro Company, USA), a 26 °C en un fotoperiodo de 16 h luz y 8 h oscuridad. Para los experimentos que requieren cultivo vegetal en condiciones asépticas, las semillas se esterilizaron superficialmente mediante su incubación en una solución NaClO al 1% (v/v), y luego se cultivaron *in vitro* en cajas de Petri con medio GM 1X [4.3 g/L de sales Murashige and Skoog con vitaminas B5 (Phytotechnology Laboratories, USA), 1% (w/v) de sacarosa y 0.8% (w/v) de phytoagar], manteniéndolas a 22 °C en un fotoperiodo de 16 h luz y 8 h oscuridad. Las líneas de inserción de T-DNA *atrabe1b* (SALK_032687C), *pbp1* (SAIL_680F04), *emb1241-1* (CS16149), *emb1241-2* (CS24098) y *Dcphsc70-1* (CS860808) se obtuvieron del Arabidopsis Biological Resource Center (www.arabidopsis.org) (Figura S3). Para los análisis de cuantificación de pigmentos y los ensayos de electroforesis no desnaturalizante en *Arabidopsis thaliana* que se describen más adelante, se escrutaron poblaciones de 30 plantas adultas (de 30 días de desarrollo) por PCR, para determinar la presencia del T-DNA en ellas; solo las plantas mutantes fueron utilizadas para los análisis.

6.2. Análisis *in silico*.

Las secuencias de las GrpE cloroplásticas se identificaron y descargaron mediante la herramienta de BLAST del repositorio UniProtKB (http://www.uniprot.org/help/uniprotkb),

utilizando la secuencia de proteína de la GrpE de *E. coli* como carnada para buscar en la base de datos Plants. En total se encontraron 136 proteínas CGE, identificadas por la presencia de peptidos de tránsito a cloroplasto. La predicción de los péptidos de tránsito se realizó con la aplicación TargetP 1.1 [Emanuelsson *et al*, 2000], y éstos fueron eliminados manualmente de las secuencias bajo análisis. El análisis involucró un total de 137 secuencias de amino ácidos (136 CGEs y la GrpE de *E. coli*) que fueron alineadas usando Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/), el alineamiento se procesó con las herramientas disponibles en el programa MEGA 7.0.18 [Kumar *et al*, 2016]. La historia evolutiva de las proteínas fue inferida usando el método de Máxima Verosimilitud basado en el modelo de matrices JTT [Jones *et al.*, 1992], se utilizó una distribución Gamma para modelar las diferencias en la tasa de evolución de las posiciones [5 categorias (+G, parameter = 1.3736)]. Se eliminaron del alineamiento todas las posiciones con menos del 70% de cobertura, es decir que solo se permitieron posiciones con menos del 30% de gaps, datos perdidos o residuos ambiguos. El árbol filogenético fue generado usando el programa vFigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/). Las CGEs se agruparon en secuencias Tipo A (48 secuencias) y Tipo B (78 secuencias) con base en la información del árbol filogenético. El análisis para identificar motivos conservados en las proteínas se realizó alimentando a la herramienta DREME del grupo MEME [Bailey, 2011] con las secuencias de CGEs Tipo A o Tipo B de manera independiente.

6.3. Clonación y construcciones génicas.

Las regiones codificantes (CDS) *PBP1* (At3g16420), *CAH1* (At3g52720), *CGE1* (At5g17710), *CGE2* (At1g36390) y *CPHSC70-1* (At4g24280) se amplificaron por PCR

usando oligonucleótidos específicos para obtener los CDS completos o para eliminarles el péptido de tránsito (Tabla S1). Los vectores de entrada para el sistema Gateway se generaron por recombinación en el vector donador pDONR^{TM}/Zeo o por clonación directa en el vector de entrada pENTR/D-TOPO®. Los vectores de expresión (Figura S2) se generaron por recombinación [LR clonasa II (Invitrogen, USA)] entre los vectores de entrada mencionados y los vectores de destino: pDEST14 (pD14), pEarleyGate103 (pEG), pSPYCE (pCE), and pSPYNE (pNE). Los vectores de expresión construidos usando los vectores pEG, pCE y pNE, asi como el marcador fluorescente de retículo endoplásmico [Nelson *et al*, 2007] se transfirieron a células de *Agrobacterium tumefaciens* C58C1, clonas individuales fueron usadas para agroinfiltrar las hojas de *N. benthamiana* para llevar a cabo lo ensayos de localización subcelular, suborganelar y de BiFC que serán descritos más adelante.

6.4. Ensayo de complementación genética en *E. coli*.

Con la intención de determinar la función de proteínas AtCGE1 y AtCGE2, la cepa de *E. coli* OD212 (*dnaK332 ΔgrpE::*Ω*-cam^{R}*) [Deloche *et al*, 1997] se transformó con los las construcciones *pD14::ΔCGE1, pD14::ΔCGE2* y *pD14::ΔccdB* (Figura S2), las células se cultivaron durante toda la noche a 25 °C en medio lisogénico (LB) sólido. Posteriormente, colonias individuales fueron cultivadas en medio LB líquido a 25 °C, 100 µL de estos cultivos fueron usados para inocular medio fresco, el cual se cultivó hasta una OD_{600nm} de 0.6. Se prepararon diluciones seriales y gotas de 5 µL se sembraron en cajas de Petri con medio LB sólido. Las cajas inoculadas se incubaron durante 20 h a 25 °C, 37 °C o 43 °C. Todo el medio LB utilizado fue suplementado con ampicilina a una concentración final de 100 µg/mL.

6.5. Localización subcelular, suborganelar, y Complementacion Bimolecular de la Fluorescencia (BiFC).

Los experimentos de localización subcelular y BiFC se llevaron a cabo en hojas de *N. benthamiana* agroinfiltradas. La localización subcelular se determinó utilizando fusiones traduccionales entre las proteínas de interés y GFP, mientras que los experimentos de BiFC se llevaron a cabo utilizando el sistema de YFP dividida [Walter *et al*, 2004] con los vectores de expresión pCE y pNE, que codifican para las mitades C-terminal y N-terminal de la YFP, respectivamente. La agroinfiltración de las hojas se hizo con clonas de *A. tumefasciens* transformadas individualmente con los plásmidos *pBN::GFP* [Voinnet *et al*, 2000], *pEG::PBP1*, *pEG::CGE1*, *pEG::CGE2*, o *pEG::CPHSC70-1* para los experimentos de localización subcelular, y con los plásmidos *pCE::CGE1, pCE::CGE2*, *pCE::CPHSC70-1*, *pNE::CGE1*, *pNE::CGE2,* o *pNE::CPHSC70-1* para los ensayos de BiFC. Todos los experimentos de agroinfiltración incluyeron a la cepa auxiliar p19 de *A. tumefasciens* [Leuzinger *et al*, 2013]. Las distintas cepas bacterianas utilizadas se cultivaron a 30 °C durante toda la noche, se cosecharon por centrifugación (1, 400 x *g*, 10 min) y se resuspendieron en una solución acuosa de 10 mM $MgCl_2$ hasta una OD_{600nm} de 1. Posteriormente y antes de inocular las plantas, las muestras de bacterias se incubaron a temperatura ambiente por 3 h en presencia de 5 µg/mL acetosyringona. Las hojas de plantas de *N. benthamiana* se infiltraron con los cultivos de bacterias y se cultivaron a 26 °C, mantenidas en un fotoperiodo de 16 h luz y 8 h oscuridad durante 96 h antes de ser analizadas. Los protoplastos de las hojas infiltradas se obtuvieron por digestión enzimática de la pared celular; las hojas se cortaron en tiras delgadas que se incubaron en una solución de enzimas digestivas [0.5 M manitol, 1% celulasa R-10, and 0.05% macerozima R-10] durante 3 h en la oscuridad, a temperatura ambiente y

agitación (30 rpm). Luego de la digestión, los restos de hojas fueron retirados y los protoplastos analizados. En caso de ser necesario, se usaron las hojas agroinfiltradas para purificar cloroplastos y sus compartimentos suborganelares según el método descrito por Salvi *et al*, 2011.

6.6. Microscopía confocal.

Las imágenes de microscopia confocal de los protoplastos de *N. benthamiana* se obtuvieron con un microscopio Olympus FV1000 (Olympus, USA), utilizando láseres de excitación de 488 nm para GFP y 515 nm para YFP. La fluorescencia de la clorofila se capturó utilizando un láser de 515 nm y un filtro BA655-755. Las proyecciones en Z se generaron utilizando el tipo de proyección de máxima intensidad en el programa Fiji [Schindelin *et al*, 2012].

6.7. Experimentos de estrés por calor y Northern blot.

Plántulas de 16 días de desarrollo de *A. thaliana* (Col-0) fueron sometidas a estrés por calor mediante incubaciones a 40 °C durante 30, 60 o 90 minutos. RNA total se preparó de las plántulas utilizando TRIzol (Ambion, Life Technologies, USA) y siguiendo el protocolo indicado por el fabricante. Para los análisis de Northern blot, 10 µg de RNA total se fraccionaron en condiciones desnaturalizantes [2% (v/v) formaldehido] en geles de agarosa al 1.5% (w/v), luego se transfirieron a membranas de nylon Hybond-N^+ (GE Healthcare Bio-Sciences, UK). Las sondas de DNA se aislaron de los CDS de cada gen (Figura S3), posteriormente fueron marcadas con la base radioactiva [α-^{32}P]-dCTP usando el sistema Megaprime DNA labeling (GE Healthcare Bio-Sciences, UK) según las indicaciones del fabricante. La hibridación y lavado de las membranas se realizaron

bajo las condiciones más astringentes de temperatura y fuerza iónica (55 °C y 0.0825 M Na^+).

6.8. Cuantificación de pigmentos.

Los pigmentos fotosintéticos de plantas adultas se extrajeron mediante incubación en una solución de acetona al 80% durante una noche. Posteriormente, la absorbancia de las muestras fue medida a 663nm, 646 nm y 470 nm. Utilizando los datos de absorbancia, la concentración de pigmentos fue calculada con las siguientes ecuaciones [Lichtenthaler and Wellburn, 1983]:

$$Clorofila\ a\ (Ca) = (12.21 \cdot A_{663}) - (2.81 \cdot A_{646})$$

$$Clorofila\ b\ (Cb) = (20.13 \cdot A_{646}) - (5.03 \cdot A_{663})$$

$$Carotenos = \frac{\big((1000 \cdot A_{470}) - (3.27 \cdot Ca) - (104 \cdot Cb)\big)}{229}$$

La significancia estadística de los datos se determinó usando una calculadora de ANOVA (http://www.socscistatistics.com/tests/anova/Default2.aspx), con un nivel de significancia de p<0.05. El numero de replicas independientes (N) para los genotipos analizados fue de 10 para las plantas Wt, 18 para *emb1241-1*, 16 para *emb1241-2* y 14 para *DcpHsc70-1*.

6.9. Microscopía electrónica.

Las muestras de hojas se fijaron durante 1 h en una solución de paraformaldehido al 4 % y glutaraldehído al 2.5 %. Posteriormente se lavaron las muestras en un buffer de 100 mM Na-Cacodilato (EMS, Hatfield, USA) pH 6.5; el tratamiento post fijación se realizó en el mismo buffer, suplementado con 1 % de OsO_4 e incubado durante 1 h en la

oscuridad a 4 °C. Las muestras se deshidrataron en una solución de etanol, seguido por un tratamiento de óxido de propileno y resina de Araldita-Epon y finalmente incubado durante 12 h a temperatura ambiente en medio de incrustación al 100 %. La polimerización de los especímenes se realizó en moldes a 60 °C durante 24 h. Las muestras semifinas (1 µm) y ultrafinas (60 nm) se cortaron en un ultramicrotomo LEICA (Wetzlar, Alemania). Las muestras semifinas se tiñeron en una solución de tinción [1 % (w/v) azul de toluidina, 1 % (w/v) borax] y se digitalizaron en un microscopio Axio Scope ZEISS (Thornwood, USA). Las muestras ultrafinas se tiñeron con una solución de tinción [4 % (w/v) de acetato de uranilo, 70 % (v/v) de etanol], seguido de un tratamiento con citrato de plomo; las imágenes se digitalizaron en un equipo JEOL 1200EXII Transmission Electron Microscope/GATAN (Warrendale, USA).

6.10. Análisis bioquímicos.

Las proteínas aisladas de las fracciones suborganelares del cloroplasto se precipitaron mediante la adición de Tris-EDTA [100 mM Tris, 10 mM EDTA, pH 8], 0.3% deoxicolato de sodio y 72% acido tricloroacético (TCA) en una proporción 1:5 (v/v) de reactivo:muestra, las muestras se incubaron 1h sobre hielo. La proteína precipitada se colectó mediante centrifugación (25, 000 x *g*), se resuspendió en una solución de acetona al 90%, y se incubó a -20 °C durante una noche. Las muestras se centrifugaron nuevamente (25, 000 x *g*) y se secaron por evaporación antes de resuspenderlas en un volumen mínimo de buffer Laemmli 2x [65 mM Tris-HCl pH 6.8, 30% (v/v) glicerol, 2% (w/v) SDS, 0.01% (w/v) azul de bromofenol, 355 mM 2-mercaptoetanol] [Laemmli, 1970]. Las muestras de proteínas (10 to 20 µL) fueron fraccionadas en geles desnaturalizantes de poliacrilamida al 12%, se transfirieron a membranas de nitrocelulosa (GE Healthcare

Bio-Sciences, UK), y se incubaron en buffer PBS-T [137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 0.1 % (v/v) Tritón X-100] suplementado con 0.5 % (w/v) de leche en polvo baja en grasa. Para los ensayos de Western blot, se usaron los anticuerpos primarios monoclonales αc-Myc (Sigma-Aldrich, México) y αHA (Santa Cruz Biotechnology Inc, USA) y los anticuerpos secundarios αMouse-HRP (Thermo Fisher Scientific, USA) y αRabbit-HRP (Thermo Fisher Scientific, USA).

Para los ensayos de BN-PAGE, se preparó proteína total de plantas de tipo silvestre y de las mutantes *emb1241-2*, y *DcpHsc70-1* utilizando tejido foliar pulverizado e incubado con el buffer de extracción BN [70 mM Tris-HCl pH 7.5, 1 mM $MgCl_2$, 25 mM KCl, 5 mM EDTA pH 8, 0.25 mM Sacarosa, 39.1 mM n-Dodecil β-D-maltosido] suplementado con el cocktail inhibidor de proteasas cOmplete™ (Sigma-Aldrich, México). Las muestras de proteínas se sometieron a BN-PAGE en un gel con un gradiente de 4 % a 14 % de acrilamida, según el protocolo de Heinemeyer *et al*, 2007. Adicionalmente, los geles de BN-PAGE se destiñeron con 10 lavados consecutivos de agua destilada hirviendo.

7. Resultados y discusión.

7.1. Objetivo 1: Caracterizar a nivel fenotípico las líneas mutantes *emb1241*, *pbp1* y *atrabe1b*.

Debido al fenotipo aberrante presentado por las plantas albinas mostradas en la Figura 3 y a su incapacidad para reproducirse de manera sexual, se cultivaron plantas verdes de las líneas vegetales *emb1241*, *pbp1* y *atrabe1b* hasta que produjeron silicuas maduras, y se analizó la progenie de estas. Nuestros resultados indican que en la progenie de las tres líneas vegetales segregan semillas verdes y semillas albinas (Figura 4A), estas últimas dan lugar a plántulas con defectos en la pigmentación al ser cultivados en condiciones *in vitro*. En el caso de las plantas de las líneas *emb1241* y *pbp1*, los embriones afectados se desarrollan en plántulas albinas (Figura 3). Sin embargo, en el caso de las plantas *atrabe1b*, los embriones afectados se desarrollan en plántulas variegadas que además presentan defectos morfológicos importantes en los órganos aéreos, como las hojas y los cotiledones (Figura 3). Además, un análisis cuantitativo de las semillas presentes en las silicuas de las plantas verdes adultas de las líneas *emb1241*, *pbp1* y *atrabe1b* reveló que las semillas verdes y albinas se presentan en una proporción de 3 semillas verdes por cada semilla albina (Figura 4B). Este tipo de segregación es característico de mutaciones monogénicas y recesivas, de hecho, el análisis de χ^2 de los datos de segregación apoya esta hipótesis con valores de $p>0.5$ (Figura 4B), es decir que la segregación de ambos fenotipos se ajusta a lo esperado para este tipo de mutaciones. Estos resultados son consistentes con los reportados por otros grupos de investigación, los cuales indican que las mutantes caracterizadas como portadoras de defectos en el desarrollo de los cloroplastos suelen ser recesivas

[Gutiérrez-Nava *et al*, 2004; Albrecht et al, 2006; Qin *et al*, 2007; Liu *et al*, 2016]. En conjunto, estas observaciones indican que las líneas mutantes bajo análisis presentan mutaciones recesivas únicas que son las causantes de los fenotipos aberrantes observados en los embriones y las plántulas, las cuales podrían tener defectos en el desarrollo de los cloroplastos.

Con la intención de demostrar esta hipótesis, se generaron micrografías electrónicas de los plástidos presentes en las células de plántulas con defectos de pigmentación de las líneas *emb1241*, *pbp1* y *atrabe1b*. En las micrografías obtenidas se puede observar que los plástidos presentes en las plantas de tipo silvestre (Wt) tienen la morfología y la estructura de membranas internas característica de los cloroplastos maduros (Figura 5). Por su parte, las plantas *atrabe1b* presentan plástidos con membranas internas estructuradas de manera similar a lo observado en las plantas Wt, pero con una morfología más esférica, no elíptica como la de los cloroplastos Wt (Figura 5). Defectos similares a estos han sido reportados en líneas mutantes afectadas en la integridad de las membranas de la envoltura y de los tilacoides; por ejemplo, la mutante *tic40* que está afectada en un componente de la membrana interna de la envoltura y las mutantes *stic1*, *stic2* y *vipp1* que tienen defectos en el importe de proteínas a los tilacoides [Chou *et al*, 2003; Vothknecht *et al*, 2012; Bédard *et al*, 2017]. En todos los casos se cree que el fenotipo esférico de los cloroplastos de las mutantes mencionadas se debe a defectos en la presión osmótica dentro de los organelos, puesto que, en el caso de *vipp1*, el fenotipo puede ser revertido modificando la concentración de osmolitos en el citoplasma [Zhang *et al*, 2012; Bédard *et al*, 2017]. Estos datos sugieren que las plantas *atrabe1b* tienen defectos en la integridad de las membranas cloroplásticas. Debido a que se

propone que ATRABE1B está involucrada en la traducción de proteínas dentro de los cloroplastos [de Luna-Valdez, 2012] y dado que el fenotipo de cloroplastos esféricos no es común entre las mutantes que afectan la traducción cloroplástica, es posible que los defectos en la integridad de las membranas cloroplásticas (que provocan la formación de cloroplastos esféricos) sean consecuencia de deficiencias en la traducción de proteínas específicas y no de defectos generales. La validez de estas hipótesis deberá ser confirmada en el futuro.

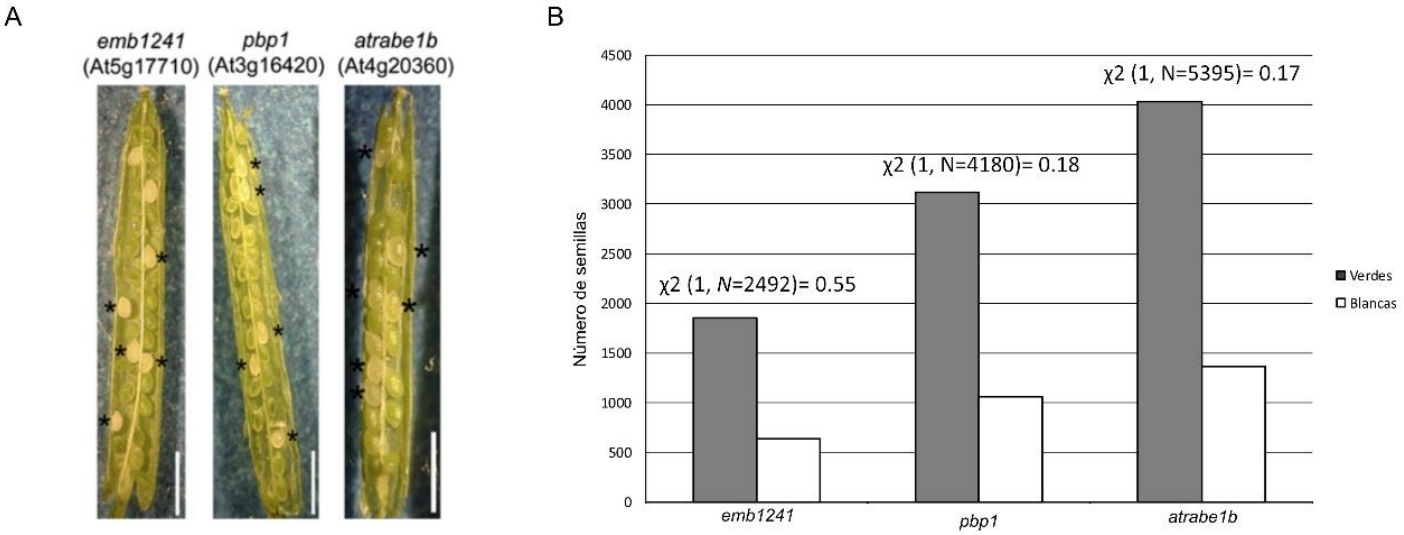


Figura 4. Análisis fenotípico de la progenie de las líneas mutantes *emb1241*, *pbp1* y *atrabe1b*. Se muestran fotografías de las semillas verdes y albinas (*) presentes en las silicuas de las plantas *emb1241*, *pbp1* y *atrabe1b* (A). Cuantificación del número de semillas verdes y albinas presentes en las plantas *emb1241*, *pbp1* y *atrabe1b*, los valores de χ^2, grados de libertad y número de individuos analizados (N), se muestran sobre las barras correspondientes; los valores de *p* asociados a las χ^2 son $p>0.5$ en los tres casos (B). Escala: 22 mm.

Por otro lado, las plantas *pbp1* presentan plástidos con una escasa acumulación de membranas internas, las cuales no logran organizarse para formar las grana de los tilacoides de los cloroplastos Wt (Figura 5), lo cual indica que estas plantas presentan defectos en el desarrollo de los cloroplastos. Por su parte, los plástidos presentes en las plantas *emb1241* carecen por completo de membranas internas (Figura 5), indicando

que estas plantas presentan defectos en el desarrollo del cloroplasto que podrían ser mucho más tempranos que los encontrados en las plantas *pbp1*.

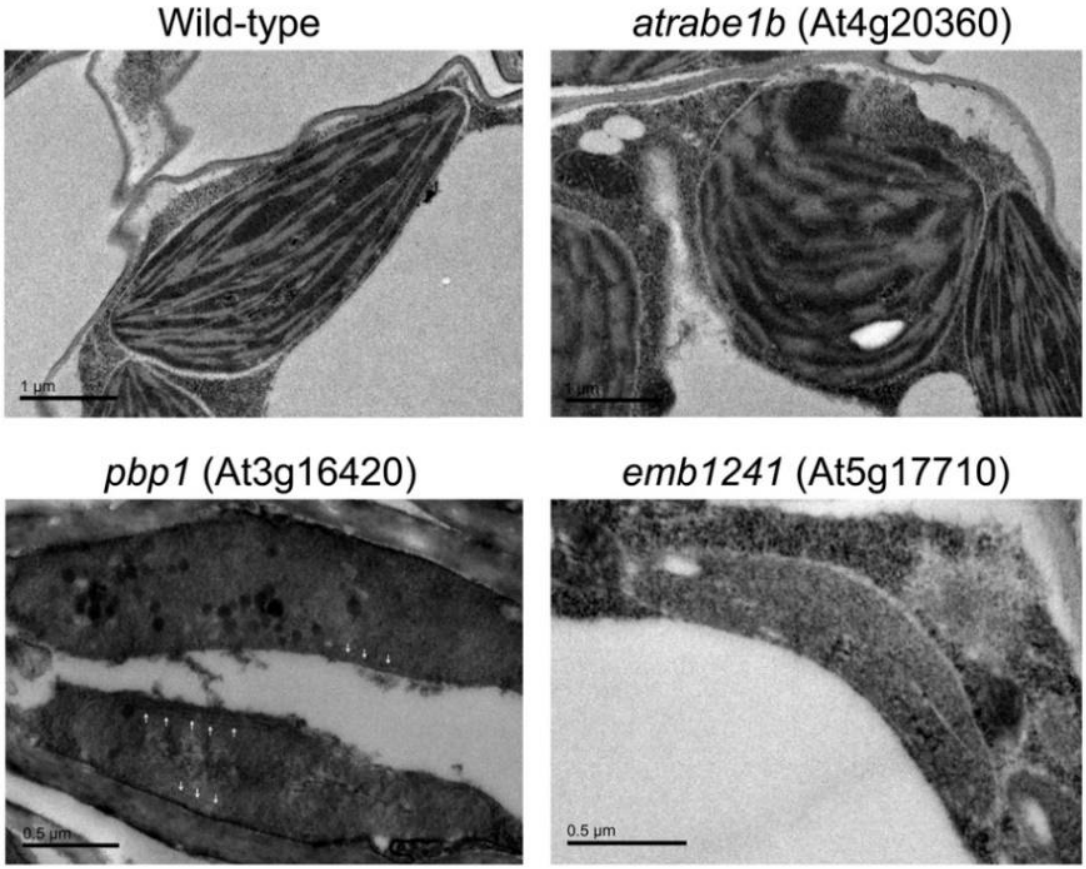


Figura 5. Micrografías de los plástidos de plantas tipo silvestre, *atrabe1b*, *pbp1* y *emb1241*. Se muestran micrografías obtenidas a partir de las hojas (*pbp1* y *emb1241*) o los cotiledones (Wt y *atrabe1b*) de plántulas de 16 días de desarrollo. Las flechas blancas marcan los sitios de acumulación de membranas internas en los plástidos de las plantas *pbp1*.

Es importante mencionar que la ultraestructura de los plástidos presentes en las líneas mutantes *pbp1* y *emb1241* es muy similar a la detectada en los plástidos de otras mutantes afectadas en componentes moleculares cuya ausencia afecta la biogénesis del cloroplasto, como el importe de proteínas en la mutante *ppi2* [Bauer *et al*, 2000], la síntesis de carotenoides en la mutante *pds3* [Qin *et al*, 2007], el importe de proteínas al tilacoide en la mutante *alb3* [Sundberg *et al*, 1997], entre otras. Además, en un estudio realizado por nuestro grupo de investigación, se determinó que las mutantes verde

pálido, amarillas y albinas de la colección *clb* presentan diversos grados de afectación en la acumulación y organización de las membranas internas [Gutiérrez-Nava *et al*, 2004]. El trabajo citado establece que existe una correlación entre los niveles de acumulación de membranas internas y la etapa de desarrollo del cloroplasto en que son necesarios los genes afectados en las mutantes: las mutantes afectadas en genes necesarios en la etapa temprana (como *clb2*, *clb3* y *clb5*) generan plástidos sin membranas internas, mientras que las mutantes afectadas en genes necesarios durante la etapa tardía (como *clb1*, *clb4* y *clb6*), generan plástidos que acumulan membranas internas pero en niveles inferiores a los cloroplastos tipo silvestre [Gutiérrez-Nava *et al*, 2004]. Los datos mencionados son consistentes con nuestras observaciones de la ultraestructura de los plástidos de las mutantes *pbp1* y *emb1241*, indicando que estas líneas mutantes tienen defectos que impactan el desarrollo del cloroplasto en etapas tempranas (*emb1241*) y tardías (*pbp1*).

7.2. Objetivo 2: Caracterizar a nivel molecular las líneas mutantes *emb1241*, *pbp1* y *atrabe1b*.

Con la intención de comprender más a detalle los defectos en el desarrollo de los cloroplastos de las plantas mutantes *emb1241*, *pbp1* y *atrabe1b*, se realizaron ensayos de Northern blot contra genes marcadores del desarrollo del cloroplasto, usando plántulas mutantes de las líneas mencionadas. A continuación, se ofrece una descripción de las características más importantes de los genes marcadores utilizados. El gen *rrn16S* codifica para el RNA ribosomal 16S, y es el gen codificado en el genoma cloroplástico con la expresión más alta en los proplástidos, por esto se usa como un marcador de la actividad transcripcional temprana dentro de los plástidos en desarrollo [Bisanz-Sayer *et*

al, 1989]. Por otro lado, *RPL21* es un gen codificado en el núcleo que codifica para una subunidad proteínica del ribosoma cloroplástico, se acumula durante la imbibición de las semillas y se expresa antes que los genes codificados en el genoma cloroplástico, por lo que se usa para monitorear el estado temprano de diferenciación de los plástidos [Harrak *et al*, 1995]. El gen *accD* codifica para la enzima acetil-CoA carboxilasa que cataliza el primer paso comprometido de la vía de síntesis de ácidos grasos en el cloroplasto. Este gen está codificado en el genoma cloroplástico y solo es transcrito por la RNA polimerasa nuclear NEP [Hajdukiewicz *et al*, 1997], por lo que se usa como marcador temprano de la transcripción en el cloroplasto. Finalmente, los genes *psbA* y *RBCS* están codificados en los genomas cloroplástico y nuclear respectivamente, estos genes codifican para componentes esenciales de la maquinaria fotosintética y son considerados marcadores tardíos de la función de los cloroplastos [Bruick y Mayfield, 1999].

En estos experimentos encontramos que las plantas mutantes *emb1241* y *pbp1* muestraron niveles reducidos de acumulación de todos los transcritos analizados. En comparación con las plantas Wt, las plantas *emb1241* presentan disminuciones en la acumulación de transcritos del 57% en *rrn16S*, 56% en *psbA*, 66% en *RBCS*, 70% en *RPL21* y solo del 5% en *accD* (Figura 6). Por su parte, las plantas *pbp1* muestran disminuciones del 40% en *rrn16S*, 67% en *psbA*, 50% en *RBCS*, 20% en *RPL21*, y 21% en *accD* (Figura 6). En contraste, las plantas *atrabe1b* solo muestran una disminución del 25% en la acumulación del transcrito de *accD*, mientras que los niveles del resto de los transcritos analizados se mantienen muy cercanos a los mostrados por las plantas tipo silvestre (Figura 6).

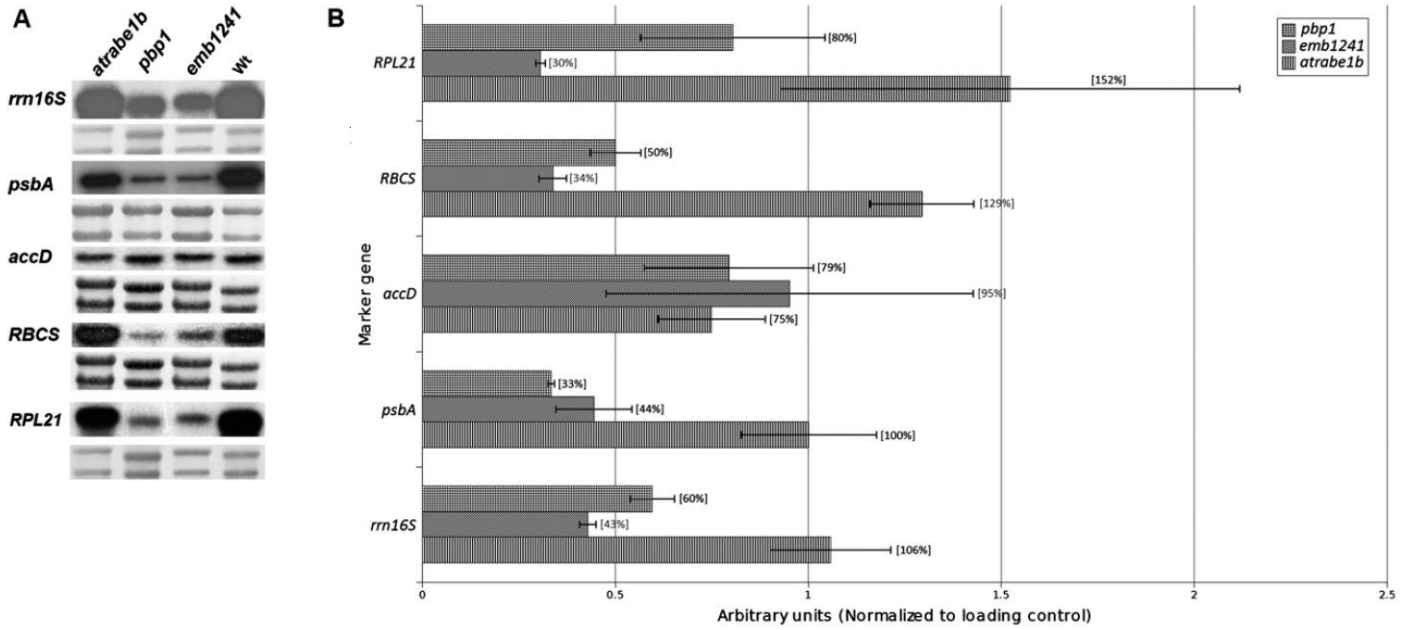


Figura 6. Análisis de expresión de genes marcadores del desarrollo y funcionamiento de los cloroplastos. Ensayos de Northern blot de diversos genes marcadores del desarrollo y funcionamiento de los cloroplastos en extractos de RNA total de plántulas Wt de 8 días de desarrollo y mutantes *emb1241*, *pbp1* y *atrabe1b* de 16 días de desarrollo. Se muestran tinciones de las membranas de RNA con azul de metileno como control de carga (A). Análisis de densitometría de las hibridaciones de Northern blot, se muestran los valores de acumulación de los transcritos detectados en cada una de las líneas mutantes analizadas, con relación a su acumulación en las plantas Wt. Los valores están normalizados con el control de carga de cada membrana utilizada. Las barras de error representan el error estándar obtenido de 5 repeticiones biológicamente independientes de cada experimento (B). *rrn16S* (rRNA cloroplástico 16S), *psbA* (subunidad D1 del centro de reacción del fotosistema II), *accD* (acetíl-CoA carboxílasa), *RBCS* (subunidad pequeña de la enzima RuBisCO), *RPL21* (proteína ribosomal cloroplástica L21).

En un análisis equivalente realizado con las mutantes de la colección CLB [Gutiérrez-Nava *et al*, 2004], se encontraron resultados similares a los nuestros, con las mutantes mostrando niveles severos de reducción en la acumulación de los transcritos de marcadores del desarrollo del cloroplasto. Nuestros resultados indican que, aunque las mutantes *emb1241* y *pbp1* tienen las características moleculares que definen a las plantas afectadas en el desarrollo del cloroplasto, sus defectos no son tan tempranos como los mostrados por las plantas *clb2* y *clb5*, las cuales muestran los defectos más

severos al carecer por completo del transcrito de RPL21, que es el marcador del desarrollo más temprano [Gutiérrez-Nava *et al*, 2004]. Al igual que el resto de las mutantes de la colección CLB, las plantas *emb1241* y *pbp1* logran superar las etapas más tempranas del desarrollo e iniciar la transcripción mediada por NEP, como lo refleja la acumulación del transcrito de *accD.* Sin embargo, las plantas *pbp1* logran iniciar la transcripción mediada por PEP (Plastid-Encoded RNA Polymerase) más eficientemente que las plántulas *emb1241*, como lo pone en evidencia la acumulación de *rrn16S* mostrado por estas plantas. Estas observaciones indican que la mutación en *emb1241* afecta el desarrollo de los cloroplastos en una etapa más temprana que la mutación de *pbp1.* A pesar de estas diferencias en las etapas tempranas, ambas mutaciones tienen un impacto similar en la expresión de los genes tardíos *RBCS* y *psbA*, lo cual pone de manifiesto la severidad de estas mutaciones.

En conjunto, nuestras observaciones indican con claridad que las tres líneas mutantes analizadas, en particular las líneas *emb1241* y *pbp1*, tienen defectos en el desarrollo de los cloroplastos que se pueden evidenciar tanto a nivel fenotípico como molecular. Los resultados descritos hasta este punto y los descritos en la tesis de Maestría citada [de Luna-Valdez, 2012] se compilaron en las publicaciones número 3 y 4 incluidas en el Apéndice II. Finalmente, a raíz de este análisis, y debido a que los defectos en el desarrollo de los cloroplastos de la línea *atrabe1b* parecen ser poco severos fenotípica y molecularmente, se decidió enfocar los análisis subsiguientes en la caracterización subcelular y funcional de las proteínas EMB1241 y PBP1.

7.3. Objetivo 3: Caracterizar a nivel funcional a la proteína PBP1.

La proteína PBP1 (PYK10-binding protein 1) es una proteína residente del retículo endoplásmico (ER) que está reportada como una interactora directa de la enzima β-glucosidasa PYK10 , la cual es el constituyente principal de los cuerpos ER del retículo endoplásmico, cataliza la hidrólisis de compuestos con enlaces O-glucosídicos y está involucrada en el establecimiento de las relaciones mutualistas entre arabidopsis y el hongo endofítico *P. indica* [Nagano *et al*, 2005; Sherameti *et al*, 2018]. Existen datos que indican que PBP1 es capaz de mejorar la actividad hidrolítica de PYK10 y que podría actuar como chaperona molecular durante el plegamiento de esta enzima en respuesta a estrés biótico [Nagano *et al*, 2005]. Según la información conocida sobre la función de PBP1, no es posible relacionarla directamente con algún proceso esencial para el desarrollo de los cloroplastos. Sin embargo, hay evidencias claras de que existe una vía de importe de proteínas al cloroplasto que está mediada por el ER, el ejemplo más claro de esto es la proteína de *A. thaliana* Anhidrasa Carbónica 1 (CAH1) [Villarejo *et al*, 2005]. Aún no se conocen detalles exactos sobre el mecanismo o los reguladores proteínicos que medían esta vía de importe, pero debido a su localización predicha y a los defectos en el desarrollo de los cloroplastos de la mutante *pbp1*, nosotros hipotetizamos que la proteína PBP1 está involucrada en la regulación de la vía de importe de proteínas al plástido que depende de la vía secretoria.

Con la intención de probar esta hipótesis, se generaron fusiones traduccionales entre PBP1 y CAH1 con el reportero GFP, estas fusiones se expresaron de manera transitoria en células de *N. benthamiana*. Estos experimentos indican que la proteína PBP1 se localiza en el ER de las células transformadas, puesto que la fluorescencia de la fusión

PBP1::GFP colocaliza con la fluorescencia del marcador fluorescente ER-CFP (Figura 7B), el cual consiste de la proteína fluorescente CFP fusionada a la señal de retención en el ER KDEL, y se distribuye en una red reticular que corresponde al ER de las células vegetales (Figura 7A).

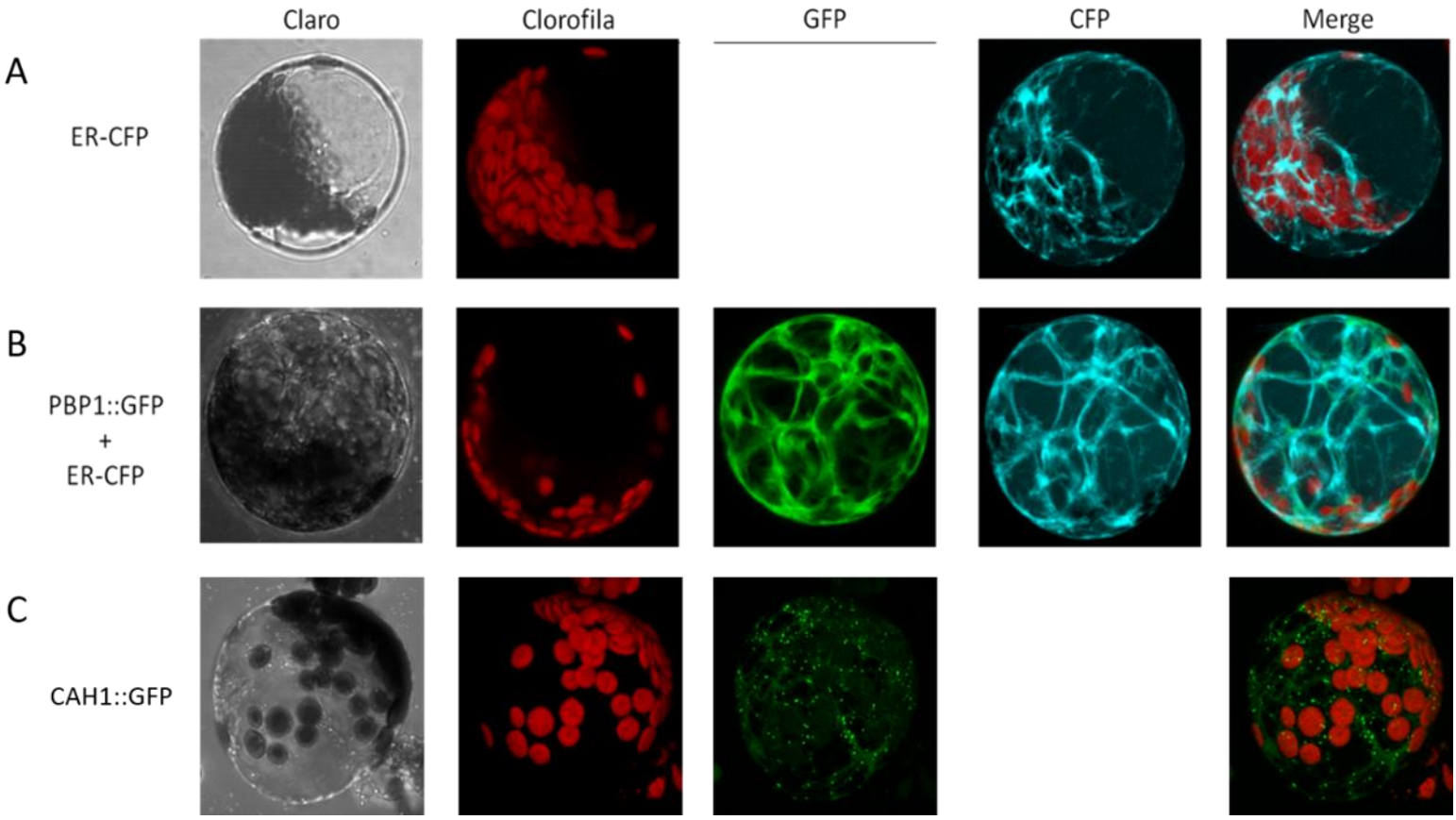


Figura 7. Localización subcelular de PBP1 y CAH1. Se muestran ensayos de expresión transitoria de las fusiones traduccionales PBP1::GFP y CAH1::GFP en protoplastos de *N.benthamiana*. Se muestran los canales del campo claro (claro), la florescencia de la clorofila (clorofila), fluorescencia de GFP (GFP), fluorescencia de CFP (CFP) y una sobre posición de los canales fluorescentes (Merge). Se incluye la localización de un marcador fluorescente que marca el ER.

De manera similar a PBP1 y al marcador ER-CFP, la fusión CAH1::GFP se distribuye a lo largo de una red parecida al ER de las células vegetales, además de que forma acumulaciones discretas que podrían ser vesículas derivadas del sistema endomembranal de las células transformadas (Figura 7C), sugiriendo que PBP1 y CAH1 se colocalizan en algunos compartimentos celulares, en particular el ER. En este sentido, es importante resaltar que los experimentos de la localización subcelular de CAH1 no están acompañados por los controles adecuados para corroborar la localización de esta proteína en el ER o algún otro compartimento del sistema endomembranal, sin embargo, es claro que la fusión bajo análisis no se localiza en los cloroplastos. Este resultado contrasta con las observaciones publicadas sobre la distribución subcelular de CAH1, las cuales indican que es una proteína del estroma del cloroplasto que solo puede ser observada en el ER cuando las células son tratadas con agentes que bloquean el transito vesicular, como la brefeldina A [Villarejo *et al*, 2005]. Esta diferencia fundamental dificulta nuestro análisis de la vía de importe de proteínas al cloroplasto a través del sistema endomembranal de las células vegetales y de la participación de PBP1 en dicho proceso, puesto que, si bien encontramos una posible colocalización parcial entre PBP1 y CAH1, está ultima no es observada en los cloroplastos, lo que impide que analicemos el tránsito completo de la proteína hasta su localización final.

Es importante mencionar que nuestras observaciones corroboran con información generada *in vivo* las evidencias bioquímicas que sugieren la localización subcelular de PBP1 en el ER [Nagano *et al*, 2005]. A pesar de esto, nuestros datos alrededor de PBP1 y CAH1 no nos permiten generar conclusiones claras que relacionen a PBP1 con el importe de proteínas al cloroplasto o cualquier otro proceso celular que involucre al ER

y a los cloroplastos. Además, actualmente no contamos con el material biológico que es necesario (como las construcciones genéticas que permitan visualizar el tránsito de CAH1 hasta el cloroplasto o plantas mutantes *cah1* para realizar análisis genéticos y bioquímicos) para seguir explorando estas alternativas. Debido a estas dificultades, la parte final de este trabajo está enfocada en la dilucidación de varias características celulares, bioquímicas y funcionales de la proteína EMB1241.

7.4. Objetivo 4: Caracterizar a nivel funcional a la proteína EMB1241.

La proteína EMB1241 tiene un dominio conocido como GrpE, que en *E.coli* se conoce como una co-chaperona con función de intercambiador de nucleótidos para la chaperona DnaK (HSP70). En las plantas, las proteínas GrpE cloroplásticas reciben el nombre de CGE (chloroplast GrpE), por lo que desde este momento el nombre de EMB1241 será sustituido por CGE1. Con la intención de expandir el conocimiento actual sobre las CGEs, realizamos un análisis filogenético en el que incluimos 136 CGEs de 71 especies diferentes (61 embriofitas: 60 traqueófitas y 1 briofita; 7 clorófitas; 1 carofita; 1 criptofita; 1 bacteria como grupo externo). Cabe resaltar que este análisis representa una gran mejora con respecto a un análisis similar realizado en el pasado, en el cual se usaron solamente 17 proteínas GrpEs (no solo CGEs) de únicamente 7 especies [Shi y Theg, 2010]. Nuestros resultados indican que las CGEs de las plantas superiores forman un grupo monofilético que se subdivide en dos clados principales, los cuales llamamos Tipo A y Tipo B (Figura S1). Estos grupos se distinguen entre sí por algunas variaciones en motivos conservados; por ejemplo, los motivos SYQGI y VKVS que caracterizan a las CGEs Tipo A (Figura 8A) y los motivos (N/D)SYQSI y MVKVS que caracterizan a las CGEs Tipo B (Figura 8B). Nuestro análisis reveló que, con respecto a la GrpE de *E. coli*,

todas las CGEs tienen modificaciones en las posiciones R74 y K82 (Figura 8C; Tabla S2), que son dos de las seis posiciones fundamentales para la actividad de intercambiador de nucleótidos de la GrpE en *E. coli* [Harrison *et al*, 1997; Gelinas *et al*, 2003; Gelinas *et al*, 2004]. Además, la posición G122 de la GrpE solamente se conserva en las CGEs de Tipo A, mientras que ha sido sustituida por Ser en el 86% de las CGEs Tipo B (Figura 8C; Tabla S2). Por su parte, el resto de las posiciones de la GrpE relacionadas con la función de intercambiador de nucleótidos, se encuentran conservadas en ambos tipos de CGEs (Figura 8C; Tabla S2). Finalmente, nuestros datos indican que los dos genes *CGE* ya conocidos en el genoma de arabidopsis [Schroda *et al*, 2001; Shi y Theg, 2010] pertenecen a cada uno de los dos grupos definidos: *AtCGE1* al Tipo B y *AtCGE2* al Tipo A (Figura S1).

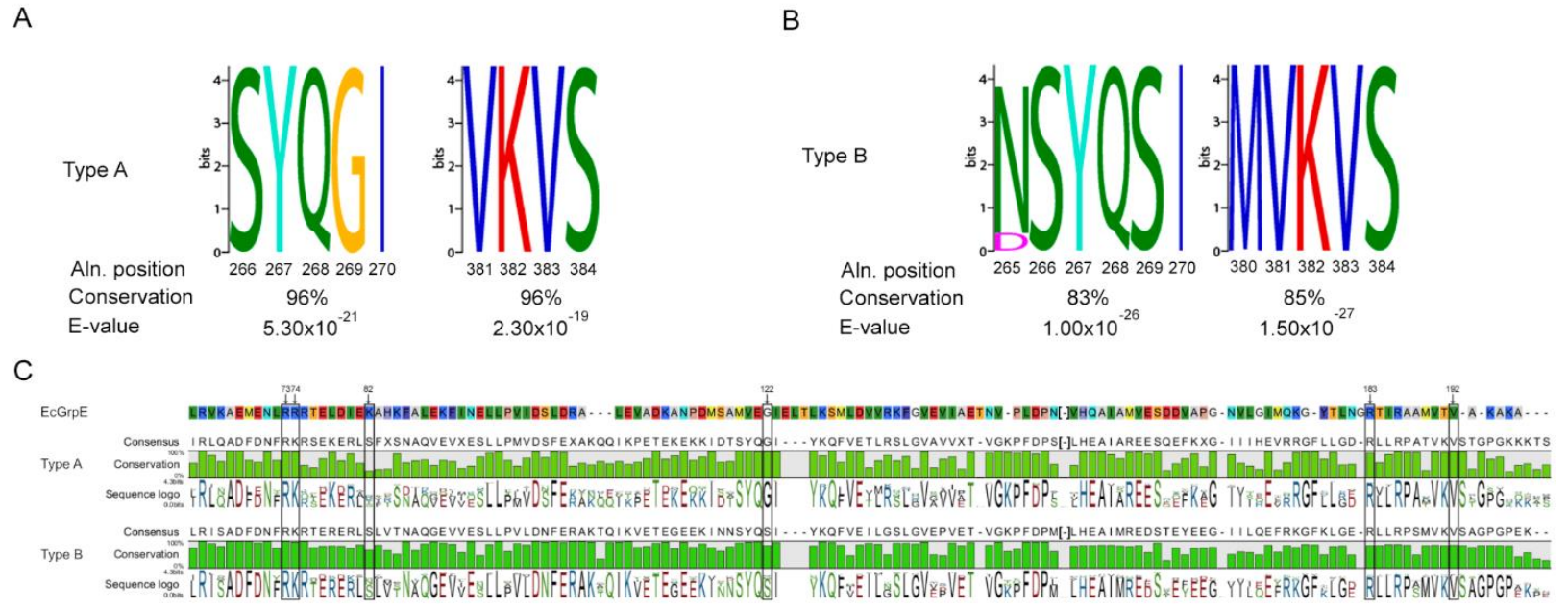


Figura 8. Identificación de motivos conservados en las CGEs. Motivos de secuencia enriquecidos en las CGEs Tipo A (A) y Tipo B (B). Alineamiento entre la GrpE de *E. coli* y los consensos de las CGEs Tipo A y Tipo B (C). Los números, flechas y cajas en las posiciones 73, 74, 82, 122, 183 y 192, marcan a los aminoácidos que son relevantes para la función de la GrpE en *E. coli*.

Con la intención de evaluar la funcionalidad de ambos tipos de CGEs como intercambiadores de nucleótidos, se utilizaron los genes de *A. thaliana AtCGE1* y *AtCGE2* para complementar a la mutante de *E. coli* OD212, la cual es knockout para el gen *grpE*. Esta mutante tiene una mutación adicional en el gen *dnak332* que suprime el fenotipo letal asociado a la mutación de *grpE*. Como resultado, esta cepa crece normalmente a 25 °C pero no a temperaturas más altas [Deloche *et al*, 1997]. Nuestros resultados indican que bajo condiciones permisivas (25 °C) no existen diferencias detectables en el crecimiento de la cepa OD212 transformada con el vector vacío y los diferentes genes analizados (Figura 9). Sin embargo, a 37 °C es posible observar que las células transformadas con las GrpEs analizadas tienen un mejor crecimiento que las transformadas con el vector vacío (Figura 9). Finalmente, bajo condiciones restrictivas (43 °C) se detectaron defectos en el crecimiento de las células transformadas con todos los genes usados; sin embargo, los defectos mostrados por las células transformadas con los genes *AtCGE1*, *AtCGE2* y *EcGrpE* mostraron un crecimiento mejor que las transformadas con el vector vacío. Además, en comparación con el gen *AtCGE1*, el gen *AtCGE2* mostró una mejor capacidad para complementar a la mutante OD212 bajo las condiciones restrictivas de crecimiento (Figura 9). Esta diferencia puede atribuirse a que, al igual que la EcGrpE, las CGEs Tipo A (como AtCGE2) conservan una glicina en la posición 122, mientras que las Tipo B (como AtCGE1) tienen una serina (Figura 8C). En *E. coli*, mutaciones en la glicina 122 se han asociado con defectos en la actividad de la EcGrpE [Harrison *et al*, 1997], lo cual indica que la sustitución de Gly por Ser en AtCGE1 podría ser la causante de la diferencia detectada entre la complementación de la cepa OD212 con AtCGE1 y AtCGE2. Resultados similares de complementación de mutantes

bacterianas se han obtenido para las GrpEs eucarióticas de *C. reinhardtii* y *S. cerevisiae* [Deloche y Georgopoulos, 1996; Schroda *et al*, 2001], los cuales corroboran nuestros resultados que indican una clara ortología entre la EcGrpE y las dos CGEs de *A. thaliana*, puesto que ambas tienen las características bioquímicas para interaccionar con las chaperonas Hsp70 y funcionar como intercambiadores de nucleótidos.

Con el objetivo de analizar algunos aspectos funcionales de las CGEs *in vivo*, se generaron fusiones traduccionales entre el reportero GFP y las proteínas AtCGE1, AtCGE2 y cpHsc70-1 (una chaperona cloroplástica de la familia Hsp70); estas fusiones se usaron para transformar células de *N. benthamiana* (Figura 10). En estos experimentos, las fusiones AtCGE1-GFP y AtCGE2-GFP se localizan dentro de los cloroplastos y también formando cúmulos discretos en la periferia de los plástidos (Figura 10B y 10C).

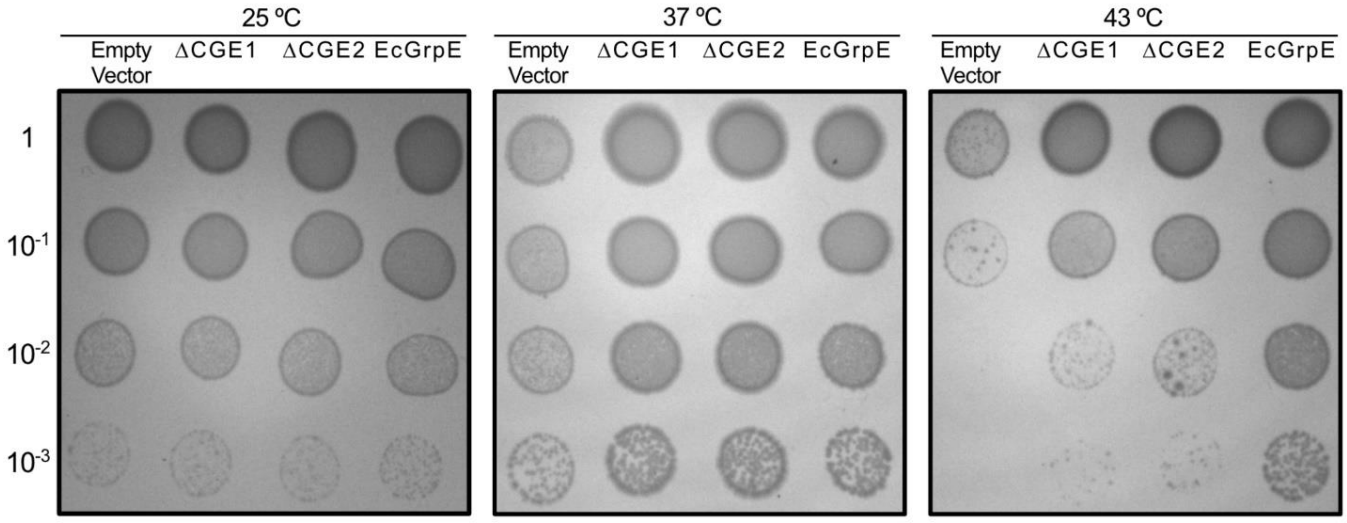


Figura 9. Complementación genética de la mutante OD212. El crecimiento de la cepa de *E. coli* OD212 transformada con el vector vacío (Empty Vector) o los vectores de expresión que contienen a los genes *AtCGE1* (ΔCGE1), *AtCGE2* (ΔCGE2) y *EcgrpE* (EcGrpE), se analizó a 25 °C , 37 °C y 43 °C. Gotas de 5 µL de los cultivos bacterianos diluidos (1, OD600nm= 0.1; 10^{-1}, OD600nm= 0.01; 10^{-2}, OD600nm= 0.001; 10^{-3}, OD600nm=0.0001) se cultivaron por 20 horas en medio LB. Al menos 3 réplicas biológicamente independientes fueron realizadas para estos experimentos.

Estos cúmulos se han descrito extensamente como deformaciones de la envoltura del cloroplasto causadas por la acumulación de proteínas de membrana, y también como vesículas derivadas de la envoltura del cloroplasto [Breuers *et al*, 2012]. Por su parte, la fusión cpHsc70-1-GFP mostró un patrón de distribución característico de las proteínas del estroma, localizándose en focos intensos dentro del cloroplasto (Figura 10D) [Farmaki *et al*, 2007; Perello *et al*, 2016]. Además, utilizando las fusiones traduccionales AtCGE1-cMyc, AtCGE2-cMyc y cpHsc70-1-cMyc, determinamos mediante técnicas de fraccionamiento suborganelar de cloroplastos purificados, que AtCGE1 y AtCGE2 se acumulan en el estroma y la envoltura, mientras que cpHsc70-1 solo se acumula en el estroma (Figura S7). En conjunto, nuestro análisis demuestra claramente que las proteínas AtCGE1 y AtCGE2 son importadas al cloroplasto, donde coexisten con la chaperona cpHsc70-1 en el estroma.

Debido a que la función descrita de las GrpEs y de algunas CGEs, como las de *C. reinhardtii* y *P. patens*, está ligada a su interacción con chaperonas de la familia Hsp70 [Schroda *et al*, 2001; Shi y Theg, 2010], nos propusimos determinar la capacidad de las proteínas AtCGE1 y AtCGE2 para interaccionar *in vivo* con la chaperona cpHsc70-1. Para esto se siguió una estrategia conocida como complementación bimolecular de la fluorescencia (BiFC), la cual consiste en fusionar una proteína con el fragmento N-terminal del reportero YFP (nYFP) mientras que otra proteína se fusiona al fragmento C-terminal (cYFP); cuando se coexpresan estas fusiones, la fluorescencia de la YFP se restaura solamente cuando las proteínas de interés forman interacciones físicas directas.

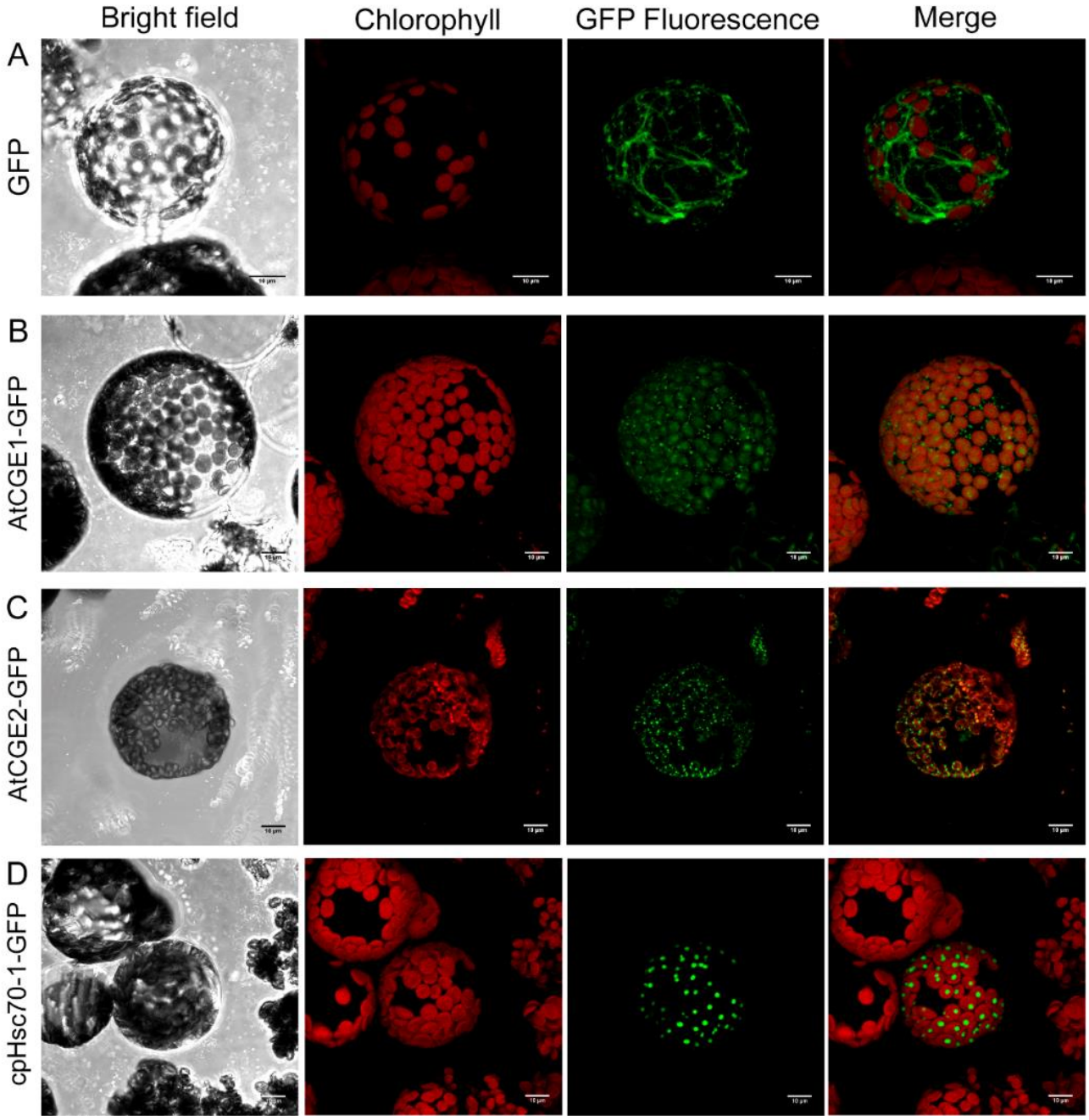


Figura 10. Localización subcelular de AtCGE1, AtCGE2 y cpHsc70-1. Se muestran micrografías de protoplastos de *N. benthamiana* que expresan a la proteína GFP (A) y a las fusiones traduccionales AtCGE1-GFP (B), AtCGE2-GFP (C) y cpHsc70-1-GFP (D). Las imágenes que corresponden al campo claro (Bright Field), fluorescencia de la clorofila (Chlorophyll), fluorescencia de GFP (GFP) y a la sobreposición de los canales de fluorescencia (Merge), se muestran en las columnas indicadas. Escala 10 µm. Se realizaron al menos 3 réplicas independientes de estos experimentos.

De manera similar a lo observado con las fusiones traduccionales AtCGE1-GFP y AtCGE2-GFP, en los experimentos de BiFC fue posible detectar fluorescencia de YFP dentro y en cúmulos en la periferia de los cloroplastos de las células cotransformadas

con las fusiones cpHsc70-1-nYFP y AtCGE1-cYFP (Figura 11A), AtCGE1-nYFP y cpHsc70-1-cYFP (Figura 11B), cpHsc70-1-nYFP y AtCGE2-cYFP (Figura 11C), y AtCGE2-nYFP y cpHsc70-1-cYFP (Figura 11D). En conjunto estas observaciones demuestran que las dos proteínas AtCGEs forman interacciones físicas directas con la chaperona cpHsc70-1 dentro de los cloroplastos *in vivo*. Es importante mencionar que las evidencias experimentales de la interacción entre CGEs y chaperonas cloroplásticas son muy escasas, estas interacciones solo se han demostrado por métodos bioquímicos para la CGE1 y la chaperona cloroplástica HSP70B del alga *C. reinhardtii* [Schroda *et al*, 2001]. En este contexto, nuestros resultados son el primer reporte de la interacción *in vivo* entre CGEs y Hsp70s de plantas vasculares.

En *E. coli* y *S. cerevisiae* se ha demostrado que la dimerización de las GrpEs es fundamental para que estas puedan establecer interacciones con las proteínas Hsp70 [Deloche y Georgopoulos, 1996; Wu *et al*, 1996; Azem *et al*, 1997; Willmund *et al*, 2007]. Mediante ensayos de BiFC, demostramos que ambas AtCGEs son capaces de formar homodímeros y heterodímeros *in vivo* (Figura 12). En estos experimentos es posible observar la reconstitución de YFP en los cloroplastos de las células cotransformadas con las fusiones traduccionales AtCGE1-cYFP y AtCGE1-nYFP (Figura 12A), AtCGE2-cYFP y AtCGE2-nYFP (Figura 12B), AtCGE1-cYFP y AtCGE2-nYFP (Figura 12C), y AtCGE2-cYFP y AtCGE1-nYFP (Figura 12D). Estos resultados demuestran que las CGEs de *A. thaliana* son capaces de formar los homodímeros necesarios para interaccionar con las chaperonas Hsp70. Por otro lado, la posible función de los heterodímeros de AtCGE1 y AtCGE2 resulta un punto interesante para investigaciones futuras, puesto que no existen reportes similares para otras GrpEs.

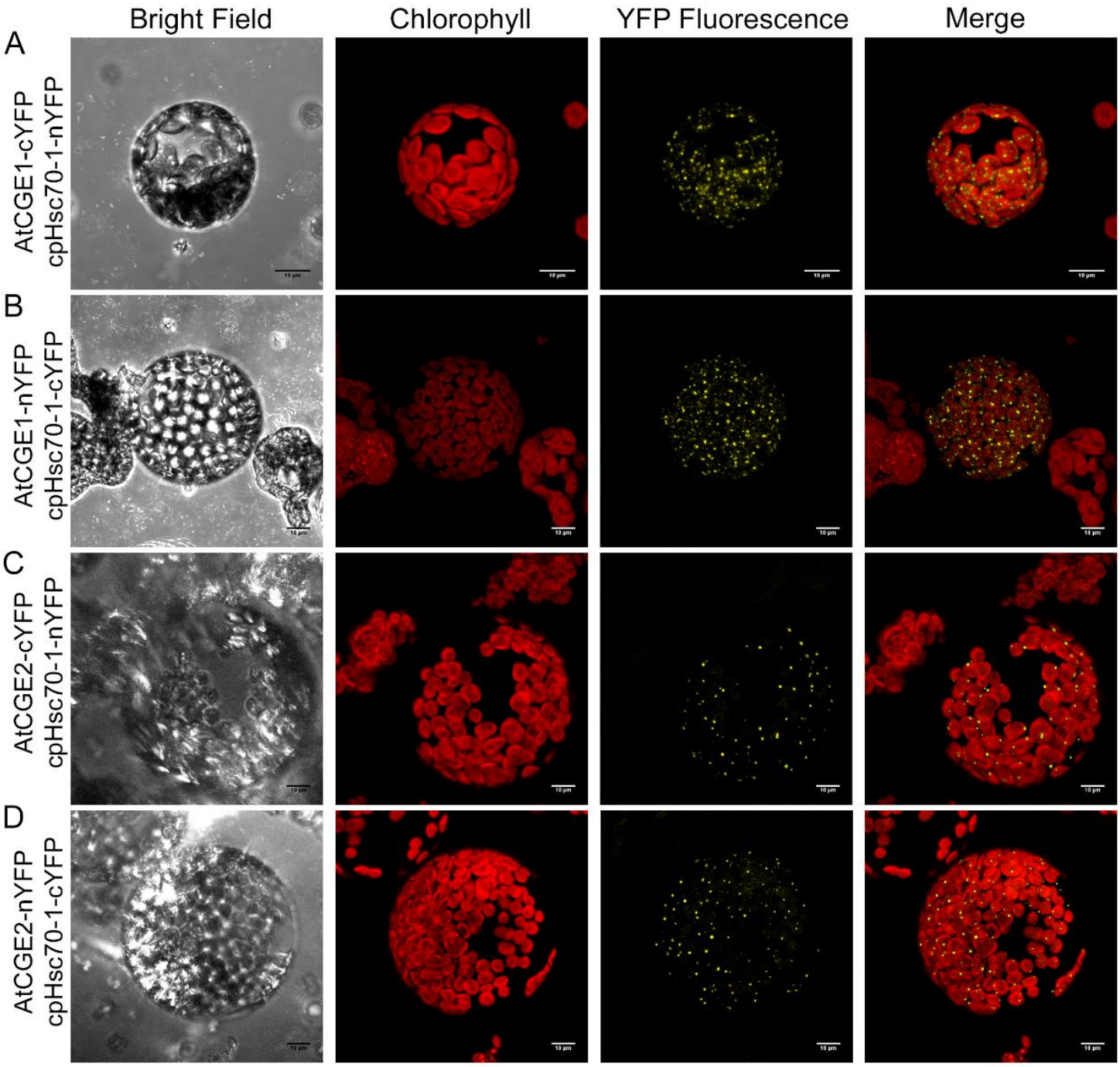


Figura 11. Análisis de las interacciones entre las proteínas AtCGE1, AtCGE2 y cpHsc70-1. Se muestran micrografías de protoplastos de *N. benthamiana* que coexpresan las fusiones traduccionales cpHsc70-1-nYFP y AtCGE1-cYFP (A), AtCGE1-nYFP y cpHsc70-1-cYFP (B), cpHsc70-1-nYFP AtCGE2-cYFP (C) y cpHsc70-1-cYFP (D).Las imágenes que corresponden al campo claro (Bright Field), fluorescencia de la clorofila (Chlorophyll), fluorescencia de YFP (YFP) y a la sobreposición de los canales de fluorescencia (Merge), se muestran en las columnas indicadas. Escala 10 µm. Se realizaron al menos 3 réplicas independientes de estos experimentos.

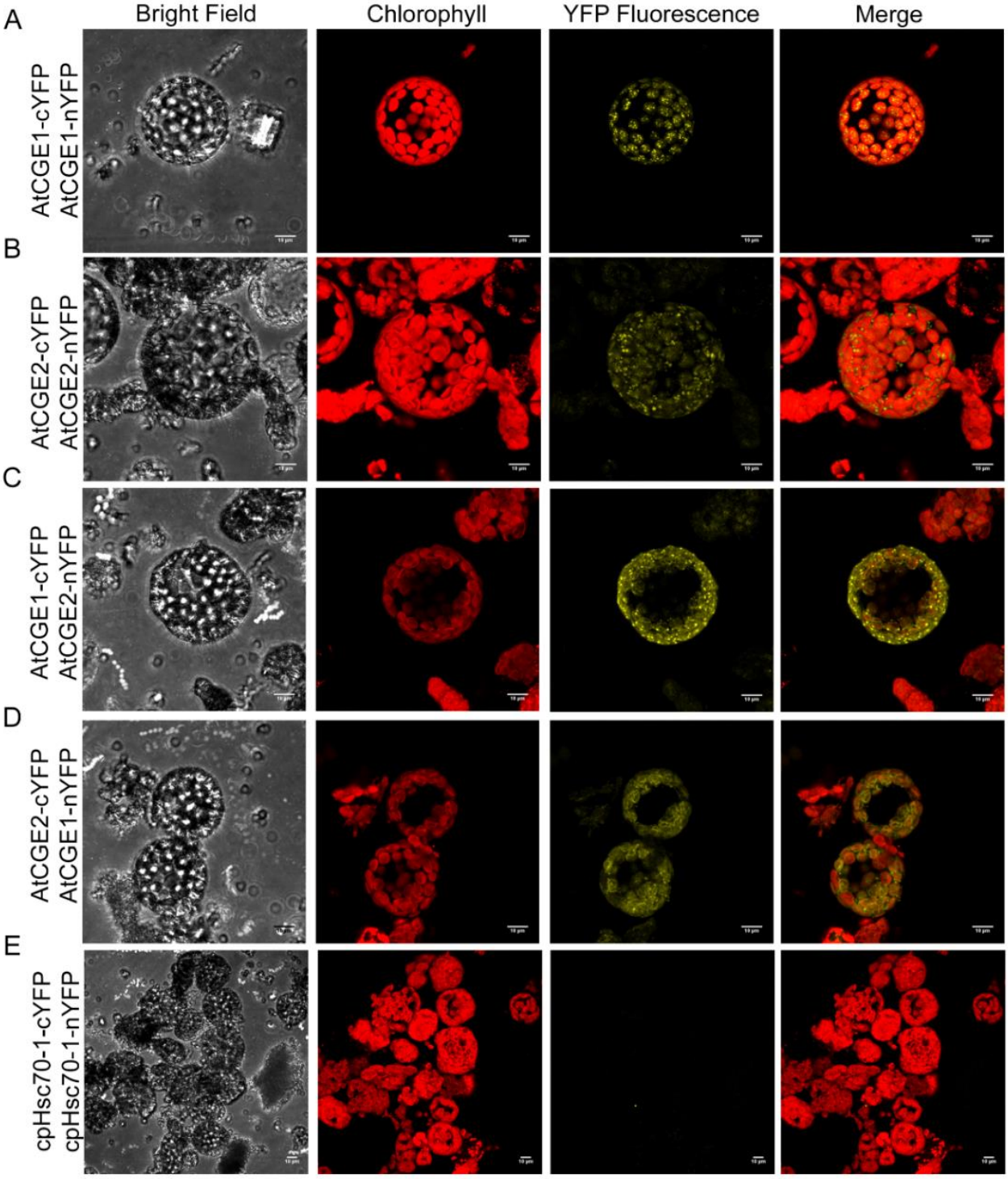


Figura 12. Análisis de la formación de dímeros de las proteínas AtCGE1, AtCGE2 y cpHsc70-1. Se muestran micrografías de protoplastos de *N. benthamiana* que coexpresan las fusiones traduccionales AtCGE1-cYFP y AtCGE1-nYFP (), AtCGE2-cYFP y AtCGE2-nYFP (B), AtCGE1-cYFP y AtCGE2-nYFP (C), AtCGE2-cYFP y AtCGE1-nYFP (D), y cpHsc70-1-cYFP y cpHsc70-1-nYFP (E).Las imágenes que corresponden al campo claro (Bright Field), fluorescencia de la clorofila (Chlorophyll), fluorescencia de YFP (YFP) y a la sobreposición de los canales de fluorescencia (Merge), se muestran en las columnas indicadas. Escala 10 µm. Se realizaron al menos 3 réplicas independientes de estos experimentos.

Adicionalmente, se utilizaron las fusiones traduccionales cpHsc70-1-cYFP y cpHsc70-1-nYFP como control negativo de todas las interacciones probadas en el trabajo, la falta de recuperación de la fluorescencia de YFP en las imágenes correspondientes (Figura 12E) demuestra que los resultados obtenidos no se deben a artefactos experimentales. Así mismo, todas las interacciones aquí demostradas se corroboraron por medio de ensayos de inmunoprecipitación (Figura S6) para asegurar su veracidad.

Con la intención de comprender el papel fisiológico de las AtCGEs, se determinó mediante Northern blot la expresión de los genes *AtCGE1*, *AtCGE2* y *CPHSC70-1* en plántulas de *A. thaliana* de 8, 10, 12 y 14 días de desarrollo. Estos ensayos indican que los tres genes analizados se coexpresan durante el desarrollo temprano de la planta (Figura 13A y 13B), dándole un contexto biológico a la colocalización y a las interacciones observadas en los experimentos ya discutidos (Figuras 10, 11 y 12). Además, debido a que en diversos modelos de estudio, como bacterias y algas, la actividad de las chaperonas Hsp70 y de las GrpEs se ha visto relacionada con las respuestas a choque térmico [Yura *et al*, 1993; Schroda *et al*, 2001], se analizó la acumulación de los transcritos de los genes AtCGE1, AtCGE2 y CPHSC70-1 en plántulas de 16 días incubadas a 40 °C durante 0, 30, 60 y 90 minutos (Figura 13C y 13D). En estos experimentos se observa que la respuesta de un gen cloroplástico no relacionado con el estrés por calor consiste en disminuir su expresión conforme el tiempo bajo estrés aumenta (Figura 13C y 13D; AtDXS1); de manera similar, la acumulación de transcrito de AtCGE2 disminuye de manera considerable en respuesta al estrés por calor (Figura 13C y 13D; AtCGE2).

Por su parte, el nivel de acumulación del transcrito de AtCGE1 disminuye en respuesta al estrés por calor, pero se mantiene en un nivel similar durante todo el tratamiento (Figura 13C y 13D; AtCGE1). En contraste, el gen CPHSC70-1 muestra una disminución inicial en abundancia, la cual es seguida por un incremento sustancial a tiempos prolongados bajo estrés por calor (Figura 13C y 13D; CPHSC70-1), esta respuesta es consistente con lo observado en otras chaperonas Hsp70 cloroplásticas [Schroda *et al*, 2001].

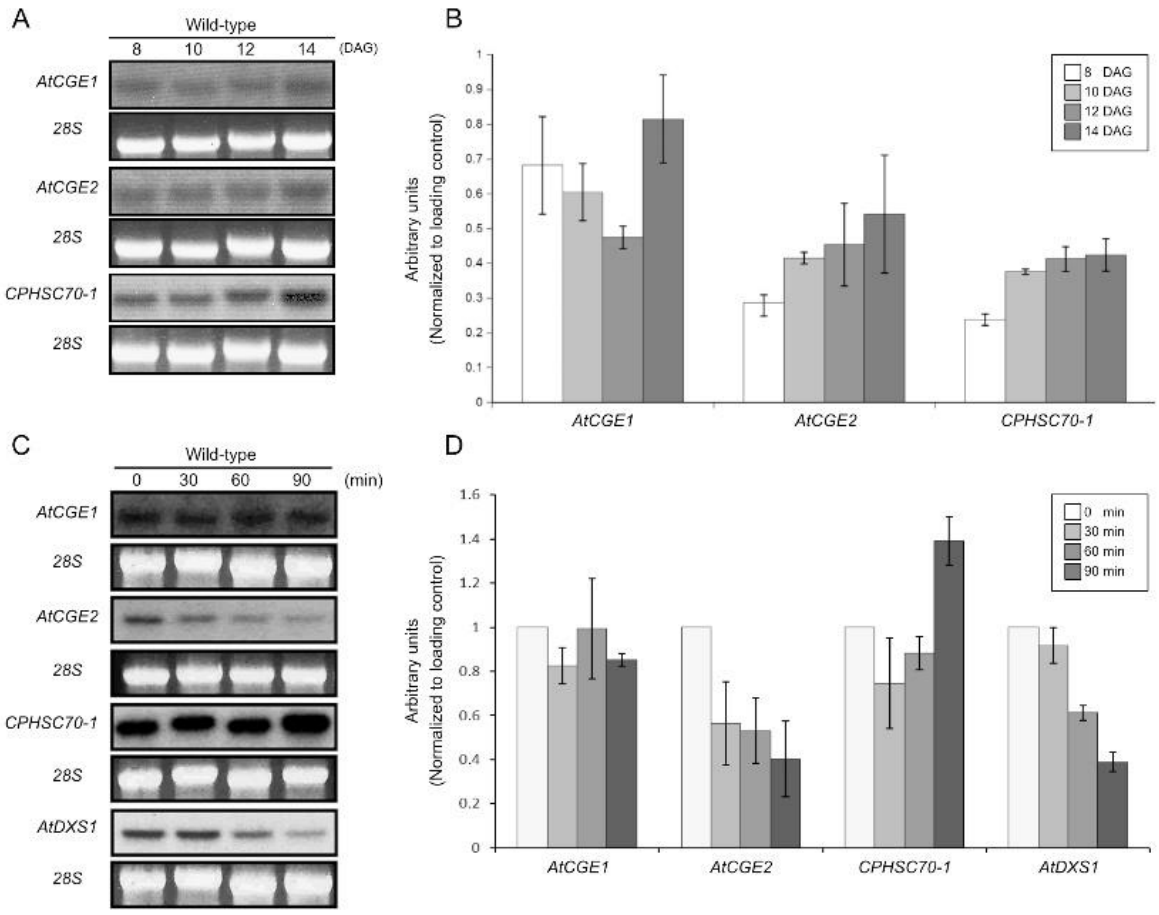


Figura 13. Análisis de la acumulación de los transcritos AtCGE1, AtCGE2 y CPHSC70-1. Se muestra un análisis por Northern blot de los transcritos AtCGE1, AtCGE2 y CPHSC70-1 en cuatro momentos del desarrollo de plántulas Wt de *A. thaliana* (A) y su densitometría correspondiente (B). Análisis por Northern blot de los transcritos AtCGE1, AtCGE2, CPHSC70-1 y DXS1 de plántulas Wt de *A. thaliana* incubadas a 40 °C durante 0, 30, 60 y 90 minutos (C), se incluye la densitometría correspondiente (D). Las barras de error representan el error estándar calculado a partir de 3 experimentos independientes.

En conjunto, estos resultados indican que, al igual que otras chaperonas de la familia Hsp70 [Yura *et al*, 1993; Schroda *et al*, 2001], la proteína cpHsc70-1 podría mediar las respuestas a estrés por calor en los cloroplastos de *A. thaliana*. A pesar de que nuestros resultados no muestran evidencias de una regulación positiva de los genes de las *AtCGEs* en respuesta al tratamiento, nosotros hipotetizamos que la AtCGE1 es probablemente la CGE involucrada en las respuestas a estrés por calor en los cloroplastos de *A. thaliana*, puesto que la abundancia de su transcrito no disminuye tan drásticamente como lo hace el transcrito de AtCGE2.

Finalmente, debido a que el fenotipo letal presentado por las plantas homocigotas de la línea mutante *emb1241* no permite realizar una caracterización bioquímica de los defectos causados por la mutación de *AtCGE1*, nos dimos a la tarea de caracterizar a nivel bioquímico a las mutantes de arabidopsis heterocigotas de las líneas *emb1241-1* y *emb1241-2*, así como a la línea homocigota *Dcphsc70-1* (Figura S3). Nuestros análisis muestran que las mutantes *emb1241-2* y *Dcphsc70-1* tienen niveles significativamente reducidos de clorofila *b* (Figura 14A), lo cual es un indicio de la existencia de alteraciones en el aparato fotosintético. Para respaldar estas observaciones, se analizó la acumulación de las proteínas PSI-D1 y PSII-D1 de los centros de reacción de los fotosistemas I y II, respectivamente. Nuestros ensayos indican que las líneas *emb1241-2* y *Dcphsc70-1* muestran niveles incrementados de acumulación de las proteínas PSI-D1 y PSII-D1 (Figura 14B) en comparación con las plantas Wt, sugiriendo que la estequiometría de los centros de reacción o de otras partes del aparato fotosintético podría estar alterada en estas plantas. Con la intención de profundizar en este fenómeno, se prepararon geles azules nativos utilizando extractos de proteínas totales de plantas

Wt, *emb1241-2* y *Dcphsc70-1* (Figura 14C). Nuestros resultados indican que, en contraste con las plantas tipo silvestre, las mutantes *emb1241-2* y *Dcphsc70-1* acumulan niveles menores del oligómero funcional del complejo captador de luz LHCII, mientras que aumentan los niveles del monómero correspondiente (Figura 14C), además de estas diferencias, no detectamos cambios significativos en ninguno de los otros complejos fotosintéticos que se resolvieron en los geles. Estos resultados indican que tanto AtCGE1 y cpHsc70-1 tienen algún papel en el establecimiento de la estequiometria de los fotosistemas, en particular en la oligomerización del complejo LHCII.

En este contexto, es importante mencionar que la actividad de la proteína cpHsc70-1 ha sido relacionada con varios procesos biológicos, entre los que destacan la termotolerancia, el desarrollo de los tilacoides, el importe de proteínas al cloroplasto y el ensamblaje del fotosistema II [Shi y Theg, 2010; Yalovsky *et al*, 1992; Schroda *et al*, 1999; Liu *et al*, 2007; Su y Li, 2010]. De igual manera, las cochaperonas cloroplásticas del tipo DnaJ que interaccionan con los blancos de las Hsp70s se han vinculado a procesos similares, como la oligomerización de la proteína VIPP1 durante la biogénesis de los tilacoides, la formación de dímeros del PSII y de los supercomplejos PSII-LHCII [Liu *et al*, 2007; Chen et al, 2010]. Esta información apoya nuestros resultados y refuerza nuestras hipótesis sobre la participación de la AtCGE1 en la oligomerización del LHCII, puesto que se sabe que los tres tipos de proteínas (Hsp70s, DnaJs y GrpEs) trabajan juntas en la homeostasis estructural de las proteínas blanco. En este sentido, nuestro trabajo representa el primer reporte que vincula a las proteínas CGEs con el LHCII, lo cual implica que la homeostasis de este complejo es mantenida por un sistema completo de chaperonas DnaK-DnaJ-GrpE en los cloroplastos de las plantas superiores.

Los resultados descritos en esta sección (7.4) se compilaron en la publicación número 8 incluida en el Apéndice II.

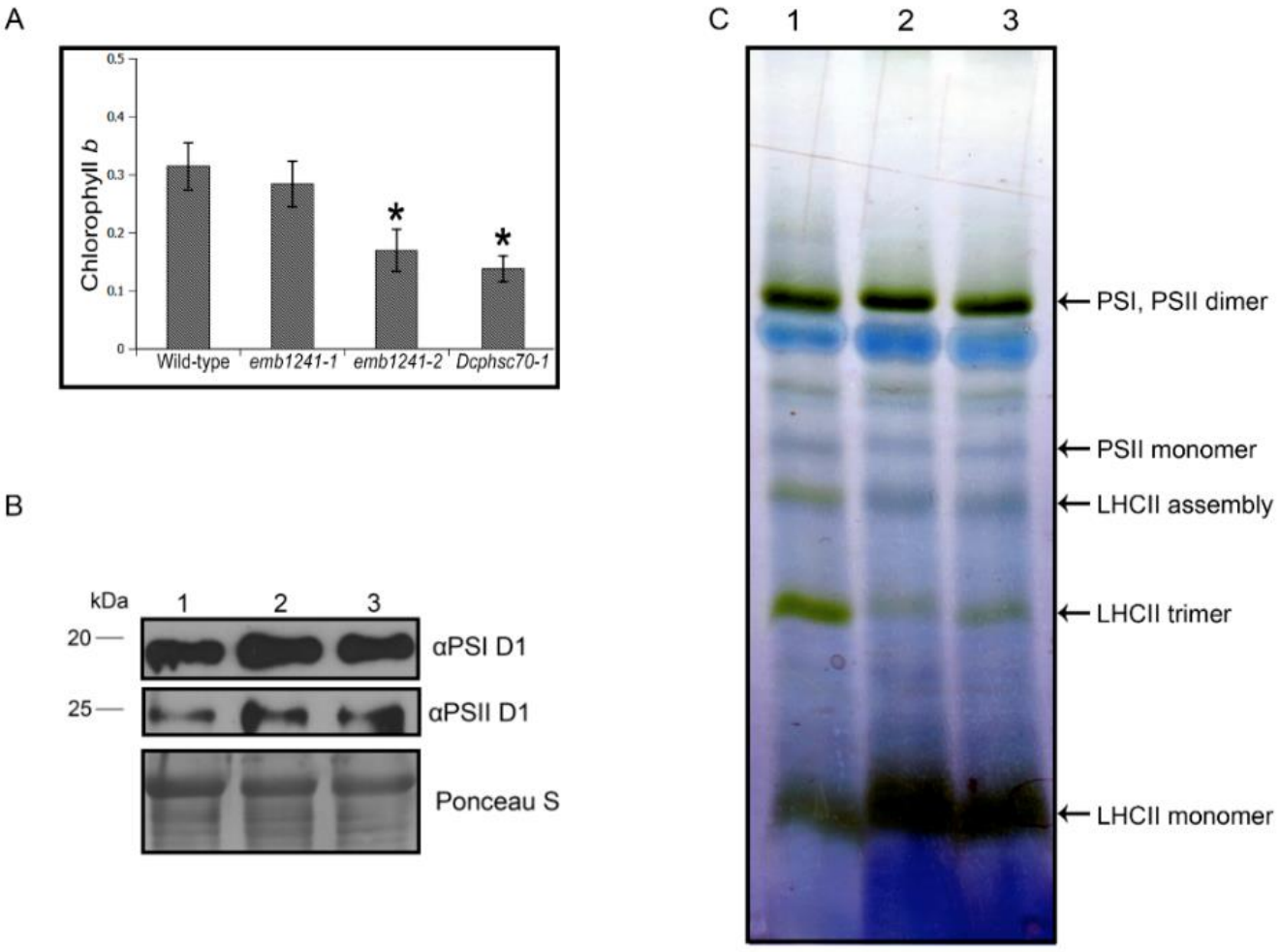


Figura 14. Análisis bioquímico de las mutantes afectadas en los genes EMB1241 y DCPHSC70-1. Se muestra un análisis de la acumulación de clorofila b en plantas tipo silvestre, *emb1241-1*, *emb1241-2* y *Dcphsc70-1* (A). Las barras de error representan el error estándar y el símbolo sobre estas (*) indica que hay diferencias significativas con un valor de $p<0.05$. También se muestran las inmunodetecciones de las proteínas PSI D1 y PSII D1 (B) en extractos de proteína total de plantas tipo silvestre (1), *emb1241-2* (2) y *Dcphsc70-1* (3). Finalmente, se muestra un gel azul nativo con extractos de proteína total de hojas de plantas tipo silvestre (1), *emb1241-2* (2) y *Dcphsc70-1* (3). Las flechas indican a las bandas que corresponden a los complejos supramoleculares visibles en las muestras.

8. Conclusión.

En conclusión, nuestros resultados demuestran que las tres líneas mutantes analizadas *atrabe1b*, *pbp1* y *emb2141* tienen defectos en el desarrollo del cloroplasto, y que estos defectos ocurren en distintas etapas del proceso en cada una de estas líneas mutantes. Por otro lado, si bien el análisis de las proteínas ATRABE1B y PBP1 no resultó concluyente debido a dificultades experimentales, nosotros demostramos que las proteínas AtCGE1 y AtCGE2 son los ortólogos funcionales de la proteína GrpE bacteriana. Además, demostramos que tanto AtCGE1 como AtCGE2 pueden interaccionar con la chaperona cloroplástica cpHsc70-1 para regular procesos biológicos independientes, puesto que nuestra evidencia indica que es la AtCGE1 y no la AtCGE2 la que está involucrada en la homeostasis del oligómero funcional del LHCII.

9. Perspectivas.

Según los resultados obtenidos y discutidos, las perspectivas principales del trabajo son las siguientes:

- Corroborar por medio de mutagénesis dirigida la funcionalidad *in planta* de los motivos conservados en las AtCGEs tipo A y B.
- Determinar la participación de los heterodímeros AtCGE1-AtCGE2 en la función de la cpHsc70-1, mediante ensayos de BiFC-FRET que permitan asegurar la formación del complejo.
- Explorar la posible participación de AtCGE1 en el importe de proteínas al cloroplasto en *arabidopsis thaliana*, por medio de ensayos de importe *in vitro* con cloroplastos aislados de la mutante *emb1241-2 (+/-)*.
- Debido a la función predicha de la proteína ATRABE1B y al fenotipo de las mutantes *atrabe1b*, el cual es similar a los defectos en el desarrollo foliar que se producen al afectar genética o químicamente la traducción en el cloroplasto, resulta interesante explorar el papel que tiene ATRABE1B en este proceso. En este contexto, es necesario estandarizar un método que permita una fenotipificación robusta de la línea *atrabe1b,* y que siente la base para experimentos con inhibidores de la traducción en el cloroplasto que permitan esclarecer el papel de ATRABE1B en este proceso. De manera similar, debido al origen bacteriano de ATRABE1B, sería posible explorar su función en la traducción mediante la complementación de mutantes en el factor de elongación *Tu* en *E. coli*, así reforzando o refutando las conclusiones derivadas de los experimentos antes propuestos.
- Nuestra hipótesis principal en torno a PBP1 señala que esta proteína está involucrada en el importe de proteínas al cloroplasto a través de una vía no canónica que involucra al sistema endomembranal de la célula. Debido a las dificultades experimentales ya descritas, no nos fue posible abordar esta hipótesis. En este sentido, es necesario diseñar estrategias experimentales que permitan abordar distintos puntos clave en para nuestra hipótesis, como la localización dual de PBP1 en el cloroplasto y el RE, la colocalización entre PBP1 y proteínas importadas al cloroplasto a través de la vía secretoria, y la determinación de la capacidad de mutantes nulas e hipomórficas de *PBP1* para importar proteínas al cloroplasto.

Referencias

Albrecht V, Ingenfeld A, Apel K. Characterization of the snowy cotyledon 1 mutant of Arabidopsis thaliana: the impact of chloroplast elongation factor G on chloroplast development and plant vitality. Plant Mol Biol. 2006; 60: 507-518.

Armbruster U, Hertle A, Makarenko E, Zühlke J, Pribil M, Dietzmann A, Schliebner I, Aseeva E, Fenino E, Scharfenberg M, Voigt C, Leister D. Chloroplast proteins without cleavable transit peptides: rare exceptions or a major constituent of the chloroplast proteome? Mol Plant. 2009; 2:1325-1335.

Arvidsson P, Sundby C. A model for the topology of the chloroplast thylakoid membrane. *Funct. Plant Biol.* 1999. 26: 687–694.

Asatsuma S, Sawada C, Itoh K, Okito M, Kitajima A, Mitsui T. Involvement of alpha-amylase I-1 in starch degradation in rice chloroplasts. Plant Cell Physiol. 2005; 46: 858-869.

Azem A, Oppliger W, Lustig A, Jenö P, Feifel B, Schatz G, Horst M. The mitochondrial hsp70 chaperone system. Effect of adenine nucleotides, peptide substrate, and mGrpE on the oligomeric state of mhsp70. J Biol Chem. 1997; 272: 20901-20906.

Bailey TL. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics. 2011; 27: 1653-1659.

Baldwin A, Wardle A, Patel R, Dudley P, Park SK, Twell D, Inoue K, Jarvis P. A Molecular-Genetic Study of the Arabidopsis Toc75 Gene Family. Plant Physiology. 2005; 138: 715-733.

Ball SG, Colleoni C, Kadouche D, Ducatez M, Arias MC, Tirtiaux C. Toward an understanding of the function of Chlamydiales in plastid endosymbiosis. Biochim Biophys Acta. 2015; 1847: 495-504.

Balsera M, Goetze TA, Kovács-Bogdán E, Schürmann P, Wagner R, Buchanan BB, Soll J, Bölter B. Characterization of Tic110, a channel-forming protein at the inner envelope membrane of chloroplasts, unveils a response to Ca(2+) and a stromal regulatory disulfide bridge. J Biol Chem. 2009; 284: 2603-2616.

Barajas-López JD, Kremnev D, Shaikhali J, Piñas-Fernández A, Strand Å. PAPP5 Is Involved in the Tetrapyrrole Mediated Plastid Signalling during Chloroplast Development. PLoS ONE. 2013; 8: e60305

Barre A, Bourne Y, Van Damme EJ, Peumans WJ, Rougé P. Mannose-binding plant lectins: different structural scaffolds for a common sugar-recognition process. Biochimie. 2001; 83: 645-651.

Bauer J, Chen K, Hiltbunner A, Wehrli E, Eugster M, Schnell D, Kessler F. The major protein import receptor of plastids is essential for chloroplast biogenesis. Nature. 2000; 403: 203-207.

Baum D. The origin of primary plastids: a pas de deux or a ménage à trois? Plant Cell. 2013; 25: 4-6.

Bédard J, Trösch R, Wu F, Ling Q, Flores-Pérez Ú, Töpel M, Nawaz F, Jarvis P.Suppressors of the Chloroplast Protein Import Mutant tic40 Reveal a Genetic Link between Protein Import and Thylakoid Biogenesis. The Plant Cell. 2017; 29: 1726-1747.

Beltrán J, Wamboldt Y, Sanchez R, LaBrant EW, Kundariya H, Virdi KS, Elowsky C, Mackenzie SA. Specialized Plastids Trigger Tissue-Specific Signaling for Systemic Stress Response in Plants. Plant Physiol. 2018; 178: 672-683.

Bisanz-Seyer C, Li YF, Seyer P, Mache R. The components of the plastid ribosome are not accumulated synchronously during the early development of spinach plants. Plant Mol Biol. 1989; 12: 201-211.

Bréhélin C, Kessler F. The plastoglobule: a bag full of lipid biochemistry tricks. Photochem Photobiol. 2008; 84: 1388-1394.

Brehmer D, Gässler C, Rist W, Mayer MP, Bukau B. Influence of GrpE on DnaK-substrate interactions. J Biol Chem. 2004; 279: 27957-27964.

Breuers FK, Bräutigam A, Geimer S, Welzel UY, Stefano G, Renna L, Brandizzi F, Weber AP. Dynamic Remodeling of the Plastid Envelope Membranes - A Tool for Chloroplast Envelope in vivo Localizations. Front Plant Sci. 2012; 3:7.

Bruick RK, Mayfield SP. Light-activated translation of chloroplast mRNAs. Trends Plant Sci. 1999; 4: 190-195.

Chan KX, Phua SY, Crisp P, McQuinn R, Pogson BJ. Learning the Languages of the Chloroplast: Retrograde Signaling and Beyond. Annu Rev Plant Biol. 2016; 67: 25-53.

Chateigner-Boutin AL, Ramos-Vega M, Guevara-García A, Andrés C, de la Luz Gutiérrez-Nava M, Cantero A, Delannoy E, Jiménez LF, Lurin C, Small I, León P. CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. Plant J. 2008; 56: 590-602.

Chen C, MacCready JS, Ducat DC, Osteryoung KW. The Molecular Machinery of Chloroplast Division. Plant Physiology. 2017; 176: 138-151.

Chen KM, Holmström M, Raksajit W, Suorsa M, Piippo M, Aro EM. Small chloroplast-targeted DnaJ proteins are involved in optimization of photosynthetic reactions in Arabidopsis thaliana. BMC Plant Biol. 2010; 10: 43.

Chen MH, Huang LF, Li HM, Chen YR, Yu SM. Signal Peptide-Dependent Targeting of a Rice α-Amylase and Cargo Proteins to Plastids and Extracellular Compartments of Plant Cells. Plant Physiology. 2004; 135: 1367-1377.

Chen X, Smith MD, Fitzpatrick L, Schnell DJ. In vivo analysis of the role of atTic20 in protein import into chloroplasts. Plant Cell. 2002; 14: 641-654.

Chen Y, Zhou B, Li J, Tang H, Tang J, Yang Z. Formation and Change of Chloroplast-Located Plant Metabolites in Response to Light Conditions. Int J Mol Sci. 2018; 19 pii: E654.

Chiu CC, Chen LJ, Su PH, Li HM. Evolution of chloroplast J proteins. PLoS One. 2013; 8: e70384.

Chou ML, Fitzpatrick LM, Tu SL, Budziszewski G, Potter-Lewis S, Akita M, Levin JZ, Keegstra K, Li HM. Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. EMBO J. 2003; 22: 2970-2980.

de la Luz Gutiérrez-Nava M, Gillmor CS, Jiménez LF, Guevara-García A, León P. CHLOROPLAST BIOGENESIS Genes Act Cell and Noncell Autonomously in Early Chloroplast Development. Plant Physiology. 2004; 135: 471-482.

de Luna-Valdez L, Guevara-García A. Análisis proteómico comparativo de mutantes de *Arabidopsis thaliana* afectadas en la biogénesis del cloroplasto. Tesis de Maestría, Universidad Nacional Autónoma de México. 2012.

de Souza A, Wang JZ, Dehesh K. Retrograde Signals: Integrators of Interorganellar Communication and Orchestrators of Plant Development. Annu Rev Plant Biol. 2017; 68: 85-108.

Deloche O, Georgopoulos C. Purification and biochemical properties of Saccharomyces cerevisiae's Mge1p, the mitochondrial cochaperone of Ssc1p. J Biol Chem. 1996; 271: 23960-23966.

Deloche O, Kelley WL, Georgopoulos C. Structure-function analyses of the Ssc1p, Mdj1p, and Mge1p Saccharomyces cerevisiae mitochondrial proteins in Escherichia coli. J Bacteriol. 1997; 179: 6066-6075.

Emanuelsson O, Nielsen H, Brunak S, von Heijne G. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J Mol Biol. 2000; 300: 1005-1016.

Estavillo GM, Crisp PA, Pornsiriwong W, Wirtz M, Collinge D, Carrie C, Giraud E, Whelan J, David P, Javot H, Brearley C, Hell R, Marin E, Pogson BJ. Evidence for a SAL1-PAP chloroplast retrograde pathway that functions in drought and highlight signaling in Arabidopsis. Plant Cell. 2011; 23: 3992-4012.

Farmaki T, Sanmartín M, Jiménez P, Paneque M, Sanz C, Vancanneyt G, León J, Sánchez-Serrano JJ. Differential distribution of the lipoxygenase pathway enzymes within potato chloroplasts. J Exp Bot. 2007; 58: 555-568.

Fink AL. Chaperone-mediated protein folding. Physiol Rev. 1999; 79: 425-449.

Galvez-Valdivieso G, Fryer MJ, Lawson T, Slattery K, Truman W, Smirnoff N, Asami T, Davies WJ, Jones AM, Baker NR, Mullineaux PM. The high light response in Arabidopsis

involves ABA signaling between vascular and bundle sheath cells. Plant Cell. 2009; 21: 2143-2162.

Gao H, Sage TL, Osteryoung KW. FZL, an FZO-like protein in plants, is a determinant of thylakoid and chloroplast morphology. Proc Natl Acad Sci U S A. 2006; 103: 6759-6764.

Gao J, Wang H, Yuan Q, Feng Y. Structure and Function of the Photosystem Supercomplexes. Front Plant Sci. 2018; 9: 357.

Gelinas AD, Toth J, Bethoney KA, Langsetmo K, Stafford WF, Harrison CJ. Thermodynamic linkage in the GrpE nucleotide exchange factor, a molecular thermosensor. Biochemistry. 2003; 42: 9050-9059.

Gelinas AD, Toth J, Bethoney KA, Stafford WF, Harrison CJ. Mutational analysis of the energetics of the GrpE.DnaK binding interface: equilibrium association constants by sedimentation velocity analytical ultracentrifugation. J Mol Biol. 2004; 339: 447-458.

Guevara-García A, San Román C, Arroyo A, Cortés ME, de la Luz Gutiérrez-Nava M, León P. Characterization of the Arabidopsis clb6 mutant illustrates the importance of posttranscriptional regulation of the methyl-D-erythritol 4-phosphate pathway. Plant Cell. 2005; 17: 628-643.

Hajdukiewicz PT, Allison LA, Maliga P. The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J. 1997; 16: 4041-4048.

Harrak H, Lagrange T, Bisanz-Seyer C, Lerbs-Mache S, Mache R. The expression of nuclear genes encoding plastid ribosomal proteins precedes the expression of chloroplast genes during early phases of chloroplast development. Plant Physiol. 1995; 108: 685-692.

Harrison C, Hayer-Hartl M, Di Liberto M, Hartl F, Kuriyan J. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science. 1997; 276:431-435.

Harrison C. GrpE, a nucleotide exchange factor for DnaK. Cell Stress Chaperones. 2003; 8: 218-224.

Heinemeyer J, Lewejohann D, Braun HP. Blue-native gel electrophoresis for the characterization of protein complexes in plants. Methods Mol Biol. 2007; 355: 343-352.

Heins L, Mehrle A, Hemmler R, Wagner R, Küchler M, Hörmann F, Sveshnikov D, Soll J. The preprotein conducting channel at the inner envelope membrane of plastids. The EMBO Journal. 2002; 21: 2616-2625.

Hernández-Verdeja T, Strand Å. Retrograde Signals Navigate the Path to Chloroplast Development. Plant Physiology. 2017; 176: 967-976.

Hinnah SC, Wagner R, Sveshnikova N, Harrer R, Soll J. The chloroplast protein import channel Toc75: pore properties and interaction with transit peptides. Biophys J. 2002; 83: 899-911.

Hirose T, Sugiura M. Functional Shine-Dalgarno-like sequences for translational initiation of chloroplast mRNAs. Plant Cell Physiol. 2004; 45: 114-117.

Hofmann NR, Theg SM. Chloroplast outer membrane protein targeting and insertion. Trends Plant Sci. 2005; 10: 450-457.

Huang J, Gogarten JP. Did an ancient chlamydial endosymbiosis facilitate the establishment of primary plastids? Genome Biol. 2007; 8: R99.

Inaba T, Li M, Alvarez-Huerta M, Kessler F, Schnell DJ. atTic110 functions as a scaffold for coordinating the stromal events of protein import into chloroplasts. J Biol Chem. 2003; 278: 38617-38627.

Inoue H, Li M, Schnell DJ. An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. Proc Natl Acad Sci U S A. 2013; 110: 3173-3178.

Inoue K, Keegstra K. A polyglycine stretch is necessary for proper targeting of the protein translocation channel precursor to the outer envelope membrane of chloroplasts. Plant J. 2003; 34: 661-669.

Jarvis P, López-Juez E. Biogenesis and homeostasis of chloroplasts and other plastids. Nat Rev Mol Cell Biol. 2013; 14: 787-802

Jones DT, Taylor WR, Thornton JM. The rapid generation of mutation data matrices from protein sequences. Comput Appl Biosci. 1992; 8: 275-282.

Keeling PJ. The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Lond B Biol Sci. 2010; 365: 729-748.

Kelly AA, Dörmann P. Green light for galactolipid trafficking. Curr Opin Plant Biol. 2004; 7: 262-269.

Kikuchi S, Bédard J, Hirano M, Hirabayashi Y, Oishi M, Imai M, Takase M, Ide T, Nakai M. Uncovering the protein translocon at the chloroplast inner envelope membrane. Science. 2013; 339: 571-574.

Kikuchi S, Oishi M, Hirabayashi Y, Lee DW, Hwang I, Nakai M. A 1-megadalton translocation complex containing Tic20 and Tic21 mediates chloroplast protein import at the inner envelope membrane. Plant Cell. 2009; 21: 1781-1797.

Kobayashi K, Kondo M, Fukuda H, Nishimura M, Ohta H. Galactolipid synthesis in chloroplast inner envelope is essential for proper thylakoid biogenesis, photosynthesis, and embryogenesis. Proc Natl Acad Sci U S A. 2007; 104: 17216-17221.

Kolodner RD, Tewari KK. Chloroplast DNA from higher plants replicates by both the Cairns and the rolling circle mechanism. Nature. 1975; 256: 708-711.

Koochak H, Puthiyaveetil S, Mullendore DL, Li M, Kirchhoff H. The structural and functional domains of plant thylakoid membranes. Plant J.; 97: 412-429.

Kouřil R, Dekker JP, Boekema EJ. Supramolecular organization of photosystem II in green plants. Biochim Biophys Acta. 2012; 1817: 2-12.

Kovacheva S, Bédard J, Patel R, Dudley P, Twell D, Ríos G, Koncz C, Jarvis P. In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. Plant J. 2005; 41: 412-428.

Kovács-Bogdán E, Benz JP, Soll J, Bölter B. Tic20 forms a channel independent of Tic110 in chloroplasts. BMC Plant Biol. 2011; 11: 133.

Kroll D, Meierhoff K, Bechtold N, Kinoshita M, Westphal S, Vothknecht UC, Soll J, Westhoff P. VIPP1, a nuclear gene of Arabidopsis thaliana essential for thylakoid membrane formation. Proc Natl Acad Sci U S A. 2001; 98: 4238-4242.

Kubis S, Patel R, Combe J, Bédard J, Kovacheva S, Lilley K, Biehl A, Leister D, Ríos G, Koncz C, Jarvis P. Functional specialization amongst the Arabidopsis Toc159 family of chloroplast protein import receptors. Plant Cell. 2004; 16: 2059-2077.

Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol. 2016; 33:1870-1874.

Laemmli UK. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature. 1970; 227: 680-685.

Lande NV, Subba P, Barua P, Gayen D, Keshava Prasad TS, Chakraborty S, Chakraborty N. Dissecting the chloroplast proteome of chickpea (Cicer arietinum L.) provides new insights into classical and non-classical functions. J Proteomics. 2017; 165: 11-20.

Lee DW, Jung C, Hwang I. Cytosolic events involved in chloroplast protein targeting. Biochim Biophys Acta. 2013; 1833: 245-252.

Lee DW, Kim JK, Lee S, Choi S, Kim S, Hwang I. Arabidopsis nuclear-encoded plastid transit peptides contain multiple sequence subgroups with distinctive chloroplast-targeting sequence motifs. Plant Cell. 2008; 20: 1603-1622.

Lee DW, Yoo YJ, Razzak MA, Hwang I. Prolines in Transit Peptides Are Crucial for Efficient Preprotein Translocation into Chloroplasts. Plant Physiol. 2018; 176: 663-677.

Lee KP, Kim C, Landgraf F, Apel K. EXECUTER1- and EXECUTER2-dependent transfer of stress-related signals from the plastid to the nucleus of Arabidopsis thaliana. Proc Natl Acad Sci U S A. 2007; 104: 10270-10275.

Leister D. Chloroplast research in the genomic age. Trends Genet. 2003; 19: 47-56.

Leuzinger K, Dent M, Hurtado J, Stahnke J, Lai H, Zhou X, Chen Q. Efficient agroinfiltration of plants for high-level transient expression of recombinant proteins. J Vis Exp. 2013; 77.

Liberek K, Marszalek J, Ang D, Georgopoulos C, Zylicz M. Escherichia coli DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. Proc Natl Acad Sci U S A. 1991; 88: 2874-2878.

Lichtenthaler H, Wellburn A. Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. Biochem. Soc. Trans. 1983; 11: 591–592.

Liu C, Willmund F, Golecki JR, Cacace S, Hess B, Markert C, Schroda M. The chloroplast HSP70B-CDJ2-CGE1 chaperones catalyse assembly and disassembly of VIPP1 oligomers in Chlamydomonas. Plant J. 2007; 50: 265-277.

Liu C, Willmund F, Whitelegge JP, Hawat S, Knapp B, Lodha M, Schroda M. J-domain protein CDJ2 and HSP70B are a plastidic chaperone pair that interacts with vesicle-inducing protein in plastids 1. Mol Biol Cell. 2005; 16: 1165-1177.

Liu D, Li W, Cheng J. The novel protein DELAYED PALE-GREENING1 is required for early chloroplast biogenesis in Arabidopsis thaliana. Sci Rep. 2016; 6: 25742.

Liu J, Lu Y, Hua W, Last RL. A New Light on Photosystem II Maintenance in Oxygenic Photosynthesis. Frontiers in Plant Science. 2019; 10: 975

Liu X, Rodermel SR, Yu F. A var2 leaf variegation suppressor locus, SUPPRESSOR OF VARIEGATION3, encodes a putative chloroplast translation elongation factor that is important for chloroplast development in the cold. BMC Plant Biology. 2010; 10: 287.

Loll B, Kern J, Saenger W, Zouni A, Biesiadka J. Towards complete cofactor arrangement in the 3.0 A resolution structure of photosystem II. Nature. 2005; 438: 1040-1044.

Lv F, Zhou J, Zeng L, Xing D. β-cyclocitral upregulates salicylic acid signaling to enhance excess light acclimation in Arabidopsis. J Exp Bot. 2015; 66: 4719-32.

Martin W, Stoebe B, Goremykin V, Hapsmann S, Hasegawa M, Kowallik KV. Gene transfer to the nucleus and the evolution of chloroplasts. Nature. 1998; 393:162-165.

Mayer MP. Gymnastics of molecular chaperones. Mol Cell. 2010; 39: 321-331.

Mochizuki N, Brusslan JA, Larkin R, Nagatani A, Chory J. Arabidopsis genomes uncoupled 5 (GUN5) mutant reveals the involvement of Mg-chelatase H subunit in plastid-to-nucleus signal transduction. Proc Natl Acad Sci U S A. 2001; 98: 2053-2058.

Morré DJ, Selldén G, Sundqvist C, Sandelius AS. Stromal low temperature compartment derived from the inner membrane of the chloroplast envelope. Plant Physiol. 1991; 97: 1558-1564.

Nagano AJ, Matsushima R, Hara-Nishimura I. Activation of an ER-body-localized beta-glucosidase via a cytosolic binding partner in damaged tissues of Arabidopsis thaliana. Plant Cell Physiol. 2005; 46: 1140-1148.

Nelson BK, Cai X, Nebenführ A. A multicolored set of in vivo organelle markers for co-localization studies in Arabidopsis and other plants. Plant J. 2007; 51: 1126-1136.

Neuhaus H, Emes M. Nonphotosynthetic metabolism in plastids. Annu. Rev. Plant Physiol. Plant Mol. Biol. 2000. 51: 111-140.

Oldenburg DJ, Bendich AJ. DNA maintenance in plastids and mitochondria of plants. Frontiers in Plant Science. 2015; 6: 883.

Ostrovsky D, Diomina G, Lysak E, Matveeva E, Ogrel O, Trutko S. Effect of oxidative stress on the biosynthesis of 2-C-methyl-D-erythritol-2,4-cyclopyrophosphate and isoprenoids by several bacterial strains. Arch Microbiol. 1998; 171: 69-72.

Page MT, McCormac AC, Smith AG, Terry MJ. Singlet oxygen initiates a plastid signal controlling photosynthetic gene expression. The New Phytologist. 2016; 213: 1168-1180.

Pape T, Wintermeyer W, Rodnina MV. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E. coli ribosome. EMBO J. 1998; 17: 7490-7497.

Parker N, Wang Y, Meinke D. Natural Variation in Sensitivity to a Loss of Chloroplast Translation in Arabidopsis. Plant Physiology. 2014; 166: 2013-2027.

Perello C, Llamas E, Burlat V, Ortiz-Alcaide M, Phillips MA, Pulido P, Rodriguez-Concepcion M. Differential Subplastidial Localization and Turnover of Enzymes Involved in Isoprenoid Biosynthesis in Chloroplasts. PLoS One. 2016; 11: e0150539.

Powikrowska M, Oetke S, Jensen PE, Krupinska K. Dynamic composition, shaping and organization of plastid nucleoids. Front Plant Sci. 2014; 5: 424.

Qin G, Gu H, Ma L, Peng Y, Deng XW, Chen Z, Qu LJ. Disruption of phytoene desaturase gene results in albino and dwarf phenotypes in Arabidopsis by impairing chlorophyll, carotenoid, and gibberellin biosynthesis. Cell Res. 2007; 17: 471-482.

Salvi D, Moyet L, Seigneurin-Berny D, Ferro M, Joyard J, Rolland N. Preparation of envelope membrane fractions from Arabidopsis chloroplasts for proteomic analysis and other studies. Methods Mol Biol. 2011; 775: 189-206.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Tabata S. Complete structure of the chloroplast genome of Arabidopsis thaliana. DNA Res. 1999; 6: 283-290.

Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. Fiji: an open-source platform for biological-image analysis. Nat Methods. 2012; 9: 676-682.

Schnell DJ, Blobel G. Identification of intermediates in the pathway of protein import into chloroplasts and their localization to envelope contact sites. J Cell Biol. 1993; 120: 103-115.

Schnell DJ, Kessler F, Blobel G. Isolation of components of the chloroplast protein import machinery. Science. 1994; 266: 1007-1012.

Schroda M, Vallon O, Whitelegge JP, Beck CF, Wollman FA. The chloroplastic GrpE homolog of Chlamydomonas: two isoforms generated by differential splicing. Plant Cell. 2001; 13: 2823-2839.
Schroda M, Vallon O, Wollman FA, Beck CF. A chloroplast-targeted heat shock protein 70 (HSP70) contributes to the photoprotection and repair of photosystem II during and after photoinhibition. Plant Cell. 1999; 11: 1165-11678.

Schröder H, Langer T, Hartl FU, Bukau B. DnaK, DnaJ and GrpE form a cellular chaperone machinery capable of repairing heat-induced protein damage. EMBO J.1993;12: 4137-4144.

Sharkey TD. Discovery of the canonical Calvin-Benson cycle. Photosynth Res. 2019; 140: 235-252.

Sherameti I, Venus Y, Drzewiecki C, Tripathi S, Dan VM, Nitz I, Varma A, Grundler FM, Oelmüller R. PYK10, a beta-glucosidase located in the endoplasmatic reticulum, is crucial for the beneficial interaction between Arabidopsis thaliana and the endophytic fungus Piriformospora indica. Plant J. 2008; 54: 428-439.

Shi LX, Theg SM. A stromal heat shock protein 70 system functions in protein import into chloroplasts in the moss Physcomitrella patens. Plant Cell. 2010; 22: 205-220.

Shih PM, Matzke NJ. Primary endosymbiosis events date to the later Proterozoic with cross-calibrated phylogenetic dating of duplicated ATPase proteins. Proc Natl Acad Sci U S A. 2013; 110: 12355-60.

Shimoni E, Rav-Hon O, Ohad I, Brumfeld V, Reich Z. Three-dimensional organization of higher-plant chloroplast thylakoid membranes revealed by electron tomography. Plant Cell. 2005; 17: 2580-2586.

Sjuts I, Soll J, Bölter B. Import of Soluble Proteins into Chloroplasts andPotential Regulatory Mechanisms. Front Plant Sci. 2017; 8: 168.

Su PH, Li HM. Arabidopsis Stromal 70-kD Heat Shock Proteins Are Essential for Plant Development and Important for Thermotolerance of Germinating Seeds. Plant Physiology. 2008; 146: 1231.

Su PH, Li HM. Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. Plant Cell. 2010; 22: 1516-1531.

Sugimoto S, Saruwatari K, Higashi C, Sonomoto K. The proper ratio of GrpE to DnaK is important for protein quality control by the DnaK-DnaJ-GrpE chaperone system and for cell division. Microbiology. 2008; 154: 1876-1885.

Suh WC, Lu CZ, Gross CA. Structural features required for the interaction of the Hsp70 molecular chaperone DnaK with its cochaperone DnaJ. J Biol Chem. 1999; 274: 30534-30539.

Sulli C, Fang Z, Muchhal U, Schwartzbach SD. Topology of Euglena chloroplast protein precursors within endoplasmic reticulum to Golgi to chloroplast transport vesicles. J Biol Chem. 1999; 274: 457-463.

Sullivan JA, Gray JC. Plastid translation is required for the expression of nuclear photosynthesis genes in the dark and in roots of the pea lip1 mutant. Plant Cell. 1999; 11: 901-910.

Sundberg E, Slagter JG, Fridborg I, Cleary SP, Robinson C, Coupland G. ALBINO3, an Arabidopsis nuclear gene essential for chloroplast differentiation, encodes a chloroplast protein that shows homology to proteins present in bacterial membranes and yeast mitochondria. Plant Cell. 1997; 9: 717-730.

Theis J, Schroda M. Revisiting the photosystem II repair cycle. Plant Signaling & Behavior. 2016; 11: e1218587

van Wijk KJ, Baginsky S. Plastid Proteomics in Higher Plants: Current State and Future Goals. Plant Physiology. 2011; 155: 1578-1588.

Viana AA, Li M, Schnell DJ. Determinants for Stop-transfer and Post-import Pathways for Protein Targeting to the Chloroplast Inner Envelope Membrane. The Journal of Biological Chemistry. 2010; 285: 12948-12960.

Villarejo A, Burén S, Larsson S, Déjardin A, Monné M, Rudhe C, Karlsson J, Jansson S, Lerouge P, Rolland N, von Heijne G, Grebe M, Bako L, Samuelsson G. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat Cell Biol. 2005; 7: 1224-1231.

Voinnet O, Lederer C, Baulcombe DC. A viral movement protein prevents spread of the gene silencing signal in Nicotiana benthamiana. Cell. 2000; 103: 157-167.

von Heijne G, Nishikawa K. Chloroplast transit peptides. The perfect random coil?. FEBS Lett. 1991; 278: 1-3.

von Heijne G, Steppuhn J, Herrmann RG. Domain structure of mitochondrial and chloroplast targeting peptides. Eur J Biochem. 1989; 180: 535-545.

Vothknecht UC, Otters S, Hennig R, Schneider D. Vipp1: a very important protein in plastids?. J Exp Bot. 2012; 63: 1699-1712.

Walter M, Chaban C, Schütze K, Batistic O, Weckermann K, Näke C, Blazevic D, Grefen C, Schumacher K, Oecking C, Harter K, Kudla J. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J. 2004; 40: 428-438.

Wang Q, Sullivan RW, Kight A, Henry RL, Huang J, Jones AM, Korth KL. Deletion of the Chloroplast-Localized Thylakoid Formation1 Gene Product in Arabidopsis Leads to Deficient Thylakoid Formation and Variegated Leaves. Plant Physiology. 2004; 136: 3594-3604.

Waters MT, Langdale JA. The making of a chloroplast. EMBO J. 2009; 28: 2861-2873.

Wei X, Su X, Cao P, Liu X, Chang W, Li M, Zhang X, Liu Z. Structure of spinach photosystem II-LHCII supercomplex at 3.2 Å resolution. Nature. 2016; 534: 69-74.

Whatley, J. A Suggested Cycle of Plastid Developmental Interrelationships. The New Phytologist. 1978. 80: 489-502.

Willmund F, Mühlhaus T, Wojciechowska M, Schroda M. The NH2-terminal domain of the chloroplast GrpE homolog CGE1 is required for dimerization and cochaperone function in vivo. J Biol Chem. 2007; 282: 11317-11328.

Woodson JD, Perez-Ruiz JM, Chory J. Heme synthesis by plastid ferrochelatase I regulates nuclear gene expression in plants. Curr Biol. 2011; 21: 897-903.

Woodson JD, Perez-Ruiz JM, Schmitz RJ, Ecker JR, Chory J. Sigma factor-mediated plastid retrograde signals control nuclear gene expression. Plant J. 2013; 73: 1-13

Wu B, Wawrzynow A, Zylicz M, Georgopoulos C. Structure-function analysis of the Escherichia coli GrpE heat shock protein. EMBO J. 1996; 15: 4806-4816.

Xiao Y, Savchenko T, Baidoo EE, Chehab WE, Hayden DM, Tolstikov V, Corwin JA, Kliebenstein DJ, Keasling JD, Dehesh K. Retrograde signaling by the plastidial metabolite MEcPP regulates expression of nuclear stress-response genes. Cell. 2012; 149: 1525-35.

Yalovsky S, Paulsen H, Michaeli D, Chitnis PR, Nechushtai R. Involvement of a chloroplast HSP70 heat shock protein in the integration of a protein (light-harvesting complex protein precursor) into the thylakoid membrane. Proc Natl Acad Sci U S A. 1992; 89: 5616-5619.

Yokthongwattana K, Chrost B, Behrman S, Casper-Lindley C, Melis A. Photosystem II damage and repair cycle in the green alga Dunaliella salina: involvement of a chloroplast-localized HSP70. Plant Cell Physiol. 2001; 42: 1389-1397.

Yoshida Y, Miyagishima SY, Kuroiwa H, Kuroiwa T. The plastid-dividing machinery: formation, constriction and fission. Curr Opin Plant Biol. 2012; 15: 714-21.

Yura T, Nagai H, Mori H. Regulation of the heat-shock response in bacteria. Annu Rev Microbiol. 1993; 47:321-350.

Zhang L, Kato Y, Otters S, Vothknecht UC, Sakamoto W. Essential role of VIPP1 in chloroplast envelope maintenance in Arabidopsis. Plant Cell. 2012; 24: 3695-3707.

Zoschke R, Bock R. Chloroplast Translation: Structural and Functional Organization, Operational Control, and Regulation. The Plant Cell. 2018; 30: 745-770.

Zubo YO, Blakley IC, Franco-Zorrilla JM, Yamburenko MV, Solano R, Kieber JJ, Loraine AE, Schaller GE. Coordination of Chloroplast Development through the Action of the GNC and GLK Transcription Factor Families. Plant Physiol. 2018; 178:130-147.

Apéndice I. Material suplementario.

Tabla S1. Secuencia de los oligonucleótidos usados en el análisis de CGE1.

	Nombre	Secuencia (5'-3')
Clonación		
	CGE1-F	caccatggccggtctactc
	CGE1-R	cctctccctcttttatctc
	CGE1-ΔTP-F	aaaaagcaggcttcgaaggagatatacatatgtcgggagaagctgaga
	CGE1-ΔTP-R	agaaagctgggttacacacgcacacaaaa
	CGE2-F	aaaaagcaggcttctttcctttgggactcact
	cge2-Rp	agaaagctgggttagcagaaggtgttatttcc
	CGE2-ΔTP-F	aaaaagcaggcttcgaaggagatatacatatggcgaattcaaagcag
	CGE2-ΔTP-R	agaaagctgggttaactcaagcagaagg
	cpHsc70-1-F2	aaaaagcaggcttcatggcatcttcagccgcccaa
	cpHsc70-1-R2	agaaagctgggttttggctgtctgtgaagtcagc
Genotipificación		
	Int2-F	gagaatcttttggctgttttgg
	Int2-R	ttcttcttctgcgcttcct
	cge2-Rp	tttttgtttccgagtgagtgg
	cge2-Lp	actcttttatggtgggggatg
	cpHsc701-F	caccatggcatcttcagccgcccaa
	cpHsc701-R	ttggctgtctgtgaagtcagc
	SAIL-LB2	gcttcctattatatcttcccaaattaccaataca
	LB4-LB	cgtgtgccaggtgcccacggaatagt
	LBb1.3	attttgccgatttcggaac

Tabla S2. Porcentaje de conservación de residuos de aminoácidos funcionalmente relevantes en las CGEs Tipo A y Tipo.

Posición	Residuo en *E. coli*	Consenso en CGEs Tipo A (% conservación)	Consenso en CGEs Tipo B (% conservación)
73	R	R(98%)	R(100%)
74	R	K(98%)	K(97%)
82	K	S(27%)	S(68%)
122	G	G(98%)	S(86%)
183	R	R(94%)	R(85%)
192	V	V(100%)	V(86%)

Tabla S3. Cuantificación de pigmentos fotosintéticos.

	Pigmento (µg/mg FW)				
	Clorofila *a*	Clorofila *b*	Clorofila *a/b*	Carotenos	n*
Col-0	0.918±0.12	0.314±0.04	2.972±0.12	0.083±0.01	10
emb1241-1(+/-)	0.957±0.13	0.284±0.04	3.365±0.09	0.130±0.02	12
emb1241-2(+/-)	0.631±0.13	0.169±0.04	3.718±0.2	0.102±0.02	13
cge2-1	1.069±0.1	0.388±0.03	2.779±0.11	0.077±0.01	13
DcpHsc70-1	0.436±0.07	0.137±0.02	3.165±0.11	0.037±0.01	14

**Número de replicas independientes.*

Los datos mostrados incluyen el error estándar calculado de la población total muestreada.

Figura S1.

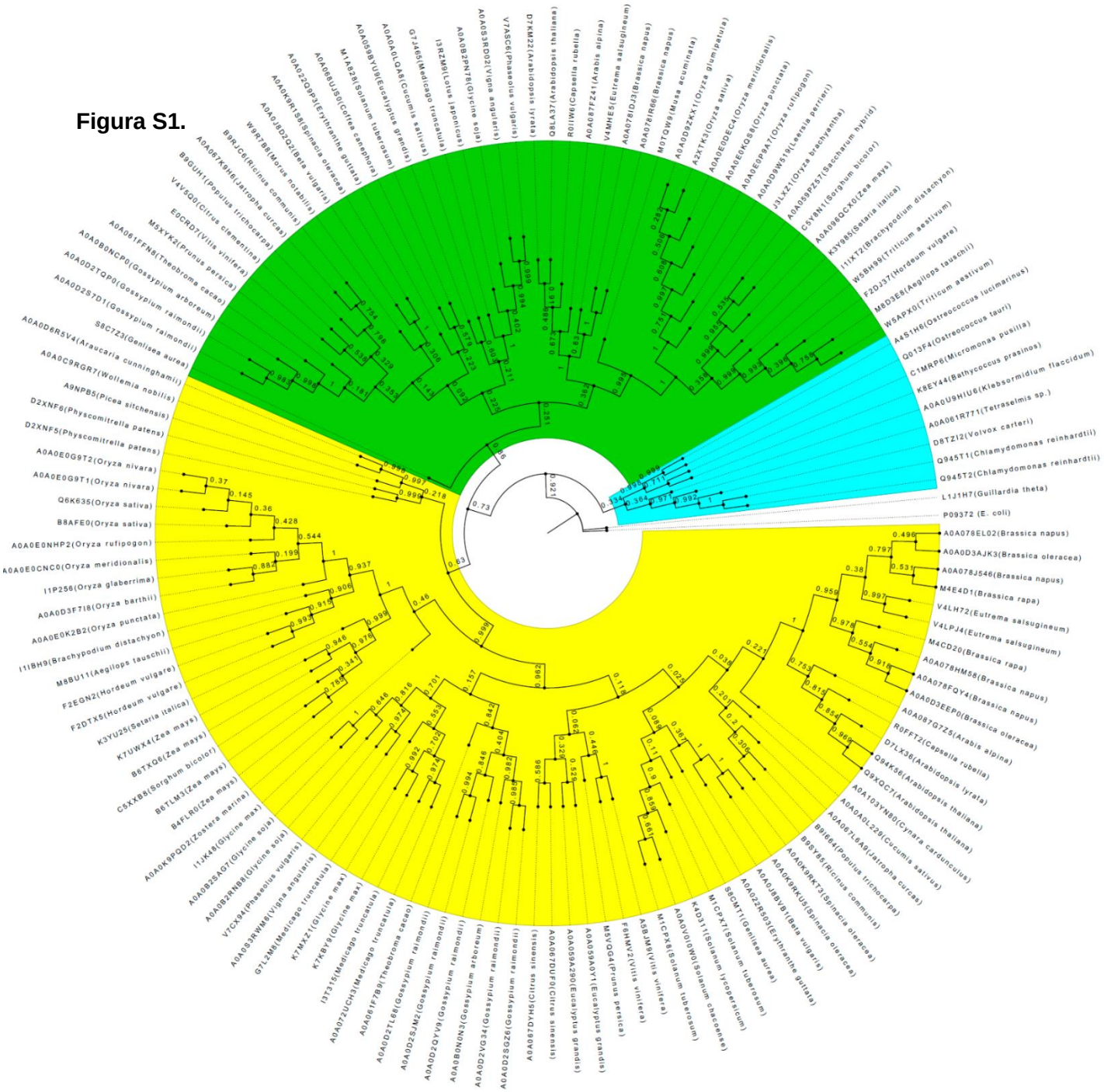


Figura S1. Análisis filogenético de las proteínas CGE. La historia evolutiva fue inferida usando el método de máxima verosimilitud basado en el modelo de matrices JTT. The evolutionary history was inferred by using the Maximum Likelihood method based on the JTT matrix-based model [Jones *et al*, 1992]. Se muestra el árbol con la mejor varosimilitud (log -17896.1536). El porcentaje de los árboles en que los taxones asociados se agruparon juntos se muestra junto a las ramas. Se muestran los tres clados principales de CGEs: Tipo algas (blue), Tipo A (verde), and Tipo B (amarillo). Las etiquetas de las ramas están en el formato: UniprotID (Especie).

Figura S2.

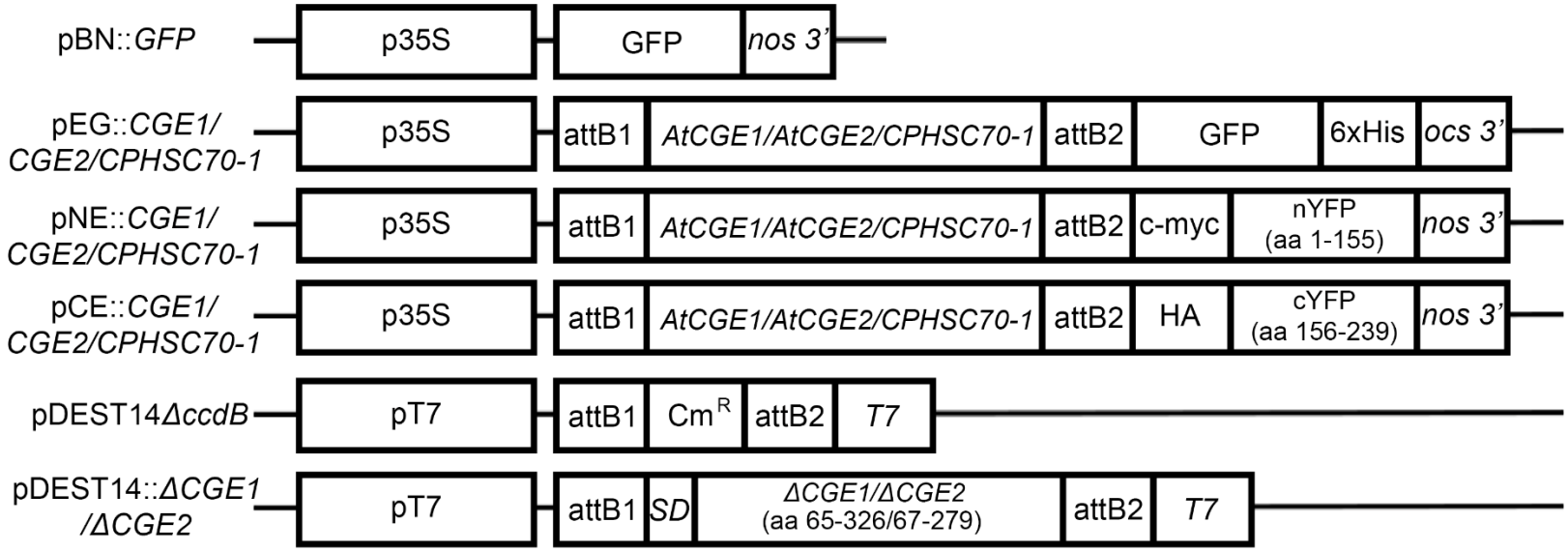


Figura S2. Construcciones genéticas usadas en el análisis de la CGE1. Diagramas de los vectores construidos para la expresión de proteinas fusionadas a la proteína verde fluorescente: GFP (pBN::*GFP*), AtCGE1-GFP (pEG::*CGE1*), AtCGE2-GFP (pEG::*CGE2*), cpHsc70-1-GFP (pEG::*CPHSC70-1*); para los ensayos de BiFC: AtCGE1-c-Myc-nYFP (pNE::*CGE1*), AtCGE2-c-Myc-nYFP (pNE::*CGE2*), cpHsc701-c-Myc-nYFP (pNE::*CPHSC70-1*), AtCGE1-HA-cYFP (pCE::*CGE1*), AtCGE2-HA-cYFP (pCE::*CGE2*), cpHsc701-HA-cYFP (pCE::*CPHSC70-1*); y para la expresión de proteínas en *E.coli*: ΔCGE1 (pD14::Δ*CGE1*), and ΔCGE2 (pD14::Δ*CGE2*). Lo elementos regulatorios incluidos son el promotor del virus del mosaico de la colifor (p35S), el promotor dependiente de la RNA polimerasa T7 (pT7), el terminador de la nopalina sintasa (nos 3'), el terminador de la octopina sintasa (ocs 3'), el terminador de la transcripción T7 (T7), la secuencia Shine-Dalgarno (SD), y los sitios de recombinacion (attB1 and attB2).

Figura S3.

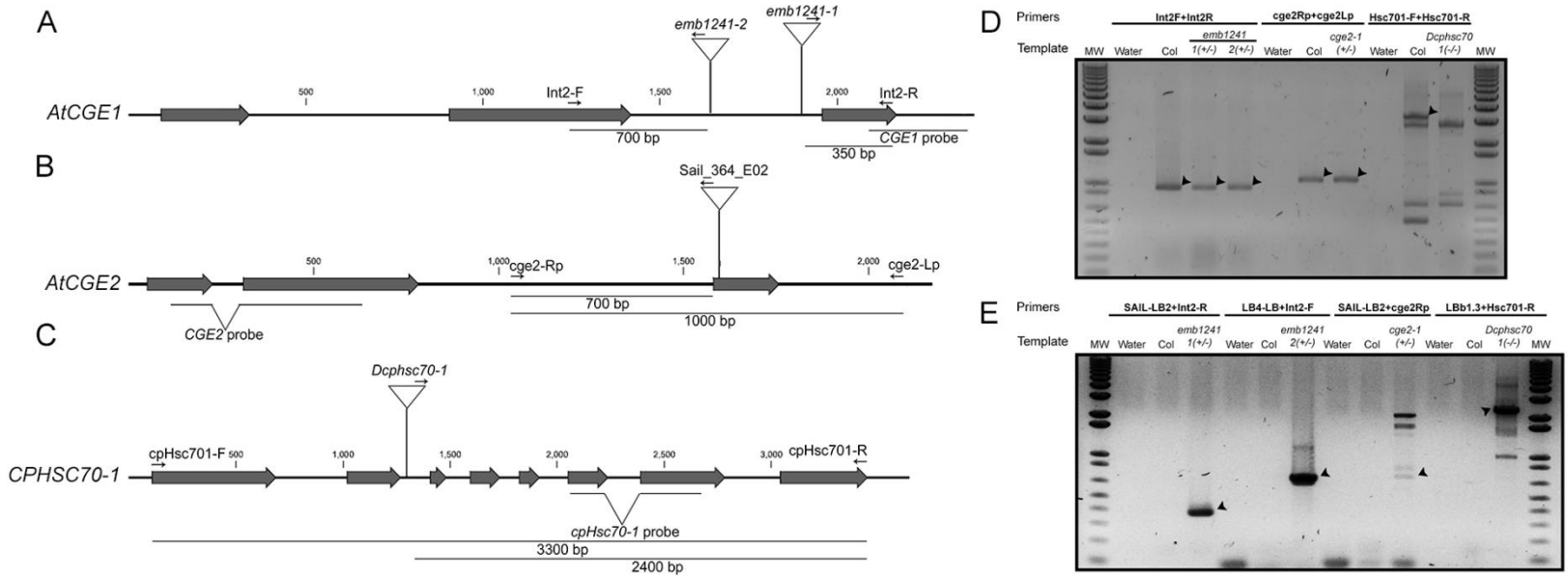


Figura S3. Genotipificación de líneas mutantes. Se muestra una representación de las regiones genómicas de los genes *AtCGE1* (A), *AtCGE2* (B), and *CPHSC70-1* (C). La posición de la inserción de T-DNA en cada una de las líneas mutantes analizadas se marca con triángulos vacíos. Las posiciones de los oligonucleótidos específicos usados para la genotipificación de las mutantes de inserción de T-DNA se marca con flechas. Los exones están representados por flechas grises y las regiones no codificantes por líneas rectas. El tamaño de los productos de PCR esperados para cada par de oligonucleótidos se indica debajo del diagrama correspondiente. Las regiones de DNA usadas para generar las sondas usadas en los Northern blots están subrayadas. Se muestra el análisis por PCR de los alelos no mutados (D) y mutados (E) de los genes *AtCGE1*, *AtCGE2*, and *CPHSC70-1* en DNA genómico extraído de plantas tipo silvestre (Col) y las líneas mutantes *emb1241-1 (+/-)*, *emb1241-2(+/-)*, *cge2-1(+/-)*, and *Dcphsc70-1.* Las puntas de flecha indican las bandas especificas obtenidas in la reacciones de PCR, se muestra un marcador de peso molecular en ambos extremos de las imágenes. Las secuencias de los oligonucleótidos están disponibles en la Tabla S1.

Figura S4.

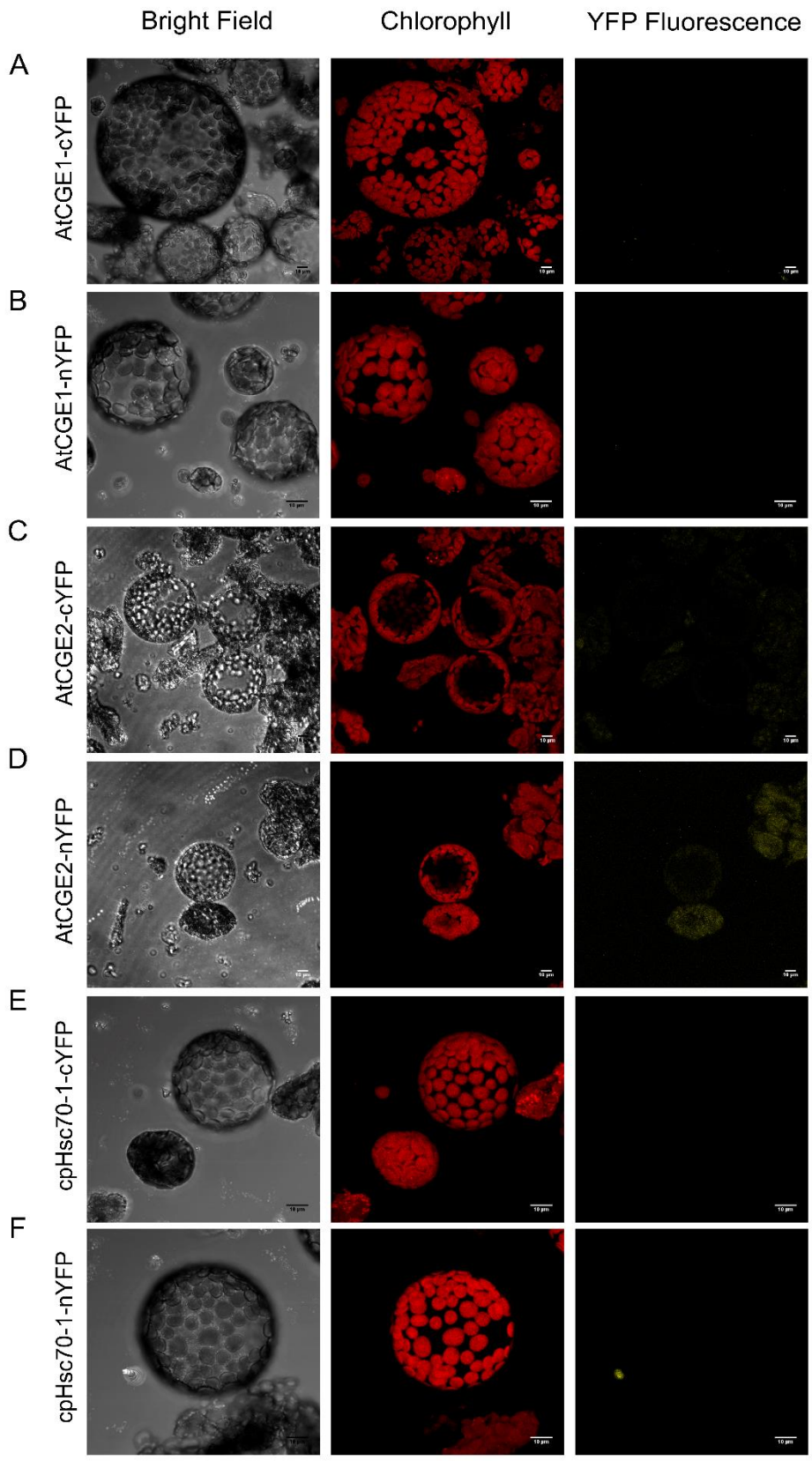

Bright Field
Chlorophyll
YFP Fluorescence
A
AtCGE1-cYFP
B
AtCGE1-nYFP
C
AtCGE2-cYFP
D
AtCGE2-nYFP
E
cpHsc70-1-cYFP
F
cpHsc70-1-nYFP

Figura S4. Controles para los ensayos de BiFC. Protoplastos de mesófilo de hojas de *N. benthamiana* transformadas con las fusiones traduccionales AtCGE1-cYFP (A), AtCGE1-nYFP (B), AtCGE2-cYFP(C), AtCGE2-nYFP, cpHsc70-1-cYFP (D), and cpHsc70-1-nYFP (E). Se muestran imágenes que corresponden a los canales del campo claro, la fluorescencia de la clorofila (Chlorophyll), y la fluorescencia de YFP (YFP fluorescence). Los protoplastos se prepararon 96 horas después de la agroinfiltración de las hojas de *N. benthamiana*. Las barras de escala corresponden a 10 µm.

Figura S5.

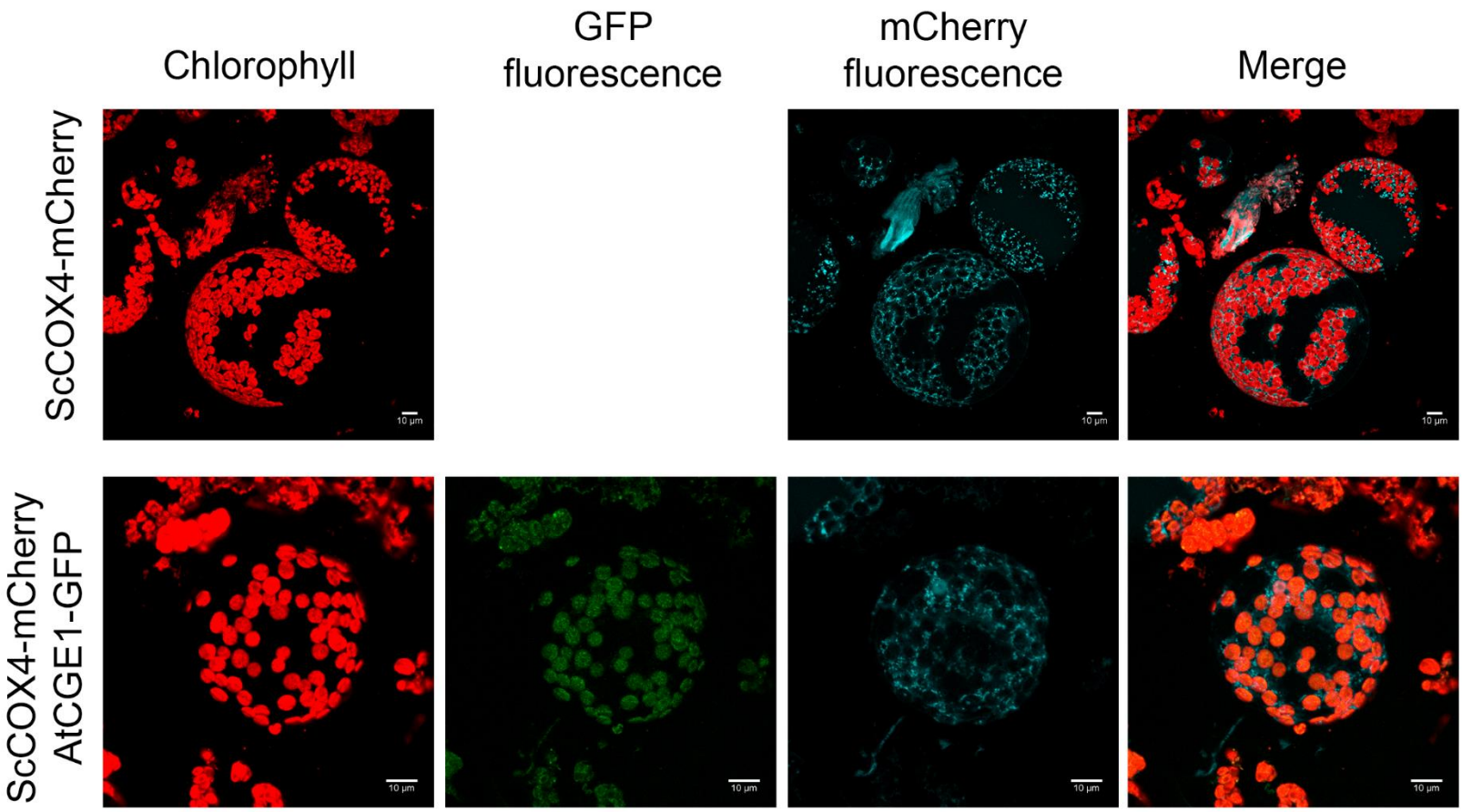


Figura S5. Coinfiltración de AtCGE1-GFP y tsCOX4-mCherry. Protoplastos de mesófilo de hojas de *N. benthamiana* transformadas con las fusiones traduccionales tsCOX4-mCherry (A), and tsCOX4-mCherry and AtCGE1-GFP (B). Se muestran imágenes que corresponden a los canales de la fluorescencia de la clorofila (Chlorophyll), la fluorescencia de GFP (GFP fluorescence), la fluorescencia de mCherry (mCherry fluorescence) pseudocoloreada en azul, y la unión entre los canales de fluorescencia de GFP y mCherry (Merge). Los protoplastos se prepararon 96 horas después de la agroinfiltración de las hojas de *N. benthamiana*. Las barras de escala corresponden a 10 µm.

Figura S6.

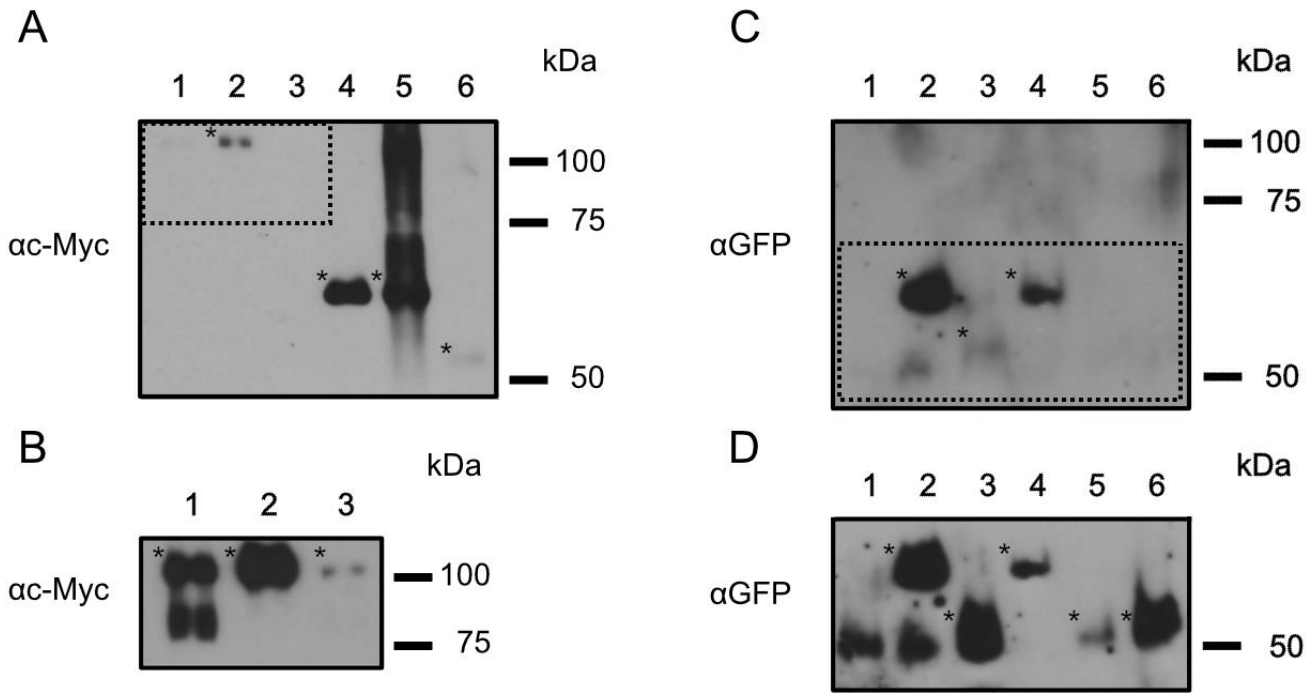


Figura S6. Experimentos de co-inmunoprecipitación. Análisis por Western-blot de las fusiones inmunoprecipitadas y abanderadas con c-Myc (A) y sus interactores co-inmunoprecipitados abanderados con GFP (C), purificados a partir de hojas co-transformadas con los plásmidos *pNE::CPHSC70-1* (c-Myc) y *pEG::CPHSC70-1* (1), *pNE::CPHSC70-1* (c-Myc) y *pEG::CGE1* (2), *pNE::CPHSC70-1* (c-Myc) y *pEG::CGE2* (3), *pNE::CGE1* (c-Myc) y *pEG::CGE1* (4), *pNE::CGE1* (c-Myc) y *pEG::CGE2* (5), y *pNE::CGE2* (c-Myc) y *pEG::CGE2* (6). Los paneles (B) y (D) muestran sobre exposiciones de las regiones marcadas con puntos en los paneles A y C, respectivamente. Los asteriscos señalan las bandas especificas esperadas en cada experimento (*).

Figura S7.

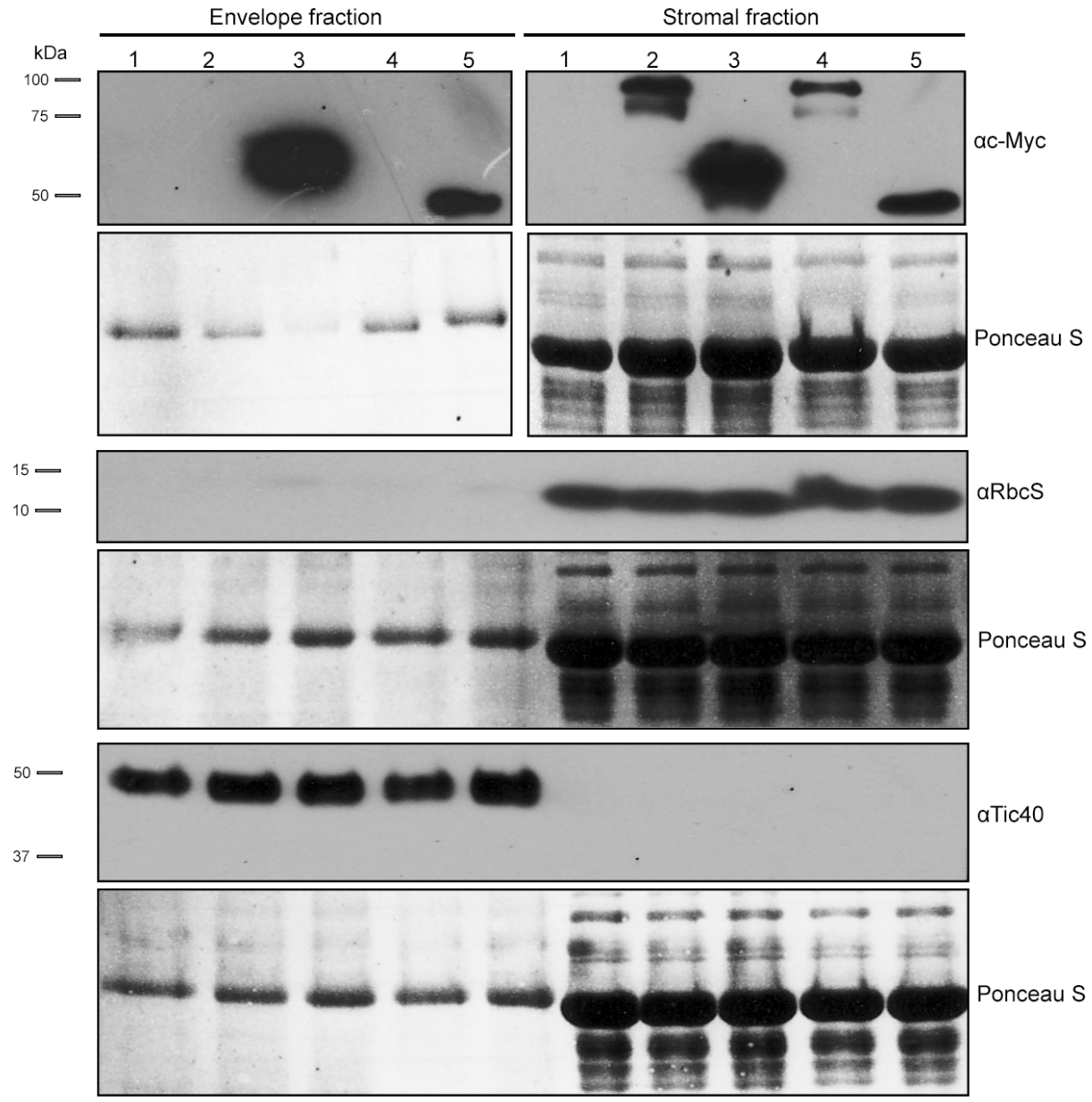


Figura S7. Localización suborganelar de AtCGE1, AtCGE2 y cpHsc70-1. Análisis por Western-blot de las fusiones traduccionales abanderadas con c-Myc, purificados a partir de hojas no transformadas (1) y transformadas con los plásmidos *pNE::CPHSC70-1* (2 y 4), *pNE::CGE1* (3), *pNE::CGE2* (5). Se muestran inmunodetecciones de las proteínas RbcS (αRbcS) y Tic40 (αTic40) como marcadores de la pureza de las fracciones de estroma y envoltura, respectivamente. Además de muestran tinciones con Ponceau S como control de carga y el marcador de peso molecular (kDa).

Apéndice II. Publicaciones en el periodo.

1. **de Luna-Valdez L**, Sepúlveda-García E, Guevara-García A. 2012. Unraveling protein-protein interactions: The yeast two-hybrid approach. In Tools to understand protein-protein interactions. Isabel Gómez-Gómez editor, *Transworld research network*.
2. López-Bucio J, Dubrovsky J, Raya-GonzálezJ, Ugartechea-Chirino Y, López-Bucio J, **de Luna-Valdez L**, Ramos-Vega M, León P and Guevara-García A. 2013. *Arabidopsis thaliana* mitogen-activated protein kinase 6 is involved in seed formation and modulation of primary and lateral root development. *Journal of Experimental Botany*, 65: 169–183.
3. **de Luna-Valdez L**, Martínez-Batallar A, Hernández-Ortiz M, Encarnación-Guevara S, Ramos-Vega M, López-Bucio J, León P, Guevara-García A. 2014. Proteomic analysis of chloroplast biogenesis (clb) mutants uncovers novel proteins potentially involved in the development of Arabidopsis thaliana chloroplasts. *Journal of Proteomics*, 111: 148-164.
4. **de Luna-Valdez L**, Martínez-Batallar A, Hernández-Ortiz M, Encarnación-Guevara S, Ramos-Vega M, López-Bucio J, León P, Guevara-García A. 2014. Data for a comparative proteomic analysis of chloroplast biogenesis (clb) mutants. *Data in Brief*, 1: 15-18.
5. **de Luna-Valdez L**, León P, Encarnación-Guevara S, Guevara-García A. 2015. Chloroplast Omics. In PlantOmics: The Omics of Plant Science. Barh D, editor. *Springer India*.
6. Villaseñor-Salmerón C. 2016. Tesis: Determinación in vivo de la interacción predicha in silico entre las proteínas EMB1241 y HSP70 de *Arabidopsis thaliana*. Dirección**: de Luna-Valdez L**, Guevara-García A*. Facultad de Ciencias Biológicas, UAEM*.
7. Saucedo A, Hernández-Domínguez E, **de Luna-Valdez L**, Guevara-García A, Escobedo-Moratilla A, Bojorquéz-Velázquez E, del Río-Portilla F, Fernández-Velasco D, Barba de la Rosa P. 2017. Insights on Structure and Function of a Late Embryogenesis Abundant Protein from *Amaranthus cruentus*: An Intrinsically Disordered Protein Involved in Protection against Desiccation, Oxidant Conditions, and Osmotic Stress. *Frontiers in Plant Science*, 8:497.
8. **de Luna-Valdez L**, Villaseñor-Salmerón C, Cordoba-Martínez E, Vera-Estrella R, León P, Guevara-García A. 2019. Functional analysis of chloroplast GrpE (CGE) proteins from Arabidopsis thaliana. *Plant Physiology and Biochemistry*, 139:293-306.
9. Chenge-Espinoza M, **de Luna-Valdez L**, López-Real G, Castillo-Ramírez S, León P. 2019. Revisiting the evolution of the DXS family of proteins. *BMC plant biology,* en revisión.

Unraveling protein-protein interactions: The yeast two-hybrid approach.

Chapter · January 2012

CITATIONS
0

READS
81

3 authors, including:

Luis Alberto De Luna-Valdez
Universidad Nacional Autónoma de México
11 PUBLICATIONS **49** CITATIONS
SEE PROFILE

Angel Guevara-Garcia
Universidad Nacional Autónoma de México
52 PUBLICATIONS **1,327** CITATIONS
SEE PROFILE

TOOLS TO UNDERSTAND PROTEIN-PROTEIN INTERACTIONS

EDITOR
ISABEL GÓMEZ

TRANSWORLD RESEARCH NETWORK

Transworld Research Network
37/661 (2), Fort P.O.
Trivandrum-695 023
Kerala, India

Tools to Understand Protein-Protein Interactions, 2012: 127-142 ISBN: 978-81-7895-552-0
Editor: Isabel Gómez

8. Unraveling protein-protein interactions: The yeast two-hybrid approach

Luis A. De Luna Valdez, E. Baldemar Sepúlveda García
and A. Arturo Guevara-García
Depto. De Biología Molecular de Plantas. Instituto de Biotecnología. Universidad Nacional Autónoma de México. Cuernavaca, Morelos, México. 62271

Abstract. All events occurring into a prokaryotic or eukaryotic cell depend on the ability of a given set of proteins to interact with each other in specific fashion. As keystones on the biological phenomena, the protein-protein interactions have been under intense study for a long time and a plethora of methods to detect and characterize those interactions have been developed. One of such methods is the so called Yeast Two-Hybrid assay, which is based on a simple concept: a reporter system split in two parts, each part fused to a different protein; upon interaction of the two proteins of interest the reporter system regains its activity, if tested proteins shall not interact, reporting activity will not occur. Since its conception, this simple yet powerful tool has been extensively modified to dissect a wide range of interactions not only between proteins, but DNA, RNA and other smaller organic molecules.

Introduction

Most, if not all, biological processes are carried out by no sole proteins but sets of proteins interacting in a specific fashion, conforming what is now

Correspondence/Reprint request: Dr. A. Arturo Guevara-García, Depto. De Biología Molecular de Plantas Instituto de Biotecnología. Universidad Nacional Autónoma de México. Cuernavaca, Morelos, 62271, México
E-mail: aguevara@ibt.unam.mx

called the interactome of a given species, tissue, or cell type under some given condition. The interactions between proteins might be stable, as those occurring within protein complexes, or might be transient as those mediating interaction between enzymes and their protein substrates, such as kinases, phosphatases or proteases. Given their importance, it is crucial to dissect protein-protein interactions to achieve a higher level of understanding of the mechanisms underlying the biological phenomena.

There are several methods available to study protein-protein interactions including protein affinity chromatography, affinity blotting, immuno-precipitation, and state-of-the-art methods such as bimolecular fluorescence complementation (1).

More than twenty years ago, in 1989 Stanley Fields and Ok-Kyu Song motivated by the development of new technologies, reported a novel genetic system designed to detect protein-protein interactions in the budding yeast *Saccharomyces cerevisiae*. The system now known as the "Yeast Two-Hybrid System" is based on the properties of the GAL4 protein. Fields and Song noticed, as it had been previously reported, that GAL4 protein could be split into two independent domains: the activation domain (AD) and the DNA-binding domain (BD) that is able to bind specific DNA sequences. The technique largely relies on the fact that neither BD nor AD alone activates transcription of the target genes. Transcription activation only occurs when AD and BD are both present and able to engage in stable interactions (2).

Fundamentally the system makes use of two hybrid proteins, one protein (bait) fused to BD and the second (prey) fused to AD of GAL4 protein. When bait-prey interactions happen, BD and AD are brought into very close contact and full GAL4 function is restored, activating transcription of a reporter gene, hence demonstrating of the existence of interactions between bait and prey proteins.

Since its very first application to detect interactions between SNF1 and SNF4 proteins (3), the yeast two-hybrid system has been used extensively and has suffered modifications in order to overcome its innate limitations.

Figure 1 shows a graphical representation of the impact the yeast two-hybrid essay has had as a method to dissect and validate protein-protein interactions, the approach has been applied successfully in a wide range of organisms and has been used as a tool in the identification of proteins that had proven very elusive in the past, such as the components of the receptor for the plant hormone abscisic acid (Pyr/Pyl/RCAR proteins) (4).

Various modifications of the classical system are now available, this derivatives from the original method have been developed to detect interactions between virtually any given pair of proteins despite its cellular localization, and have gone far beyond the original application with the development of the"Yeast three-hybrid essay" to detect RNA-protein interactions.

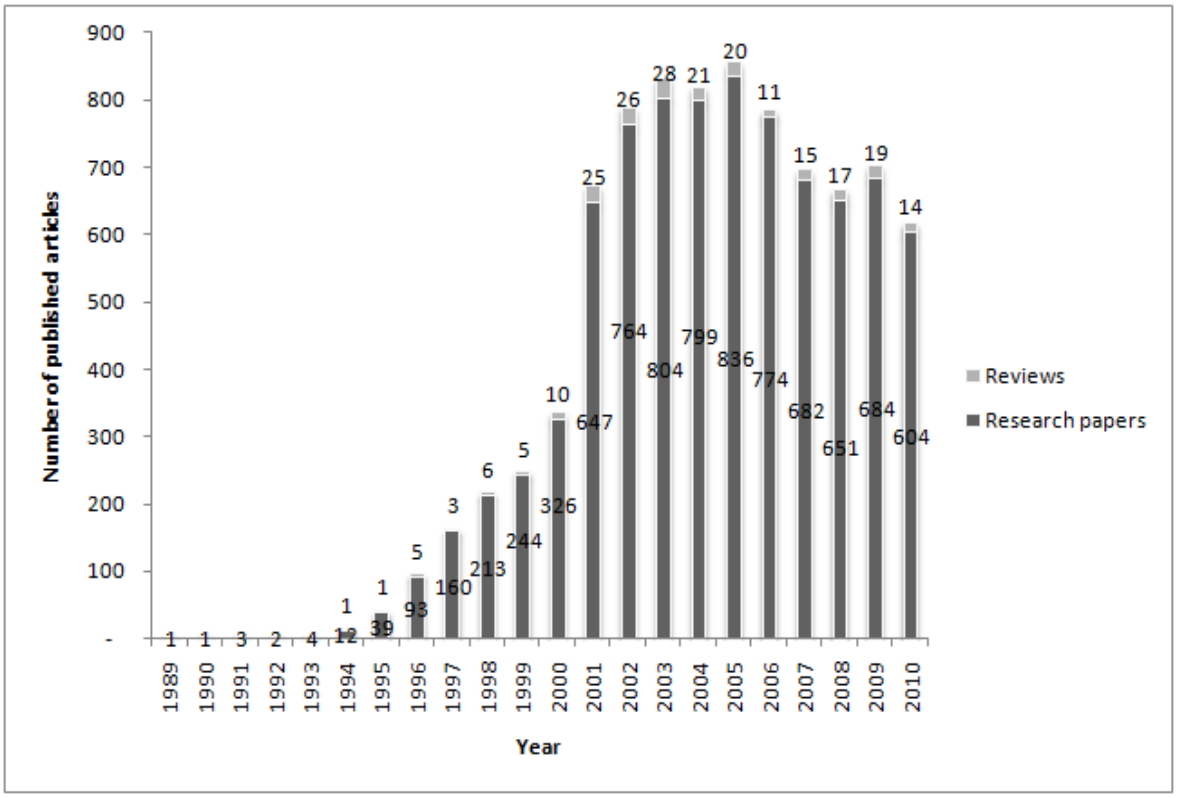


Figure 1. Usage of the yeast two-hybrid system since its publication in 1989. Above the bars, average number of PubMed hits obtained with the key-words "Yeast two-hybrid", "Yeast two-hybrid system", "Yeast two-hybrid assay" and "Yeast two-hybrid screen".

The classical two-hybrid assay

In 1986 Ptashne *et al.* discovered that it was possible to separate the DNA binding function and the transcriptional-activating function of the GAL4 protein from *Saccharomyces cerevisiae.* They found that GAL4 binds to a specific DNA sequence (CGG-N_{11}-CCG) called UAS (Upstream Activation Sequence) to activate gene transcription in the presence of galactose. When separated, the N-terminal portion of the protein was still able to bind the UAS, but could not activate transcription (Fig. 2B), this activity lying in the C-terminal fragment of the protein that did not bind to DNA (Fig. 2C). The two domains of the protein where named DNA-Binding Domain (BD) and Activation Domain (AD), and even when separated the two domains could interact to reconstitute a functional GAL4 protein that was able to activate transcription in the presence of galactose (2).

Soon after the report of the independent function of the two GAL4 protein domains, Fields and Song (3) reported their genetic system in which they fused one protein (SNF1) called bait to the BD, and a second protein

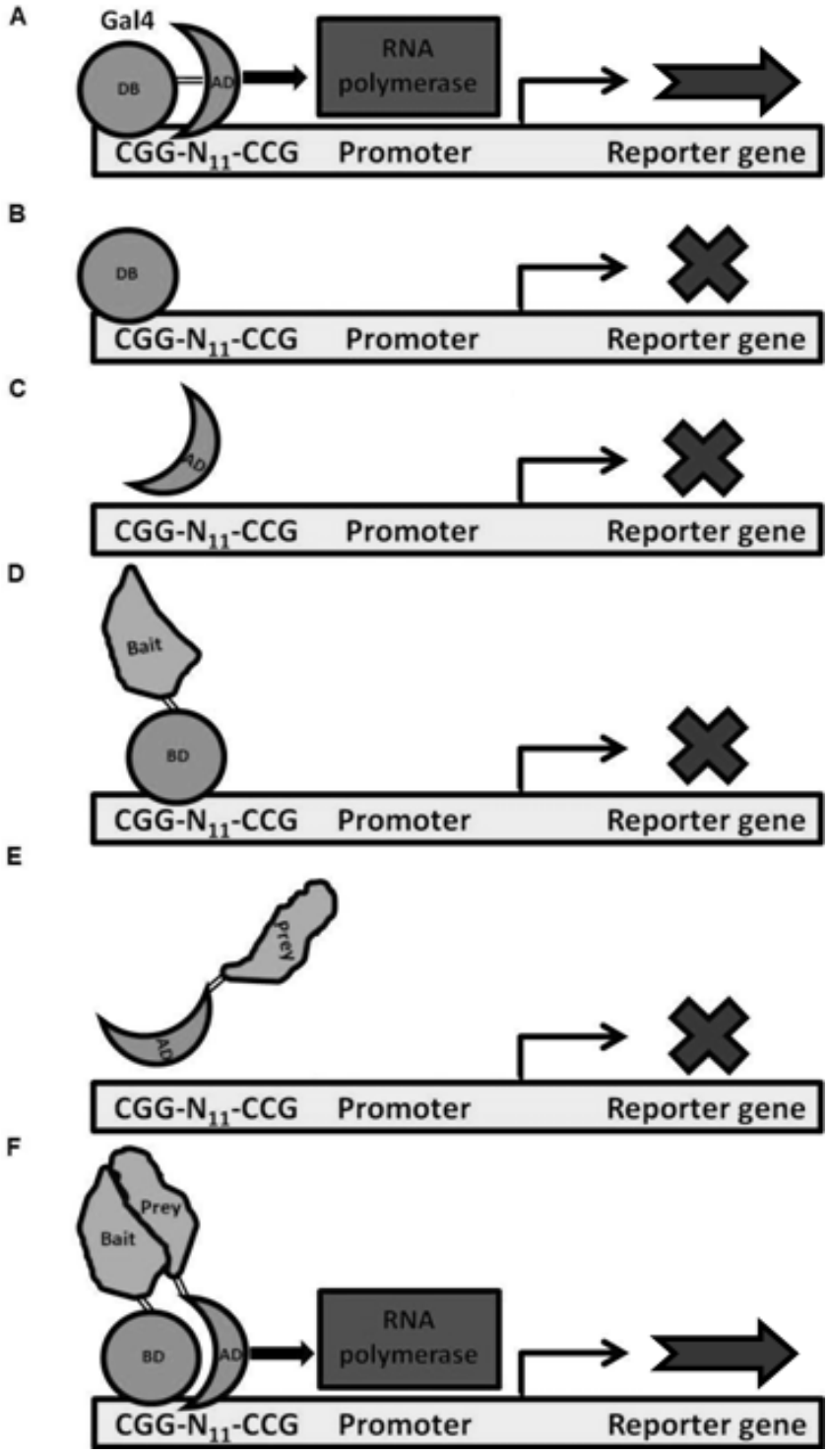


Figure 2. Summary of the classical yeast two-hybrid experiments. GAL4 protein binds to UAS via BD (circle) and activates transcription by RNA polymerase recruitment via AD (moon shape) (A). Binding of sole BD to UAS does not activate transcription (B). Expression of AD alone cannot activate reporter gene transcription (C). Fusion proteins Bait-BD and Prey-AD expressed independently cannot activate reporter gene transcription (D, E). Co-expression of Bait-BD and Prey-AD fusion proteins reconstitute BD-AD interactions and reporter gene expression.

(SNF4) to the AD of GAL4, this latter protein named prey. Co-transformation of a yeast strain with plasmids coding for each of the two hybrid proteins, resulted in successful reconstitution of GAL4 transcriptional activation of a reporter gene (*lacZ*). The recovery or GAL4 function could only be explained by the fact that bait and prey proteins were interacting so tightly, that BD and AD were brought together in such close proximity that it was possible for the two GAL4 moieties to interact with each other as if they were native proteins, hence activating transcription of the reporter gene by recruitment of an RNA polymerase (Fig. 2F) (3).

Nowadays, the basic characteristics of the two-hybrid system are used not only to test interactions between two specific proteins, but to screen for unknown interacting partners at whole-genome scale. These screens are carried out by making prey libraries, in which a cDNA pool is cloned in a particular AD-containing vector, and then the whole library is tested for interactions with some given bait of interest (5). The two basic approaches described up to now involve a specific bait protein whose interactors are to be found, but there is one more experimental design to be discussed, the matrix approach. In the matrix approach all ORFs (Open Reading Frames) of a given system are cloned as baits and preys in either of the two yeast mating types *a* or α. A matrix is constructed (Fig. 3) where using both ORF libraries bait and prey proteins are brought to interaction by the mating of the yeast mating types. This way, all the possible combinations are examined and interactome maps can be constructed (6; 7; 8).

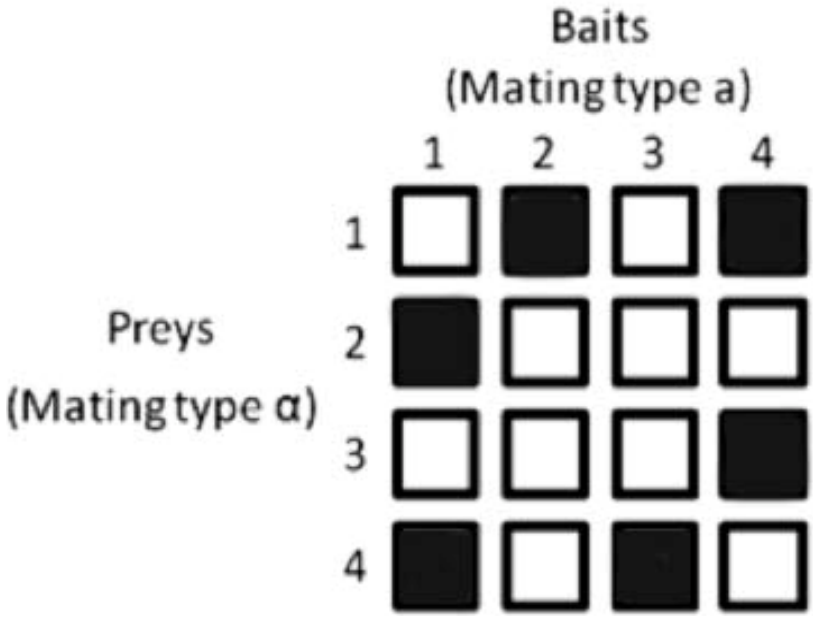


Figure 3. The matrix approach. Full ORFs are cloned as baits and preys in yeast mating types a or α. Black boxes indicate interaction between two ORFs.

In real laboratory life, there are restrictions on the amount of ORFs that can be used in a matrix approach, so this method has a high false negative rate, compared to the cDNA library approach, which has a lower false negative rate, although it has increased false positive rate.

Actually, this high false positive rate represents the main problem of the classical two-hybrid assay. When testing non-yeast proteins, the proteins are still expressed in a heterologous system that may lack specific posttranslational modification mechanisms to proper protein folding and non-specific yet detectable interactions may occur. Also, the assay has a high false negative rate because it is designed to test for interactions occurring into the cell nucleus, so in addition to being expressed in a heterologous system and the mentioned folding problems, loads of information are lost if cytoplasmic, and membrane or organelle proteins are trying to be tested (9).

An usual strategy to make the assay less prone to false positives due to non-specific interactions, is to use two reporter genes at the same time. This way, the stringency of the method is increased, but weak and yet specific interactions may be discriminated against (false negative rate increases).

Regardless of the classical yeast two-hybrid assay inherent drawbacks, the system has been successfully used in many occasions, as in the identification of PYR/PYL/RCAR proteins as interacting partners of PP2C proteases in the plant abscisic acid signaling pathway (4). Additional examples are the identification of the interacting WEB1 and PMI2 proteins as mediators of the accumulation and avoidance chloroplast responses in *Arabidopsis,* and the determination of the domains in Auxin Response Factors essential for the interaction with the auxin receptor Aux/IAA in rice (10; 11). Also, by means of yeast two-hybrid methods it has been possible to determine partial interactomes of various organisms such as human, yeast, and fruit fly (8; 12; 13; 14).

Even though the yeast two-hybrid approach has strong limitations, several methods have been developed to overcome them, nowadays there are about a dozen variants of the original method that target protein interactions occurring between proteins in the cytoplasm, at the membrane and even between extracellular proteins.

Derivatives of the yeast two-hybrid system idea

Since its first application the two-hybrid system has proved to be easy to perform and its principles easy to understand and bend in order to face new challenges. A few deviations of the classical system were developed to not only search for protein-protein interactions, but to answer different questions such as the identification of receptors for non-protein compounds (Three-

hybrid system: Fig. 4C) (15), comparing interaction specificities between two different proteins and a prey (Dual-bait system: Fig. 4B-1, B-2) (16), and to investigate protein-DNA interactions (One-hybrid system: Fig. 4A) (17; 18).

In the one-hybrid system, a DNA sequence upstream a reporter gene is used as bait, and a single hybrid protein is used in which the AD is the same as the one used in the two-hybrid approach, whereas BD (prey) is constructed from a library containing different DNA-binding domains. Transcriptional activation of the reporter gene shall only occur when a proper BD interacts with the bait DNA sequence.

The Dual-bait system consists of two different bait proteins fused one to the BD of LexA protein, and the other bait fused to BD of cI repressor from bacteriophage λ. The prey protein can be a full protein or a fragment of a protein fused to an AD, preys are peptides whose interaction-specificity with the two baits is to be tested. As both baits are fused to different BDs, then each bait interaction with prey is followed by a different reporter gene activity, and activation of one reporter gene over the second implies that more specific interactions occur between the bait fused to the particular BD targeted to the higher expressing reporter gene, and the prey protein. Interactions between prey and LexA AD-bait-1 are usually followed via *LacZ* (β-galactosidase activity) and *LEU2* (leucine prototrophy) expression, while prey and cI AD-bait-2 interactions are tested via *uidA* (β-glucuronidase activity) and *LYS2* (lysine prototrophy) gene expression.

The three-hybrid approach is used when it is suspected that stable interaction between two proteins may depend on the presence of a third molecule. If the third interacting molecule happens to be another protein, the suspected interacting partner is transformed into yeast cells under the control of the methionine-repressed promoter Met25. Bait and prey proteins are expressed as in the classical two-hybrid assay, and involvement of the third protein is revealed when reporter gene is expressed in culture medium without methionine and no reporter expression is detected when methionine is added to the culture medium.

Another possibility is to use a RNA-hybrid molecule to look for RNA interacting proteins. In this case the RNA-hybrid is made out of one RNA fragment that interacts with a protein fused to the AD, and the second RNA fragment can be used as a bait to catch proteins (preys) able to interact with that specific RNA fragment. A library of preys can be constructed fusing the preys to a particular BD such as LexA BD. Interaction between prey proteins and the bait RNA sequence trigger reporter gene transcription, flagging that specific bait-RNA interaction as positive (19). This experimental scheme has been used to identify interactions between the HIV trans-activator protein to

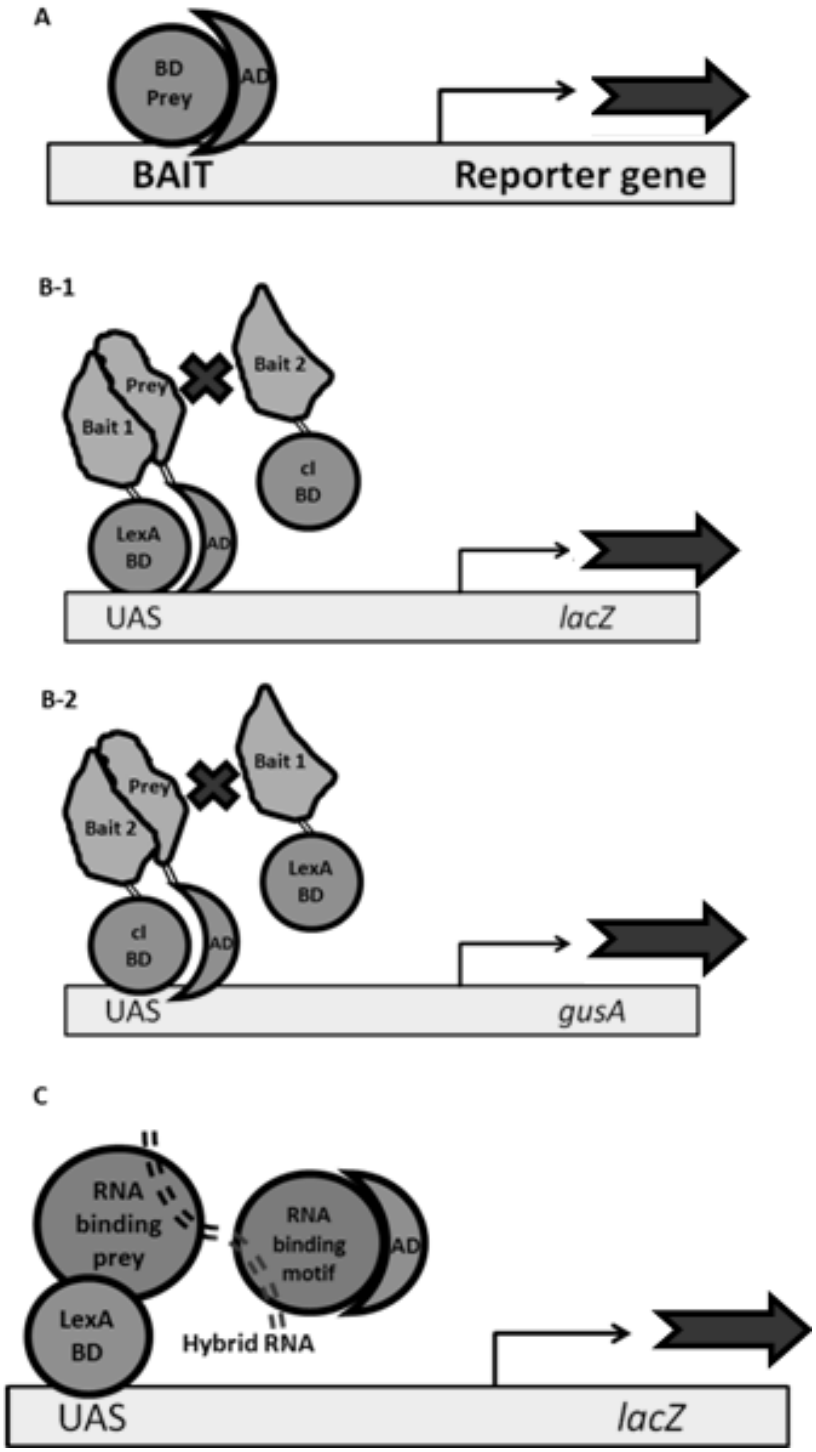


Figure 4. Deviations from the classical system. One hybrid system uses only a DNA binding protein fused to the canonical AD domain (A). Dual-bait system unravels different specificities in the bait-prey interactions by tracking different reporter genes activated upon specific activation of the prey with one of the two available baits (B-1 and B-2). Three-hybrid system makes use of a hybrid RNA molecule, one fragment of which interacts with a known protein fused to AD, the second part (bait) interacts with a second RNA binding protein (prey) fused to BD.

the HIV trans-activation response element, and binding of iron regulatory protein 1 to the iron response element (20).

Variants of the classical yeast two-hybrid system

Nowadays, there are several variants of the classical yeast two-hybrid assay that allow researchers access to almost the entire proteome of any organism. Even if system variants are designed to target different kind of proteins and to knock over different limitations of the classical two-hybrid assay, all of them have something in common, they all rely on the property of the interaction reporting system to be split in two non-fully functional independent parts, and the capability of those parts to reconstitute a fully functional reporting system upon bait-prey interaction. In the Table 1 a summary of the different variants of the two-hybrid method developed up to date, highlights the types of proteins that can be tested by each variant is showed.

1. Target 1: Cytosolic proteins

Two similar experimental designs that look for protein interactions occurring in the cytosol are the SOS and Ras recruitment systems (SRS and RRS respectively). These two systems use the Ras signaling pathway as interaction reporter, this particular pathway is conserved in both yeast and mammals and depends on the GTPase activity of the Ras protein which must be localized to the plasma membrane to undergo GDP to GTP exchange to be activated, via interaction with the guanyl exchange factors Cdc25 in yeast and Son Of Sevenless (SOS) in mammals. The set up of these two systems includes a temperature conditional yeast strain deficient in Cdc25, which is inactivated at temperatures over 36 °C, so yeast cells cultured over 36 °C would die for the lack of Ras signaling.

In the SRS, cytosolic baits are fused to mammalian SOS protein, and the prey is localized to the membrane by myristoylation. Upon bait-prey interaction, and incubation of yeast cultures at restricting temperature conditions, the SOS protein is brought to the membrane where it interacts with yeast Ras-GDP and activates the Ras protein by GDP to GTP exchange, allowing Ras-dependent downstream signaling events occur to rescue the temperature sensitive phenotype (21).

The RRS is similar to the SRS set up, but in this case the bait is fused to a constitutively active mammalian Ras protein; as Ras is already active, it only needs to be targeted to the membrane in order to activate downstream signaling. It is important to note that this system requires neither yeast Cdc25

Table I. Summary of yeast two-hybrid system variants.

Y2H method	Possible baits	Response	Cellular compartment	Yeast genotype	Promoter/ reporter
Classic Y2H system	Non-transactivating proteins capable of entering nucleus	Transcriptional activation	Nucleus	ura3, his3, ade2, lys2, trp1 gal4, gal80	UAS
SOS recruitment system (SRS)	Transactivating, cytosolic proteins	Ras signalling	Membrane periphery	*cdc25-2*	Not required
Split-ubiquitin system (SUS)	Nuclear, membrane and cytosolic proteins	Uracil auxotrophy and 5-FoA resistance	Cytosol	Classic selection plasmid	Not required
Membrane splitubiquitin system (MbY2H)	Membrane proteins	Transcriptional activation	Membrane periphery	Classic selection plasmid	LexA
Ras recruitment system (RRS)	Transactivating, cytosolic proteins	Ras signaling	Membrane periphery	*cdc25-2*	Not required
Dual bait system	Two non-transactivating proteins capable of entering nucleus	Transcriptional activation	Nucleus	Classic selection plasmid	LexA and λcI
G-protein fusion system	Membrane proteins	Inhibition of protein G signalling	Membrane periphery	ste18	FUS1 (Pheromone-inducible promoter)
RNA polymerase III based system (RPS)	Transactivating proteins	Transcriptional Activation	Nucleus	snr6	UAS
Repressed transactivator system (RTA)	Transactivating proteins capable of entering nucleus	Inhibition of transcriptional activation	Nucleus	Classic selection plasmid	UAS
Reverse Ras recruitment system (rRRS)	Membrane proteins	Ras signalling	Membrane periphery	*cdc25-2*	Not required
SCINEX-P system	Extracellular and transmembrane proteins	Downstream signalling and transcriptional activation	Endoplasmic reticulum (ER)	ire1	KAR2 (unfolded protein stress element)
Cytosolic split-ubiquitin system (CyMbY2H)	Transactivating, cytosolic proteins	Transcriptional activation	ER membrane periphery	Classic selection plasmid	LexA

URA3(Orotidine-5'-phosphate decarboxylase); HIS3(Imidazoleglycerol-phosphate dehydratase); ADE2(phosphoribosylamino-imidazole-carboxylase); LYS2(α-aminoadipate reductase); TRP1 (Phosphoribosylanthranilate isomerase); UAS(Upstream activation sequence); CDC25(guanyl nucleotide exchange factor); STE18(G-protein γ subunit); SNR6(U6 snRNA).

nor mammalian SOS proteins. Upon interaction between membrane-bound prey and Ras-fused bait, the Ras signaling pathway is able to rescue the temperature sensitive phenotype of the Cdc25 defective yeast strain, in a slightly different way than SRS. Instead of complementing the Cdc25 conditional mutation with SOS, as it is done in SRS, the RRS bypasses the guanyl exchange factors to rescue the phenotype (22).

Yet another method to test cytosolic interacting partners is the Split-Ubiquitin System (SUS). This particular system was developed to unravel interactions occurring in the cytosol. The ubiquitin protein can be separated into its N-terminal fragment (N_{ub}) and C-terminal fragment (C_{ub}); specific point mutations on N_{ub} avoid spontaneous reassembly of a full ubiquitin protein (23). As in other yeast two-hybrid assays, bait and prey are fused to only one moiety of the reporting system (either N_{ub} or C_{ub}); however, in this case the reporter system is quite different and depends on the activity of UBiquitin-specific Proteases (UBP), which catalyze the cleavage of ubiquitin from proteins flagged to proteasome-mediated degradation. In this system, C_{ub} is fused to a specific protein whose activity or lack of activity is exploited as method to report the interaction between bait and prey proteins. Upon bait-prey interaction, N_{ub} and C_{ub} reassemble into an ubiquitin form that can be recognized by UBP, the latter is then able to cleave the reporter protein fused to C_{ub}, and bait-prey interaction can be inferred from activity of that reporter protein.

There are several proteins available to be used as interaction reporters, one of the most reliable is Ura3p protein which is involved in the synthesis of uracil, and the production of a toxic compound when cells are fed with 5-FOA (5-fluoroorotic acid). This experimental set up uses an Ura3p protein modified for rapid degradation via the N-end rule pathway (24), and the whole experiment must be carried out in uracil auxotroph yeast strains. Once released from C_{ub} via UBP, Ura3p is rapidly degraded rendering uracil auxotroph and 5-FOA resistant cells. If bait and prey interaction does not occur, Ura3p is not degraded and cells remain prototrophic and 5-FOA sensitive (25).

Up to now, the split-ubiquitin system has proved so useful that it has been adapted for matrix and cDNA-library large scale screens (26; 27; 28); also two extensions of the system have been developed, one to look for interaction between membrane proteins (MbY2H), the other, named CyMbY2H, is an adaptation of the MbY2H system (discussed in the next section).

The CyMbY2H system is based on the split-ubiquitin technique and detects protein-protein interactions in the cytoplasm. A bait of interest is at the same time anchored to the endoplasmic reticulum membrane via fusion with Ost4p (small membrane protein) and fused to a reporter cassette composed of C_{ub} and the artificial transcription factor LexA-VP16 (*E. coli* LexA protein and *Herpes simplex* virus VP16 protein). Anchoring to the endoplasmic reticulum membrane allows testing putative transcriptionally active baits (such as transcription factors) by preventing their translocation to the nucleus. As in regular split-ubiquitin system, preys are fused to N_{ub}. If

the two proteins interact, C_{ub} and N_{ub} are forced into proximity and re-associate to form a full ubiquitin protein, which is then recognized by UBPs. The UBPs cleave off the artificial LexA-VP16 transcription factor from C_{ub}, releasing it from the membrane and allowing translocation to the nucleus, where it activates the LexA-responsive reporter genes (*HIS3*, *ADE2*, and *lacZ*). Detection of the protein interaction is achieved by measuring the output of the activated reporter genes (29).

One of the major drawbacks of the basic yeast two-hybrid screen, is that they cannot be used to detect interactions when one of the interacting partners is a protein capable to translocate to the nucleus and activate transcription (transactivation), such as transcription factors. In addition to some of the systems previously described (SRS, RRS, SUS and CyMY2H) that just happened to work fine for transcription factors, two different methods were developed to deal specially with transactivating proteins: the RNA Polymerase III based System (RPS), and the Repressed TransActivator system (RTA).

The RPS is suitable to test proteins interacting with transcription factors that activate RNA polymerase II-dependent transcription. This technique makes use of SNR6 as a reporter gene, this gene codes for the U6 small nuclear RNA (a component of the splicing apparatus) whose promoter was modified to harbor five copies of the GAL4 binding sequence (UAS) (30). Bait protein is fused to the BD of GAL4 and prey is fused to τ-138p protein, a subunit of the multimeric transcription factor TFIIIC that is necessary for RNA polymerase III-mediated transcription. The system is set up in a yeast strain carrying a conditional mutation in the endogenous SNR6 gene, that cannot grow when incubated at temperatures over 36 °C (31). If bait and prey interact with each other, the TFIIIC complex recruits another transcription factors able to recruit RNA polymerase III and initiate transcription of the temperature resistant U6 snRNA, that way allowing yeast growth over 36 °C.

In the RTA system, the bait-BD fusion must be able to activate transcription of the *URA3* reporter gene rendering yeast cells sensitive to the compound 5-FOA. On the other hand, prey is fused to the N-termini of the Repression Domain (RD) of Tup1 transcription repressor. When interactions between tested proteins occur, the RD domain represses transcription of *URA3*, allowing growth of yeast cells even in the presence of 5-FOA. This way the key event to detect interaction is not activation of the reporter gene, but its deactivation (32; 33).

2. Target 2: Membrane proteins

As mentioned previously, one modification of the split-ubiquitin system was developed in order to look for interactions between membrane proteins,

the method is known as Membrane split-ubiquitin system (MbY2H), and has two main differences compared to SUS. First, bait and prey are both membrane-anchored proteins, and second, instead of an enzyme, the reporter protein is a transcription factor.

The basic principle underlying the MbY2H is exactly the same as it is for the normal SUS. Membrane bait and prey proteins are fused to either N_{ub} and C_{ub} and when interactions between bait and prey proteins exist, full ubiquitin protein is reassembled and the UBP proteases are then able to cut off the transcription factor fused to C_{ub}, the free transcription factor then goes into the nucleus to activate transcription of a reporter gene. The transcription factor used in this assay is usually the artificial transcription factor LexA-VP16 that is able to activate transcription of *LacZ* and *HIS3* reporter genes. Upon LexA-VP16 release from C_{ub}, the transcription factor activates its gene targets and provides means to select the colonies on which bait-prey interaction is taking place, yeast cells are rendered able to grow on minimal medium containing no histidine, and white/blue color colony selection can be performed if X-gal is added on the growth medium (34; 35).

A second method to detect interactions occurring at the membrane is a variant of the Ras Recruitment System, the reverse Ras Recruitment System (rRRS). In this system prey is fused to the constitutively active mammalian Ras, and bait is a trans-membrane or membrane-anchored protein (arrangement of bait and prey fusions is reverse to that of the RRS). Upon interaction at the membrane, Ras signaling provides the reporter signal by colony growth at temperatures over 36 °C. This system can be used to detect interactions between cytosolic and membrane-associated proteins, but its major drawback comes into view when prey is a membrane protein itself. Since the Ras protein used in the assay only needs to localize at the membrane to undergo GDP to GTP exchange, a membrane associated prey would allow Ras signaling with no bait-prey interactions, leading to a high false positive rate (36).

Another way to test interactions between membrane proteins is the G-protein fusion system. The G-protein-coupled receptors can respond to various extracellular cues such as hormones, neurotransmitters, and light. Receptor activation triggers a conformational change in the G-protein α-subunit (exchange of GDP to GTP), and dissociation of Gα from β and γ subunits (37; 38). In yeast cells, pheromone stimulation leads to activation of a G-protein composed of the products of the *GPA1* (Gα), *STE4* (Gβ), and *STE18* (Gγ) genes. Gβγ subunit in turn activates a kinase dependent signaling cascade that culminates in growth arrest, gene transcription, cell fusion, and mating (39).

In this particular system, bait is a membrane protein (like baits used in rRRS) and the prey is fused to the γ subunit of the G-protein. When the

membrane bait interacts with the prey protein, G-protein γ subunit is taken away from α and β subunits thus preventing signaling events depending on the full heterotrimeric G-protein. When exposed to a specific pheromone (pheromone factor α) yeast cells stop growing, via a G-protein mediated signaling pathway. Upon interaction of bait and prey proteins, pheromone-induced growth arrest is prevented by the lack of γ subunit in the G-protein complex, resulting in cell growth. Additionally, a reporter gene controlled by the pheromone-inducible promoter of *FUS1* gene can be used to monitor interaction, **decreased** levels of reporter gene activity (in the presence of pheromone factor α) being the signal of positive bait-prey interactions (40).

3. Target 3: Extracellular proteins

The SCINEX-P system (SCreening for INteractions Between EXtracellular Proteins) is based on the properties of Ire1p, a transmembrane protein that is localized into the endoplasmic reticulum (ER), and controls the yeast unfolded protein response (UPR). The N-terminal lumenal domain (NLD) of Ire1p acts as a sensor for the state of ER proteins and controls Ire1p dimerization. The C-terminal cytosolic part of Ire1p harbors a Ser/Thr kinase domain, which is activated upon dimerization, and a RNase domain, regulated by phosphorylation. The accumulation of misfolded proteins in the ER activates UPR by Ire1p dimerization-dependent signaling pathways, which up regulate Hac1p transcription factor inducing the transcription of chaperones (41; 42).

Two mutant forms of Ire1p protein have been described that are affected in the kinase domain, but are still able to complement each other. The Ire1pK702R mutant contains a point mutation in the kinase domain, and ΔIre1p is a truncated form that lacks 133 amino acids of the C-terminal (43). The proteins of interest are fused each to one of the reciprocally complementing Ire1p derivatives Ire1pK702R and ΔIre1p, replacing the Ire1p N-terminal domain. Interaction of bait and prey proteins brings the Ire1p moieties in close proximity allowing complementation of Ire1pK702R and ΔIre1p, and enabling UPR signaling and enhanced expression of Hac1p protein. The system uses a yeast strain that has *LacZ* and *HIS3* as reporter genes under the control of a specific synthetic promoter containing Hac1p binding sequences, when bait and prey interactions allow Hac1p expression via UPR, *LacZ* and *HIS3* reporter genes are expressed (44).

The Yeast Two-Hybrid assay and its multiple versions described in this chapter could be considered the most powerful experimental strategy to identify protein-protein interactions. Since its publication in 1989 this methodology has been used successfully to find interactors of numerous

proteins from different organisms. Moreover, its great potential and versatility strongly suggest that it will continue being used for long time to establish and confirm diverse interactomes, as it is one of the principal goals of this post genomic era.

Bibliography

1. **Kerppola, T.** 2008, *Methods Cell Biology, 85*: 431-470.
2. **Keegan, L., Gill, G., & Ptashne, M.** 1986, *Science, 231*: 699-704.
3. **Fields, S., & Song, O.** 1989, *Nature, 340*: 245-246.
4. **Park, S.-Y., Fung, P., Nishimura, N., Jensen, D., Hiroaki, F., Yang, Z., *et al.*** 2009, *Science, 324*: 1068-1071.
5. **Auerbach, D., Thaminy, S., Hottiger, M., & Stagljar, I.** 2002, *Proteomics, 2*: 611-623.
6. **Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., & Sakaki, Y.** 2001, *Proccedings of the National Academy of Sciences USA, 98*: 4569-4574.
7. **Uetz, P., Giot, L., Cagney, G., Mansfield, T., Judson, R., Knight, J., *et al.*** 2000, *Nature, 403*: 623-627
8. **Fromont-Racine, M., Mayes, A., Brunet-Simon, A., Rain, J., Colley, A., Dix, I., *et al.*** 2000, *Yeast, 17*: 95-110.
9. **Brückner, A., Polge, C., Lentze, N., Auerbach, D., & Schlattner, U.** 2009, *International Journal of Molecular Sciences, 10*: 2763-2788.
10. **Kodama, Y., Suetsugu, N., Kong, S., & Wada, M.** 2010, *Proccedings of the National Academy of Sciences USA, 107*: 19591-19596.
11. **Shen, C., Wang, S., Bai, Y., Wu, Y., Zhang, S., Chen, M., et al.** 2010, *Journal of Experimental Botany, 61*: 3971-3981.
12. **Formstecher, E., Aresta, S., Collura, V., Hamburguer, A., Meil, A., Trehin, A., *et al.*** 2005, *Genome Research, 15*: 376-384.
13. **Stelzl, U., Worm, U., Lalowski, M., Haenig, C., Brembeck, F., Goehler, H., *et al.*** 2005, *Cell, 122*: 957-968.
14. **Rual, J., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., *et al.*** 2005, *Nature, 437*: 1173-1178.
15. **Licitra, E., & Liu, J.** 1996, *Proccedings of the National Academy of Sciences, 93*: 12817-12821.
16. **Serebriiskii, I., Khazak, V., & Golemis, E.** 1999, *Journal of Biological Chemistry, 274*: 17080-17087.
17. **Feng, S., Ota, K., & Ito, T.** 2010, *Nucleic Acids Research, 38*: e189.
18. **Chen, G., DenBoer, L., & Shin, J.** 2008, *Biotechniques, 45*: 295-304.
19. **Hook, B., Bernstein, D., Zhang, B., & Wickens, M.** 2005, *RNA, 11*: 227-233.
20. **Sengupta, J., Zhang, B., Kraemer, B., Pochart, P., Fields, S., & Wickens, M.** 1996, *Proccedings of the National Academy of Sciences USA, 93*: 8496-8501.
21. **Aronheim, A., Engelberg, D., Li, N., al-Alawi, N., Schlessinger, J., & Karin, M.** 1994, *Cell, 78*: 949-961.

22. **Broder, Y., Katz, S., & Aronheim, A.** 1998, *Current Opinion in Plant Biology, 8*: 1121-1124.
23. **Johnsson, N., & Varshavsky, A.** 1994, *Proccedings of the National Academy Sciences USA, 91*: 10340-10344.
24. **Varshavsky, A.** 1997, *Genes to Cells, 2*: 13-28.
25. **Wittke, S., Lewke, N., Muller, S., & Johnsson, N.** 1999, *Molecular Cell Biology, 10*: 2519-2530.
26. **Thaminy, S., Auerbach, D., Arnoldo, A., & Stagljar, I.** 2003, *Genome Research, 13*: 1744-1753.
27. **Felkl, M., & Leube, R.** 2008, *Neuroscience, 156*: 344-352.
28. **Pasch, J., Nickelsen, J., & Schunemann, D.** 2005, *Applied Microbiology and Biotechnology, 69*: 440-447.
29. **Möckli, N., Deplazes, A., Hassa, P., Zhang, Z., Peter, M., Hottiger, M., *et al*.** 2007, *Biotechniques, 42*: 725-730.
30. **Marsolier, M., Prioleau, M., & Sentenac, A.** 1997, *Journal of Molecular Biology, 268*: 243-249.
31. **Petrascheck, M., Castagna, F., & Barberis, A.** 2001, *Biotechniques, 20*: 296-302.
32. **Wafa, L., Cheng, H., Rao, M., Nelson, C., Cox, M., Hirst, M., *et al*.** 2003, *Biochemical Journal, 375*: 373-383.
33. **Hirst, M., Ho, C., Sabourin, L., Rudnicki, M., Penn, L., & Sadowski, I.** 2001, *Procceding of the National Academy of Sciences USA, 98*: 8726-8731.
34. **Stagljar, I., Korostensky, C., Johnsson, N., & te Heesen, S.** 1998, *Proccedings of the National Academy of Sciences USA, 95*: 5187-5192.
35. **Hooker, B., Bigelow, D., & Lin, C.** 2007, *Biochemical and Biophysical Research Communications, 363*: 457-461.
36. **Hubsman, M., Yudkovsky, G., & Aronheim, A.** 2001, *Nucleic Acids Research, 29*, e18.
37. **McCudden, C., Hains, M., Kimple, R., Siderovski, D., & Willard, F.** 2005, *Cell and Molecular Life Sciences, 65*: 551-577.
38. **Milligan, G., & Kostenis, E.** 2006, *British Journal of Pharmacology, 147*: S46–S55.
39. **Sprang, S.** 1997, *Annual Reviews, 66*: 639-678.
40. **Ehrhard, K., Jacoby, J., Fu, X., Jahn, R., & Dohlman, H.** 2000, *Nature Biotechnology, 18*: 1075-1079.
41. **Kozutsumi, Y., Segal, M., Normington, K., Gething, M., & Sambrook, J.** 1988, *Nature, 332*: 462-464.
42. **Cox, J., Shamu, C., & Walter, P.** 1993, *Cell, 73*: 1197-1206.
43. **Chapman, R., Sidrauski, C., & Walter, P.** 1998, *Annual Reviews Cell and Developmental Biology, 14*: 459-485.
44. **Urech, D., Lichtlen, P., & Barberis, A.** 2003, *Biochimica et Biophysica Acta, 1622*: 117-127.

Journal of Experimental Botany, Vol. 65, No. 1, pp. 169–183, 2014
doi:10.1093/jxb/ert368 Advance Access publication 11 November, 2013
This paper is available online free of all access charges (see http://jxb.oxfordjournals.org/open_access.html for further details)

Journal of Experimental Botany
www.jxb.oxfordjournals.org

RESEARCH PAPER

Arabidopsis thaliana mitogen-activated protein kinase 6 is involved in seed formation and modulation of primary and lateral root development

J. S. López-Bucio[1], J. G. Dubrovsky[1], J. Raya-González[2], Y. Ugartechea-Chirino[3], J. López-Bucio[2], L. A. de Luna-Valdez[1], M. Ramos-Vega[1], P. León[1] and A. A. Guevara-García[1,*]

[1] Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, 62250 Cuernavaca, Morelos, México
[2] Instituto de Investigaciones Químico-Biológicas, Universidad Michoacana de San Nicolás de Hidalgo, Edificio A-1´, CP 58030 Morelia, Michoacán, México
[3] Departamento de Ecología Funcional, Instituto de Ecología, Universidad Nacional Autónoma de México, Ciudad Universitaria, 3er circuito exterior SN, Del. Coyoacán, México D.F. 04510, México

* To whom correspondence should be addressed. E-mail: aguevara@ibt.unam.mx

Received 13 June 2013; Revised 10 September 2013; Accepted 2 October 2013

Abstract

Mitogen-activated protein kinase (MAPKs) cascades are signal transduction modules highly conserved in all eukaryotes regulating various aspects of plant biology, including stress responses and developmental programmes. In this study, we characterized the role of MAPK 6 (MPK6) in *Arabidopsis* embryo development and in post-embryonic root system architecture. We found that the *mpk6* mutation caused altered embryo development giving rise to three seed phenotypes that, post-germination, correlated with alterations in root architecture. In the smaller seed class, mutant seedlings failed to develop the primary root, possibly as a result of an earlier defect in the division of the hypophysis cell during embryo development, but they had the capacity to develop adventitious roots to complete their life cycle. In the larger class, the MPK6 loss of function did not cause any evident alteration in seed morphology, but the embryo and the mature seed were bigger than the wild type. Seedlings developed from these bigger seeds were characterized by a primary root longer than that of the wild type, accompanied by significantly increased lateral root initiation and more and longer root hairs. Apparently, the increment in primary root growth resulted from an enhanced cell production and cell elongation. Our data demonstrated that MPK6 plays an important role during embryo development and acts as a repressor of primary and lateral root development.

Key words: *Arabidopsis*, embryo development, MAP kinases, MPK6, plant signalling, root development.

Introduction

Mitogen-activated protein kinase (MAPK) cascades are signal transduction modules that are highly conserved in eukaryotes (Zhang *et al.*, 2006). A MAPK module consists of at least three kinases: a MPKKK, a MPKK, and a MPK, which activate downstream targets by phosphorylation. The last kinase of the module, a MPK, is able to phosphorylate several substrates, including transcription factors, to regulate gene expression (Andreasson and Ellis, 2009). MAPKs are known regulators of biotic and abiotic stress responses, hormone perception, and developmental programmes (Colcombet and Hirt, 2008; Suarez-Rodriguez *et al.*, 2010).

Abbreviations: CPR, cell production rate; DAG, days after germination; IAA, indole-3-acetic acid; LR, lateral root; LRP, lateral root primordium; MAPK, mitogen-activated protein kinase; MPK6, *Arabidopsis thaliana* mitogen-activated protein kinase 6; NO, nitric oxide; PD, proliferation domain; PR, primary root; RAM, root apical meristem; RH, root hair; TD, transition domain.

The *Arabidopsis* genome encodes 20 different MPKs (MAPK Group, 2002), from which MPK3, MPK4, and MPK6 play important roles both in stress and developmental responses (Colcombet and Hirt, 2008). In particular, MPK6 has been found to participate in bacterial and fungal resistance (Nuhse *et al.*, 2000; Asai *et al.*, 2002; Menke *et al.*, 2004; Wan *et al.*, 2004; Zhou *et al.*, 2004; Zhang *et al.*, 2007), in mutualistic interactions (Vadassery *et al.*, 2009), in priming of stress (Beckers *et al.*, 2009), and in regulation of plant architecture (Bush and Krysan, 2007; Müller *et al.*, 2010).

Functional redundancy is common among MAPKs. Particularly, MPK3 and MPK6 participate in biotic and abiotic stress resistance as well as in developmental processes (Lee and Ellis, 2007; Hord *et al.*, 2008; Lampard *et al.*, 2009; Liu *et al.*, 2010). MPK4/MPK6 and even MPK3/MPK4/MPK6 have been shown to act redundantly in osmotic, touch, wounding, and defence responses (Ichimura *et al.*, 2000; Droillard *et al.*, 2004; Meszaros *et al.*, 2006; Brader *et al.*, 2007). MPKs are proposed to act through common downstream targets and upstream activators (Feilner *et al.*, 2005; Merkouropoulos *et al.*, 2008; Andreasson and Ellis, 2009; Popescu *et al.*, 2009), but the interaction of these pathways is poorly understood. The MPK6 loss-of-function mutant displays alterations in the embryo and early root development, indicating that, at least for these processes, the function of this kinase cannot be substituted by any other MPK (Bush and Krysan, 2007; Müller *et al.*, 2010; Wang *et al.*, 2010).

The first evidence demonstrating that MPK6 (and/or MPK3) is involved in embryo development was reported by Wang *et al.* (2007), who showed that *mpk3*$^{-/-}$ *mpk6*$^{-/-}$ double mutants die at the embryo stage and a viable double mutant (*mpk6*$^{-/-}$ MPK3RNAi) is developmentally arrested at the cotyledon stage. In a different study, Bush and Krysan (2007) reported that several development programmes are influenced by MPK6. In that work, it was observed that *mpk6* null mutant alleles had defects in anther and embryo development, and displayed reduced male fertility. The observed *mpk6* phenotypes display variable penetrance, probably influenced by the growth conditions. Additionally, mutations in the *MPK6* gene have been linked to protrusion of the embryo detected in about 7% of the seeds from an *mpk6* homozygous population (Bush and Krysan, 2007).

Post-embryonic root development is regulated by multiple plant hormones, nutrient availability, and environmental signals (Fukaki and Tasaka, 2009; López-Bucio *et al.*, 2003). The primary root (PR) originates from the embryo and gives rise to many lateral roots (LRs) during vegetative growth, and each of these will produce more LRs. The quantity and placement of these structures among other factors determine the root system architecture (RSA), and this in turn plays a major role in determining whether a plant will survive in a particular environment (Casimiro *et al.*, 2003; Dubrovsky and Forde, 2012). A further adaptation to increase water and nutrient absorption is performed by root hairs (RHs). These are long tubular-shaped epidermal cell extensions covering roots and increase their total absorptive surface (Datta *et al.*, 2011). Auxin (indole-3-acetic acid, IAA) is recognized as the key hormone controlling both RSA and RH development, whereas cytokinins are regulators of PR growth and LR formation (Fukaki and Tasaka, 2009; De Smet *et al.*, 2012).

Current challenges are focused on determining the signalling events for which cell identity regulators are connected with hormone receptors to coordinate stress and development responses. Recently, MPK6 was proposed to be involved in early root development, possibly through altering cell division plane control and modulating the production of second messengers, such as nitric oxide (NO) in response to hydrogen peroxide (Müller *et al.*, 2010; Wang *et al.*, 2010). It was observed that *mpk6-2* and *mpk6-3* mutants produced more and longer LRs than wild-type seedlings after application of a NO donor or H_2O_2 (Wang *et al.*, 2010). However, the hormonal responses underlying these root alterations and the role of MPK6 in these processes are still unknown. Thus, independent data support the participation of MPK6 in both shoot and root development, but no relationship has been established between embryo and root phenotypes in *mpk6* mutants, nor the impact of earlier root development alterations in the configuration of post-embryonic root architecture.

In this study, we provided physiological and molecular evidence that seedlings defective in two independent *mpk6* mutant alleles showed three distinct classes of seed phenotype, which correlated with alterations in cell division and elongation processes that affected root architecture. These alterations were independent of MPK3. These data indicate that MPK6 is an essential component of early signalling processes linked to proper embryo development and maintenance of *Arabidopsis* RSA.

Materials and methods

Additional details are available in Supplementary Methods at *JXB* online.

Plant material and growth conditions

Arabidopsis thaliana Heyhn wild-type and mutant plant lines were in the Columbia-0 (Col-0) ecotype. *MPK6* (At2g43790) T-DNA insertion lines (SALK_073907 and SALK_127507) were obtained from the Salk T-DNA collection (Alonso *et al.*, 2003) and provided by TAIR (http://arabidopsis.org). Both mutant lines were described previously as *mpk6-2* and *mpk6-3* (Liu and Zhang, 2004). The *MPK3* T-DNA insertion line (SALK_151594), was kindly donated by Dr Shuqun Zhang from Missouri University, USA (Wang *et al.*, 2007). The transgenic line *ABI4*::*GUS* (Söderman *et al.*, 2000) was kindly provided by Dr Ruth Finkelstein from the University of California, USA. This marker gene was introduced into the *mpk6-2* background by crossing homozygous plants. Surface-sterilized seeds were incubated at 4 °C for 3 d to break dormancy and then grown on agar (0.8%, w/v, Bacto™ Agar, BD Difco, Sparks, MD, USA) solidified 0.2× MS medium (Caisson, Laboratories, Noth Logan, UT, USA) with 1% (w/v) sucrose. Kinetin and IAA were purchased from Sigma (Sigma-Aldrich, St Louis, MO, USA) and added to the medium at the indicated concentration. Seedlings were grown on vertically oriented Petri dishes maintained in growth chambers at 21 °C under a 16:8 h light:darkness photoperiod under 105 µmol m^{-2} s^{-1} light intensity. For seed production, plants were grown in Metro-Mix 200 (Grace Sierra, Milpitas, CA, USA) in a growth room at 23 °C under a 16/8 h photoperiod and a light intensity of 230 µmol m^{-2} s^{-1}.

Embryo analysis

Wild-type and *mpk6-2* mutant embryos were processed as described previously (Ugartechea-Chirino *et al.*, 2010). Briefly, ovules were dissected from the silique and punched with a needle in order to favour contact between the embryos and the staining solutions. Embryos were fixed for 1–7 d with 50%, v/v, methanol, 10%, v/v, acetic acid. They were rinsed and incubated at room temperature for 30–45 min in 1% periodic acid. After a second rinse, they were incubated for 2 h in pseudo-Schiff's reagent (1.9 g of sodium metabisulfite in 97 ml of H_2O and 3 ml of 5 M HCl with 0.1 mg ml^{-1} of propidium iododide). Embryos were rinsed again and transferred to a drop of chloral hydrate (80 g of chloral hydrate in 30 ml of H_2O) on a microscope slide. Excess chloral hydrate was removed, and the embryos were mounted in Hoyer's solution (30 g of gum arabic, 200 g of chloral hydrate, 20 g of glycerol in 50 ml of H_2O). Mounted embryos were cleared in Hoyer's solution for at least a week before confocal imaging.

Seed size analysis

Dry seeds were measured individually using ImageJ (http://rsb.info.nih.gov/ij) with a set scale tool to establish a 1 mm reference on a micrometer image taken with a Nikon SMZ1500 stereomicroscope equipped with a digital SIGHT DS-Fi1c camera. Seed stereomicroscope images were then analysed with ImageJ using the 1 mm reference. Seed weight was obtained by weighting a batch of 100 seeds placed in Eppendorf tubes in an analytical balance.

Growth analysis

Photographs of representative seedlings were taken with an EOS REBEL XSi digital camera (Canon, Tokyo, Japan). The growth of PRs was registered using a ruler. LR number was determined counting all LRs emerged from the PRs under the Nikon SMZ1500 stereomicroscope. LR density, LR primordium (LRP) density, length of cortical cells, LR initiation index, length of root apical meristem (RAM), length of proliferation domain (PD), length of transition domain (TD), and number of cells in the PD (NC_{PD}) were determined on cleared roots as described previously (Dubrovsky *et al.*, 2009; Dubrovsky and Forde, 2012; Ivanov and Dubrovsky, 2013). Position of the most distal (rootward) LRP and the most distal LR as well the number of LRPs in the LR formation and the branching zones was determined on cleared root preparations under a Zeiss Axiovert 200M microscope (Zeiss, Oberkochen, Germany), equipped with differential interferential contrast optics. Cortical cell length was determined for 10 cells per root on cleared preparations using an ocular micrometer. Images of RHs and etiolated seedling images were taken under a Nikon SMZ1500 stereomicroscope equipped with a digital SIGHT DS-Fi1c camera. RH density (number of RHs mm^{-1}) and RH length were determined from roots mounted in H_2O on microscope slides and observed under a Zeiss Axiovert 200M microscope. Cell production rate (CPR) was calculated with the equation $CPR = Vl_e^{-1}$, where V (μm h^{-1}) is the rate of root growth during the last 24 h before the termination of the experiment and l_e (μm) is the length of fully elongated cortical cells, whereas cell-cycle duration (T, hours) was calculated with the equation $T = (NC_{PD} \times l_e \times \ln 2)/V^{-1}$ in accordance with Ivanov and Dubrovsky (1997). This method is applicable to steady-state growing roots. One condition of steady-state growing roots is a linear increase in the root length. We analysed root growth during the last 24 h in seedling samples 5 and 8 d after germination (DAG) and found that at both time points the growth in both the mutant and the wild-type was stabilized (see Results). Another condition was a constant number of cells in the meristem (Ivanov and Dubrovsky, 1997). As the transition domain of the RAM has not been defined previously, the number of meristematic cells in the cited work corresponds to the NC_{PD} in the current study. To verify if the NC_{PD} was constant during the analysed time periods, we estimated this parameter in samples at t_0 (24 h prior to termination of the experiments) and found no statistical differences in the NC_{PD} within the same genotype at t_0 and at final time points. This preliminary analysis permitted us to apply the above equation for estimation of average cell-cycle time in the PD. Criteria for defining the PD and TD have been described (Ivanov and Dubrovsky, 2013). Briefly, the PD comprises cells that maintain proliferation activity and the TD comprises cells that have very low probability of cell proliferation but grow at the same rate as cells in the PD and have not yet started rapid elongation. As no marker lines have yet been proposed to identify these domains, we determined the domains based on relative changes in the cell lengths analysed on cleared root preparations. In the PD, the cell length commonly varies no more than 2-fold, and in the TD cells are longer than the longest cells in the PD. The distance from the quiescent centre to the point where a cortical cell becomes longer than the longest cell in the PD was considered to be the border between the PD and the TD. In the elongation zone, the cell length starts steadily to increase simultaneously in all tissues. The point where this increase can be observed was defined as the distal (rootward) border between the TD and elongation zone. All measurements were done with an ocular micrometer.

All experimental data were analysed statistically with SigmaPlot 11 (Systat Software, San Jose, CA, USA). Student's *t*-test or Tukey's post-hoc test were used for testing differences in growth and root developmental responses, as indicated. The number of independent experiments in each case is indicated in the corresponding figure legend.

Results

Mutation of the MPK6 *gene causes three distinct and stable seed phenotypes*

Through a careful phenotypic analysis of two independent *mpk6* T-DNA insertion null mutant lines (SALK_073907 and SALK_127507) (Supplementary Fig. S1A at *JXB* online), we corroborated that the protruding embryo phenotype, previously described by Bush and Krysan (2007), was present in the homozygous seed populations from both mutant alleles. Closer inspection of the seeds from these mutants showed three segregating phenotypically distinctive classes. In the larger class (~70%, *mpk6wb*/*w*ild-type-like *b*igger seeds) the lack of MPK6 did not cause any evident alteration in seed morphology, but the seeds were significantly bigger than those from wild-type plants (Figs 1C and 2). The second class (~23%, *mpk6rs*/raisin-like *s*eed phenotype) included seeds with rough coats (Fig. 1D). Finally, the smaller class (~7%, *mpk6bs*/*b*urst *s*eed phenotype) included seeds with protruding embryos from the seed coat (Fig. 1E). It is important to point out that the rough coat phenotype was not uniform, as we observed some seeds that looked more affected than others (Fig. 1D). However, in this study, all of them were pooled together within the same class. In contrast to the heterogeneous phenotype of the rough coat seeds, the other two seed phenotypes were clearly recognized. To determine whether all three *mpk6* seed phenotypes were linked to the *MPK6* mutation, we performed crosses between a homozygous *mpk6-2* mutant with pollen from wild-type (Col-0) plants. In the F1 progeny of these crosses, no phenotypic alterations were evident. Interestingly, in seedlings from all three different seed classes obtained from *mpk6* homozygous mutant populations, MPK6 activity was absent, and this was observed consistently in at least three subsequent generations of homozygous *mpk6*

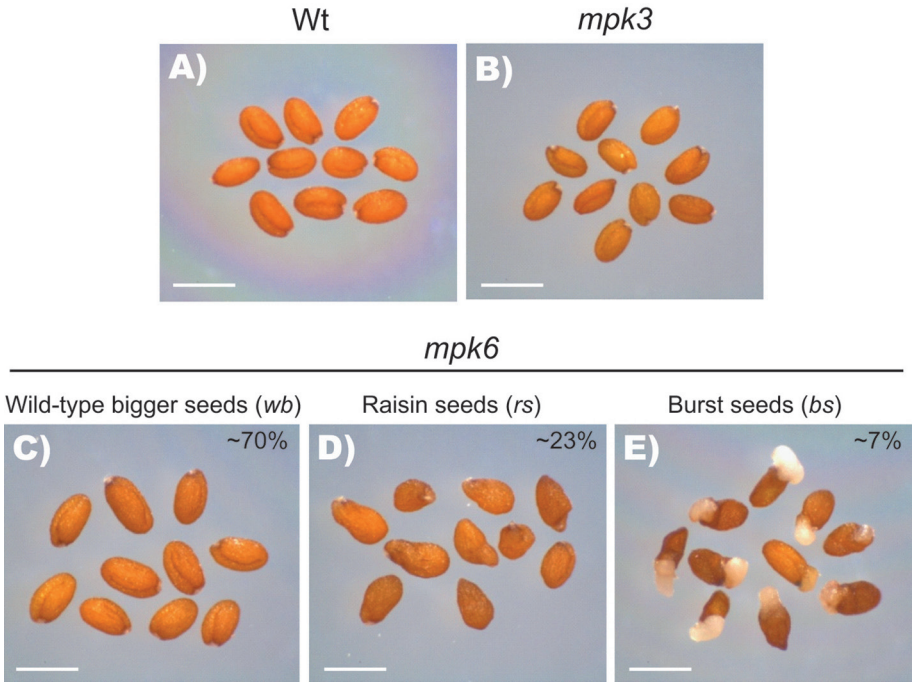


Fig. 1. *MPK6* mutation causes three different seed phenotypes. (A, B) Seeds from wild-type plants (Wt, Col-0) (A) and *mpk3* mutant (B) are shown for comparison to stable and distinguishable *mpk6* mutant seed phenotypes. (C) Seeds resembling the wild type but with a bigger seed phenotype (*mpk6wb*). (D) Seeds displaying a rough coat raisin-like seed phenotype (*mpk6rs*). (E) Embryos protruding from the seed coat burst seed phenotype (*mpk6bs*), described previously by Bush and Krysan (2007). For the pictures, seeds from each class were pooled, but the proportion of each phenotype, obtained from 1000 analysed *mpk6* seeds through several generations, is indicated. Bars, 500 µm.

seedlings from the referred seed phenotypes (Supplementary Fig. S1B).

For a better inspection of seed structure, a *pABI4::GUS* transgene encoding β-glucuronidase (GUS), expressed in embryos (Bossi *et al.*, 2009; Söderman *et al.*, 2000), was introduced into the *mpk6-2* homozygous line. We found that in homozygous *mpk6 pABI4::GUS* F3 populations, all three seed phenotypes were present (Supplementary Fig. S2, at *JXB* online). Previous studies have demonstrated redundancy between MPK6 and MPK3 (Lee and Ellis, 2007; Hord *et al.*, 2008; Lampard *et al.*, 2009; Liu *et al.*, 2010). However, a null *mpk3* mutant allele (SALK_100651) did not display any distinguishable seed phenotype when grown side by side with wild-type or *mpk6* seedlings (Figs 1B and 2), nor was the exacerbated MPK3 activity on *mpk6* seedlings (Supplementary Fig. S1B) able to compensate the *mpk6* phenotypes. Therefore, we concluded that the defects observed in seed morphology were caused specifically by a mutation in *MPK6* and they were apparently independent of *MPK3*.

mpk6 *seed phenotypes are linked to root developmental alterations*

To analyse whether the observed alterations in *mpk6* seeds affected post-germination development, we compared the early seedling growth of wild-type and *mpk6* homozygous mutant populations. Initially, we included *mpk3* seeds in our analysis, but we did not find any phenotypic alteration in *mpk3* mutant seeds or root seedlings (data not shown). Inspection of seedlings at 2 DAG demonstrated that it was possible to differentiate three different root phenotypes within the *mpk6* seedlings. Around 70% of the population analysed displayed PRs of greater length than WT seedlings (*l*onger *r*oot; *mpk6lr*). Additionally, roughly 20% of the seedlings displayed *s*hort *r*oots (*mpk6sr*), whereas around 10% of the seedlings did not develop PRs (*m*inus *r*oots; *mpk6mr*) (Fig. 3A, B). Although a previous report has already documented defects in root formation in the *mpk6-2* (SALK_073907) mutant (Müller *et al.*, 2010), no further analysis of these morphological alterations in the root architecture or their relationship with earlier embryonic alterations was performed. Interestingly, the proportion of each of the three root phenotypes correlated with those proportions observed from the seed phenotypes described previously (Figs 1C–E and 3A), suggesting that they may be related.

To analyse if *mpk6* mutant seed phenotypes had any effect on the post-germination development, the different classes of seed from this mutant were separated and germinated independently. The root morphology from each seed population was then compared with that of wild-type seedlings. Surprisingly, a high proportion of the *mpk6bs* seeds germinated *in vitro*, indicating that, in spite of the protrusion from the seed coat, these embryos were viable (Fig. 3C). However, around 80% of these germinated seedlings failed to develop PRs and most of them died a few days after germination. Those seedlings that survived all developed adventitious roots (Fig. 3C, inset) and were able to complete their life cycle and produce seeds. The progeny from these *mpk6bs* seedlings segregated again

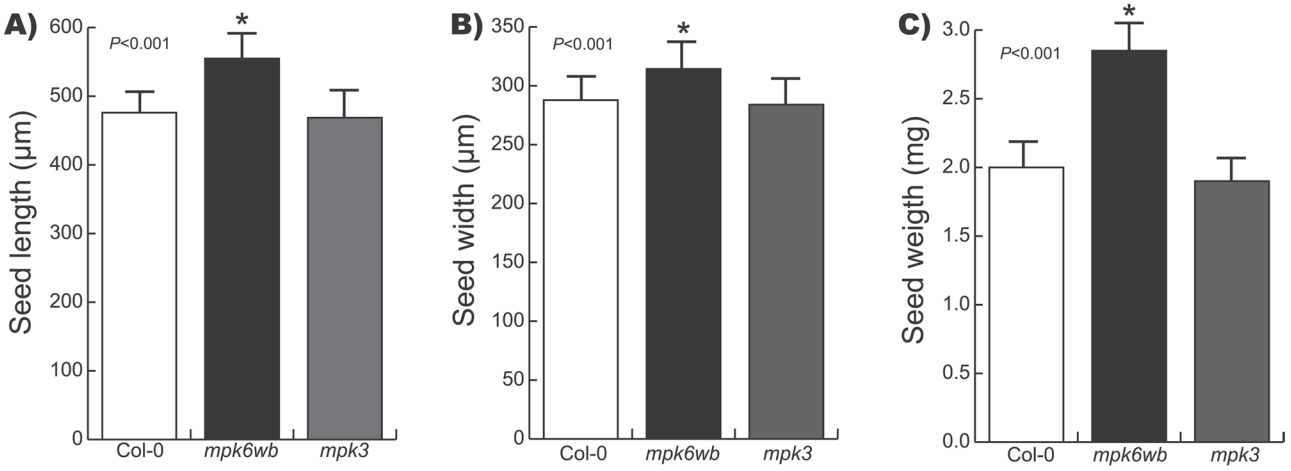


Fig. 2. *mpk6* mutant produces seeds bigger than the wild-type. The *mpk6* seeds were longer (A), wider (B), and heavier (C) than wild-type (Col-0) and *mpk3* seeds. Error bars represent the mean ±standard error (SE) from 500 seedlings analysed at each line. Asterisk indicates Student's *t*-test statistically significant differences at *P* values indicated.

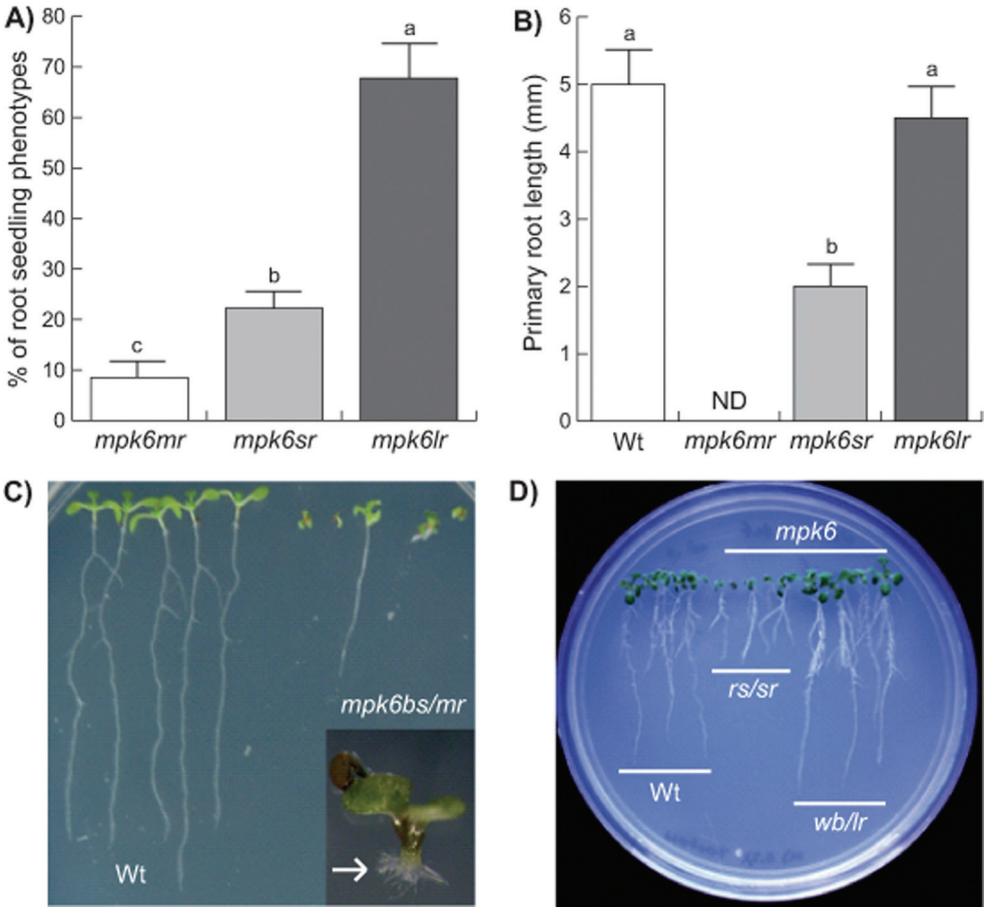


Fig. 3. *mpk6* mutant displays three different root phenotypes. The *mpk6* mutant displayed three root phenotypes each related to the seed morphology: seedlings lacking PR (minus root; *mpk6mr*), seedlings with short roots (*mpk6sr*), and seedlings with PR longer than wild-type root (*mpk6lr*). (A) Proportion of 6 DAG seedlings in each *mpk6* root mutant class.(B) PR length in 2 DAG wild-type (Wt) and the three *mpk6* root mutant classes seedlings. Notice that in this developmental stage the root length of the later *mpk6lr* phenotype is similar to that of the wild-type. Error bars represent the mean±SE from data obtained from three independent experiments, each performed with 120 (A) or 100 (B) seedlings. Different letters on the bars indicate Tukey's post-hoc test statistical difference at $P \leq 0.001$. (C) Seedlings lacking PR (minus root) developed from *mpk6bs* seeds (*mpk6bs/mr*); some of these seedlings were able to produce adventitious roots (white arrow). (D) Seeds at 6 DAG *mpk6rs* (*rs/sr*) and *mpk6wb* (*wb/lr*) develop shorter and longer PRs compared with the wild-type roots.

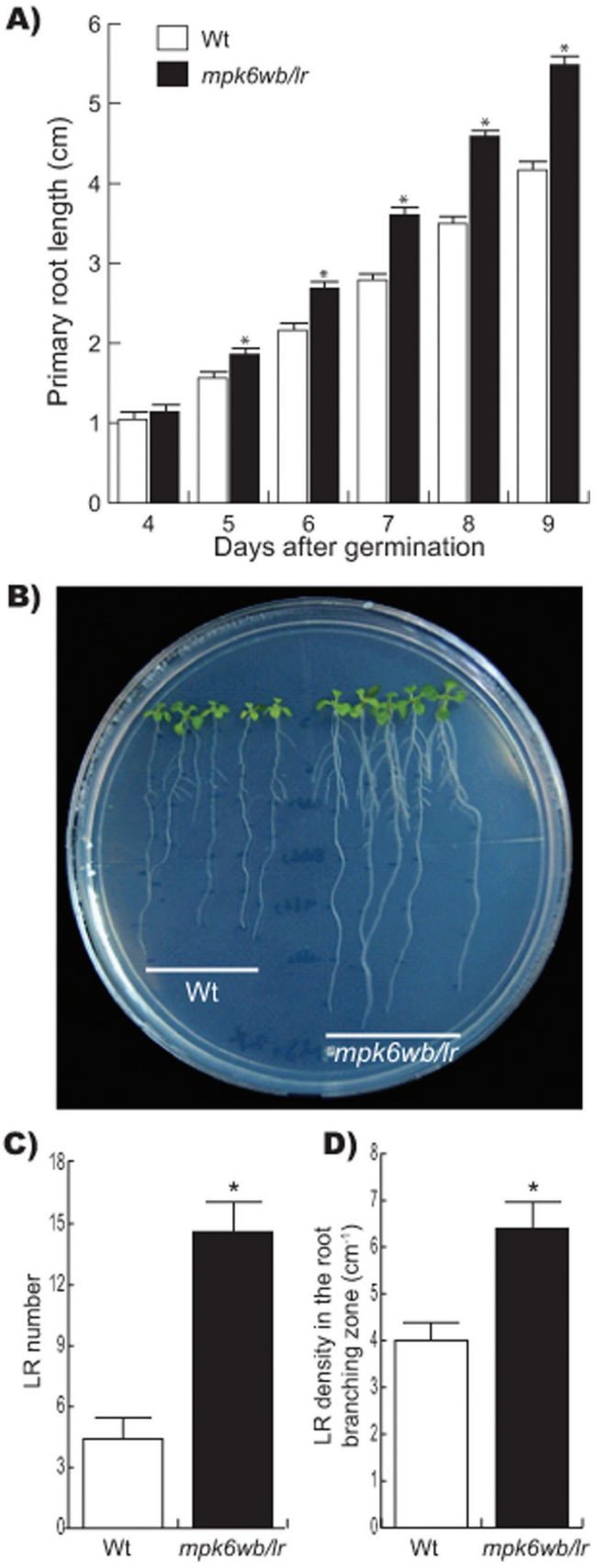


Fig. 4. *mpk6* mutant has altered primary root development, more LRs and higher LR density. (A) Primary root length changes with time. Starting from 5 DAG, statistically significant differences were observed between wild-type (Wt) and *mpk6wb/lr* PR length. Error bars represent the mean ±SE from 30 seedlings analysed at each indicated DAG. The experiment was repeated three times with similar results. Asterisk indicates Student's *t*-test statistically significant differences at $P \leq 0.001$. (B) Representative photograph of wild-type and *mpk6wb/lr* 8 DAG seedlings. Notice that *mpk6wb/lr* seedlings had longer PRs, and more and longer LRs than the wild-type seedlings. (C, D) LR number (C) and LR density (D) were obtained from the root branching zone of wild-type and *mpk6wb/lr* mutant. Error bars in (C) represent the mean ±SE from 60 seedlings analysed at 8 DAG from three independent experiments and in (D) represents the mean ±SE from 12 seedlings at 6 DAG from two independent experiments. Asterisk indicates Student's *t*-test statistically significant differences at $P \leq 0.001$.

into all three seed phenotypes shown in Fig. 1 with similar proportions (data not show). Marked differences were also observed in root development of the seedlings derived from the *mpk6wb* and *mpk6rs* seed types when compared with wild-type seedlings. Seedlings at 6 DAG derived from *mpk6rs* displayed shorter roots than the wild-type, and those derived from *mpk6wb* had longer roots (Fig. 3D). In particular, analysing the rate of growth of the PR of *mpk6wb/lr*, we found that, starting from 5 DAG, it was greater in the mutant than in wild-type seedlings (Fig. 4A). The data described so far clearly demonstrated that MPK6 plays an important role in root development and that these root phenotypes are linked with particular seed phenotypes. Besides a longer root, *mpk6wb/lr* seedlings grown *in vitro* also clearly developed more LRs (Fig. 4B). We next decided to explore the participation of MPK6 in LR development, quantifying the number of LRs and LR density in *mpk6wb/lr* and wild-type plants. These analyses demonstrated that *mpk6wb/lr* seedlings contained a higher number of LRs and a greater LR density in the root branching zone (Fig. 4C, D). These data indicated that MPK6 acts as a negative regulator of LR formation.

mpk6 *mutant has embryo development defects*

The longer root phenotype of the *mpk6wb/lr* mutant could be a result of differences in germination time compared with that of wild-type. We found that this was not the case, as both wild-type and *mpk6wb* seeds had similar germination times (data not shown) and similar root length during the first days after germination (Figs 3B and 4A). Thus an alternative hypothesis is that the short-root phenotype and the inability to form PRs are associated with defects during embryo development and do not represent alterations in the vegetative root developmental programme. To test this idea, we analysed the morphology of a total of 239 *mpk6* embryos representing all stages of embryonic development from two independent experiments (Fig. 5). During early embryogenesis, the suspensor uppermost cell is recruited to the embryo proper and acquires hypophyseal identity (Jürgens, 2001). Mutant lines defective in generating this cell lineage often produce rootless seedlings (Peris *et al.*, 2010; Jeong *et al.*, 2011). Microscopic analyses of early *mpk6* embryos showed ectopic divisions in the suspensor at the time when

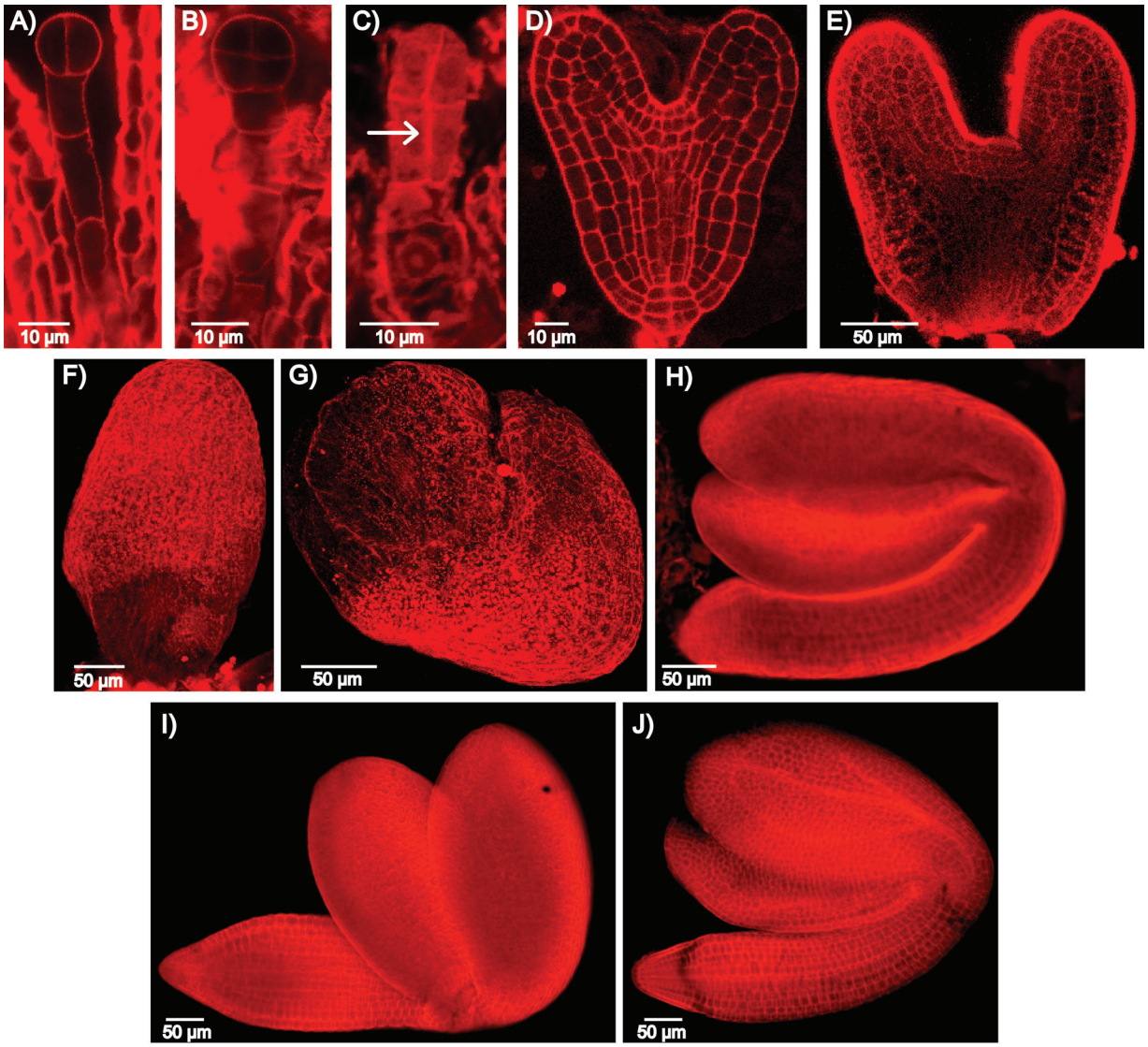


Fig. 5. *mpk6* cell organization is affected throughout embryonic development. Cell organization in wild-type (Col-0) and *mpk6* embryos. (A, B) Representative two- to four-cell (A) and eight-cell (B) wild-type embryos during their first three rounds of cell divisions. (C) Two-cell *mpk6* embryo showing ectopic periclinal cell division in the uppermost suspensor cell (arrow). This embryo failed to establish the transversal cell division plane necessary to generate an eight-cell embryo proper. (D) Wild-type heart-stage embryo. (E–G) Immature *mpk6* embryos showing arrested development at the heart (E), torpedo (F), or bent cotyledon (G) stages. (H) *mpk6* embryo with complete embryonic organogenesis at the bent cotyledon stage. (I, J) Representative *mpk6* (I) and wild-type (J) embryos at mature stage. Propidium iodide pseudo-Schiff staining of the cell wall (red) was carried out according to Ugartechea-Chirino *et al.* (2010). Bars are as indicated.

the hypophyseal cell should be specified. Seven out of 23 embryos observed between the one-cell and globular stages had ectopic divisions either in the suspensor or in the embryo proper (Fig. 5C). Later during development, 24% of the embryos were arrested at the heart stage embryo and did not proceed to develop hypocotyl and root (Fig. 5E), while 18% showed arrested development but developed hypocotyl and root (Fig. 5F–G) and 71% achieved complete embryonic organogenesis by the bent cotyledon stage (Fig. 5H). Interestingly, at the mature stage, *mpk6* embryos seemed to be bigger than the wild-type (Fig. 5I, J). Considering that *mpk6* developed several short siliques, with few seeds and many abortion events (Supplementary Fig. S3 at *JXB* online), we estimated that the frequency of these *mpk6* embryo developmental patterns correlated roughly with the frequencies observed for the burst, the raisin-like and the bigger phenotypes of mature seeds, respectively.

MPK6 is involved in the control of RSA

RSA is an important trait determining plant productivity. At present, little is known about the intrinsic mechanisms that control root growth and branching. To analyse the role that MPK6 has over RSA, we performed experiments to compare PR growth, LR formation, and RH development in wild-type and *mpk6* mutants. As they apparently do not have embryo alterations that can affect the post-germination development, to do this analysis we used only the big seed phenotype that

was also associated with long roots (*mpk6wb/lr*). LR number and length were important determinants of RSA and both were affected in the *mpk6wb/lr* mutant (Fig. 4C, D). LRs develop from the parent root through the specification of pericycle founder cells. After activation, these cells undergo repeated rounds of cell division leading to the formation of a LRP that eventually emerges as a new LR (Laskowski *et al.*, 1995; Malamy and Benfey, 1997; Dubrovsky *et al.*, 2001). A strict analysis of LR development must take into account all LR initiation events (Dubrovsky and Forde, 2012). Thus, the densities of all LR initiation events (LR and LRP) in the branching zone and in the branching zone plus LR formation zone (the latter comprises the root portion from the most rootward primordium to the most rootward emerged root) were also analysed. As shown in Fig. 6A, B, both the LR and LRP densities were significantly higher in the *mpk6wb/lr* mutant than in wild-type roots. As the fully elongated cortical cells in the *mpk6wb/lr* mutant could be longer than those from the wild type, the cell length in the LR formation zone (Fig. 6C) and the LR initiation index (Fig. 6D) were also evaluated. The latter parameter permits evaluation of LR initiation on a cellular basis and estimates the number of LR initiation events per root portion comprising 100 cortical cells of average length in a file (Dubrovsky *et al.*, 2009). This analysis also confirmed that LR initiation was significantly higher in the *mpk6wb/lr* mutant compared with wild-type seedlings. These data together strongly supported the conclusion that MPK6 acts as a negative regulator of LR initiation in *Arabidopsis*.

RHs differentiate from the root epidermal cells in a cell-position-dependent manner, increasing the total surface of roots (Tominaga-Wada *et al.*, 2011). Based on our previous results, we were interested to determine the effect of the *MPK6* mutation on RH differentiation. As shown in Fig. 7, the total number of RHs was significantly increased in the mutant compared with wild-type seedlings (Fig. 7A, B). These data indicated that the loss of function of MPK6 resulted in more and longer RHs. Moreover, we also found that the length of RHs in two different zones of the PR (2–3 and 5–7 mm root portions from the root tip) was also increased in the *mpk6wb/lr* mutant (Fig. 7C). These data also showed an important role of MPK6 in the differentiation and subsequent growth of RHs.

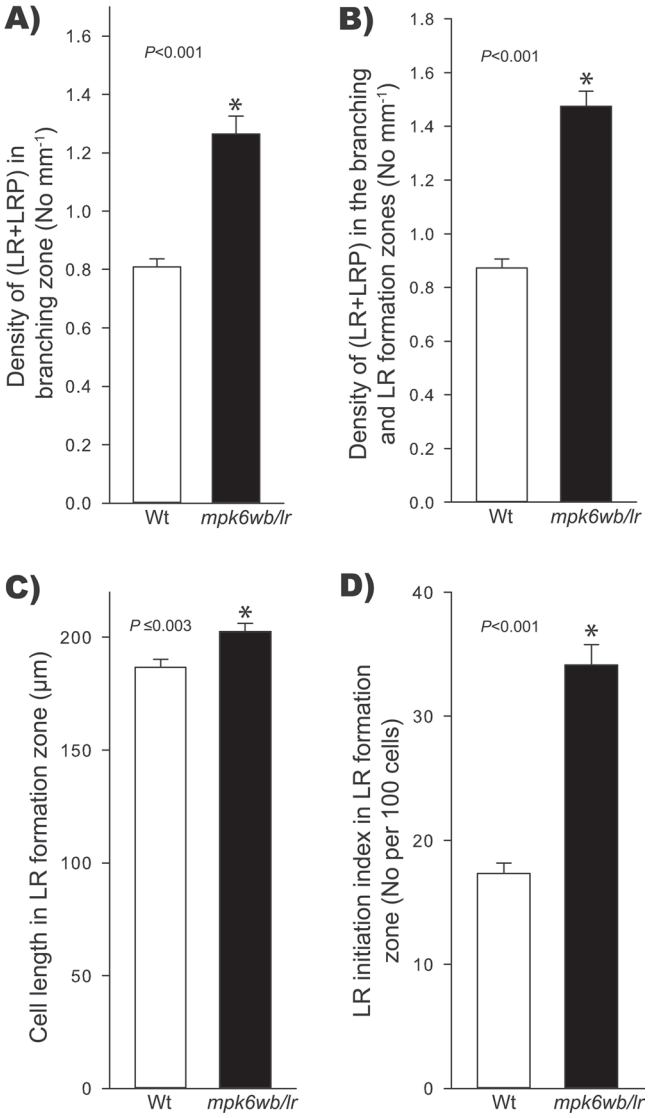


Fig. 6. The *mpk6wb/lr* mutant has increased LR initiation. LR formation in wild-type (Wt) and *mpk6wb/lr* seedlings. (A) Density of LRs and LRP in the primary root branching zone. (B) Density of LRs and LRP in branching and in LR formation zones. (C) Cortical cell length in the LR formation zone. For each individual root, 10 fully elongated cortical cells were measured. (D) LR initiation index in the LR formation zone. Error bars represent mean ±SE of 23 roots from two independent experiments. Asterisks mark Student's *t*-test significant differences at the *P* values indicated.

mpk6 *primary root growth alterations are multifactorial*

Cell division, elongation, and differentiation are closely linked cellular processes that determine PR growth. The RAM comprises two different zones: the PD, where high cell proliferation activity and a relatively slow growth takes place, and the TD, where cell proliferation probability is low but cell growth is maintained at a similar level to that found in the PD (Ivanov and Dubrovsky, 2013). After cells leave the RAM, they enter into the elongation zone, where rapid cell elongation takes place. To elucidate the contribution of cell division and elongation to the longer root phenotype of the *mpk6wb/lr* mutant, a detailed morphometric analysis of both PD and TD was conducted on 5 DAG plants. We observed that, while the size of the TD in the *mpk6wb/lr* mutant was 45% greater than that in the wild-type, the PD was 9% greater in the *mpk6wb/lr* mutant (Table 1). We also found that the number of cells, the rate of root growth, the fully elongated cell length, and cell production were also increased in *mpk6wb/lr* PRs compared with those of wild-type (Table 1 and Fig. 8A–C). A significant decrease (13%) in cell-cycle duration over time (5–8 DAG) was found in mutant seedlings (Student's *t*-test at $P \leq 0.001$), whereas in the wild type, no changes in cell-cycle duration were found during the same period (Fig. 8D). These results indicated that both cell production and cell elongation have a significant impact on the accelerated root growth found in the *mpk6wb/lr* mutant.

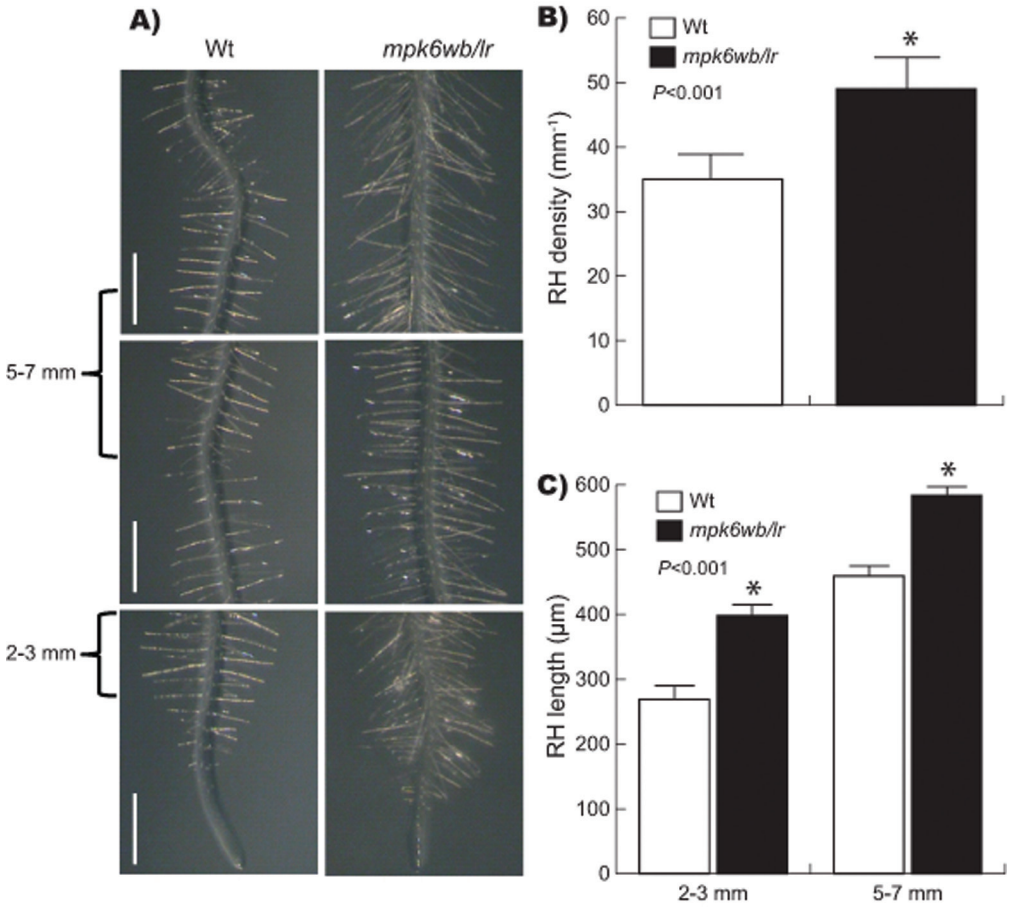


Fig. 7. The *mpk6wb/lr* mutant develops more and longer RHs. (A) RHs formed along ~1 cm from the tip of the PR from representative 6 DAG seedlings. Bars, 1 mm). (B) The RH density in 6 DAG seedlings from 2–3 and 5–7 mm root segments measured from the root tip. (C) Comparative quantification of RH length in the same root segments as in (B). Error bars represent the mean ±SE from 30 seedlings in three independent experiments. Asterisks mark Student's *t*-test significant differences at the *P* values indicated.

Table 1. *Wild-type (Col-0) and mpk6wb/lr RAM comparative analysis*

Genotype	RAM length (µm)	Difference (%)	TD length (µm)	Difference (%)	PD length (µm)	Difference (%)	NC_{PD}	Difference (%)
Col-0	355 ± 39	–	142 ± 20	–	213 ± 19	–	37.3 ± 2.7	–
mpk6wb/lr	438 ± 47*	23	206 ± 18*	45	232 ± 29**	9	45.7 ± 4.6*	22

Combined data were used from two independent experiments (n=24).
*Student's *t*-test significant differences at $P < 0.001$.
**Student's *t*-test significant differences at $P \leq 0.019$.

Discussion

MPK6 is essential for embryogenesis and root development

The central role that MAPK signalling has over different aspects of plant development is well accepted (Andreasson and Ellis, 2009; Suarez-Rodriguez *et al.*, 2010). However, the dissection of the particular function of each of the MAPK proteins has been difficult due to the partial redundancy among them. Using a genetic strategy, MPK6 has been associated with pathogen resistance (Menke *et al.*, 2004) and anther, inflorescence, embryo, and root development (Bush and Krysan, 2007; Müller *et al.*, 2010; Wang *et al.*, 2010). In particular, with respect to embryo and root development, previous analysis focused on embryo protruding seeds (Bush and Krysan, 2007), short-root seedlings (Müller *et al.*, 2010), and LR development in response to NO treatment (Wang *et al.*, 2010). In this work, we performed a detailed analysis of seed morphology and its correlation with root development in *mpk6* mutants. Three phenotypic classes of seed were identified in the progeny of homozygous *mpk6* plants, including seeds with a normal appearance but bigger than wild-type seeds, seeds with rough coats, and seeds with protruding embryos, each giving rise to seedlings with totally different root growth and development patterns (Figs 1–4). A previous work reported that the *mpk6* mutant displayed a

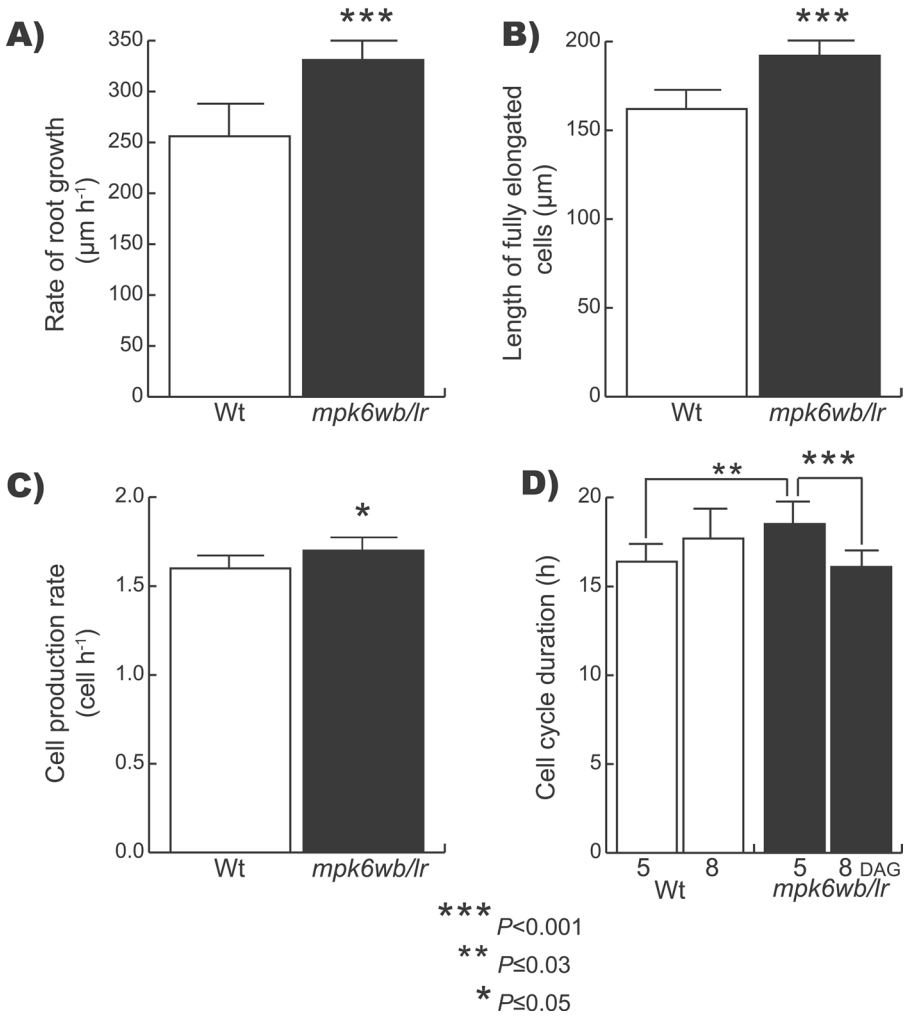


Fig. 8. Quantitative analysis of the wild-type (Wt) and *mpk6wb/lr* primary root growth and development. (A–C) Comparison between wild-type and *mpk6wb/lr* of root growth rate (A), length of fully elongated cells (B) and cell production rate (C) from 5 DAG seedlings. Error bars represent the mean ±SE from 24 seedlings in two independent experiments. (D) Cell-cycle duration (*T*) in wild-type and *mpk6wb/lr* at 5 and 8 DAG. In 5 DAG seedlings, *T* is the mean ±SE from 24 roots in two independent experiments. In 8 DAG seedlings, *T* is the mean ±SE from 12 roots in one experiment. Asterisks indicate Student's *t*-test significant differences at the *P* values indicated.

reduced fertility phenotype with variable penetrance depending on growth conditions, but the environmental variable affecting that phenotype remained to be determined (Bush and Krysan, 2007). However, the seed and root phenotypes reported here were reproducible in at least four generations of the progenies of homozygous *MPK6* from two independent null alleles (SALK_073907 and SALK_127507) grown under conditions of 21 °C, long days (16/8 h light/dark), 105 µmol m^{-2} s^{-1} of light intensity, and 45–60% of relative humidity at ~1580 m above sea level.

Between 500 and 1000 *Arabidopsis* loci have been related to embryo-defective phenotypes (Meinke *et al.*, 2009), and several of these include members of the MAPK cascade. For example, mutations in the *MPKKK4* (*YDA*) protein kinase gene cause defects in embryo development (http://www.seedgenes.org). Since the identification of *YDA* as a gene required for embryonic cell fates, it has been suggested that a MAPK signalling pathway is involved in *Arabidopsis* embryogenesis (Lukowitz *et al.*, 2004). Interestingly, the *yda/emb71* ethyl methanesulfonate heterozygous mutant, affected in the *YDA* gene (At1g63700), displays a similar embryo-protruding phenotype to that observed in *mpk6* (Lukowitz *et al.*, 2004; Meinke *et al.*, 2009). *MPKKK4* and *MPK6* are components of a common MAPK cascade involved in regulation of the embryo (Bush and Krysan, 2007) and in stomata developmental programmes (Wang *et al.*, 2007). The *yda* mutant has defects in hypophysis development similar to that observed in *mpk6bs* mutants (Fig. 5) (Lukowitz *et al.*, 2004; Meinke *et al.*, 2009). The phenotypes observed during early embryogenesis suggest that MPK6 acts as a repressor of cell proliferation involved in the establishment of embryonic polarity (Fig. 5C). This MPK6 role seems to be maintained throughout development because the *mpk6* mature embryos that achieved complete organogenesis were larger than their wild-type counterparts (Fig. 5I, J). The molecular and cellular mechanisms regulating seed development and size are complex (Sun *et al.*, 2010). Potential targets (transcription factors) and activators (leucine-rich repeat receptor kinases),

but not MAPKs were previously involved in that process (Sun *et al.*, 2010). The data from this study further confirm that both YDA and MPK6 are components of a MAPK cascade involved in the regulation of the embryo developmental programme, as already proposed (Bush and Krysan, 2007). The components acting up- and downstream of these MAPKs remain to be identified. In addition, our data support the suggestion that the failure to form a PR and the short-root phenotypes are consequences of *mpk6* mutant embryo development defects. Therefore, without considering the pleiotropic effects caused by embryo defects, the PR phenotype that can be directly associated with the loss of *MPK6* function is a long PR. Notably, this *mpk6* long PR was observed previously (Takahashi *et al.*, 2007), and in the *Arabidopsis* Hormone Database (http://ahd.cbi.pku.edu.cn) *MPK6* is included as one of the 79 genes related to a long-root phenotype.

Previous analysis made on *mpk6* short-root seedlings showed that the loss of *MPK6* function resulted in a slight but significant reduction in the number of LRs, suggesting that MPK6 is a positive regulator of LR formation (Müller *et al.*, 2010). In contrast, the data shown here using only *mpk6* mutants germinated from seeds without embryo damage demonstrated that the longer PRs had increased numbers of LRs and RHs (Figs 4 and 7). These apparent contradictory results could be explained from the different seed classes produced in the *mpk6* mutant progenies. These observations highlight how critical is to perform detailed analyses of the phenotypes associated with a gene mutation in different organs and under strictly controlled growth conditions.

MPK6 is a negative regulator of primary root growth

The comparisons of the RAM TDs and PDs and the fully elongated cell lengths between wild-type and *mpk6* mutant roots revealed that both cell division and elongation are altered in PRs of *mpk6* mutant (Table 1 and Fig. 8). Moreover, the number of cells in the *mpk6wb/lr* PD was also higher compared with that in the wild type (Fig. 6) and correlated with lower cell-cycle duration in this mutant (Fig. 8), supporting an important role of MPK6 in controlling cell proliferation and suggesting that its loss of function has a direct consequence in the long-root phenotype. For more than a decade, experimental evidence has supported the involvement of MAPKs in the regulation of cell-cycle progression in yeast (Gustin *et al.*, 1998; Strickfaden *et al.*, 2007), animals (Aliaga *et al.*, 1999; Rodríguez *et al.*, 2010), and plants (Calderini *et al.*, 1998; Jonak *et al.*, 2002; Suarez-Rodriguez *et al.*, 2010).

Regulation of plant cell division and growth is associated with microtubule reorganization, which is assisted with the action of microtubule-associated proteins. Previous reports have shown that some microtubule-associated proteins are regulated by reversible phosphorylation through MAPK cascades (Komis *et al*, 2011). For example, a MAPK from *M. sativa* (MKK3) is indispensable for spindle microtubule reorganization during mitosis (Bögre *et al.*, 1999) and the *Arabidopsis* MPK4 has been shown to be essential for the correct organization of microtubules through the phosphorylation of microtubule-associated protein 65-1 (Beck *et al.*, 2010). Additionally, the *Arabidopsis* MPK18 has been demonstrated to interact physically with a dual-specificity MAPK phosphatase (PROPYZAMIDE HYPERSENSITIVE 1/PHS1) to conform to a reversible phosphorylation/dephosphorylation switch that regulates cortical microtubule formation (Walia *et al.*, 2009). Phosphorylation of a MAP3K (NPK1) by cyclin-dependent kinases has been proposed to be critical for the appropriate cytokinesis progression in *Arabidopsis* (Sasabea *et al.*, 2011). Expression of *MPK6* has been shown to be strong in both the RAM PD and TD, specifically during the pre-prophase band and in the phragmoplast, where it controls cell division plane specification (Müller *et al.*, 2010). Epigenetic modifications like methylation or deacetylation of histones have also been suggested to regulate root development (Fukaki *et al.*, 2006; Krichevsky *et al.*, 2009). Interestingly, in animal systems, MAPK mediates histone phosphorylation, which in turns drives acetylation of histone H3, impacting on gene transcription (Clayton *et al.*, 2000). It remains to be addressed whether a similar regulatory mechanism operates in plant systems.

MPK6 regulates LR initiation

RSA is determined primarily by the spatio-temporal regulation of lateral root initiation events (Bielach *et al.*, 2012). Mutants having increased number of LRs are relatively infrequent compared with those with reduced number of LRs (De Smet *et al.*, 2006), although an increased number of LRs does not necessary indicate an increase in LR initiation (Dubrovsky and Forde, 2012). The participation of MAPK cascades in LR formation was documented by the phenotypes observed in mutants of MPK4 and its upstream activator MEKK1-1, both displaying from a slight to a severe reduction in LR density (Nakagami *et al.*, 2006; Su *et al.*, 2007). Previous studies have also shown that MPK3 and MPK6 are activated in response to the same signals as MEKK1/MPK4, supporting a possible role of these kinases in the LR development programme (Ichimura *et al.*, 2006; Suarez-Rodriguez *et al.*, 2007). However, our results demonstrated that the role of MPK6 in LR development is opposite to that of MPK4 (and MEKK1). The observation that the *mpk6wb/lr* long PR phenotype is accompanied by an increase in LR initiation (Figs 4C, D and 6), fully demonstrated that MPK6 acts as a negative regulator of LR initiation. The clearest examples of increased LR initiation are the mutants related to auxin homeostasis and signalling (Zhao, 2010). CEGENDUO, a subunit of SCF E3 ligase, has a negative role in auxin-mediated LR formation (Dong *et al.*, 2006). MAPK cascades have been found to directly or indirectly affect auxin signalling (Mockaitis and Howell, 2000; Takahashi *et al.*, 2007), which could alter LR formation. Cytokinin is a negative regulator of LR initiation. Decreased endogenous cytokinin concentration in protoxylem-adjacent pericycle cells results in increased LR initiation; in contrast, when the cytokinin concentration in these cells is increased, LR initiation is repressed (Laplaze *et al.*, 2007). In this context, the *mpk6* mutant shows a phenotype of decreased

endogenous cytokinin content. Altogether, these findings highlight the complexity of the MAPK cascades in root morphogenesis. However, an increase in cell production in the PR meristem and increased LR initiation in *mpk6wb/lr* both indicate a link between cell proliferation and its regulation by MPK6. Stress and development responses are tightly coordinated by MPKs, but their interaction is still poorly understood. As few research studies have focused on the interplay between development and environmental stresses, our findings highlight the power of studying root processes in terms of unravelling MPK signalling interactions.

MPK6 is important for RH formation

Our data also demonstrated that MPK6 is a negative regulator of RH development, as its mutation resulted in an increase in the number and size of RHs (Fig. 7). Multiple cellular factors regulate RH growth and development, including vesicle exocytosis, calcium (Ca^{2+}) homeostasis, reactive oxygen species and cytoskeleton modifications (Cardenas, 2009). Ca^{2+} is a universal second messenger, which, through interactions with Ca^{2+} sensor proteins, performs important roles in plant cell signalling (Batistič and Kudla, 2012). These sensor proteins include calmodulins, calmodulin-like proteins, Ca^{2+}-dependent protein kinases, calcineurin B-like proteins, and their interacting kinases, among others. Several genes implicated in RH differentiation have been identified; one of them, *OXIDATIVE SIGNAL INDUCIBLE 1* (*OXI1*) from *Arabidopsis*, is required for MPK6 activation by reactive oxygen species (Rentel *et al.*, 2004). The MPKKK1 (MEKK1) also has been involved in reactive oxygen species homeostasis (Nakagami *et al.*, 2006) and apparently, together with MKK2 and MPK4/MPK6, constitutes a MAPK cascade that participates in several stress responses (Ichimura *et al.*, 2000). In alfalfa, stress-induced MAPK (SIMK), an *Arabidopsis MPK6* orthologue, performs an important role in RH tip growth (Šamaj *et al.*, 2002). The alfalfa SIMK protein seems to be a positive regulator of RH growth, as treatment of plants with the MAPK inhibitor UO126 resulted in aberrant RHs, whereas the overexpression of SIMK in tobacco induced a rapid growth of these cells (Šamaj *et al.*, 2002). These results contrast with our observations of the function of MPK6 and highlight a specific role of each member of the MAPK cascade in a particular developmental process.

Possible role of MPK6 in hormone responses

As the precise mechanism underlying the root developmental alterations in *mpk6* seedlings is still unknown, we hypothesized that *mpk6wb/lr* root architectural phenotypes might result from altered responses to auxins or cytokinins, as these hormones control RSA (Perilli *et al.*, 2012). Thus, to determine whether MPK6 could affect PR responses to auxins or cytokinin, we evaluated the PR growth of wild-type and *mpk6wb/lr* mutant seedlings in response to the exogenous addition of IAA and kinetin. We observed that, at low concentrations of IAA (0.03–0.125 µM) and kinetin (0.25–2 µM), *mpk6wb/lr* was slightly insensitive and slightly hypersensitive, respectively, to the inhibitory effects of these hormones on PR growth. However, these differences were not clear at higher concentrations (0.25–0.5 µM IAA and 4–16 µM kinetin) of both hormones (Supplementary Fig. S4 at *JXB* online). These assays showed that *mpk6wb/lr* PR is not insensitive to the exogenous addition of these two plant growth regulators, suggesting that the observed root length differences in the *mpk6* mutant cannot be explained by different sensitivities to auxin or cytokinins. This observation agrees with the finding that the root growth-inhibition response to several hormones of the MPKKK mutant *yda*, which acts upstream of MPK6, is normal (Lukowitz *et al.*, 2004).

In summary, the results presented here indicate that MPK6 is a negative regulator of at least three developmental programmes in the root, namely PR growth, LR formation, and RH development, which probably occurs through regulation of cell division and elongation processes. Understanding the signalling events regulated by MPK6 activity during root development will ultimately require identification of the up- and downstream components, as well as the signal (or combination of signals) turning on and off phosphorylation of the MAPK cascade and impacting on RAM behaviour.

Supplementary data

Supplementary data are available at *JXB* online.

Supplementary Methods.

Supplementary Figure S1. *mpk6* is a null mutant.

Supplementary Figure S2. *mpk6* seed phenotypes are stable.

Supplementary Figure S3. *mpk6* siliques are shorter than wild type and contain many aborted seeds.

Supplementary Figure S4. Effect of auxin and cytokinins on primary root growth.

Acknowledgements

The authors thank Shuqun Zhang and Ruth Finkelstein for providing seeds of the homozygous *mpk3* mutant line and *pABI4::GUS* line, respectively. We also thank Patricia Jarillo for technical support, Karina Jiménez Durán (USAI, Fac. Química, UNAM), and Selene Napsucialy-Mendivil for expert technical assistance; Alma Lidia Martínez, Juan Manuel Hurtado, Roberto Rodríguez Bahena, and Arturo Ocádiz for computer support; and Paul Gaytan and Eugenio López for oligonucleotide synthesis. This work was supported by UNAM-DGAPA-PAPIIT (grants IN217111 to AGG, IN208211 to PL, and IN204312 to JGD) and CONACYT-México (CB-129266 to AGG, 127546 to PL, and 127957 to JGD).

References

Aliaga JC, Deschênes C, Beaulieu JF, Calvo EL, Rivard N. 1999. Requirement of the MAP kinase cascade for cell cycle progression and differentiation of human intestinal cells. *American Journal of Physiology—Gastrointestinal and Liver Physiology* **277,** G631–G641.

Alonso JM, Stepanova AN, Leisse TJ, *et al.* 2003. Genome-wide insertional mutagenesis of *Arabidopsis thaliana. Science* **301,** 653–657.

Andreasson E, Ellis B. 2009. Convergence and specificity in the Arabidopsis MAPK nexus. *Trends in Plant Science* **15,** 106–113.

Asai T, Tena G, Plotnikova J, Willmann MR, Chiu WL, Gomez-Gomez L, Boller T, Ausubel FM, Sheen J. 2002. MAP kinase signalling cascade in Arabidopsis innate immunity. *Nature* **415,** 977–983.

Batistič O, Kudla J. 2012. Analysis of calcium signaling pathways in plants. *Biochimica et Biophysica Acta* **1820,** 1283–1293.

Beck M, Komis G, Muller J, Menzel D, Šamaj J. 2010. Arabidopsis homologs of nucleus- and phragmoplast-localized kinase 2 and 3 and mitogen-activated protein kinase 4 are essential for microtubule organization. *Plant Cell* **22,** 755–771.

Beckers GJ, Jaskiewicz M, Liu Y, Underwood WR, He SY, Zhang S, Conrath U. 2009. Mitogen-activated protein kinases 3 and 6 are required for full priming of stress responses in *Arabidopsis thaliana. Plant Cell* **21,** 944–953.

Bielach A, Podlešáková K, Marhavý P, Duclercq J, Cuesta C, Müller B, Grunewald W, Tarkowski P, Benková E. 2012. Spatiotemporal regulation of lateral root organogenesis in Arabidopsis by cytokinin. *Plant Cell* **24,** 1–15.

Bögre L, Calderini O, Binarova P, *et al.* 1999. A MAP kinase is activated late in plant mitosis and becomes localized to the plane of cell division. *Plant Cell* **11,** 101–113.

Bossi F, Cordoba E, Dupré P, Santos-Mendoza M, San Román C, León P. 2009 The Arabidopsis ABA-INSENSITIVE (ABI) 4 factor acts as a central transcription activator of the expression of its own gene, and for the induction of ABI5 and SBE2.2 genes during sugar signaling. *The Plant Journal* **59,** 359–374.

Brader G, Djamei A, Teige M, Palva ET, Hirt H. 2007. The MAP kinase kinase MKK2 affects disease resistance in Arabidopsis. *Molecular Plant–Microbe Interaction* **20,** 589–596.

Bush SM, Krysan PJ. 2007. Mutational evidence that the Arabidopsis MAP kinase MPK6 is involved in anther, inflorescence, and embryo development. *Journal of Experimental Botany* **58,** 2181–2191.

Calderini O, Bogre L, Vicente O, Binarova P, Heberle-Bors E, Wilson C. 1998. A cell cycle regulated MAP kinase with a possible role in cytokinesis in tobacco cells. *Journal of Cell Science* **111,** 3091–3100.

Cardenas L. 2009. New findings in the mechanisms regulating polar growth in root hair cells. *Plant Signaling and Behavior* **4,** 4–8.

Casimiro I, Beeckman T, Graham N, Bhalerao R, Zhang H, Casero P, Sandberg G, Bennett MJ. 2003. Dissecting Arabidopsis lateral root development. *Trends in Plant Science* **8,** 165–171.

Clayton AL, Rose S, Barratt MJ, Mahadeva LC. 2000. Phosphoacetylation of histone H3 on c-fos- and c-jun-associated nucleosomes upon gene activation. *EMBO Journal* **17,** 3714–3726.

Colcombet J, Hirt H. 2008. Arabidopsis MAPKs: a complex signalling network involved in multiple biological processes. *Biochemical Journal* **413,** 217–226.

Datta S, Kim CM, Pernas M, Pires N, Proust H, Tam T, Vijayakumar P, Dolan L. 2011. Root hairs: development, growth and evolution at the plant–soil interface. *Plant and Soil* **346,** 1–14.

De Smet I, Vanneste S, Inzé D, Beeckman T. 2006. Lateral root initiation or the birth of a new meristem. *Plant Molecular Biology* **60,** 871–887.

De Smet I, White PJ, Bengough AG, *et al.* 2012. Analyzing lateral root development: how to move forward. *Plant Cell* **24,** 15–20.

Dong L, Wang L, Zhang Y, Zhang Y, Deng X, Xue Y. 2006. An auxin-inducible F-Box protein CEGENDUO negatively regulates auxin-mediated lateral root formation in Arabidopsis. *Plant Molecular Biology*, **60,** 599–615.

Droillard MJ, Boudsocq M, Barbier-Brygoo H, Lauriere C. 2004. Involvement of MPK4 in osmotic stress response pathways in cell suspensions and plantlets of Arabidopsis thaliana: activation by hypoosmolarity and negative role in hyperosmolarity tolerance. *FEBS Letters* **574,** 42–48.

Dubrovsky JG, Forde BG. 2012. Quantitative analysis of lateral root development: pitfalls and how to avoid them. *Plant Cell* **24,** 4–14.

Dubrovsky JG, Rost TL, Colon-Carmona A, Doerner P. 2001. Early primordium morphogenesis during lateral root initiation in *Arabidopsis thaliana*. *Planta* **214,** 30–36.

Dubrovsky JG, Soukup A, Napsucialy-Mendivil S, Jeknic Z, Ivanchenko MG. 2009. The lateral root initiation index: an integrative measure of primordium formation. *Annals of Botany* **103,** 807–817.

Feilner T, Hultschig C, Lee J, *et al.* 2005. High throughput identification of potential Arabidopsis mitogen-activated protein kinases substrates. *Molecular and Cellular Proteomics* **4,** 1558–1568.

Fukaki H, Taniguchi N, Tasaka M. 2006. PICKLE is required for SOLITARY-ROOT/IAA14-mediated repression of ARF7 and ARF19 activity during Arabidopsis lateral root initiation. *The Plant Journal* **48,** 380–389.

Fukaki H, Tasaka M. 2009. Hormone interactions during lateral root formation. *Plant Molecular Biology* **69,** 437–449.

Gustin MC, Albertyn J, Alexander M, Davenport K. 1998. MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. *Microbiology and Molecular Biology Reviews* **62,** 1264–1300.

Hord CL, Sun YJ, Pillitteri LJ, Torii KU, Wang H, Zhang S, Ma H. 2008. Regulation of Arabidopsis early anther development by the mitogen-activated protein kinases, MPK3 and MPK6, and the ERECTA and related receptor-like kinases. *Molecular Plant* **1,** 645–658.

Ichimura K, Casais C, Peck SC, Shinozaki K, Shirasu K. 2006. MEKK1 is required for MPK4 activation and regulates tissue-specific and temperature-dependent cell death in Arabidopsis. *Journal of Biological Chemistry* **281,** 36969–36976.

Ichimura K, Mizoguchi T, Yoshida R, Yuasa T, Shinozaki K. 2000. Various abiotic stresses rapidly activate Arabidopsis MAP kinases ATMPK4 and ATMPK6. *The Plant Journal* **24,** 655–665.

Ivanov VB, Dubrovsky JG. 1997. Estimation of the cell-cycle duration in the root meristem: a model of linkage between cell-cycle duration, rate of cell production, and rate of root growth. *International Journal of Plant Sciences* **158,** 757–763.

Ivanov VB, Dubrovsky JG. 2013. Longitudinal zonation pattern in plant roots: conflicts and solutions. *Trends in Plant Science* **18,** 237–243.

Jeong S, Palmer MT, Lukowitz W. 2011. The RWP-RK factor GROUNDED promotes embryonic polarity by facilitating YODA MAP kinase signaling. *Current Biology* **21,** 1268–1276.

Jonak C, Okresz L, Bogre L, Hirt H. 2002. Complexity, cross talk and integration of plant MAP kinase signalling. *Current Opinion in Plant Biology* **5,** 415–424.

Jürgens G. 2001. Apical–basal pattern formation in Arabidopsis embryogenesis. *EMBO Journal* **20,** 3609–3616.

Komis G, Illés P, Beck M, Šamaj, J. 2011. Microtubules and mitogen-activated protein kinase signalling. *Current Opinion in Plant Biology* **14,** 650–657.

Krichevsky A, Zaltsman A, Kozlovsky SV, Tian GW, Citovsky V. 2009. Regulation of root elongation by histone acetylation in Arabidopsis. *Journal of Molecular Biology* **385,** 45–50.

Lampard GR, Lukowitz W, Ellis BE, Bergmann DC. 2009. Novel and expanded roles for MAPK signaling in Arabidopsis stomatal cell fate revealed by cell type-specific manipulations. *Plant Cell* **21,** 3506–3517.

Laplaze L, Benkova E, Casimiro I, *et al.* 2007.Cytokinins act directly on lateral root founder cells to inhibit root initiation. *Plant Cell* **19,** 3889–3900.

Laskowski MJ, Williams ME, Nusbaum HC, Sussex IM. 1995. Formation of lateral root meristems is a two-stage process. *Development* **121,** 3303–3310.

Lee JS, Ellis BE. 2007. Arabidopsis MAPK phosphatase 2 (MKP2) positively regulates oxidative stress tolerance and inactivates the MPK3 and MPK6 MAPKs. *Journal of Biological Chemistry* **282,** 25020–25029.

Liu XM, Kim KE, Kim KC, *et al.* 2010. Cadmium activates Arabidopsis MPK3 and MPK6 via accumulation of reactive oxygen species. *Phytochemistry* **71,** 614–618.

Liu Y, Zhang S. 2004. Phosphorylation of 1-aminocyclopropane-1-carboxylic acid synthase by MPK6, a stress-responsive mitogen-activated protein kinase, induces ethylene biosynthesis in Arabidopsis. *Plant Cell* **16,** 3386–3399.

López-Bucio J, Cruz-Ramírez A, Herrera-Estrella L. 2003. The role of nutrient availability in regulating root architecture. *Current Opinion in Plant Biology* **6,** 280–287.

Lukowitz W, Roeder A, Parmenter D, Somerville C. 2004. A MAPKK kinase gene regulates extra-embryonic cell fate in Arabidopsis. *Cell* **116,** 109–119.

Malamy JE, Benfey P. 1997. Organization and cell differentiation in lateral roots of *Arabidopsis thaliana*. *Development* **124,** 33–44.

MAPK Group. 2002. Mitogen-activated protein kinase cascades in plants: a new nomenclature. *Trends in Plant Science* **7,** 301–308.

Meinke D, Sweeney C, Muralla R. 2009. Integrating the genetic and physical maps of *Arabidopsis thaliana*: identification of mapped alleles of cloned essential (EMB) genes. *PloS One* **4,** e7386.

Menke FL, van Pelt JA, Pieterse CM, Klessig DF. 2004. Silencing of the mitogen-activated protein kinase MPK6 compromises disease resistance in Arabidopsis. *Plant Cell* **16,** 897–907.

Merkouropoulos G, Andreasson E, Hess D, Boller T, Peck SC. 2008. An Arabidopsis protein phosphorylated in response to microbial elicitation, AtPHOS32, is a substrate of MAP kinases 3 and 6. *Journal of Biological Chemistry* **283,** 10493–10499.

Meszaros T, Helfer A, Hatzimasoura E, *et al.* 2006. The Arabidopsis MAP kinase kinase MKK1 participates in defence responses to the bacterial elicitor flagellin. *The Plant Journal* **48,** 485–498.

Mockaitis K, Howell SH. 2000. Auxin induces mitogenic activated protein kinase (MAPK) activation in roots of Arabidopsis seedlings. *The Plant Journal* **24,** 785–796.

Müller J, Beck M, Mettbach U, Komis G, Hause G, Menzel D, Samaj J. 2010. Arabidopsis MPK6 is involved in cell division plane control during early root development, and localizes to the pre-prophase band, phragmoplast, trans-Golgi network and plasma membrane. *The Plant Journal* **61,** 234–248.

Nakagami H, Soukupova H, Schikora A, Zarsky V, Hirt H. 2006. A Mitogen-activated protein kinase kinase kinase mediates reactive oxygen species homeostasis in Arabidopsis. *Journal of Biological Chemistry* **281,** 38697–38704.

Nuhse TS, Peck SC, Hirt H, Boller T. 2000. Microbial elicitors induce activation and dual phosphorylation of the *Arabidopsis thaliana* MAPK 6. *Journal of Biological Chemistry* **275,** 7521–7526.

Perilli S, Di Mambro R, Sabatini S. 2012. Growth and development of the root apical meristem. *Current Opinion in Plant Biology* **15,** 17–23.

Peris CI, Rademacher EH, Weijers D. 2010. Green beginnings—pattern formation in the early plant embryo. *Current Topics in Developmental Biology* **91,** 1–27.

Popescu SC, Popescu GV, Bachan S, Zhang Z, Gerstein M, Snyder M, Dinesh-Kumar SP. 2009. MAPK target networks in *Arabidopsis thaliana* revealed using functional protein microarrays. *Genes and Development* **23,** 80–92.

Rentel MC, Lecourieux D, Ouaked F, *et al.* 2004. OXI1 kinase is necessary for oxidative burst-mediated signalling in Arabidopsis. *Nature* **427,** 858–861.

Rodríguez J, Calvo F, González JM, Casar B, Andrés V, Crespo, P. 2010. ERK1/2 MAP kinases promote cell cycle entry by rapid, kinase-independent disruption of retinoblastoma–lamin A complexes. *Journal of Cell Biology* **191,** 967–979.

Šamaj J, Ovecka M, Hlavacka A, *et al.* 2002. Involvement of the mitogen-activated protein kinase SIMK in regulation of root hair tip growth. *EMBO Journal* **21,** 3296–3306.

Sasabea M, Boudolfb V, De Veylderb L, Inzéb D, Genschikd P, Machida Y. 2011. Phosphorylation of a mitotic kinesin-like protein and a MAPKKK by cyclin-dependent kinases (CDKs) is involved in the transition to cytokinesis in plants. *Proceedings of the National Academy of Sciences, USA* **108,** 17844–17849.

Söderman EM, Brocard IM, Lynch TJ, Finkelstein RR. 2000. Regulation and function of the Arabidopsis *ABA-insensitive4* gene in seed and abscisic acid response signaling networks. *Plant Physiology* **124,** 1752–1765.

Strickfaden SC, Winters MJ, Ben-Ari G, Lamson RE, Tyers M, Pryciak PM. 2007. A Mechanism for cell cycle regulation of MAP kinase signaling in a yeast differentiation pathway. *Cell* **128,** 519–531.

Su SH, Suarez-Rodriguez MC, Krysan P. 2007. Genetic interaction and phenotypic analysis of the Arabidopsis MAP kinase pathway mutations *mekk1* and *mpk4* suggests signaling pathway complexity. *FEBS Letters* **581,** 3171–3177.

Suarez-Rodriguez MC, Adams-Phillips L, Liu Y, Wang H, Su SH, Jester PJ, Zhang S, Bent AF, Krysan PJ. 2007. MEKK1 is required for flg22-induced MPK4 activation in Arabidopsis plants. *Plant Physiology* **143,** 661–669.

Suarez-Rodriguez MC, Petersen M, Mundy J. 2010. Mitogen-activated protein kinase signaling in plants. *Annual Review of Plant Biology* **61,** 621–649.

Sun X, Shantharaj D, Kang X, Ni M. 2010. Transcriptional and hormonal signaling control of Arabidopsis seed development. *Current Opinion in Plant Biology* **13,** 611–620.

Takahashi F, Yoshida R, Ichimura K, Mizoguchi T, Seo S, Yonezawa M, Maruyama K, Yamaguchi-Shinozaki K, Shinozaki K. 2007. The mitogen-activated protein kinase cascade MKK3–MPK6 is an important part of the jasmonate signal transduction pathway in Arabidopsis. *Plant Cell* **19,** 805–818.

Tominaga-Wada R, Ishida T, Wada T. 2011. New insights into the mechanism of development of Arabidopsis root hairs and trichomes. *International Review of Cell and Molecular Biology* **286,** 67–105.

Ugartechea-Chirino Y, Swarup R, Swarup K, Péret B, Whitworth M, Bennett M, Bougourd S. 2010. The AUX1 LAX family of auxin influx carriers is required for the establishment of embryonic root cell organization in *Arabidopsis thaliana*. *Annals of Botany* **105,** 277–289.

Vadassery J, Ranf S, Drzewiecki C, Mithofer A, Mazars C, Scheel D, Lee J, Oelmuller R. 2009. A cell wall extract from the endophytic fungus *Piriformospora indica* promotes growth of Arabidopsis seedlings and induces intracellular calcium elevation in roots. *The Plant Journal* **59,** 193–206.

Walia A, Lee JS, Wasteneys G, Ellis B. 2009. Arabidopsis mitogen-activated protein kinase MPK18 mediates cortical microtubule functions in plant cells. *The Plant Journal* **59,** 565–675.

Wan J, Zhang S, Stacey G. 2004. Activation of a mitogen-activated protein kinase pathway in Arabidopsis by chitin. *Molecular Plant Pathology* **5,** 125–135.

Wang H, Ngwenyama N, Liu Y, Walker JC, Zhang S. 2007. Stomatal development and patterning are regulated by environmentally responsive mitogen-activated protein kinases in Arabidopsis. *Plant Cell* **19,** 63–73.

Wang, P, Du Y, Li Y, Ren D, Song CP. 2010. Hydrogen peroxide-mediated activation of MAP kinase 6 modulates nitric oxide biosynthesis and signal transduction in Arabidopsis. *Plant Cell* **22,** 2981–2998.

Zhang T, Liu Y, Yang T, Zhang L, Xu S, Xue L, An L. 2006. Diverse signals converge at MAPK cascades in plant. *Plant Physiology and Biochemistry* **44,** 274–283.

Zhang X, Dai Y, Xiong Y, DeFraia C, Li J, Dong X, Mou Z. 2007. Overexpression of Arabidopsis MAP kinase kinase 7 leads to activation of plant basal and systemic acquired resistance. *The Plant Journal* **52,** 1066–1079.

Zhao Y. 2010. Auxin biosynthesis and its role in plant development. *Annual Review of Plant Biology* **61,** 49–64.

Zhou F, Menke FL, Yoshioka K, Moder W, Shirano Y, Klessig DF. 2004. High humidity suppresses *ssi4*-mediated cell death and disease resistance upstream of MAP kinase activation, H_2O_2 production and defense gene expression. *The Plant Journal* **39,** 920–932.

JOURNAL OF PROTEOMICS 111 (2014) 148–164

Available online at www.sciencedirect.com

ScienceDirect

www.elsevier.com/locate/jprot

Proteomic analysis of chloroplast biogenesis (*clb*) mutants uncovers novel proteins potentially involved in the development of *Arabidopsis thaliana* chloroplasts☆

L.A. de Luna-Valdez[a], *A.G. Martínez-Batallar*[b], *M. Hernández-Ortiz*[b], *S. Encarnación-Guevara*[b], *M. Ramos-Vega*[a], *J.S. López-Bucio*[a], *P. León*[a], *A.A. Guevara-García*[a,*]

[a]*Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, 62214 Cuernavaca, Morelos, Mexico*
[b]*Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad 565, Chamilpa, 62210 Cuernavaca, Morelos, Mexico*

ARTICLE INFO

Available online 19 August 2014

Keywords:
Chloroplast development
Comparative proteomics
clb mutants
Arabidopsis thaliana

ABSTRACT

Plant cells outstand for their ability to generate biomass from inorganic sources, this phenomenon takes place within the chloroplasts. The enzymatic machinery and developmental processes of chloroplasts have been subject of research for several decades, and this has resulted in the identification of a plethora of proteins that are essential for their development and function. Mutant lines for the genes that code for those proteins, often display pigment-accumulation defects (e.g., albino phenotypes). Here, we present a comparative proteomic analysis of four chloroplast-biogenesis affected mutants (*cla1-1*, *clb2*, *clb5*, *clb19*) aiming to identify novel proteins involved in the regulation of chloroplast development in *Arabidopsis thaliana*. We performed 2D-PAGE separation of the protein samples. These samples were then analyzed by computational processing of gel images in order to select protein spots with abundance shifts of at least twofold, statistically significant according to Student's *t*-test ($P < 0.01$). These spots were subjected to MALDI-TOF mass-spectrometry for protein identification. This process resulted in the discovery of three novel proteins potentially involved in the development of *A. thaliana* chloroplasts, as their associated mutant lines segregate pigment-deficient plants with abnormal chloroplasts, and altered mRNA accumulation of chloroplast-development marker genes.

Biological significance
This report highlights the potential of using a comparative proteomics strategy for the study of biological processes. Particularly, we compared the proteomes of wild-type seedlings and four mutant lines of *A. thaliana* affected in chloroplast biogenesis. From this proteomic

☆ This article is part of a Special Issue entitled: Proteomics, mass spectrometry and peptidomics, Cancun 2013. Guest Editors: César López-Camarillo, Victoria Pando-Robles and Bronwyn Jane Barkla.

DOI of original article: http://dx.doi.org/10.1016/j.dib.2014.07.001.

* *Corresponding author.*
E-mail addresses: ldeluna@ibt.unam.mx (L.A. de Luna-Valdez), angelmb@ccg.unam.mx (A.G. Martínez-Batallar), magda@ccg.unam.mx (M. Hernández-Ortiz), encarnac@ccg.unam.mx (S. Encarnación-Guevara), mramos@ibt.unam.mx (M. Ramos-Vega), lopbucio@ibt.unam.mx (J.S. López-Bucio), patricia@ibt.unam.mx (P. León), aguevara@ibt.unam.mx (A.A. Guevara-García).

http://dx.doi.org/10.1016/j.jprot.2014.07.003

analysis it was possible to detect common mechanisms in the mutants to respond to stress and cope with heterotrophy. Notably, it was possible to identify three novel proteins potentially involved in the development or functioning of chloroplasts, also it was demonstrated that plants annotated to carry T-DNA insertions in the cognate genes display pigment-deficient phenotypes, aberrant and underdeveloped chloroplasts, as well as altered mRNA accumulation of chloroplast biogenesis marker genes.

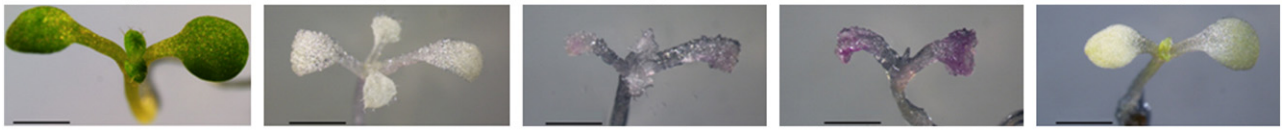
This article is part of a Special Issue entitled: Proteomics, mass spectrometry and peptidomics, Cancun 2013. Guest Editors: César López-Camarillo, Victoria Pando-Robles and Bronwyn Jane Barkla.

1. Introduction

Among eukaryotes, plant cells are notable for their ability to generate biomass from CO_2 and sunlight via the process known as photosynthesis. This process takes place in the membranes and compartments of a remarkable organelle, the chloroplast. Given their ability to perform photosynthesis, chloroplasts represent the main source of food (sugars) for plant cells and are the keystone for the entire food chain sustaining animal life. The chloroplasts are not only responsible for the production of carbon resources, they are also the biosynthesis site of many other important metabolites, such as amino acids, lipids, hormones, vitamins, and isoprenoids [1–3]. These functions make chloroplasts an essential organelle for the development and survival of plants. Thus, abnormal development of chloroplasts often results in lethality [4–7].

Thus far, several mutant lines harboring defects in chloroplast development have been isolated, and the mutated genes have been found to encode for several different components of the plastid machinery for protein import, isoprenoid biosynthesis, RNA processing, protein maturation, plastid gene expression, thylakoid biogenesis, chloroplast to nucleus signaling, and other important processes [8–19]. Examples include the chloroplast-biogenesis (*clb*) mutants, which were originally isolated based on the phenotypes of impaired pigment accumulation, rendering plants albino, yellow, or pale green [19]. In vitro cultured *clb* mutants contain underdeveloped chloroplasts, showing reduced organelle diameter, low accumulation of thylakoid membranes, low levels of photosynthetic pigments, and deficiencies in the expression of several nuclear and chloroplast encoded genes. Several *clb* mutants have been fully characterized, and the genes affected in such mutants are now known. For instance, *cla1-1*, *clb4*, and *clb6* mutant plants are affected in the expression of deoxy-xylulose-5-phosphate synthase (DXS1), hydroxy-3-methylbut-2-enyl diphosphate synthase (HDS), and hydroxy-3-methylbut-2-enyl diphosphate reductase (HDR), respectively; these three enzymes each catalyze a different step in the chloroplastic pathway for isoprenoid biosynthesis (Fig. 1A, B) [9,10,19]. Other *clb* mutants, such as *clb19*, affect the expression of a PPR domain-containing protein known to be involved in the editing of the *ClpP* and *rpoA* plastid mRNAs (Fig. 1A, B), which code for the catalytic subunit of the main protease complex in plastids and the α-subunit of the plastid-encoded RNA polymerase, respectively [19]. Another characterized *clb* mutant is *clb2*, whose affected locus was very recently assigned in our laboratory using next-generation sequencing methods, to a homogentisate prenyltransferase gene, which is known to be involved in the biosynthesis of plastoquinone 9 (Fig. 1A, B; unpublished data) [19,20]. Finally, during the time of production of this research, the *clb5* mutant line was found to be affected in the expression of ZDS (ζ-carotene desaturase) enzyme, which is responsible of the biosynthesis of the essential carotenoid lycopene [21]. Despite all the information that has been obtained through the characterization of *clb* mutants, there are still some *clb* lines (*clb1 and clb3*) whose characterization is still under investigation.

A



B

Locus	Mutant	Description	Phenotype	Reference
At4g15560	*cla1-1*	1-deoxy-D-xylulosa 5-phosphate synthase; chloroplastic pathway for isoprenoid synthesis.	Albino	Mandel M *et al*, 1996.
At1g05750	*clb19*	Pentatricopeptide repeat protein; editing of clpP and rpoA plastid transcripts.	Yellow	Chateigner-Boutin A *et al*, 2009.
At3g11945	*clb2*	Homogentisate prenyl transferase; plastoquinone-9 biosynthesis.	Albino	Gutiérrez-Nava M *et al*, 2004. Tian L *et al*, 2007.
At3g04870	*clb5*	ζ-carotene desaturase; carotenoid biosynthesis.	Albino	Gutiérrez-Nava M *et al*, 2004. Avendaño-Vázquez A *et al*, 2014.

Fig. 1 – Seedlings used in this analysis. A) Wild-type (Wt; 8 DAG) and mutant (16 DAG) seedling phenotypes. Plants were germinated and grown in vitro and collected after the emergence of the first pair of true leaves. B) Loci affected by the *clb* mutations under analysis; a short description of the affected proteins is included. Scale bar represents 10 μm.

Since these chloroplasts are essential for plant development, an understanding of the precise mechanisms underlying their development has been the subject of intense research by several groups around the world, and a plethora of experimental strategies have been applied to explore this phenomenon; this has been extensively described in recent reviews [22,23]. In the past few years, proteomics has been successfully used as the main tool to identify key proteins for many different biological processes, such as cancer, human fertility, and abiotic stress responses in crops [24–29]. Given the degree of success of proteomic approaches, it is not surprising to find that several efforts have been made to determine chloroplast proteomes for different plant species, including Arabidopsis, rice, maize, and wheat [30–36]. In Arabidopsis, studies have been conducted to compare the proteomes of several albino mutants, such as *ppi2*, *ffc1-2*, and *clpr4*, against the proteome of wild-type plants [37–39].

In 2012, proteomic studies took one step further in plant biology, and a study was conducted to compare the proteomes of three independent pigment-defective mutants; the results were insight into the molecular phenomena affected in such albino plants and a better understanding of several key chloroplastic processes [40]. However, from these studies it was not possible to identify specific proteins involved in chloroplast biogenesis. In this work, we perform a similar comparative analysis using four pigment-defective *clb* mutants. To accomplish this, protein extracts from mutant seedlings of four *clb* mutant lines (*cla1-1*, *clb2*, *clb5*, and *clb19*) were resolved in a 2-D PAGE then digitized to generate a super-image. A robust set of comparisons was conducted to select those protein spots that showed differences in their abundance of at least twofold and were statistically significant according to Student's *t*-test ($P < 0.01$). Mass spectrometry (MS) was then used to identify the protein associated with each selected spot. Several of the identified proteins were chosen as candidates for important functions in chloroplast biogenesis. The progeny of the available mutant lines of the candidate proteins were screened for pigment-deficient phenotypes. With this strategy, we identified three mutant lines that segregate pigment-defective seedlings. The initial characterization of those mutants demonstrated that they affect the expression of specific chloroplastic and nonchloroplastic proteins that are allegedly involved in plastid biogenesis.

2. Materials and methods

2.1. Plant material and growth conditions

Arabidopsis thaliana heterozygous mutant lines corresponding to *cla1-1* (At4g15560; [9]), *clb2* (At3g11945; [19]), *clb19* (At1g05750; [19]), *clb5* ([19,21]), *emb1241* (SALK_045238), *pbp1* (SAIL_773_D06), and *atrabe1b* (SALK_069644) were used in this study. Seeds were surface-sterilized using the solutions of 100% C_2H_6O and 1% NaClO, then cultured on 0.5× Murashige & Skoog media supplemented with 0.05 g/l 2-(N-morpholino)ethanosulfonic acid, 0.5 g/l sucrose, 100 mg/l myo-inositol, 1 mg/l nicotinic acid, 1 mg/l pyridoxine-HCl, 10 mg/l thiamine-HCl, and 8 g/l phyto agar. Seedlings from the four mutant lines that presented the wild-type phenotype and the first pair of true leaves were harvested after 8 days of culture. These were then pooled for processing as the wild-type protein samples used in this study. In order to minimize the effect of using plants in different developmental stages (detection of development-related proteins), pigment-deficient plants were collected after 16 days of culture; at this time, all the seedlings display at least the first pair of true leaves. Three biologically independent seedling batches were generated for further processing.

2.2. Extraction and quantitation of total protein

Total protein extracts were prepared according to the phenol extraction protocol reported by Hurkman & Tanaka [41]; the adjustments made to the original protocol are described here. Briefly, the starting plant material was 1 g of mutant or wild-type seedlings grown in vitro on GM medium, seedlings were ground in a mortar using liquid nitrogen and re-suspended in extraction buffer (0.7 M sucrose; 0.5 M Tris; 30 mM HCL; 50 mM EDTA; 0.1 MKCl, 12 mg/ml PVPP (Polyvinylpolypyrrolidone) and 2% β-mercaptoethanol). An equal volume of water-saturated phenol was added followed by centrifugation (6000 g for 10 min) to separate the phases. The phenol phase was re-extracted with one volume of extraction buffer then precipitated with 5 volumes of 0.1 M ammonium acetate in methanol at −20 °C overnight. The protein precipitate was washed three times with 0.1 M ammonium acetate in methanol and once with 80% acetone at −20 °C. The pellets were air dried under vacuum and re-suspended in lysis buffer (8 M Urea; 2 M Thiourea, 4% (w/v) CHAPS; 2% ampholines (1.5% pH range 5–7 and 0.5% pH range 3–10) and 60 mM DTT). Determination of protein concentration in the extracts was determined by colorimetric assays as reported by Encarnación et al., 2005 [41].

2.3. 2-D PAGE and protein visualization

500 μg (analytical gels) or 750 μg (preparative gels) of total protein extracts were separated in 12% acrylamide gels under denaturating conditions. The first dimension was run using ampholytes in the range of pH 3 to 10 and enriched in pH 4 to 8. After 2-D electrophoresis, gels were fixed and stained using colloidal Coomassie brilliant blue, following [42]. The stained gels were digitalized using a GS-800 densitometer (Bio-Rad Hercules, CA, EUA) and the image analysis software PD-Quest 8.0.1 (Bio-Rad Hercules, CA, EUA).

2.4. In silico processing of gel images

Images from 2-D gels of three biologically independent protein extracts from mutants (*cla1-1*, *clb2*, *clb5*, and *clb19*) and wild-type plants were generated and processed using the PD-Quest 8.0.1 software (Bio-Rad, Hercules CA). Protein spots in all replicates were detected automatically by the software, and the detection was then improved by the manual addition of missing spots and the removal of erroneously detected spots. Normalization of gel images was performed using the local regression model normalization method provided by PD-Quest software. Furthermore, in order to properly compare the samples, the gel images were adjusted to fit a common distortion model; this was done by matching spots

that were common to all the gel images. The gel images from the different protein samples were compared to each other in order to generate a robust data set containing all the spots represented in the samples with 98% statistical confidence ($P < 0.01$) in a Student's *t*-test. Finally, the protein spots in the statistical data set displaying ± twofold abundance change were selected as candidates for the MS analysis.

2.5. MALDI-TOF mass spectrometry and protein identification

The selected protein spots were manually excised from the preparative gels. The samples were alkylated, reduced, and trypsin-digested prior to their elution and MS analysis. Samples of digested protein spots were automatically transferred to MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization-Time of Flight; Autoflex, Bruker Daltonics, Billerica, MA, USA) using Proteineer SP and SPII systems (software SPcontrol 3.1.48.0v; Bruker Daltonics, Breme, Germany). The Bruker Daltonics Autoflex system was operated in the delayed extraction and reflectron mode, and the resolution threshold was set to a signal-to-noise ratio of 1500. The specific protocols can be accessed in [42]. The m/z spectra were searched against the *A. thaliana* NCBInr (http://www.ncbi.nlm.nih.gov/guide/proteins/), SwissProt (http://www.isb-sib.ch/), and IPI (http://www.ebi.ac.uk/IPI/IPIarabidopsis.html) databases, using the Mascot (http://www.matrixscience.com) and Profound (http://prowl.rockefeller.edu) search engines. The Mascot engine was used to query NCBInr and SwissProt databases, while Profound was used to query NCBInr and IPI databases; both search engines were operated using a mass tolerance of 200 ppm, with cysteine carbamidomethylation as the constant modification and methionine oxidation as the variable modification. The significance threshold was set to $P < 0.05$ for the Mascot search.

2.6. In silico analysis of the identified proteins

Subcellular localizations were determined by manual query of the SUBA database (http://suba.plantenergy.uwa.edu.au/). Functional clustering of the identified proteins was performed using the Functional annotation tool available at DAVID (http://david.abcc.ncifcrf.gov/home.jsp) [43], and clustering was carried out using the annotations available at the Protein Information Resource (http://pir.georgetown.edu/pirwww/index.shtml) and Gene Ontology (http://www.geneontology.org/) databases. Stringency of the classification was set on medium and the rest of the options were set as default. Reconstruction of metabolic pathways was achieved using the metabolism overview pathways in the MapMan 3.5.1 software (http://mapman.gabipd.org/web/guest/mapman) with the Ath_AGI_TAIR9_Jan2010 mappings. MapMan was fed an array of data containing, for each protein, the $\log_2$ of the ratio of the detected abundance in each mutant over that registered in wild-type plants.

2.7. Northern blot, gene expression analysis

Total RNA was prepared from 12-day-old wild-type plants (Wassilewskija ecotype) and 16-day-old mutants cultured on GM media by extraction with TRIzol® Reagent (Life technologies, Carlsbad, CA USA). 5 μg of total RNA were subjected to electrophoresis in 1.2% (w/v) gels and transferred to Hybond-N + nylon membrane (GE Healthcare Life Sciences, Pittsburgh, PA, USA). RNA-Probe Hybridization was carried out under high-stringency conditions according to a previously described protocol [19]. Molecular probes were radiolabeled using ^{32}P with the Megaprime DNA labeling system (GE Healthcare Life Sciences). DNA probes used to monitor expression of *rrn16S* (AtCg00920), *accD* (AtCg00500), *psbA* (AtCg00020), and *RBCS* (At1g67090) genes were specific fragments from the corresponding genes; an *RPL21* (At1g35680) probe was obtained by DNA digestion (*SalI-NotI*) of the 146E8 clone from the ABRC stock center (http://www.arabidopsis.org).

2.8. Electron microscopy

Leaf or cotyledon samples were fixed for 1 h each in 4% paraformaldehyde and 2.5% glutaraldehyde. They were then washed in 100 mM Na-cacodylate buffer (EMS, Hatfield, PA, USA) at pH 6.5. Postfixation was performed in 1% OsO4 (EMS) in the same buffer at 4 °C for 1 h in the dark at room temperature. After the aldehyde fixation, the samples were washed with Na-cacodylate buffer and distilled water. Samples were dehydrated in a graded series of ethanol solutions followed by two changes in anhydrous propylene oxide, infiltrated in a graded series of epon-araldita resin (EMS) in propylene oxide, and finally left in a 100% embedding medium for 6 to 12 h at room temperature. Polymerization of specimens was made in molds at 60 °C for at least 24 h. Semithin (1 μ) and ultrathin (60 nm) specimens were sectioned in a LEICA (Wetzlar, Germany) ultramicrotome. Semithin sections were stained with 1% w/v toluidine blue made in 1% w/v aqueous borax solution for the digitalization of images in an Axio Scope ZEISS (Thornwood, NY, USA) microscope. Ultrathin sections were stained with 4% w/v uranyl acetate in 70% v/v ethanol followed by Reynold's lead citrate for digitalization of images in a JEOL 1200EXII Transmission Electron Microscope/GATAN (Warrendale, PA. USA).

3. Results

3.1. Sample comparison and protein identification

Total protein extracts (500 μg) from three biologically independent batches of wild-type plants and *cla1-1*, *clb2*, *clb5*, and *clb19* mutant lines were fractionated in 2-D PAGE gels; they were then stained with Coomassie blue. The gel images were processed as stated in the Material and Methods section. In this analysis, it was possible to detect more than 700 protein spots for each sample (Fig. 2A); with very few exceptions, most of the spots detected were shared between the different samples analyzed (Fig. 2A). Reproducibility of the experiments was assessed by calculating the mean coefficient of variation (CV) for each replicate group. Gel images from the *clb2* and Wt protein extracts displayed the lowest CVs, which were 29.41 and 29.52, respectively. On the other hand, images from the *cla1-1*, *clb5*, and *clb19* protein samples scored slightly higher CVs (30.3, 35.19, and 33.89, respectively; [44]).

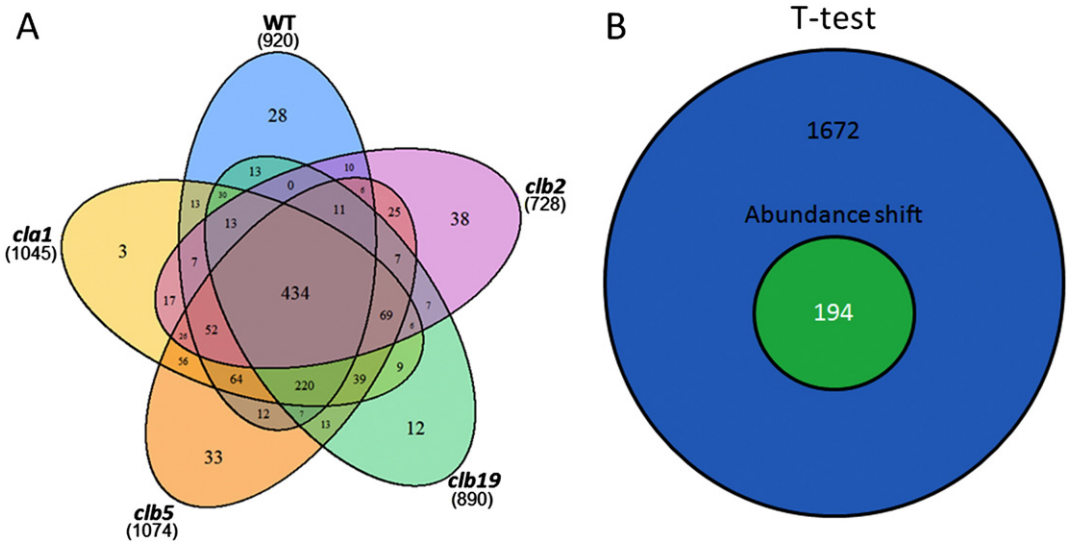


Fig. 2 – Venn diagrams depicting relationships existing between the detected spots. A) Venn diagram showing the total peptide spots detected on each mutant line and wild-type (Wt) under analysis, numbers indicate the total amount of spots detected in each protein sample; the intersections between the different subsets of spots are indicated by color. B) A Venn diagram is shown depicting the number of spots that displayed an abundance change of at least ±twofold (smaller green circle) and the total number of spots represented with a statistical confidence level of 98% (Student's *t*-test; $P < 0.01$) (larger blue circle). Null hypothesis for the Student's *t*-test states that the means of the tested populations (spots) are equal.

The total number of protein spots detected in this study is in agreement with the number of spots detected in a similar analysis of albino plants [45]. A *t*-test ($P < 0.01$) applied to the detected spots revealed that the presence in the gels of 1672 individual protein spots was statistically significant among all the samples (Fig. 2B), while only 194 spots displayed ± twofold differential abundance (Fig. 2B). Considering that the changes in the abundance of the protein spots between the analyzed mutants and wild-type protein samples might be related to differential regulation of chloroplast biogenesis, the dataset of the 194 protein spots identified to have changed their abundance in the samples was further analyzed using mass spectrometry.

Mass spectrometry analysis and subsequent database queries using different search engines resulted in the identification of 136 spots. A major constraint in the identification of proteins was the quality of the spectra. In some cases, different spots were identified as the same protein, which was probably due to posttranslational modifications, and the MS and bioinformatic analysis resulted in the identification of 96 unique proteins. Table 1 lists the identified proteins along with the data related to the search engine results and to their abundance in each mutant line.

3.2. The identified proteins have different subcellular localizations

Putative subcellular localization of the 96 MS-identified proteins was determined using the information available at the SUBA database (http://suba.plantenergy.uwa.edu.au/). As shown in Fig. 3, the list includes proteins with several subcellular localizations, such as the endoplasmic reticulum, cytoplasm, peroxisome, vacuole, nucleus, mitochondrion, plasma membrane, and chloroplast. We note that few of these proteins were predicted to have extracellular localization. However, the vast majority (37%; Fig. 3) of the analyzed proteins are presumably targeted to chloroplasts. This result supports the conclusion that the multicomparison method used in the present study efficiently detects changes in the abundance of chloroplast proteins, and similar results have been found in other proteomics studies of pigment-deficient mutants [45]. It is important to note that using the information available at SUBA, some proteins are predicted to be targeted to several cellular compartments, and further discrimination between the localizations was not possible due to lack of relevant data. The information related to the subcellular localization of each identified protein is available in Table 1.

3.3. The identified proteins cluster into stress-related groups and positively affect catabolic pathways

In order to discover proteins whose abundance shifted in all the mutants analyzed, the abundance of all the identified proteins in mutant plants was compared to their abundance in wild-type plants. Through this comparative analysis, 43 out of the 96 proteins were found down- and 42 up-regulated in at least three *clb* mutants. Inconsistent results were found among the mutants for eleven proteins [44]. The three protein sets (down-regulated, up-regulated, and inconsistent) were independently subjected to DAVID's (http://david.abcc.ncifcrf.gov/home.jsp) functional annotation tool (as described in the Material and Methods section). This method resulted in the generation of several functional clusters, as shown in Fig. 4.

The up-regulated protein subset formed only two groups with enrichment scores over the threshold (ES > 5), these groups represented functionalities related to abiotic stress

Table 1 – Proteins identified by mass spectrometry.

ID	Uniprot ID	Function	Subcellular localization	Mutant lines available	Spot abundance in WT (O. D.)	Spot abundance in *clb* mutants (O. D.)				Coverage (%)	Matched spectra	Unmatched spectra
						cla1-1	*clb5*	*clb19*	*clb2*			
At1g67090	P10795	RBCS-1A; RuBisCO small subunit 1A.	Chloroplast	Yes	2031.00	501	310	141	n. d.	29	6	9
At1g13950	Q9XI91	EIF-5A; traduction initiation factor 5A.	Cytoplasm/nucleus	Yes	624	996	2459	448	1852	29	5	28
At3g52960	Q949U7	Probable peroxiredoxin.	Chloroplast	Yes	n. d.	428	86	353	n. d.	31	7	13
At3g16640	P31265	TCPTP; translationally controlled tumour protein.	Chloroplast/nucleus/ membrane	Yes	264	239	261	223	232	60	8	13
At5g06290[a]	Q9C5R8	2-Cys peroxiredoxin A; reactive oxygen species detoxification.	Chloroplast	Yes	89	70	28	79	n. d.	32	5	8
At3g11630[a]	Q96291	2-Cys peroxiredoxin B; reactive oxygen species detoxification.	Chloroplast	Yes	186	158	62	203	42	32	5	6
At5g20720	O65282	CPN20; chaperonin 20.	Chloroplast	Yes	450	748	350	812	213	46	8	12
At1g06680	Q42029	PSBP-1; photosystem II P-1 subunit.	Chloroplast	No	715	51	n. d.	90	409	45	6	16
At1g19570	Q9FWR4	DHAR1; dehydroascorbate reductase.	Chloroplast/mitochondrion/ peroxisome/membrane	Yes	156	152	182	68	135	43	7	19
At3g01500[a]	Q56X90	Carbonic anhydrase 1 enzyme.	Chloroplast	No	470	110	n. d.	217	691	34	8	19
At1g07890[a]	C0Z2H6	Ascorbate peroxidase 1.	Chloroplast/membrane	Yes	624	996	2459	448	1852	42	6	16
At2g37220[a]	Q9ZUU4	Probable RNA binding protein cp29.	Chloroplast	Yes	233	258	30	223	42	21	6	8
At3g50820	Q9S841	PSBO2; photosystem II O-2 subunit.	Chloroplast	Yes	787	259	1234	187	1081	23	6	21
At2g37660	C0Z300	NADP binding protein.	Chloroplast	Yes	177	180	113	271	n. d.	36	8	15
At5g14740[a]	P42737	CA2; carbonic anhydrase 2 enzyme.	Cytoplasm	Yes	469	169	117	296	30	41	10	22
At5g17710[a]	Q94K56	EMB1241; Embryo deficient 1241; GrpE domain containing protein.	Chloroplast	Yes	65	43	38	42	n. d.	20	7	16
At1g20020	Q8W493	FNR2: ferredoxin NADP(+) oxidoreductase 2.	Chloroplast	Yes	305	91	70	142	n. d.	24	12	10
At5g66190	B9DI26	Ferredoxina NADP(H) oxidoreductase, leaf isoform.	Chloroplast	Yes	305	117	112	198	n. d.	37	11	16
At1g66430	Q9C524	PfkB-like carbohydrate kinase.	Chloroplast	Yes	69	42	83	52	44	21	7	15
At1g09340[a]	Q9SA52	CSP41B; RNA binding protein. Chloroplast RNA 23S metabolism.	Chloroplast/peroxisome	Yes	316	79	40	173	n. d.	32	10	18
At3g26650	P25856	Glyceraldehyde-3-phosphate dehydrogenase A subunit.	Chloroplast/membrane	Yes	304	120	156	352	n. d.	19	8	8
At4g04640	Q9SUI9	Chloroplast ATP synthase γ subunit precursor.	Chloroplast/membrane	No	405	65	73	91	121	21	7	9
At3g04120	P25858	Glyceraldehyde-3-phosphate dehydrogenase C subunit 1.	Membrane/nucleus/ mitochondrion/cytoplasm	Yes	221	126	212	219	295	37	10	10
At3g63140[a]	Q9LYA9	CSP41A; RNA binding protein. Chloroplast RNA 23S metabolism.	Chloroplast	Yes	307	115	87	120	n. d.	20	7	11
At5g65670[a]	D3K0E6	Indol-3-acetic acid induced protein 9.	Chloroplast/nucleus	Yes	65	29	n. d.	52	13	30	6	7
At1g04410	P93819	Malate dehydrogenase enzyme.	Chloroplast/nucleus/Vacuole	Yes	437	863	1136	775	969	26	7	12
At3g52930	Q9LF98	Probable fructose bisphosphate aldolase enzyme.	Chloroplast/nucleus/Vacuole	Yes	440	702	1120	524	669	48	12	18
At2g43750	P47999	OASB; o-acetylserine (thiol) liase B.	Chloroplast/mitochondrion	Yes	70	48	31	92	52	27	7	6
At4g38970	Q944G9	Probable fructose bisphosphate aldolase enzyme.	Chloroplast	Yes	932	607	210	987	192	22	7	12
At2g21330	Q9SJU4	Probable fructose bisphosphate aldolase enzyme.	Chloroplast	Yes	499	253	26	383	31	24	7	16
At1g35720[a]	Q9SYT0	Annexin; calcium-dependent membrane binding protein.	Chloroplast/membrane/ peroxisome	Yes	119	292	862	147	807	38	11	22
At1g79550	Q9SAJ4	Phosphoglycerate kinase enzyme.	Chloroplast/nucleus/ membrane/vacuole	Yes	207	348	314	293	469	25	7	8
At2g39730	P10896	RuBisCO activase enzyme.	Chloroplast/membrane	Yes	732	601	496	776	113	22	9	17

(continued on next page)

Table 1 (*continued*)

ID	Uniprot ID	Function	Subcellular localization	Mutant lines available	Spot abundance in WT (O. D.)	Spot abundance in *clb* mutants (O. D.)				Coverage (%)	Matched spectra	Unmatched spectra
						cla1-1	*clb5*	*clb19*	*clb2*			
At4g20360[a]	P17745	ATRABE1B; translation elongation factor. Homologous to RAB E1B.	Chloroplast/nucleus	Yes	622	843	284	896	183	22	8	13
At1g43670	Q9MA79	Fructose-1,6-bisphosphatase enzyme.	Chloroplast	Yes	163	56	52	97	97	27	7	9
At1g32060	P25697	Phosphoribulose kinase enzyme.	Chloroplast	Yes	516	239	102	633	49	25	11	17
At5g35630	B9DGD1	GS2; glutamin synthetase 2 enzyme.	Chloroplast/mitochondrion	Yes	873	585	216	1223	235	28	8	14
At4g05390	Q9M0V6	Ferredoxin NADP + reductase enzyme.	Chloroplast	Yes	234	52	42	150	34	24	6	17
At5g55220	Q8S9L5	Trigger factor-family protein.	Chloroplast	Yes	138	41	23	n. d.	n. d.	24	13	18
At4g01050	Q9M158	TROL; thylakoid rhodanasa-like protein. Flavoprotein anchor to thylakoid membrane.	Chloroplast	Yes	133	19	n. d.	50	n. d.	18	7	5
At4g24280	Q9STW6	cpHsc70-1; chloroplast heat shock 70-1 protein. Protein import.	Chloroplast/nucleus/ membrane	Yes	314	512	160	507	178	17	13	18
At5g49910	Q9LTX9	cpHsc70-2; chloroplast heat shock 70-2 protein. Protein import.	Chloroplast/membrane	Yes	232	153	82	82	58	13	9	19
AtCg00120	P56757	ATP synthase CF1 α subunit.	Chloroplast/nucleus/ membrane	Yes	121	39	n. d.	46	9	20	8	8
At3g17390	Q9LUT2	MTO3; methionine adenosyltransferase enzyme.	Nucleus/membrane	Yes	398	519	822	419	536	20	5	4
At3g09260[a]	Q9SR37	PYK10; β glucosidase.	Endoplasmic reticulum/ nucleus/membrane	Yes	585	1517	3693	1654	2450	16	8	16
At1g50480	Q9SPK5	THFS; formyltetrahydrofolate synthase enzyme.	Chloroplast/membrane	Yes	94	143	152	156	282	16	9	11
At1g29880[a]	O23627	Probable glycyl-tRNA synthetase enzyme.	Cytoplasm/mitochondrion/ membrane	Yes	95	85	60	75	44	13	9	5
At3g48870[a]	Q9SXJ7	HSP93-III; homologous to the regulatory subunit of Clp protease.	Chloroplast/membrane/ mitochondrion	Yes	89	98	36	97	33	11	7	9
At1g62750	Q9SI75	SCO1; snowy cotyledon. translation elongation factor G.	Chloroplast/mitochondrion	Yes	149	73	19	75	27	20	12	17
At3g56700	B9TSP7	Fatty acid reductase 6 enzyme.	Chloroplast/mitochondrion	Yes	179	15	42	105	34	19	7	32
At5g02500	P22953	HSC70-1; heat shock 70-1 cognate protein.	Chloroplast/nucleus/ peroxisome	Yes	504	316	379	167	n. d.	20	9	14
AtCg00490	O03042	RuBisCO large subunit.	Chloroplast	Yes	1326	865	404	184	173	19	9	11
AtCg00480	P19366	ATP synthase CF1 β subunit.	Chloroplast/membrane	Yes	104	85	n. d.	205	n. d.	21	8	10
At4g23670	Q9SUR0	Major latex protein-related.	Vacuole/membrane	Yes	624	996	2459	448	1852	43	12	18
At3g16450[a]	O04311	Probable lectin; mannose binding protein.	Nucleus/extracellular	Yes	183	99	220	55	223	46	11	14
At3g16420[a]	O04314	PBP1; PYK10 binding protein.	Nucleus/extracellular	Yes	76	107	279	177	219	30	7	14
At1g16080	C0Z2K9	Unknown.	Chloroplast	Yes	74	129	158	76	156	22	5	8
At5g55190[a]	Q8H156	RAN GTPase.	Membrane	Yes	260	471	153	732	223	20	5	19
At4g37930	Q9SZJ5	SHM1; serine transhydroxymethyl transferase 1 enzyme.	Chloroplast/nucleus/ mitochondrion	Yes	390	114	125	380	n. d.	22	12	8
At4g35090[a]	P25819	CAT2; catalase 2 enzyme.	Chloroplast/mitochondrion/ peroxisome	Yes	242	226	304	624	413	31	15	12
At3g22200	Q9LIE2	HER1; aminotransferase enzyme.	Mitochondrion/vacuole	Yes	86	96	134	101	242	13	5	8
At5g17920	O50008	METS; cobalamine-independent methionine synthase enzyme.	Chloroplast/membrane/ cytoplasm/vacuole	Yes	966	1908	1991	1609	1210	17	10	13
At2g26080	O80988	GLDP2; glycine decarboxylase P-2 enzyme.	Chloroplast/mitochondrion	Yes	23	n. d.	43	n. d.	n. d.	10	9	12
At3g60750	Q9LZY8	Transketolase enzyme.	Chloroplast	Yes	499	472	173	661	478	13	7	14
At3g12580[a]	Q9LHA8	HSP70; heat shock 70 protein.		Yes	89	244	154	251	259	12	7	8

			Mitochondrion/vacuole/ membrane									
At3g15950[a]	Q712H0	TSA1; TSK-associated protein like.	Endoplasmic reticulum/ peroxisome	Yes	41	101	693	64	742	19	15	29
At1g42970	P25857	Glyceraldehyde-3-phosphate dehydrogenase B subunit.	Chloroplast	Yes	393	286	157	611	117	17	7	16
At1g65930	Q9SRZ6	Isocitrate dehydrogenase enzyme.	Chloroplast/membrane	Yes	338	709	1083	719	780	23	9	13
At5g07440	Q38946	GDH2; glutamate dehydrogenase subunit 2.	Mitochondrion/vacuole	Yes	75	164	162	280	140	22	11	19
At2g36530	P25696	LOS2; enolase.	Extracellular	Yes	1654	1487	1757	1058	1149	27	10	14
At5g08670	Q8H135	ATP synthase β subunit.	Mitochondrion/peroxisome	Yes	1275	1011	1133	941	1031	33	14	22
At1g64190	Q9SH69	6-phosphogluconate dehydrogenase enzyme.	Chloroplast/membrane	Yes	n. d.	137	163	228	n. d.	25	10	10
At3g16400[a]	Q9SDM9	Mirosinase binding protein.	Extracellular	Yes	157	378	313	216	902	24	9	24
At3g03250	Q9M9P3	UDP-glucose pyrophosphorylase enzyme.	Membrane	Yes	187	161	199	188	153	20	8	13
At2g24200	P30184	Aminopeptidase enzyme.	Chloroplast/membrane/ vacuole	Yes	52	213	190	180	283	20	7	7
At2g28000	P21238	CPN60; 60 kDa chaperonin α.	Chloroplast/mitochondrion	Yes	335	614	178	591	403	53	10	23
At1g77510	Q9SRG3	ATPDIL1-2; disulfide isomerase enzyme.	Endoplasmic reticulum/ vacuole	Yes	60	79	85	125	97	15	7	7
At3g19170[a]	Q9LJL3	PREP1; zinc-dependent protease. Transit peptide degradation.	Chloroplast/mitochondrion	Yes	83	95	71	149	54	14	12	7
At1g22530[a]	Q56ZI2	Patellin II; membrane-trafficking events during cytokinesis.	Plasma membrane	Yes	14.7	9.5	7.7	11.9	11.3	18	10	8
At1g78900	O23654	VHA-A; catalytic subunit of the peripheral V1 complex of vacuoler ATPase.	Plasma membrane/vacuole	Yes	173.2	310.9	257.7	347.9	198.1	58	37	44
At3g25230	Q38931	FKBP62; co-chaperone that modulates thermotolerance by interacting with HSP90.	Plasma membrane/ cytoplasm	Yes	86.1	33.9	67.4	46.4	41.2	15	8	12
At3g16460	O04310	JACALIN-RELATED LECTIN 34; mannose-binding lectin superfamily protein.	Peroxisome	Yes	n. d.	71.1	316.5	n. d.	907.7	35	21	36
At1g76030	P11574	VHA-B1; non-catalytic subunit of the peripheral V1 complex of vacuoler ATPase.	Chloroplast/membrane/ vacuole	Yes	120.4	231.7	190.4	219	613.3	34	13	25
At3g09440	O65719	HSC70-3; heat shock 70-3 cognate protein.	Membrane	Yes	240.6	523.5	704.6	402.6	736.1	41	23	18
At2g05710	Q9SIB9	ACO2; catalyzes the isomerization of citrate to isocitrate via cis-aconitate.	Chloroplast/mitochondrion/ membrane	Yes	227.3	151.8	287.1	200.7	452.7	37	27	30
At3g02090	Q42290	Beta-MPP; cleaves (transit peptides) from Mitochondrionl protein precursors.	Mitochondrion	Yes	150	190.9	267.8	178.7	526	48	20	40
Atmg01190	P92549	ATPA; mitochondrionl membrane ATP synthase.	Mitochondrion	Yes	392.7	716.8	1036.7	608.5	929	42	18	34
At1g59720[a]	Q0WQW5	CCR28; pentatricopeptide repeat protein. Editing of multiple plastid transcripts.	Chloroplast	Yes	59.3	91.2	167	82.5	126.3	13	9	16
At3g16430[a]	O04313	Jacalin-related lectin 31.	Extracellular	Yes	n. d.	n. d.	n. d.	n. d.	87.9	44	13	29
At3g19160	Q9LJL4	IPT8; adenylate dimethylallyltransferase 8 enzyme.	Chloroplast	Yes	n. d.	18.5	n. d.	n. d.	130.2	15	3	14
At1g48090	NA	Calcium-dependent lipid-binding protein.	Chloroplast/membrane	Yes	50.1	n. d.	62.9	n. d.	90.6	13	36	125
At2g39310	O80950	Myrosinase-binding protein-like.	Membrane	Yes	n. d.	51.3	n. d.	n. d.	510.9	32	12	10
At1g65840	Q8H191	PAO4; flavoenzyme. Oxidation of the secondary amino group of spermine.	Peroxisome	Yes	n. d.	14.3	45.4	n. d.	64.4	29	8	26
At5g04110	Q8VY11	GYRB3; DNA gyrase B3.	Membrane	Yes	208.2	257	330.9	204.9	423	15	7	44
At4g34050	O49499	CCOAOMT1; caffeoyl-CoA O-methyltransferase 1 enzyme.	Unknown	Yes	73.9	123.7	148	89.2	195.9	32	9	16
At1g71330	Q9FVV9	NAP5; probable non-intrinsic ABC protein 5.	Membrane	Yes	21	57.4	59.6	61.3	107	13	5	25

n. d. (not detected).
O. D. (optic density).
[a] CLB-related protein candidates.

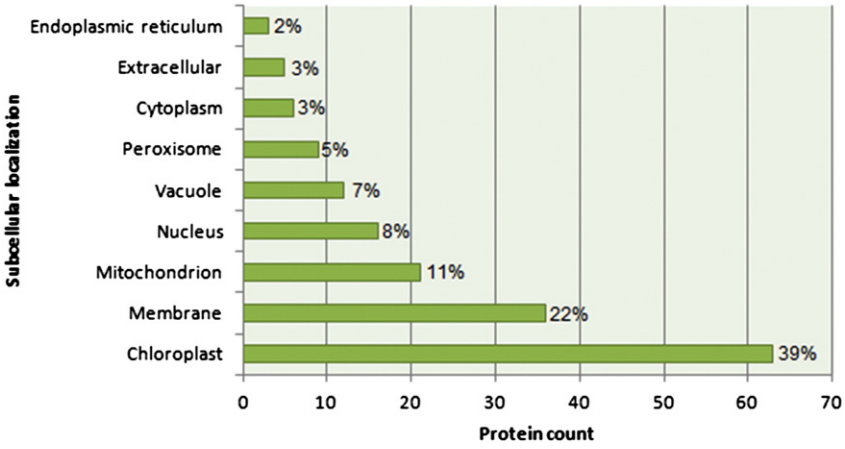


Fig. 3 – Predicted subcellular localization of proteins. The graph shows the number of proteins annotated for each subcellular compartment. The fraction of proteins in each localization is also indicated. Information was retrieved from the SUBA database (http://suba.plantenergy.uwa.edu.au/).

responses and included peroxidases such as CATALASE 2 and ANNEXIN D1 (Fig. 4A, B). In contrast, the down-regulated protein subset clustered in at least five groups with significant enrichment scores ([44]; ES > 5). Three of the most significant clusters include a chloroplast-related group (Fig. 4C) and two stress-related functional groups (Fig. 4D, E). Finally, the subset of proteins whose abundance shift was not consistent among the mutant plants formed two small clusters related to chloroplasts (Fig. 4F) and to abiotic stress (Fig. 4G); both clusters displayed enrichment scores under the significance threshold. These subsets displayed enrichment of different GO terms that link the proteins in each subset to common molecular processes.

Further analysis using MapMan 3.5.1R2 software (http://mapman.gabipd.org/web/guest) revealed that several metabolic processes are clearly affected in the *clb* mutant plants. For instance, it is clear that functions related to sucrose synthesis, photosynthetic light reactions and the Calvin cycle display strong down-regulation in the mutant lines compared to their behavior in the wild-type plants (Fig. 5). While functions related to catabolic processes, such as glycolysis, mitochondrial electron transport, the citric acid cycle, metabolism of some amino acids, and NH_3 homeostasis showed evident up-regulation in the different mutant backgrounds (Fig. 5). Details about the specific functions represented by each bin in Fig. 5 can be found in [44].

From the presented data, it is possible to speculate that the lack of fully functional plastids in pigment-deficient mutants such as *clb* and *apg* dramatically alters the plants' stress responses and the affected plants are capable of modulating their metabolism in order to adapt to heterotrophy.

3.4. *Identification of putative novel chloroplast biogenesis regulators*

All the proteins identified in this analysis (listed in Table 1) were manually annotated for their function (Description, Table 1), subcellular localization and availability of mutant lines using databases such as TAIR (http://www.arabidopsis.org/), UniProtKB (http://www.uniprot.org/), and SUBA. The list of 96 proteins was screened in the search for novel chloroplast biogenesis-related proteins. In that screening, those proteins with no mutant lines available, as well as those with annotated roles related to primary metabolism, chloroplast development, or chloroplast function, were discarded (Fig. 6A). We also excluded proteins whose predicted subcellular localization was not chloroplastic, except for some endoplasmic reticulum proteins, since recent evidence has demonstrated that at least one protein is transported to chloroplasts via an endoplasmic reticulum–Golgi pathway [46]. Furthermore, the specific function of each protein was assessed to identify proteins participating in important processes of the development or functioning of the chloroplasts, such as mRNA editing, chloroplast gene expression, and importation of proteins.

Taking into account the information presented in Table 1 and the screening parameters described above, we selected 26 candidate proteins to investigate as possible chloroplast biogenesis-related proteins. Available mutant lines for the 26 selected candidates were obtained from the Arabidopsis Biological Resource Center (Table 1), and a phenotypic screening was conducted in the search of plants displaying a pigment-deficient phenotype, which has been unequivocally associated to defects in the development or function of chloroplasts. Three mutant lines from this collection segregated pigment-deficient seedlings in their progeny. The *emb1241* (TAIR: At5g17710; SALK_045238) and *pbp1* (TAIR: At3g16420; SAIL_773_D06) mutants segregated albino and yellow phenotypes, respectively, while the *atrabe1b* (TAIR: At4g20360; SALK_069644) mutant segregated seedlings with slight variegation in cotyledons (Fig. 6B). Interestingly, none of these mutants have been previously characterized for their defects in the biogenesis or function of chloroplasts. Along with these putative novel chloroplast biogenesis regulators, several proteins were identified with known functions in chloroplast biogenesis or function, such as RBCS, CPN60, SCO1, and RuBisCO activase (Table 1).

3.5. *Expression levels of chloroplast development marker genes on mutant lines*

To obtain a more detailed characterization of the chloroplast defects in each of the selected mutants, the expression of several chloroplast- and nuclear-encoded marker genes related

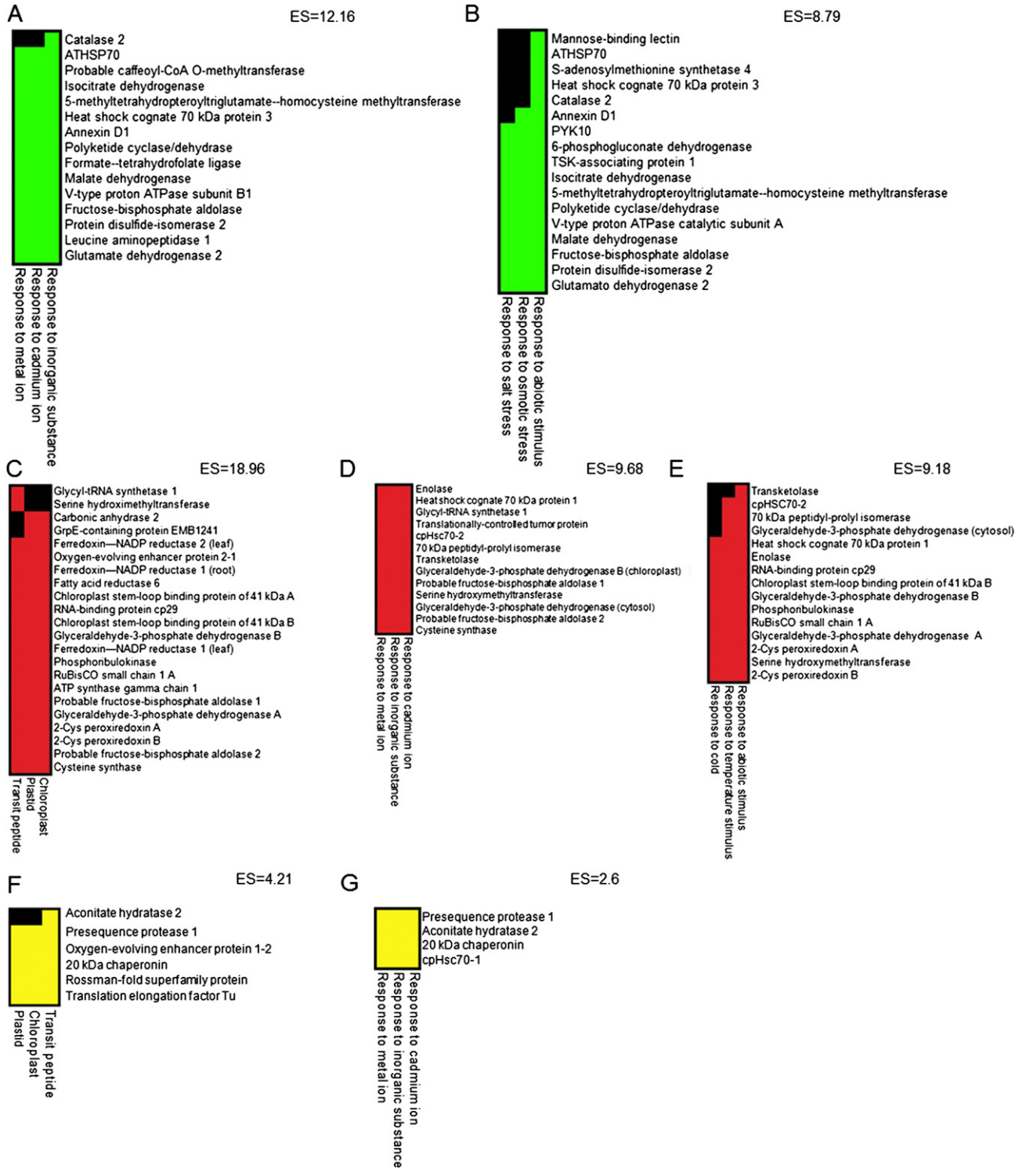


Fig. 4 – Functional clustering of the identified proteins. Significant clusters are shown of up- (A, B) and down-regulated (C, D, E) proteins, according to DAVID's functional annotation tool (http://david.abcc.ncifcrf.gov/home.jsp). Two nonsignificant clusters formed by the proteins with inconsistent abundance (F, G) are also displayed. The protein names on each cluster are shown at the right side, the corresponding GO terms are shown at the bottom of the corresponding panel, and the enrichment scores (ES) are shown at the top. The Enrichment Score is the geometric mean (in −log scale) of member's p-values in a corresponding annotation cluster.

to chloroplast functionality or development was determined. For this analysis, we choose the plastid- and the nuclear-encoded genes associated with chloroplast development, as shown in Table 2. As can be seen in Fig. 7, the different mutant lines under analysis display different expression levels of each marker gene; the *emb1241* and *pbp1* mutants clearly displayed

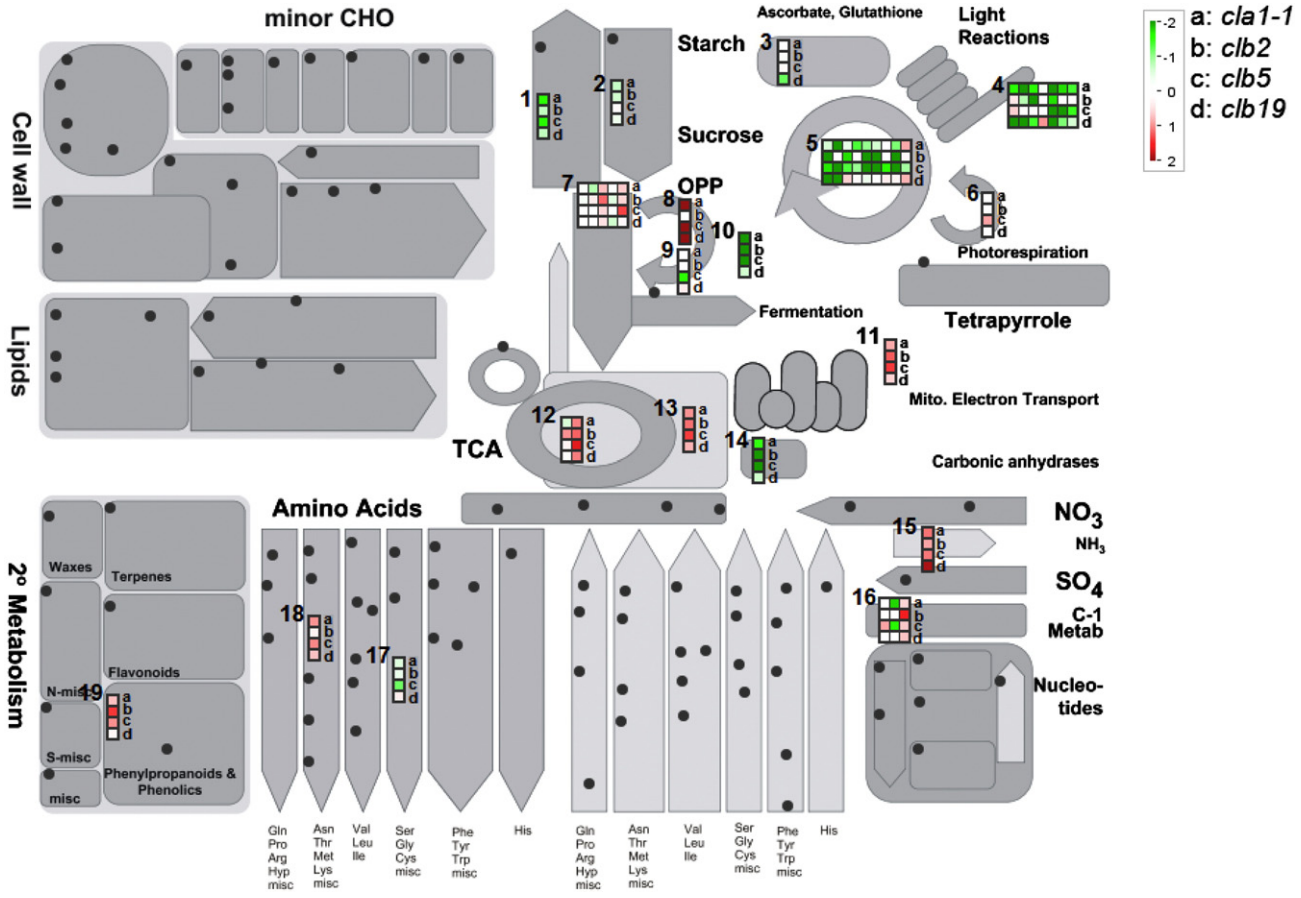


Fig. 5 – Mapping the identified proteins to metabolic pathways. The protein list (Table 1) was mapped to a database containing all the metabolic pathways of *Arabidopsis thaliana* using MapMan 3.5.1R2 software (http://mapman.gabipd.org/web/guest); matches between functions annotated in the database for a given pathway and the protein functionalities of the analyzed protein list are shown as a bin or group of bins (colored squares), whose color corresponds to its abundance level in each mutant compared to the wild-type levels. Abundance data is plotted using a $\log_2$ scale.

the strongest down-regulation of most marker genes, while the *atrabe1b* mutant showed expression levels closer to those of the wild-type (Fig. 7A, B). Compared to the wild-type expression levels, the *emb1241* mutant plants showed strong down-regulation of *psbA* (44%), *RBCS* (34%), *rrn16S* (43%), and *RPL21* (30%) genes; in a similar way, the expression levels of *psbA* (33%), *RBCS* (50%), and *rrn16S* (60%) marker genes displayed severe down-regulation in the *pbp1* mutant. On the other hand, mild down-regulation of *RPL21* (80%) was detected in *pbp1* mutants, as well as mild down-regulation of *accD* in both *emb1241* (95%) and *pbp1* (79%) plants.

Finally, the *atrabe1b* mutants displayed only small differences, if any, in the expression levels of most marker genes compared to those from the wild-type; *psbA* transcript levels did not display changes, compared to the wild-type plants, while *rrn16S* (106%) and *accD* (75%) genes showed a very slight increase and mild down-regulation of mRNA accumulation, respectively. Interestingly, in this mutant a substantial increase in *RBCS* (129%) and *RPL21* (152%) transcript levels was observed. These results indicate that the *embl241* and *pbp1* mutant lines are most affected for chloroplast biogenesis, while the *atrabe1b* line is least affected.

3.6. Chloroplast ultrastructure of mutant lines

Finally, to further access the developmental defects associated with each of these mutations, the ultrastructure of the plastids in each mutant line was analyzed using electron microscopy (Fig. 8). Compared to the wild-type plant, the *atrabe1b* mutant was found to be heteroplastidic, containing both wild-type-like and malformed plastids (Fig. 8). In the case of *emb1241* and *pbp1*, both mutants contain plastids of small size in which the development was arrested at an early stage (Fig. 8). Plastids in the *pbp1* mutant contain undeveloped internal membranes with long, single, linear membranes and have dense vesicles. In the case of *emb1241*, the plastids seemed to be arrested earlier since they lacked almost all internal membranes (Fig. 8).

4. Discussion

In this study, several different yet related mutant plants were compared using a classical proteomics workflow coupled with

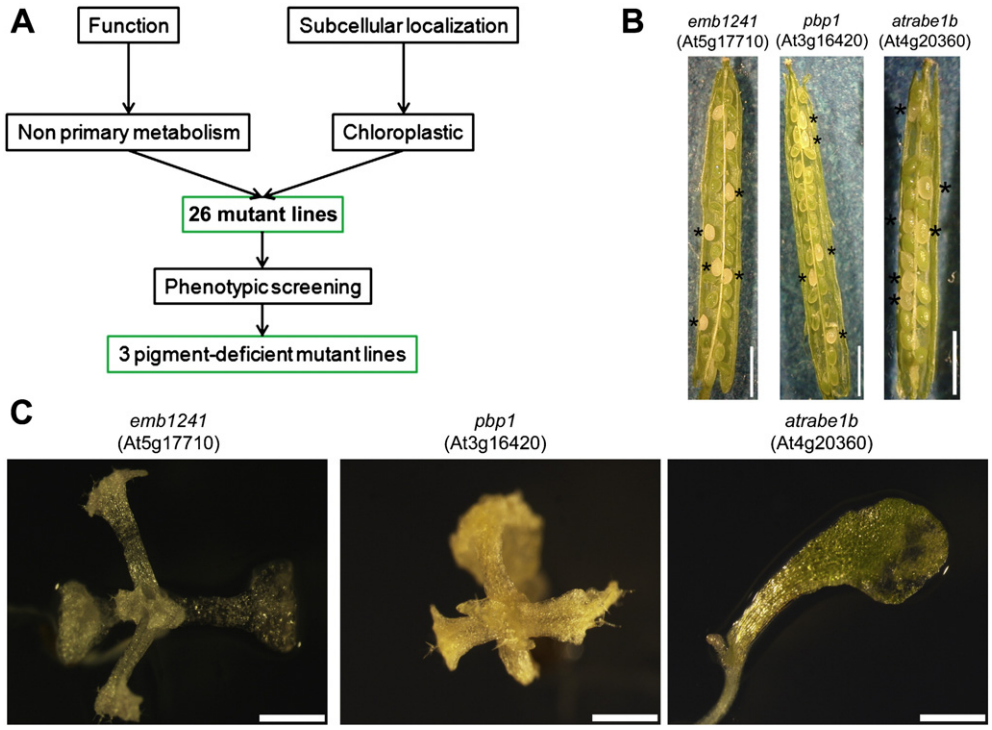


Fig. 6 – Screening the mutants for CLB-related proteins. The pipeline followed for the screening of the protein list is depicted (A). Criteria used for the candidate selection are shown in boxes. The 26 acquired mutant lines subjected to phenotypic screening are shown in Table 1. Siliques (B) and 16 day-old seedling (C) from the three mutant lines found to segregate plants with pigment accumulation defects (*emb1241*: albino; *pbp1*: yellow; *atrabe1b*: variegated) are shown. Scale bars represent 25 mm (B) and 10 μm (C).

phenotype screenings and a human-driven functional assessment; this resulted in the successful identification of proteins related to the development or functions of chloroplasts. Several of the identified proteins have known functions in the chloroplast machinery for protein import, photosynthesis, and other important chloroplast-related processes; such is the case for RBCS, CPN60, SCO1, and RuBisCO proteins (Table 1) [47–50]. Since these proteins are known to affect chloroplast development or function, we concluded that their presence among the 96 identified proteins validates our strategy. Furthermore, it was also possible to identify three proteins with mutant lines that displayed pigment-defective phenotypes and whose functions in chloroplast function have not yet been explored.

The up-regulated proteins cluster into two abiotic stress-related groups, one containing proteins related to ion stress (Fig. 4A; including response to Cd^{2+} ion) and the second one related to salt and osmotic stress (Fig. 4B); also, the down-regulated and inconsistent protein sets formed clusters that contain proteins related to the same ion stress (Fig. 4D, G). Similar results were reported by Motohashi et al. in 2012, when they compared the proteomes of 3 pigment-deficient mutants (*apg1–3*) and wild-type plants, one of the enriched GO terms consistently found to be 5 to 10 times up-regulated in

Table 2 – Chloroplast development molecular marker probes.

Monitored events	Molecular probe	Gene ID	Coding genome	Description
Early transcriptional activity				
	rrn16S	AtCg00920	Plastid	16S rRNA; highly expressed in proplastids.
	rpl21	At1g35680	Nucleus	Ribosomal protein L21; expressed early after seed imbibition.
Late transcriptional activity				
	accD	AtCg00500	Plastid	Acetyl-CoA carboxylase; chloroplast lipid biosynthesis; only transcribed by NEP[a]
Late chloroplast function				
	psbA	AtCg00020	Plastid	Photosystem II reaction center D1 protein.
	RBCS	At1g67090	Nucleus	RuBisCO small subunit.

[a] Nuclear-encoded RNA polymerase.

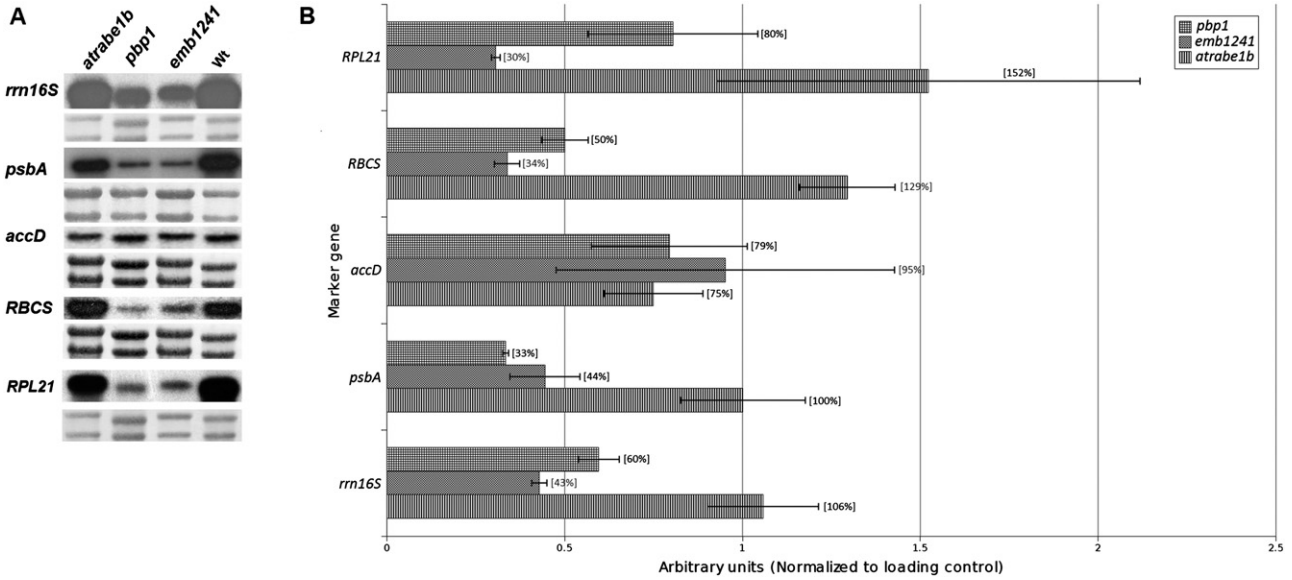


Fig. 7 – CLB marker gene expression analysis. A) Northern-blot hybridization of molecular probes for indicated marker gene on total RNA extracts from *atrabe1b*, *pbp1*, *emb1241*, and the wild-type (Wt) Wassilewskija (WS) ecotype. B) Densitometry analysis of Northern-blot hybridization, the ratio of the expression level of each probe in the mutants against the wild-type plants is shown (normalized to the corresponding loading controls; methylene blue staining; panels under each probe signal). For all mutants, the percentage of expression of each molecular probe compared to the wild-type plants is shown inside brackets: *rrn16S* (16S rRNA), *psbA* (Photosystem II reaction center D1 protein), *accD* (Acetyl-CoA carboxylase), *RBCS* (RuBisCO small subunit), *RPL21* (Ribosomal protein L21). A more detailed description of molecular probes is shown in Table 2. Error bars represent the mean ± standard error (SE) from data obtained from five independent experiments.

the *apg1–3* mutants was the response to Cd^{2+} ion [40]. There may not be a straightforward reason why all these pigment-deficient mutants display similar regulation of the response to ion stress, but it is possible that chloroplasts have molecular mechanisms that sense divalent cations (essential or toxic) in the cellular environment, such as Cd^{2+} and Mg^{2+}, and modulate the expression of proteins whose function might be affected by the intracellular levels of such ions. At the transcriptional level,

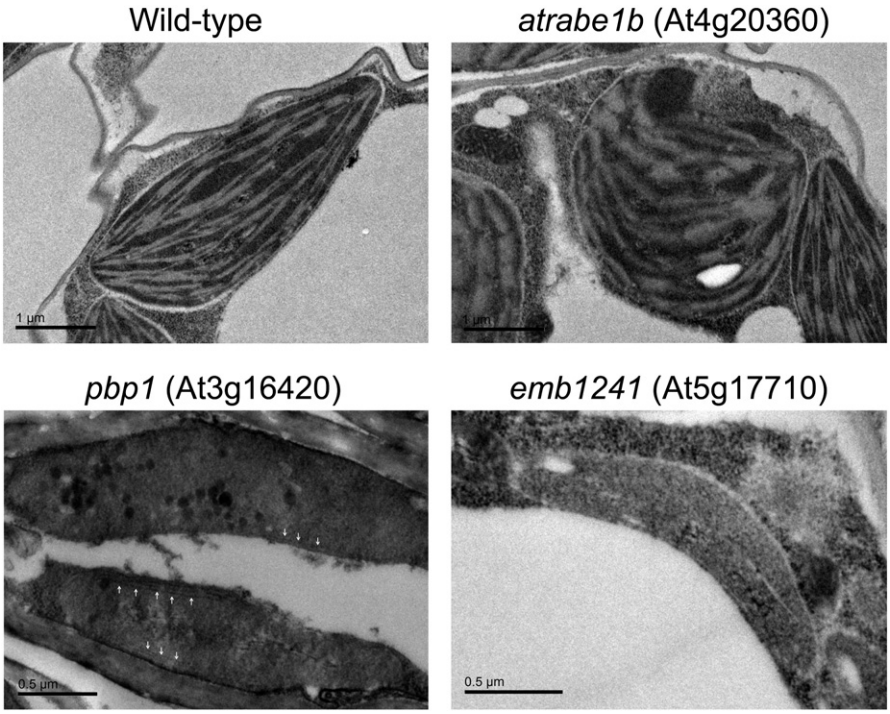


Fig. 8 – Ultrastructure of mutant plastids. Ultrastructure of mutant plastids was resolved by electron microscopy of leaves (*pbp1* and *emb1241*) or cotyledons (wild-type like and *atrabe1b*) from 16-day-old plants. Arrows indicate endomembranes.

evidence has been found that Mg^{2+} regulates the stability of chloroplastic mRNAs, such as *psbA*, *rbcL*, and *rrn16S* [51], strongly suggesting the existence of ion-sensing molecular mechanisms that might regulate protein accumulation in chloroplasts. Moreover, the down-regulated protein subset contains proteins that cluster due to temperature stress; also, Motohashi et al. reported the down-regulation of proteins related to responses to fungus and water deprivation in the *apg1–3* mutants [40]. Considering that data, it seems likely that chloroplasts function as stress-processing components, gathering information about the cellular conditions and modulating the expression of cpDNA- and nucleus-encoded genes. In the absence of properly developed chloroplasts (as in the case of the *clb* and *apg* mutants), the correct regulation of such genes and their products would be lost.

Further bioinformatic analysis of the protein list revealed that all four mutant lines display strong down-regulation of photosynthetic functions, both light-dependent and light-independent reactions (Fig. 5); at the same time, molecular processes involved in catabolic pathways show up-regulation in the mutant lines, clearly indicating that pigment-deficient plants (including *apg* mutants [40]) are forced to adopt a heterotrophic lifestyle. These results correlate with similar data reported by Motohashi et al. and directly reflect at the molecular level the impairment in chloroplast assembly displayed by the *clb* and *apg* mutants. Other processes consistently affected in the *clb* mutants are those related to amino acid metabolism and NH_3 homeostasis; the catabolism of glycogenic amino acids such as methionine is enhanced in mutant plants (Fig. 5, [44] Table S2 probably due to the possibility that its carbon skeleton can be assimilated into glucose, making this amino acid a valuable source of energy in the heterotrophic cellular environment; similar results were obtained by Motohashi et al., through the analysis of *apg1–3* mutants [40].

The data presented in Figs. 7 and 8 corroborate at the molecular and cellular levels the pipeline presented in Fig. 6; this validates assessment of function as a critical step when screening for proteins implicated in certain phenomena. Our gene expression and microscopic analyses demonstrated that *emb1241* and *pbp1* mutant lines had altered (earlier) chloroplast biogenesis, compared to *atrabe1b*. The *atrabe1b* mutant is able to assemble both normal and malformed plastids, even though the malformed plastids are able to generate considerable amounts of endomembranes and probably remain functional (Fig. 8). Given the low expression levels of the two early transcriptional activity marker genes (Fig. 7) and the complete lack of endomembranes in the plastids (Fig. 8), we concluded that the *emb1241* mutant is affected earliest in the chloroplast developmental processes, followed by the *pbp1* mutant. The slight down-regulation of *accD* transcript levels in *pbp1* mutants, along with the incipient formation of endomembranes in their plastids (Fig. 8), indicate that plastids in the *pbp1* mutants are able to progress to later developmental stages, which correlates with the slightly yellow pigmentation of the *pbp1* plants, indicating low levels of chlorophyll biosynthesis.

As shown in Table 1, the *EMB1241* gene codes for a chloroplastic GrpE-containing protein. GrpE proteins are bacterial co-chaperones that function as nucleotide-exchange factors in the chaperone/co-chaperone system DnaK/DnaJ/GrpE (reviewed in [52,53]). They enhance the activity of DnaK chaperones (such as Hsp70s) by improving the rate of release of the hydrolyzed ATP from the chaperon's ATPase domain [54,55]. It is likely that the EMB1241 protein plays a role in the import of proteins into the chloroplast to the stroma, given that bacterial-like chaperones such as Hsp70 and DnaJ exist in association with the TOC/TIC import machinery [56–61]. Moreover, it has been demonstrated that mutant lines lacking GrpE in *Physcomitrella patens* are defective in translocating proteins into chloroplasts [62]. Our data strongly support the conclusion that EMB1241 plays an important but not yet characterized role in the importation of proteins into *Arabidopsis* chloroplasts.

In the case of the *pbp1* mutant, we observed important defects in the endomembrane system, which contained only a few scattered thylakoids. At the transcription level, we found a slight down-regulation of the early transcriptional marker *rrn16S*, but strong down-regulation of late marker genes, which indicate that these plastids are able to progress to later developmental stages, compared to the *pbp1* mutant. This is supported by the yellow pigmentation of these mutants. Interestingly, the PBP1 protein is the major interacting protein of the inactive form of PYK10, a very well-characterized β-glucosidase enzyme that is involved in the interactions of plant roots and the mutualistic fungus *Piriformospora indica* [63]. PBP1 and PYK10 are predicted to have endoplasmic reticulum localization, and it is thought that PBP1 functions as a chaperone that is necessary for the proper folding of PYK10 that is required for activity [63,64]. Despite the fact that PBP1 has been studied in the past, the pigment-deficient phenotype displayed by *pbp1* mutant plants has not been previously reported. Interestingly, we observed only mild down-regulation (79%) of *accD* transcript levels in *pbp1* plants. The *accD* gene encodes for the Acetyl-CoA Carboxylase, an enzyme that catalyzes the synthesis of Malonil-CoA, which is the first committed metabolite of the chloroplastic fatty acid synthesis pathway [65]. This observation, in addition to the existence of contact sites between ER and chloroplast membranes that mediate lipid trafficking between both organelles and thus impacts thylakoid membrane composition [66–69], suggest that *pbp1* plants might contain altered lipid levels that affect the composition and function of thylakoid membranes. This possibility will be addressed in a future analysis. Furthermore, in 2005, Villarejo et al. reported that the plastid localized carbon anhydrase 1 (CA1) is imported through a secretory pathway. This finding provided evidence for the existence of a novel protein import pathway to the chloroplasts that depend on the endoplasmic reticulum and Golgi [46]. Taking into account the strong molecular and cellular chloroplast-development defects displayed by the *pbp1* mutant plants, the extensive evidence showing PBP1 endoplasmic reticulum localization and the existence of ER-dependent pathways to deliver proteins and lipids into the chloroplasts, it is possible that the PBP1 protein plays a role mediating one or both transport systems.

Finally, the ATRABE1B protein is a chloroplast resident bacterial-like Tu-type translation elongation factor (Table 1). Participation of putative bacterial-like translation elongation factors in chloroplast development has been previously documented; the SVR3 (suppressor of variegation 3) protein is a chloroplastic putative bacterial-like typA translation elongation

factor, and mutant lines affected in the expression of the *SVR3* gene display a pale green phenotype under normal conditions. Moreover, this phenotype is accentuated when mutants are challenged to cold stress [70]. ATRABE1B is part of the GTP-binding elongation-factor protein family that is composed of 6 members of *Arabidopsis*; given the slight variegation phenotype shown by the *atrabe1b* mutant (Fig. 6B), it is possible that functional redundancy exists within other members of the protein family, attenuating an otherwise strong phenotype.

5. Concluding remarks

In this study, using four pigment deficient mutants, we were able to identify proteins that were commonly or specifically affected among these lines. Further analysis allowed the identification of three novel proteins whose mutant lines displayed pigment-deficient phenotypes; transcriptional and electron microscopy analysis of these mutant plants showed that the identified proteins modulate the development or function of chloroplasts. Also, our analysis revealed clear evidence of heterotrophy as the *clb* mutant energy production strategy, as well as their severe impairment on photosynthesis.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.jprot.2014.07.003.

Transparency document

The Transparency document associated with this article can be found, in the online version.

Acknowledgments

The authors thank Patricia Jarillo for the technical support, Alma Lidia Martínez, Juan Manuel Hurtado, Roberto Rodríguez Bahena and Arturo Ocádiz for computer support, Paul Gaytan and Eugenio López for oligonucleotide synthesis and Guadalupe Zavala Padilla (UME, Instituto de Biotecnolgía, UNAM) and Fernando García Hernández (UM, Instituto de Fisiología Celular, UNAM) for chloroplast electron microscopy analysis. This work was supported by UNAM-DGAPA-PAPIIT (grants IN217111 to AGG and IN208211 to PL) and CONACYT-México (CB-129266 to AGG and 127546 to PL).

REFERENCES

[1] Kannangara C, Henningsen K, Stumpf P, Appelqvist L, von Wettstein D. Lipid biosynthesis by isolated barley chloroplasts in relation to plastid development. Plant Physiol 1971;48:526–31.
[2] Lichtenthaler H. The 1-deoxy-d-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Plant Mol Biol 1999;50:47–65.
[3] Kirk PR, Leech RM. Amino acid biosynthesis by isolated chloroplasts during photosynthesis. Plant Physiol 1972;50:228–34.
[4] Baldwin A, Wardle A, Patel R, Dudley P, Park SK, Twell D, et al. A molecular–genetic study of the Arabidopsis Toc75 gene family. Plant Physiol 2005;138:715–33.
[5] Hust B, Gutensohn M. Deletion of core components of the plastid protein import machinery causes differential arrest of embryo development in *Arabidopsis thaliana*. Plant Biol 2006;7:18–30.
[6] Chou ML, Fitzpatrick LM, Tu SL, Budziszewski G, Potter-Lewis S, Akita M, et al. Tic40, a membrane-anchored co-chaperone homolog in the chloroplast protein translocon. EMBO J 2003;22:2970–80.
[7] Kovacheva S, Bedard J, Patel R, Dudley P, Twell D, Rios G, et al. In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. Plant J 2005;41:412–28.
[8] Constan D, Froehlich JE, Rangarajan S, Keegstra K. A stromal Hsp100 protein is required for normal chloroplast development and function in *Arabidopsis*. Plant Physiol 2004;136:3605–15.
[9] Mandel MA, Feldmann KA, Herrera-Estrella L, Rocha-Sosa M, Leon P. CLA1, a novel gene required for chloroplast development, is highly conserved in evolution. Plant J 1996;9:649–58.
[10] Guevara-Garcia A, San Román C, Arroyo A, Cortés ME, De La Luz Gutiérrez-Nava M, León P. Characterization of the *Arabidopsis* clb6 mutant illustrates the importance of posttranscriptional regulation of the methyl-D-erythritol 4-phosphate pathway. Plant Cell 2005;17(2):628–43.
[11] Chateigner-Boutin AL, Ramos-Vega M, Guevara-García A, Andrés C, De La Luz Gutiérrez-Nava M, Cantero A, et al. CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts. Plant J 2008;56(4):590–602.
[12] Koussevitzky S, Stanne TM, Peto CA, Giap T, Sjögren LL, Zhao Y, et al. An *Arabidopsis thaliana* virescent mutant reveals a role for ClpR1 in plastid development. Plant Mol Biol 2007;63(1):85–96.
[13] Schweer J, Türkeri H, Kolpack A, Link G. Role and regulation of plastid sigma factors and their functional interactors during chloroplast transcription—recent lessons from *Arabidopsis thaliana*. Eur J Cell Biol 2010;89:940–6.
[14] Ishizaki Y, Tsunoyama Y, Hatano K, Ando K, Kato K, Shinmyo A, et al. A nuclear-encoded sigma factor *Arabidopsis* SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. Plant J 2005;42:133–44.
[15] Kroll D, Meierhoff K, Bechtold N, Kinoshita M, Westphal S, Vothknecht UC, et al. VIPP1, a nuclear gene of *Arabidopsis thaliana* essential for thylakoid membrane formation. Proc Natl Acad Sci U S A 2001;98(7):4238–42.
[16] Zhang L, Kato Y, Otters S, Vothknecht UC, Sakamoto W. Essential role of VIPP1 in chloroplast envelope maintenance in *Arabidopsis*. Plant Cell 2012;24(9):3695–707.
[17] Wang Q, Sullivan RW, Kight A, Henry RL, Huang J, Jones AM, et al. Deletion of the chloroplast-localized thylakoid formation1 gene product in *Arabidopsis* leads to deficient thylakoid formation and variegated leaves1. Plant Physiol 2004;136(3):3594–604.
[18] Larkin RM, Alonso JM, Ecker JR, Chory J. GUN4, a regulator of chlorophyll synthesis and intracellular signaling. Science 2003;299:902–6.
[19] Gutierrez-Nava M, Gillmor CS, Jiménez LF, Guevara-García A, León P. CHLOROPLAST BIOGENESIS genes act cell and noncell autonomously in early chloroplast development. Plant Physiol 2004;135(1):471–82.
[20] Tian L, DellaPenna D, Dixon R. The pds2 mutation is a lesion in the *Arabidopsis* homogentisate solanesyltransferase gene involved in plastoquinone biosynthesis. Planta 2007;226(4):1067–73.
[21] Avendaño-Vázquez AO, Cordoba E, Llamas E, San Román C, Nisar N, de la Torre S, et al. An uncharacterized apocarotenoid-derived signal generated in ζ-carotene

desaturase mutants regulates leaf development and the expression of chloroplast and nuclear genes in *Arabidopsis*. Plant Cell 2014;26:2524–37.
[22] Waters MT, Langdale JA. The making of a chloroplast. EMBO J 2009;28(19):2861–73.
[23] Pogson BJ, Albrecht V. Genetic dissection of chloroplast biogenesis and development: an overview. Plant Physiol 2011; 155(4):1545–51.
[24] Rho J, Mead JR, Wright WS, Brenner DE, Stave JW, Gildersleeve JC, et al. Discovery of sialyl Lewis a and Lewis X modified protein cancer biomarkers using high density antibody arrays. J Proteomics 2013;96:291–9.
[25] Kim PY, Tan O, Diakiw SM, Carter D, Sekerye EO, Wasinger VC, et al. Identification of plasma complement C3 as a potential biomarker for neuroblastoma using a quantitative proteomic approach. J Proteomics 2013;96:1–12.
[26] Bianchi L, Gagliardi A, Campanella G, Landi C, Capaldo A, Carleo A, et al. A methodological and functional proteomic approach of human follicular fluid en route for oocyte quality evaluation. J Proteomics 2013;90:61–76.
[27] Wang G, Guo Y, Zhou T, Shi X, Yu J, Yang Y, et al. In-depth proteomic analysis of the human sperm reveals complex protein compositions. J Proteomics 2013;79:114–22.
[28] Abreu IA, Farinha AP, Negrão S, Gonçalves N, Fonseca C, Rodrigues M, et al. Coping with abiotic stress: proteome changes for crop improvement. J Proteomics 2013;93:145–68.
[29] Sobhanian H, Aghaei K, Komatsu S. Changes in the plant proteome resulting from salt stress: toward the creation of salt-tolerant crops? J Proteomics 2011;74:1323–37.
[30] Kleffmann T, Russenberger D, von Zychlinski A, Christopher W, Sjolander K, Gruissem W, et al. The *Arabidopsis thaliana* chloroplast proteome reveals pathway abundance and novel protein functions. Curr Biol 2004;14:354–62.
[31] Friso G, Giacomelli L, Ytterberg A, Peltier J, Rudella A, Sun Q, et al. In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. Plant Cell 2004; 16:478–99 [Online].
[32] Sun Q, Emanuelsson O, van Wijk K. Analysis of curated and predicted plastid subproteomes of *Arabidopsis*. Subcellular compartmentalization leads to distinctive proteome properties. Plant Physiol 2004;135(2):723–34.
[33] Kleffmann T, von Zychlinski A, Russenberger D, Hirsch-Hoffmann M, Gehrig P, Gruissem W, et al. Proteome dynamics during plastid differentiation in rice. Plant Physiol 2007;143:912–23.
[34] Majeran W, Cai Y, Sun Q, van Wijk K. Functional differentiation of bundle sheath and mesophyll maize chloroplasts determined by comparative proteomics. Plant Cell 2005;17:3111–40.
[35] Majeran W, Zybailov B, Ytterberg AJ, Dunsmore J, Sun Q, van Wijk K. Consequences of C4 differentiation for chloroplast membrane proteomes in maize mesophyll and bundle sheath cells. Mol Cell Proteomics 2008;7:1609–38.
[36] Kamal AH, Cho K, Choi JS, Bae KH, Komatsu S, Uozumi N, et al. The wheat chloroplastic proteome. J Proteomics 2013;93: 326–42.
[37] Bischof S, Baerenfaller K, Wildhaber T, Troesch R, Vidi PA, Roschitzki B, et al. Plastid proteome assembly without Toc159: photosynthetic protein import and accumulation of N-acetylated plastid precursor proteins. Plant Cell 2011;11: 3928–39113.
[38] Kim J, Rudella A, Ramirez-Rodriguez V, Zybailov B, Olinares PDB, van Wijk K. Subunits of the plastid ClpPR protease complex have differential contributions to embryogenesis, plastid biogenesis, and plant development in *Arabidopsis*. Plant Cell 2009;21(6):1669–92.
[39] Rutschow H, Ytterberg AJ, Friso G, Nilsson R, van Wijk KJ. Quantitative proteomics of a chloroplast SRP54 sorting mutant and its genetic interactions with CLPC1 in *Arabidopsis*. Plant Physiol 2008;148(1):156–75.
[40] Motohashi R, Rodiger A, Agne B, Baerenfaller K, Baginsky S. Common and specific protein accumulation patterns in different albino/pale-green mutants reveals regulon organization at the proteome level. Plant Physiol 2012;160(4): 2189–201.
[41] Hurkman WJ, Tanaka CK. Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis. Plant Physiol 1986;81(3):802–6.
[42] Encarnación S, Hernández M, Martínez-Batallar G, Contreras S, Del Carmen Vargas M, Mora J. Comparative proteomics using 2-D gel electrophoresis and mass spectrometry as tools to dissect stimulons and regulons in bacteria with sequenced or partially sequenced genomes. Biol Proced 2005;7(1):117–35 [online].
[43] Huang DW, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat Protoc 2009;4(1):44–57.
[44] de Luna-Valdez LA, Martínez-Batallar AG, Hernández-Ortiz M, Encarnación-Guevara S, Ramos-Vega M, López-Bucio JS, et al. Data for proteomic analysis of chloroplast biogenesis(clb) mutants uncovers novel proteins potentially involved in the development of Arabidopsis thaliana chloroplasts 2014. http://dx.doi.org/10.1016/j.dib.2014.07.001 [In Press, Accepted Manuscript, Available online 12 August 2014].
[45] Li Q, Huang J, Liu S, Li J, Yang X, Liu Y, et al. Proteomic analysis of young leaves at three developmental stages in an albino tea cultivar. Proteome Sci 2011;9(1):44.
[46] Villarejo A, Burén S, Larsson S, Déjardin A, Monné M, Rudhe C, et al. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat Cell Biol 2005;7(12):1224–31.
[47] Izumi M, Tsunoda H, Suzuki Y, Makino A, Ishida H. RBCS1A and RBCS3B, two major members within the Arabidopsis RBCS multigene family, function to yield sufficient RuBisCO content for leaf photosynthetic capacity. J Exp Bot 2012;63(5):2159–70.
[48] Suzuki K, Nakanishi H, Bower J, Yoder D, Osteryoung K, Miyagishima S. Plastid chaperonin proteins Cpn60α and Cpn60β are required for plastid division in *Arabidopsis thaliana*. BMC Plant Biol 2009;9(38). http://dx.doi.org/10.1186/1471-2229-9-38.
[49] Albrecht V, Ingenfeld A, Apel K. Characterization of the snowy cotyledon 1 mutant of *Arabidopsis thaliana*: the impact of chloroplast elongation factor G on chloroplast development and plant vitality. Plant Mol Biol 2006;60(4):507–18.
[50] Somerville C, Portis A, Ogren W. A mutant of *Arabidopsis thaliana* which lacks activation of RuBP carboxylase in vivo. Plant Physiol 1982;70(2):381–7.
[51] Horlitz M, Klaff P. Gene-specific trans-regulatory functions of magnesium for chloroplast mRNA stability in higher plants. J Biol Chem 2000;275(45):35638–45.
[52] Flores-Pérez Ú, Jarvis P. Molecular chaperone involvement in chloroplast protein import. Biochim Biophys Acta 2013; 1833(2):332–40.
[53] Fink A. Chaperone-mediated protein folding. Physiol Rev 1999;70(2):425–42.
[54] Liberek K, Marszalek J, Ang D, Georgopoulos C, Zylicz M. *Escherichia coli* DnaJ and GrpE heat shock proteins jointly stimulate ATPase activity of DnaK. Proc Natl Acad Sci 1991; 88(7):2874–8.
[55] Harrison CJ, Hayer-Hartl M, Di Liberto M, Hartl FU, Kuriyan J. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science 1997;276:431–5.
[56] Becker T, Hritz J, Vogel M, Caliebe A, Bukau B, Soll J, et al. Toc12, a novel subunit of the intermembrane space preprotein translocon of chloroplasts. Mol Biol Cell 2004;15: 5130–44.

[57] Sohrt K, Soll J. Toc64, a new component of the protein translocon of chloroplasts. J Cell Biol 2000;148:1213–21.
[58] Ivey III RA, Bruce BD. In vivo and in vitro interaction of DnaK and a chloroplast transit peptide. Cell Stress Chaperones 2000;5:62–71.
[59] Ivey III RA, Subramanian C, Bruce BD. Identification of a Hsp70 recognition domain within the RuBisCO small subunit transit peptide. Plant Physiol 2000;122:1289–99.
[60] Rial DV, Arakaki AK, Ceccarelli EA. Interaction of the targeting sequence of chloroplast precursors with Hsp70 molecular chaperones. Eur J Biochem 2000;267:6239–48.
[61] Zhang XP, Glaser E. Interaction of plant mitochondrial and chloroplast signal peptides with the Hsp70 molecular chaperone. Trends Plant Sci 2002;7:14–21.
[62] Shi LX, Theg SM. A stromal heat shock protein 70 system functions in protein import into chloroplasts in the moss *Physcomitrella patens*. Plant Cell 2010;22(1):205–20.
[63] Ahn YO, Shimizu B, Sakata K, Gantulga D, Zhou Z, Bevan DR, et al. Scopolin-hydrolyzing-glucosidases in roots of *Arabidopsis*. Plant Cell Physiol 2010;51(1):132–43.
[64] Nagano AJ, Matsushima R, Hara-Nishimura I. Activation of an ER-body-localized β-glucosidase via a cytosolic binding partner in damaged tissues of *Arabidopsis thaliana*. Plant Cell Physiol 2005;46(7):1140–8.
[65] Li-Beisson Y, Shorrosh B, Beisson F, Andersson MX, Arondel V, Bates PD, et al. Acyl-lipid metabolism. Arabidopsis Book 2013;11:e0161.
[66] Wang Z, Benning C. Chloroplast lipid synthesis and lipid trafficking through ER–plastid membrane contact sites. Biochem Soc Trans 2012;40(2):457–63.
[67] Andersson MX, Goksor M, Sandelius AS. Optical manipulation reveals strong attracting forces at membrane contact sites between endoplasmic reticulum and chloroplasts. J Biol Chem 2006;282(2):1170–4.
[68] Roughan PG, Holland R, Slack CR. The role of chloroplasts and microsomal fractions in polar-lipid synthesis from [1-14C] acetate by cell-free preparations from spinach (*Spinacia oleracea*) leaves. Biochem J 1980;188(1):17–24.
[69] Benning C. Mechanisms of lipid transport involved in organelle biogenesis in plant cells. Annu Rev Cell Dev Biol 2009;25(1):71–91.
[70] Liu X, Rodermel SR, Yu F. A var2 leaf variegation suppressor locus, SUPPRESSOR OF VARIEGATION3, encodes a putative chloroplast translation elongation factor that is important for chloroplast development in the cold. BMC Plant Biol 2010; 10(1):287.

Data in Brief 1 (2014) 15–18

Contents lists available at ScienceDirect

Data in Brief

journal homepage: www.elsevier.com/locate/dib

Data Article

Data for a comparative proteomic analysis of chloroplast biogenesis (*clb*) mutants

L.A. de Luna-Valdez [a], A.G. Martínez-Batallar [b], M. Hernández-Ortiz [b], S. Encarnación-Guevara [b], M. Ramos-Vega [a], J.S. López-Bucio [a], P. León [a], A.A. Guevara-García [a,*]

[a] *Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, 62214 Cuernavaca, Morelos, México*
[b] *Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Av. Universidad 565, Chamilpa, 62210 Cuernavaca, Morelos, Mexico*

ARTICLE INFO

Article history:
Received 17 July 2014
Received in revised form
28 July 2014
Accepted 28 July 2014
Available online 12 August 2014

Keywords:
Chloroplast development
Comparative proteomics
clb mutants
Arabidopsis thaliana

ABSTRACT

This data article contains data related to the research article titled **Proteomic analysis of chloroplast biogenesis (clb) mutants uncovers novel proteins potentially involved in the development of *Arabidopsis thaliana* chloroplasts** (de Luna-Valdez et al., 2014) [1]. This research article describes the 2-D PAGE-based proteomic analysis of wild-type and four mutant lines (*cla1-1*, *clb2*, *clb5* and *clb19*) affected in the development of *Arabidopsis thaliana* chloroplasts. The report concludes with the discovery of three proteins potentially involved in chloroplast biogenesis. The information presented here represent the tables and figures that detail the processing of the raw data obtained from the image analysis of the 2-D PAGE gels.

* Corresponding author.
E-mail addresses: ldeluna@ibt.unam.mx (L.A. de Luna-Valdez), angelmb@ccg.unam.mx (A.G. Martínez-Batallar), magda@ccg.unam.mx (M. Hernández-Ortiz), encarnac@ccg.unam.mx (S. Encarnación-Guevara), mramos@ibt.unam.mx (M. Ramos-Vega), lopbucio@ibt.unam.mx (J.S. López-Bucio), patricia@ibt.unam.mx (P. León), aguevara@ibt.unam.mx (A.A. Guevara-García).

http://dx.doi.org/10.1016/j.dib.2014.07.001

Specifications table

Subject area	Biology
More specific subject area	*Plant proteomics*
Type of data	*Tables and fiures*
How data was acquired	***Electron microscopy***: *Images were extracted from* [2,3,4] ***2-D PAGE and image analysis***: *GS-800 densitometer* (*Bio-Rad Hercules, CA, EUA*); *image analysis software PD-Quest 8.0.1* (*Bio-Rad Hercules, CA, EUA*) ***Mass Spectrometry***: Matrix-Assisted Laser Desorption/Ionization-Time of Flight; Autoflex, Bruker Daltonics, Billerica, MA, USA
Data format	*Processed.*
Experimental factors	*No pretreatment of samples was performed.*
Experimental features	*Total protein was extracted from mutant and wild-type plants by triplicate. 2-D PAGE gel images were generated and compared in order to discover reliable (T-test $P < 0.01$) spots with abundance shift of at least ± 2-fold.*
Data source location	*NA.*

Value of the data

- The data further validate the information presented in de Luna-Valdez et al. (2014) [1].
- The data present alternative ways of visualizing the abundance of the proteins under study.
- The data provide specifics on the biochemical processes affected in all the analyzed clb mutants.

Direct link to deposited data in public repository
The data is directly available in this article and related to de Luna-Valdez et al. (2014) [1].

1. Experimental design

Total protein was extracted from 16-days old mutant and 8-days old wild-type plants by triplicate. 2-D PAGE gel images were generated and compared in order to discover reliable (*T*-test $P < 0.01$) spots with abundance shift of at least ± 2-fold. Protein identification was performed using MALDI-TOF Mass spectrometry.

2. Material and methods [1]

2.1. Plant material and growth conditions

Arabidopsis thaliana heterozygous mutant lines corresponding to *cla1-1* (At4g15560) [2], *clb2* (At3g11945) [3], *clb19* (At1g05750) [4], *clb5* (At3g04870) [3,5], *emb1241* (SALK_045238), *pbp1* (SAIL_773_D06), and *atrabe1b* (SALK_069644) were used in this study (Fig. S1, S2). Seeds were surface-sterilized using solutions of 100% C_2H_6O and 1% NaClO, then cultured on 0.5X Murashige & Skoog media supplemented with 0.05 g/l 2-(N-morpholino)ethanosulfonic acid, 0.5 g/l sucrose, 100 mg/l myo-inositol, 1 mg/l nicotinic acid, 1 mg/l pyridoxine-HCl, 10 mg/l thiamine-HCl, and 8 g/l phyto agar. Seedlings from the four mutant lines that presented the wild-type phenotype and the first pair of true leaves were harvested after 8 days of culture. These were then pooled for processing as the wild-type protein samples used in this study. In order to minimize the effect of using plants in different developmental stages (detection of development-related proteins), pigment-deficient plants were collected after 16 days of culture; at this time, all the seedlings display at least the first pair of true leaves. Three biologically independent seedling batches were generated for further processing.

2.2. Extraction and quantification of total protein

Total protein extracts were prepared according to the phenol extraction protocol reported by Hurkman and Tanaka [6]; the adjustments made to the original protocol are described here. Briefly, the starting plant material was 1 g of mutant or wild-type seedlings grown in vitro on GM medium, seedlings were ground in a mortar using liquid nitrogen and re-suspended in extraction buffer (0.7 M sucrose; 0.5 M Tris; 30 mM HCL; 50 mM EDTA; 0.1 MKCl, 12 mg/ml PVPP (Polyvinylpolypyrrolidone) and 2% ⋄-mercaptoethanol). An equal volume of water-saturated phenol was added followed by centrifugation (6000*g* for 10 min) to separate the phases. The phenol phase was re-extracted with one volume of extraction buffer then precipitated with 5 volumes of 0.1 M ammonium acetate in methanol at −20 °C overnight. The protein precipitate was washed three times with 0.1 M ammonium acetate in methanol and once with 80% acetone at −20 °C. The pellets were air dried under vacuum and re-suspended in lysis buffer (8 M Urea; 2 M Thiourea, 4% (w/v) CHAPS; 2% ampholines (1.5% pH range 5–7 and 0.5% pH range 3–10) and 60 mM DTT). Determination of protein concentration in the extracts was determined by colorimetric assays as reported by Encarnación et al. (2005) [7].

2.3. 2-D PAGE and protein visualization

500 μg (analytical gels) or 750 μg (preparative gels) of total protein extracts were separated in 12% acrylamide gels under denaturating conditions. The first dimension was run using ampholytes in the range of pH 3–10 and enriched in pH 4–8. After 2-D electrophoresis, gels were fixed and stained using colloidal Coomassie brilliant blue, following [7]. The stained gels were digitalized using a GS-800 densitometer (Bio-Rad Hercules, CA, EUA) and the image analysis software PD-Quest 8.0.1 (Bio-Rad Hercules, CA, EUA) (Fig. S1).

2.4. In silico processing of gel images

Images from 2-D gels of three biologically independent protein extracts from mutants (*cla1-1*, *clb2*, *clb5*, and *clb19*) and wild-type plants were generated and processed using the PD-Quest 8.0.1 software (Bio-Rad, Hercules CA) (Fig. S1). Protein spots in all replicates were detected automatically by the software, and the detection was then improved by the manual addition of missing spots and the removal of erroneously detected spots. Normalization of gel images was performed using the local regression model normalization method provided by PD-Quest software. Furthermore, in order to properly compare the samples, the gel images were adjusted to fit a common distortion model; this was done by matching spots that were common to all the gel images. The gel images from the different protein samples were compared to each other in order to generate a robust data set containing all the spots represented in the samples with 98% statistical confidence ($P < 0.01$) in a Student's *t*-test. Finally, the protein spots in the statistical data set displaying ±2-fold abundance change were selected as candidates for the MS analysis (Table S1).

2.5. MALDI-TOF mass spectrometry and protein identification

The selected protein spots were manually excised from the preparative gels. The samples were alkylated, reduced, and trypsin-digested prior to their elution and MS analysis. Samples of digested protein spots were automatically transferred to MALDI-TOF (Matrix-Assisted Laser Desorption/Ionization-Time of Flight; Autoflex, Bruker Daltonics, Billerica, MA, USA) using Proteineer SP and SPII systems (software SPcontrol 3.1.48.0v; Bruker Daltonics, Breme, Germany). The Bruker Daltonics Autoflex system was operated in the delayed extraction and reflectron mode, and the resolution threshold was set to a signal-to-noise ratio of 1500. The specific protocols can be accessed in [7]. The *m*/*z* spectra were searched against the Arabidopsis thaliana NCBInr (http://www.ncbi.nlm.nih.gov/guide/proteins/), SwissProt (http://www.isb-sib.ch/), and IPI (http://www.ebi.ac.uk/IPI/IPIarabidopsis.html) databases, using the Mascot (http://www.matrixscience.com) and Profound (http://prowl.rockefeller.edu) search engines. The Mascot engine was used to query NCBInr and SwissProt

databases, while Profound was used to query NCBInr and IPI databases; both search engines were operated using a mass tolerance of 200 ppm, with cysteine carbamidomethylation as the constant modification and methionine oxidation as the variable modification. The significance threshold was set to $P < 0.05$ for the Mascot search.

2.6. In silico analysis of the identified proteins

Functional clustering of the identified proteins was performed using the Functional annotation tool available at DAVID (http://david.abcc.ncifcrf.gov/home.jsp) [8], and clustering was carried out using the annotations available at the Protein Information Resource (http://pir.georgetown.edu/pirwww/index.shtml) and Gene Ontology (http://www.geneontology.org/) databases. Stringency of the classification was set on medium and the rest of the options were set as default (Fig. S3). Reconstruction of metabolic pathways was achieved using the metabolism overview pathways in the MapMan 3.5.1 software (http://mapman.gabipd.org/web/guest/mapman) with the Ath_AGI_TAIR9_-Jan2010 mappings. MapMan was fed an array of data containing, for each protein, the $\log_2$ of the ratio of the detected abundance in each mutant over that registered in wild-type plants (Table S2).

Acknowledgments

The authors thank Patricia Jarillo for the technical support, Alma Lidia Martínez, Juan Manuel Hurtado, Roberto Rodríguez Bahena and Arturo Ocádiz for computer support, Paul Gaytan and Eugenio López for oligonucleotide synthesis and Guadalupe Zavala Padilla (UME, Instituto de Biotecnolgía, UNAM) and Fernando García Hernández (UM, Instituto de Fisiología Celular, UNAM) for chloroplast electron microscopy analysis. This work was supported by UNAM-DGAPA-PAPIIT (Grants IN217111 to AGG and IN208211 to PL) and CONACYT-México (CB-129266 to AGG and 127546 to PL).

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at http://dx.doi.org/10.1016/j.dib.2014.07.001.

References

[1] L.A. de Luna-Valdez, A.G. Martínez-Batallar, M. Hernández-Ortiz, S. Encarnación-Guevara, M. Ramos-Vega, J.S. López-Bucio, P. León, A.A. Guevara-García, Proteomic analysis of chloroplast biogenesis (clb) mutants uncovers novel proteins potentially involved in the development of Arabidopsis thaliana chloroplasts, J. Proteomics (2014), http://dx.doi.org/10.1016/j.jprot.2014.07.003.

[2] M.A. Mandel, K.A. Feldmann, L. Herrera-Estrella, M. Rocha-Sosa, P. Leon, CLA1, a novel gene required for chloroplast development, is highly conserved in evolution, Plant J. 9 (1996) 649–658.

[3] M. Gutierrez-Nava, C.S. Gillmor, L.F. Jiménez, A. Guevara-García, P. León, Chloroplast biogenesis genes act cell and noncell autonomously in early chloroplast development, Plant Physiol. 135 (1) (2004) 471–482.

[4] A.L. Chateigner-Boutin, M. Ramos-Vega, A. Guevara-García, C. Andrés, M. De La Luz Gutiérrez-Nava, A. Cantero, E. Delannoy, L.F. Jiménez, C. Lurin, I. Small, P. León, CLB19, a pentatricopeptide repeat protein required for editing of rpoA and clpP chloroplast transcripts, Plant J. 56 (4) (2008) 590–602.

[5] A.O. Avendaño-Vázquez, E. Cordoba, E. Llamas, C. San Román, N. Nisar, S. De la Torre, M. Ramos-Vega, M.D. Gutiérrez-Nava, C.I. Cazzonelli, B.J. Pogson, P. León, An uncharacterized apocarotenoid-derived signal generated in ζ-carotene desaturase mutants regulates leaf development and the expression of chloroplast and nuclear genes in arabidopsis, Plant Cell 26 (2014) 291–299.

[6] W.J. Hurkman, C.K. Tanaka, Solubilization of plant membrane proteins for analysis by two-dimensional gel electrophoresis, Plant Physiol. 81 (3) (1986) 802–806.

[7] S. Encarnación, M. Hernández, G. Martínez-Batallar, S. Contreras, M. Del Carmen Vargas, J. Mora, Comparative proteomics using 2-D gel electrophoresis and mass spectrometry as tools to dissect stimulons and regulons in bacteria with sequenced or partially sequenced genomes, Biol. Proced. Online 7 (1) (2005) 117–135.

[8] D.W. Huang, B.T. Sherman, R.A. Lempicki, Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources, Nat. Protoc. 4 (1) (2009) 44–57.

Chloroplast Omics

Chapter · February 2015

DOI: 10.1007/978-81-322-2172-2_18

4 authors:

Luis Alberto De Luna-Valdez
Universidad Nacional Autónoma de México
11 PUBLICATIONS **49** CITATIONS
SEE PROFILE

Patricia León
Universidad Nacional Autónoma de México
55 PUBLICATIONS **4,786** CITATIONS
SEE PROFILE

Sergio Encarnacion
Universidad Nacional Autónoma de México
163 PUBLICATIONS **1,512** CITATIONS
SEE PROFILE

Angel Guevara-Garcia
Universidad Nacional Autónoma de México
52 PUBLICATIONS **1,327** CITATIONS
SEE PROFILE

Chloroplast Omics

L.A. de Luna-Valdez, P. León-Mejía,
S. Encarnación-Guevara, and A.A. Guevara-García

Contents

Abstract

The chloroplast is the most remarkable organelle of plant cells; it is the site of a myriad of different chemical reactions; among chloroplast's many functionalities is photosynthesis, perhaps the most fundamental biological process on the biosphere. The chloroplast has been subject of a plethora of research efforts that try to understand the molecular mechanisms that regulate its biochemical capabilities, development, and evolutionary origin. Omic technologies have provided researchers with tools to study different aspects of biology from a global perspective, and, not surprisingly, chloroplast research has taken advantage of them. This chapter explores how chloroplasts organize their genomes and regulate their transcriptomes, proteomes, and metabolomes, trying to focus on classical knowledge and reviewing new datasets obtained through large-scale research projects that shed light on chloroplast functionality.

Keywords

Chloroplast • Chlorogenomes • Chlorotranscriptomes • Chloroproteomes • Chlorometabolomes • Chloroplast's system biology

L.A. de Luna-Valdez, Ph.D. • P. León-Mejía, Ph.D.
A.A. Guevara-García, Ph.D. (✉)
Instituto de Biotecnología, Universidad Nacional Autónoma de México, Apartado Postal 510-3, 62250 Cuernavaca, Morelos, Mexico
e-mail: aguevara@ibt.unam.mx

S. Encarnación-Guevara, Ph.D.
Centro de Ciencias Genómicas, Universidad Nacional Autónoma de México, Cuernavaca 62271, Morelos, Mexico

D. Barh et al. (eds.), *PlantOmics: The Omics of Plant Science*,
DOI 10.1007/978-81-322-2172-2_18, © Springer India 2015

Introduction

Chloroplasts are semiautonomous organelles that came in the eukaryotic life scene around 1.6 billion years ago, through a process of primary endosymbiosis between cyanobacteria and ancient heterotrophic mitochondria-bearing eukaryotes (Yoon et al. 2004). From that moment on, chloroplasts have integrated themselves in almost every aspect of the biology of plants; chloroplasts perform functions that impact a wide range of processes like ecological traits (volatile emissions), specific molecular events (chloroplast-to-nucleus signaling), and biomass production.

As several other well-studied endosymbionts, at the morphological level, chloroplasts are surrounded by two lipid bilayers that enclose a fluid compartment called stroma (Keeling 2010). The stroma contains a most distinctive net of internal membranes known as thylakoids, whose lumen is also fluid. The high degree of compartmentalization existing in chloroplasts is essential for the proper production of the plethora of metabolites and chemical processes that find source in the membranes and compartments of the chloroplasts. For instance take the carbon fixation process, commonly known as photosynthesis; this process is a series of energy-driven chemical reactions that mediate the assimilation of inorganic carbon atoms into sugar molecules. Photosynthesis is composed of two conceptually independent sets of reactions, the light-dependent and light-independent reactions; light-dependent reactions take place in the thylakoid membrane and involve the conversion of light energy to chemical energy through a series of electron donors and acceptors that are associated to the thylakoid membrane. Starting from antenna and reaction center pigments (such as chlorophylls and carotenes), the electrons travel through proteins like plastoquinones and cytochromes to finally reduce NADPH (nicotinamide adenine dinucleotide phosphate). Along the way from the antenna to the NADPH, electron carriers mediate the pumping of H^+ from the chloroplast stroma to the lumen of the thylakoids, this way forming an electrochemical gradient that is used by an ATP (adenosine triphosphate) synthase complex oriented towards the plastid stroma to generate ATP in that compartment. During light-independent reactions, enzymes in the stroma use the ATP and NADPH produced during the light-dependent reactions to generate sugars, from inorganic CO_2 molecules (Blankenship 2002); it is important to note that compartmentalized changes in pH are key to regulate enzymes important for the light-independent reactions to take place, such as ribulose-bisphosphate carboxylase-oxygenase whose activation is facilitated by the generation of a pH gradient through the thylakoid membrane (Campbell and Ogren 1990; Chen et al. 2010). In this simple example, it results evident how the existence of several compartments in the chloroplasts is essential to the performance of one given function.

Since photosynthesis is the source of all the organic carbon found in every single metabolite, chloroplast function results essential to almost every process in the cell, and synthesis pathways of major metabolites and the photosynthetic activity of the plant tissues are fine tuned (Paul and Pellny 2003). Even though dependent on carbon fixed during photosynthesis, other important chemical reactions take place inside chloroplasts such as lipids, isoprenoids, amino acids, proteins, and complex carbohydrate synthesis (Kannangara et al. 1971; Lichtenthaler 1999; Kirk and Leech 1972; Siddell and Ellis 1975), making the chloroplast a source of energy, structural components, and a plethora of signals that regulate several aspects of plant development.

Given the importance of chloroplasts for life, several efforts have been made worldwide in order to grasp understanding of the intricate regulatory networks underlying its origin, function and development. Here we present a review of the current knowledge on how chloroplasts organize their genomes and regulate their transcriptomes, proteomes, metabolomes, and even some new and not so well-studied fields, such as lipidomics.

Chloroplast Genomes

In 1920 Hans Winkler adapted the term genome to refer to the haploid chromosome set of a given organism. More than 50 years later, in 1976, Walter Fiers reported the complete sequence of the bacteriophage MS2 RNA genome (Fiers et al. 1976); in 1977 Frederick Sanger reported the first sequence of a DNA (deoxyribonucleic acid) genome (Phage Φ X-174; Sanger et al. 1977). In the following years, the genomic sequences of representative organisms from the three different domains of life were reported, being *Haemophilus influenzae*, *Saccharomyces cerevisiae*, and *Methanococcus jannaschii* the firsts on their respective domains (Fleischmann et al. 1995; Goffeau et al. 1996; Bult et al. 1996). Furthermore, in 1986 the first two chloroplast genomes came into view when two different Japanese research teams reported full sequences for the chloroplast genomes of *Marchantia polymorpha* and *Nicotiana tabacum* (Ohyama et al. 1986; Shinozaki et al. 1986). Nowadays, the next generation of DNA sequencing technologies has advanced in precision, time consumption and cost, to the point that genomic information for 10,904 organisms is now publicly available; among them only 197 genomes correspond to plants (http://www.ncbi.nlm.nih.gov/genome/browse/).

It is now common knowledge that eukaryotic organelles like mitochondria and plastids originated by means of a process called endosymbiosis, in which prokaryotic cells were engulfed by pro-eukaryotic cells and then underwent a process of functional specialization. Organelles generated this way have a very unique trait compared to other cell organelles: they have their own DNA genome, which is a reminiscent of their past prokaryotic genomes. The chloroplast genome (cpDNA) varies little in topology and content among plants, it is generally accepted that cpDNA molecules are made of quadripartite circular molecules of around 145 kbps in size, this molecules are organized in two single-copy segments of different length (long single copy, LSC; short single copy, SSC) which are separated by

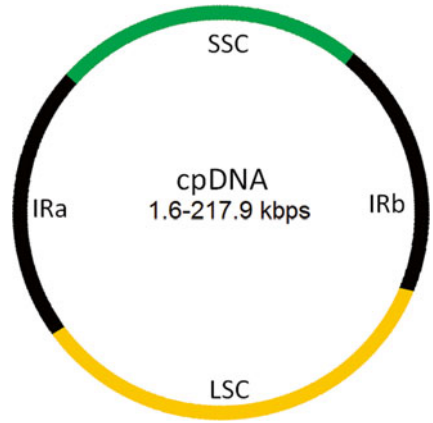


Fig. 1 Map of a standard cpDNA molecule. Graphical representation of a standard circular cpDNA molecule. Annotated LSC (large single copy), SSC (short single copy), and IR (inverted repeat) elements

two inverted repeat segments (IR) (Fig. 1) (Palmer 1991). Even if it is widely accepted that cpDNA molecules are circular molecules, over the years evidence has been accumulated showing that the circle does not define the most common conformation of cpDNA molecules. As a matter of fact, early evidence showed that circular cpDNA molecules represent only the 37 % of the total cpDNA molecules extracted from pea chloroplasts (Kolodner and Tewari 1972). For decades the view of circular cpDNA molecules dominated, and the theory suggested that most of the cpDNA molecules suffer random breakage during the extraction process and then concatenated, this way explaining the oligomeric linear forms of cpDNA molecules found in all the experiments reported. In contrast, in 2004, Oldenburg and Bendich reported high-resolution digital images of ethidium-stained plastid DNA extracted from maize meristematic tissue, showing the cpDNA as complex branched linear molecules whose size was several times larger than the genome size, it was also demonstrated that the ends of such molecules are specific rather than random, as expected from molecules randomly broken during the extraction process (Oldenburg and Bendich 2004). Furthermore, it is now proposed that cpDNA replication does not proceed by the canonical cairn structure and rolling circle mechanism; instead, a recombination-

dependent mechanism is thought to mediate replication, since this mechanism better explains the formation of the reported head-to-tail-branched concatemers (Bendich 2004). Evidence of these linear complexes of cpDNA has been reported for several species such as *A. thaliana*, maize, pea, and watermelon (Bendich 1991, 2004; Rowan et al. 2004). Even though there is plenty of evidence to show the linear nature of cpDNA molecules, to this date, chloroplast genomes are still reported as circular molecules, whether the idea of a circular cpDNA molecule is very well rooted in the scientific community or it is used as a way to homogenize the way chloroplast genomes are reported, future literature must adapt to accommodate the findings about linear cpDNA molecules.

Despite the controversial information regarding the shape of cpDNA molecules, it is very well known that the standard cpDNA molecule has around 128 different genes organized in several clusters; these cpDNA-encoded genes include protein (84), tRNA (37), and rRNA (8) coding sequences (Saski et al. 2005; Wu et al. 2009). Table 1 compiles publicly available information on genome size and gene content for chloroplastic genomes of several economically important plants. To this date, 555 entries exist that report sequences for cpDNA of different species (32 Chlorophyta and 523 Streptophyta) in the Viridiplantae clade (http://www.ncbi.nlm.nih.gov/genomes/ORGANELLES/organelles.html). From the Streptophyta information available at NCBI, *Conopholis americana (American cancer-root)* and *Pelargonium hortorum (Garden geranium)* represent the smallest and the largest cpDNA sequences, respectively. The cpDNA of *Conopholis americana* has only 45.63 kbps, while the cpDNA of *Pelargonium hortorum* is 217.9 kbps in length (Fig. 2). In accordance to the great difference existing in genome length, differences exist in the number of genes coded by the cpDNA of each species, the cpDNA of *Pelargonium hortorum* contains 181 genes (131 proteins, 40 tRNAs, 10 rRNAs), while *Conopholis americana* cpDNA codes only for 44 genes (21 proteins, 18 tRNAs, 4 rRNAs); functional characterization of the proteins coded by each gene reveals that *Pelargonium hortorum* codes for proteins with a wide set of functionalities, including photosynthesis, transcription, translation, and energy metabolism (Fig. 2). On the other hand, *Conopholis americana* cpDNA lacks most of the plastid-encoded photosynthetic and energy metabolism genes, presenting genes coding for the plastid translation machinery; this information correlates with the parasitic lifestyle of this plant species (Fig. 2a). The information regarding the specific genes encoded by all the cpDNA sequences available to this date is compiled in the cpBase: The Chloroplast Genome Database (http://chloroplast.ocean.washington.edu/). Chloroplast genome information has seeded the development of one particular field of plant biology, the chloroplast phylogenomics; this approach tries to uncover the phylogenetic relationships existing between different plant species by analyzing their cpDNA sequences. Chloroplast phylogenomics, aided by the available high-throughput next-generation technologies of DNA sequencing, has been applied successfully to resolve the controverted hypothesis dealing with the proper placement of the genus ginkgo in the phylogenetic tree of land plants, strongly supporting the previously proposed monophyly between ginkgo and cycad groups (Wu et al. 2013). Moreover, cpDNA sequences generated by state-of-the-art sequencing methods have been used to explore the angiosperm phylogenetic tree; it was found that the most basal lineage of angiosperms is represented by aquatic and herbaceous plants, whose surviving relatives are represented by plants of the genuses *Trithuria* and *Amborella* and the family Nymphaeaceae, instead of the previous belief that pointed the plants in the genus *Amborella* as the only surviving organisms related to the root of the angiosperm phylogenetic tree (Goremykin et al. 2013).

Developing photosynthetic tissues contain around one thousand cpDNA molecules per plastid; though the functionality of such a high copy number during development is not clear, two major hypotheses try to explain this phenomenon. First, high copy number may compensate random sorting of cpDNA molecules during plastid division; second, the increased gene dosage

Table 1 Gene content of chloroplastic genomes of several crop species

Species	Accession	Size (Kbps)	Genes		
			Protein	tRNA	rRNA
Adiantum capillus-veneris	NC_004766	150.568	87	35	8
Agrostis stolonifera	NC_008591	136.584	85	40	8
Anthriscus cerefolium	NC_015113	154.719	85	37	8
Brassica napus	NC_016734	152.86	87	37	8
Capsella bursa-pastoris	NC_009270	154.49	85	37	8
Capsicum annuum	NC_018552	156.781	86	38	8
Carica papaya	NC_010323	1.601	84	37	8
Castanea mollissima	NC_014674	160.799	83	37	8
Cedrus deodara	NC_014575	119.299	75	35	4
Cicer arietinum	NC_011163	125.319	75	29	4
Citrus sinensis	NC_008334	160.129	87	45	8
Coffea arabica	NC_008535	155.189	85	45	8
Coix lacryma-jobi	NC_013273	140.745	104	40	8
Colocasia esculenta	NC_016753	162.424	86	37	8
Cucumis melo subsp. melo	NC_015983	156.017	88	37	8
Cucumis sativus	NC_007144	155.293	85	37	8
Elaeis guineensis	NC_017602	156.973	86	38	8
Festuca arundinacea	NC_011713	136.048	80	38	8
Glycine max	NC_007942	152.218	83	37	8
Gossypium hirsutum	NC_007944	160.301	83	37	8
Helianthus annuus	NC_007977	151.104	85	43	8
Hevea brasiliensis	NC_015308	161.191	84	37	8
Hordeum vulgare subsp. vulgare	NC_008590	136.462	83	48	8
Lactuca sativa	NC_007578	152.765	84	37	7
Lathyrus sativus	NC_014063	121.02	74	30	4
Liriodendron tulipifera	NC_008326	159.886	84	37	8
Nicotiana tabacum	NC_001879	155.943	98	37	8
Olea europaea	NC_013707	155.872	85	37	8
Oryza sativa Japonica	NC_001320	134.525	108	38	8
Phaseolus vulgaris	NC_009259	150.285	83	36	8
Ricinus communis	NC_016736	163.161	86	37	8
Sesamum indicum	NC_016433	153.324	87	37	8
Solanum tuberosum	NC_008096	155.296	84	45	8
Sorghum bicolor	NC_008602	140.754	84	48	8
Spinacia oleracea	NC_002202	150.725	96	37	8
Theobroma cacao	NC_014676	160.619	81	37	8
Triticum aestivum	NC_002762	134.545	83	42	8
Vigna unguiculata	NC_018051	152.415	84	38	8
Vitis vinifera	NC_007957	160.928	84	45	8
Zea mays	NC_001666	140.384	111	38	8

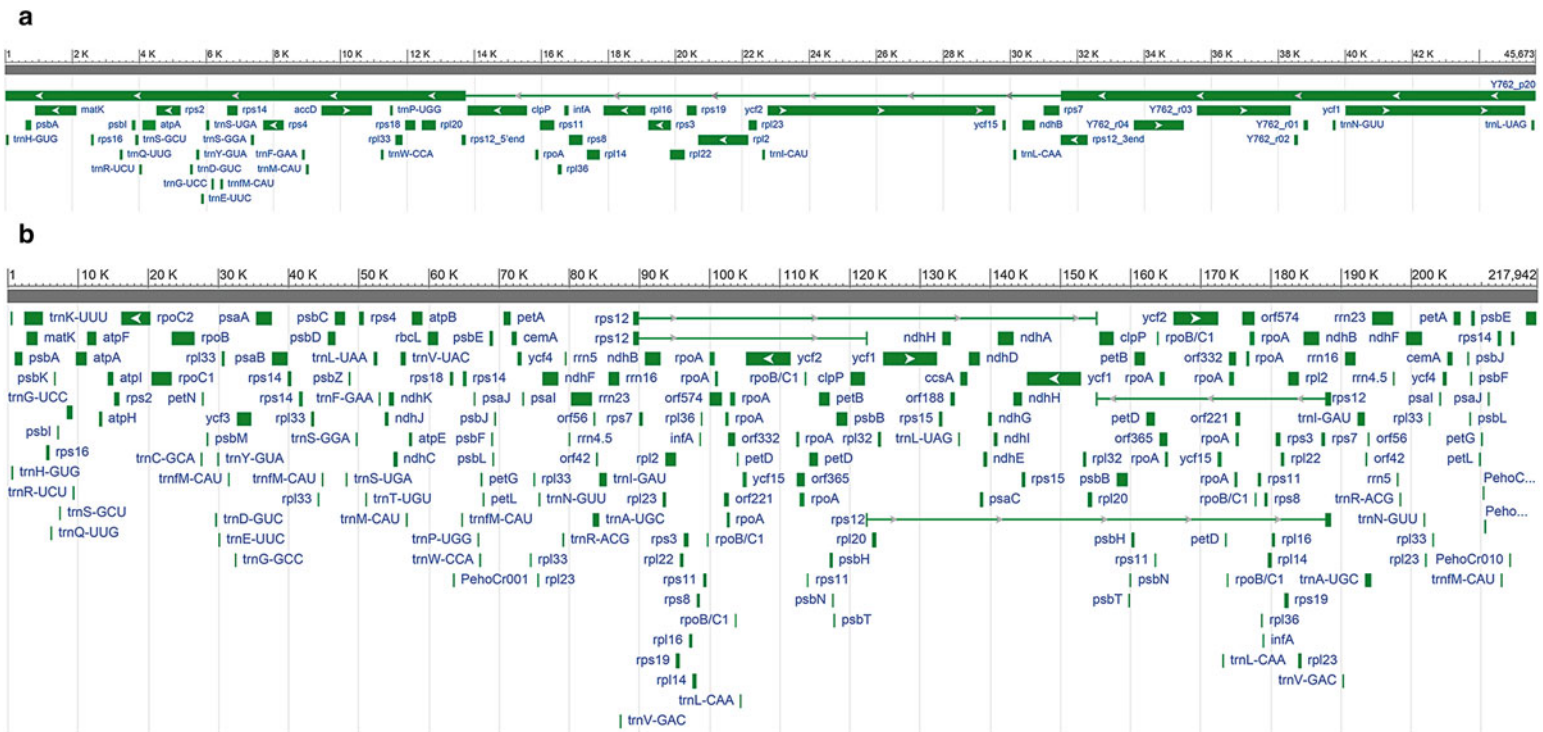


Fig. 2 cpDNA maps for two plant species. Graphical representation of the cpDNA maps for *Conopholis americana* (**a**) and *Pelargonium hortorum* (**b**). The genes coded by each genome are represented by *green bars*

may be necessary in early developmental stages to maintain the high demand of photosynthetic proteins necessary for the establishment of photoautotrophic metabolism (Bendich 1987). Recently, analysis of the maize albino mutant *w2* (defective in a chloroplast-specific DNA polymerase) has proven the functional importance of gene dosage in maize chloroplasts; it is shown that protein subunits of photosynthetic enzymes are affected to levels not matching to the decrease in their respective mRNA levels, suggesting the lowered availability of rRNAs (and hence plastid ribosomes) and tRNAs affects the rate of protein synthesis and directly relating the defective plastid function to gene dosage (Udy et al. 2012).

Besides the knowledge available related to cpDNA-encoded genes, there is also data showing that there is plenty of noncoding regions in cpDNA molecules. Around 42.9 % of the total length of the *Nicotiana* LSC and SSC regions is noncoding DNA (10.6 % introns and 32.3 % intergenic regions) (Shaw et al. 2007), and the precise knowledge of the main function of these noncoding sequences and the coding sequences they might regulate is pivotal for the development of transplastomic plant lines, a field that has been successful in overcoming specific crop-related problems.

In 1988 Boynton et al. reported a method based on the bombardment of *Chlamydomonas reinhardtii* cells with cpDNA-coated tungsten microprojectiles, which made possible the transformation of the chloroplastic genome of *C. reinhardtii*. Only a few years later in 1990 and 1993, respectively, the concept was applied and improved for tobacco cpDNA by Svab et al., successfully generating the first tobacco transplastomic line resistant to the antibiotic spectinomycin (Svab and Maliga 1993).

Transplastomic plant lines have several advantages over the classical transgenic plants:

- Since chloroplasts are maternally inherited, transgenic plastids are not disseminated into the environment by pollen, allowing the propagation of transgenic crops without risking naturally occurring varieties (Svab et al. 1990).
- High levels of protein accumulation when transgenes are stably integrated in cpDNA, caused by the elevated copy number of cpDNA molecules (De Cosa et al. 2001).
- Lack of position effect. Integration of transgenes into cpDNA is mediated by a site-specific recombination mechanism, this way preventing random undesired effects produced by stochastic transgene insertions (Svab et al. 1990).
- No transgene silencing has been reported so far.
- Multigene cloning in a single transformation event (Quesada-Vargas et al. 2005).

Recently, the information available on the functionality of the noncoding regions of cpDNA sequences of several plant species and the polycistronic nature of the chloroplastic RNAs, has led to the optimization of chloroplast genetic engineering tools. For instance, expression cassettes have been optimized to direct the insertion of entire operons in cpDNA molecules via recombination with different species-specific segments of cpDNAs or highly conserved IR segments, under the regulation of the promoters, 5′ and 3′ regulatory regions of different chloroplast-encoded genes, which allow high expression levels of the genes of interest in transplastomic plants. RNA-processing sites, translation signals, and amino acid sequences that affect protein turnover are also added to the expression cassettes to enhance translation and protein accumulation (Verma and Daniell 2007). Figure 3 shows a diagram depicting the basic elements necessary for the construction of cpDNA transformation cassettes, detailed information about the different cpDNA transformation vectors, and the regulatory elements contained within them which will be discussed in further chapters of this book.

Several years ago in 1998, the idea of a universal chloroplast transformation vector was proposed by Danielle et al., using the sequences of the IR-encoded genes *trnA* and *trnI* as flanking sequences for the recombination process and the conserved noncoding spacer region between those genes as insertion site. This system was designed using tobacco cpDNA as model, and it

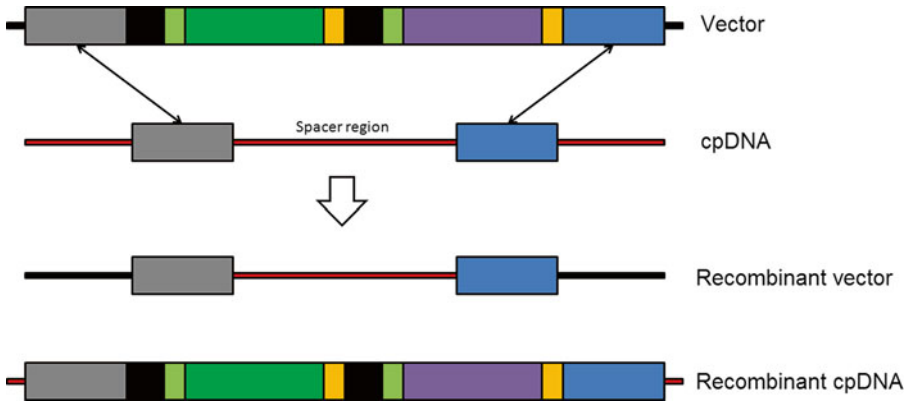


Fig. 3 cpDNA transformation cassette. Representation of the different cis-elements included in chloroplast DNA transformation cassettes. Flanking sequences (*blue* and *gray boxes*) for recombination reaction with the cpDNA; promoter sequences (*black boxes*); 5′- and 3′-UTR regions (*green* and *yellow boxes*) to ensure proper RNA processing; gene of interest (*purple box*) and spacer region (*white box*)

worked for transformation of potato and tomato chloroplasts, but at a lower transformation efficiency than the achieved for tobacco plastid transformation (Sidorov et al. 1999; Ruf et al. 2001). Later studies comparing chloroplast genomes of different groups of phylogenetically related plants, found only a few or no identical spacer regions to be used in the design of vectors to allow the cpDNA transformation of several different yet related species, making clear that the lack of enough complete cpDNA sequences still poses a great problem for cpDNA genetic engineering (Saski et al. 2007).

At the present time, several transplastomic lines are available for different species displaying economically important agronomical traits, such as insect resistance (McBride et al. 1995; Kota et al. 1999; De Cosa et al. 2001; Hou et al. 2003; Dufourmantel et al. 2005; Chakrabarti et al. 2006), herbicide resistance (Danielle et al. 1998), pathogen resistance (Jin et al. 2012), drought tolerance (Lee et al. 2003), salt tolerance (Kumar et al. 2004), vitamin production (Yabuta et al. 2012), bioplastics production (Bohmert-Tatarev et al. 2011), and carotenoid production (Apel and Bock 2009). Furthermore, several efforts have been made to improve crop yield through the engineering of RuBisCO (ribulose-1,5-bisphosphate carboxylase/oxygenase) enzyme and are extensively reviewed by Parry et al. (2012) and Hanson et al. (2012).

Along with the data provided by transcriptomics and proteomics, genomics has been useful for the development of genetically modified organisms, a technological keystone that has boosted research in virtually all fields of biology and is now impacting on economically important subjects.

Chloroplast Transcriptomes

The transcriptome is best defined as the entire set of RNA transcripts found in a cell, tissue, or organ at one specific developmental stage or physiological condition. Hence, the field arising to understand the transcriptome is called transcriptomics; the main goals of transcriptomics consist on (1) the identification of all the RNA species that exist on a given transcriptome, be it coding, noncoding, and small RNA molecules; (2) the quantitation of the changes in transcript abundance levels under different conditions or developmental stages; and (3) the determination of transcript structure, including 5′ and 3′ ends, splicing patterns, and other posttranscriptional modifications such as editing.

Chloroplast transcription is mediated by two different kinds of RNA polymerases, a single-subunit viral-like nuclear-encoded RNA polymerase (NEP) and a multi-subunit bacterial-like plastid-encoded RNA polymerase (PEP) (Allison et al. 1996). These enzymes have different target promoters, and based on the RNA polymerase in charge of its transcription, chloroplast genes can be classified into three groups: genes only transcribed by PEP are class I, genes transcribed by both PEP and NEP are called class II, and those genes transcribed only by NEP are class III (Hajdukiewicz et al. 1997). Even though the PEP core subunits are plastid-encoded, several other noncore subunits of PEP are nuclear encoded, and evidence suggests that PEP might interact with up to 50 accessory subunits. Furthermore, specificity of PEP binding to its target promoters depends on the interaction with several different nucleus-encoded σ factors, these factors recognize bacterial-like promoters harboring −10 and −35 boxes (Gruissem and Zurawski 1985). In contrast to what is seen in bacteria, where essential and nonessential σ factors exist, there is no evidence that proves the existence of essential σ factors in higher plants. For instance, *A. thaliana* nuclear genome contains six genes that code for functional σ factors, but the analysis of several insertional mutants and antisense lines for factors *AtSig1-5* only showed weak defects on plant development (pale-green pigmentation) (Schweer et al. 2010); however, *AtSig6* mutant lines display a strong pigment-accumulation phenotype during cotyledon stage, but this phenotype is however not present in later developmental stages, suggesting that expression of this factor is essential only during early developmental stages (Ishizaki et al. 2005; Loschelder et al. 2006; Schweer et al. 2006). Nevertheless, these observations strongly suggest an overall nonessential role for σ factors in plastid transcription and indicate redundancy between the σ factors studied. Furthermore, it has been recently proposed that phosphorylation of several amino acid residues on σ factors is important for regulation of their function, positive or negative effects on promoter specificity and transcription depending on the amino acids phosphorylated (Link 2003; Baginsky and Link 2005). Also, it has been reported that PEP- and NEP-dependent gene expression is mainly regulated in a development-dependent fashion, being NEP target genes expressed during earlier stages of chloroplast development, followed by PEP expression of target genes in later developmental stages (Courtois et al. 2007; Swiatecka-Hagenbruch et al. 2008). According to the current knowledge, most housekeeping genes are class II, while the genes coding for photosystem I and II proteins are all class I, and finally only a few housekeeping genes are class III (Hajdukiewicz et al. 1997; Swiatecka-Hagenbruch et al. 2007).

As it was previously stated, chloroplast genes are organized as operons and then transcribed as polycistronic RNAs; around 60 operons have been identified in tobacco cpDNA (Sugita and Sugiura 1996). To this day, the most studied chloroplastic operon is the *psbB* operon of *Chlamydomonas* and tobacco, which contains five different genes that code for important photosynthetic proteins. The genes *petB* and *petD* contained in this operon have one group II intron each, with the particularity that these two introns require a set of six proteins (APO1, APO2, CAF1, CAF2, CRS2, and CFM3) to undergo proper splicing (Watkins et al. 2011; Barkan 2011; Asakura et al. 2008). Furthermore, specific intercistronic stabilization PPR (pentatricopeptide repeat)-like proteins (HCF107, Mbb1, and HCF152) are needed in order to protect the UTR (untranslated region)-contained sequences that enhance RNA stabilization and later translation (Vaistij et al. 2000; Felder et al. 2001; Hammani et al. 2012; Zhelyazkova et al. 2012a). Figure 4 shows a diagram depicting the basic processes involved in RNA maturation of *psbB* operon genes. Taking all this data in consideration, it is clear that several different mechanisms are involved in the metabolism of native RNAs, providing a complex regulatory network for fine-tuning mature RNA stability and/or translation.

In addition, most chloroplast genomes of angiosperms are reported to have around 20 group II introns and around 30–40 editing sites (Schmitz-Linneweber and Barkan 2007). RNA editing is a transcript maturation step that

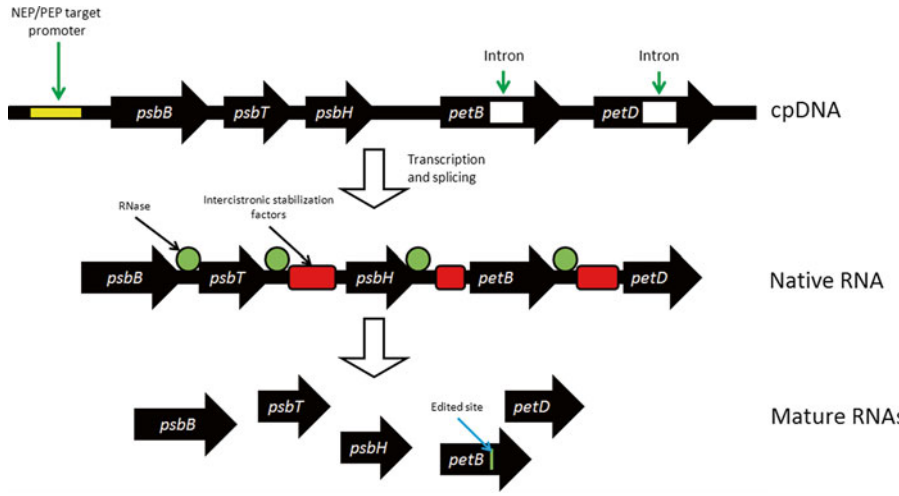


Fig. 4 Structure and posttranscriptional processing of the *psbB* operon. Organization of the *psbB*, *psbT*, *psbH*, *petB*, and *petD* genes (linked *black arrows*) in the *psbB* operon, under the regulation of the same promoter region (highlighted in *yellow*). *petB* and *petD* group II introns (*white spaces*) are posttranscriptionally spliced, and several different RNA-binding proteins (*green circles* and *red boxes*) stabilize and further process the polycistronic native RNA into individual mRNAs (*single black arrows*) that may undergo editing (highlighted in *green*)

involves the deamination of specific cytidine, generating C-to-U changes in mature mRNAs; the C-to-U changes usually lead to change in the amino acid coded by the modified codon (Chateigner-Boutin and Small 2010). It is known that the degree of site editing changes when plants are exposed to different stressing conditions; however, recent studies using novel technologies for transcriptome sequencing (RNA-seq) have found that, under normal growth conditions, most of the chloroplast editing sites are edited with efficiencies over the 74 %. However, as much as 45 % of the total mRNA count was found unedited for the translation start codon of the *ndhD* gene (Ruwe et al. 2013). Despite the importance of editing for protein synthesis, very little is known about the molecular mechanism underlying the process. Very recently, pentatricopeptide repeat proteins (PPR) have been implicated in the recognition of the cytosine bases to be edited, since it was demonstrated that PPR proteins bind to RNA segments of 10–20 nucleotides long that are upstream to the editing site (Chateigner-Boutin and Small 2010; Okuda et al. 2006). To this date, controversy exists about the biological role of transcript editing; however, some observations point that it might be a mechanism to cope with mutations, since replacement of tobacco chloroplast gene *psbF* (not edited in tobacco) with the spinach orthologue generated plants with slow growth and low chlorophyll accumulation (Bock et al. 1994).

Providing yet another RNA metabolism regulation module are chloroplastic noncoding RNAs (ncRNAs). Noncoding RNAs are widely distributed among eukaryotes and prokaryotes; they are involved in several different steps of RNA metabolism, from transcription regulation by chromatin modification to gene silencing by directed mRNA degradation. In plastids, it was thought that only a few ncRNAs were present, but this idea is now being challenged (thanks to new technologies such as RNA-seq) by recent findings that point out the accumulation of several ncRNAs derived from cpDNA in *A. thaliana*, rice, barley, and Chinese cabbage (Hotto et al. 2011; Wang et al. 2011; Chen et al. 2006; Zhelyazkova et al. 2012b). Figure 5 depicts the sites that give rise to ncRNAs in the plastid genomes of *A. thaliana* and barley. It is important to note that the precise functions and biogenesis mechanisms of the newly discovered ncRNAs are still largely unknown, though there are experiments showing that chloroplastic ncRNAs may

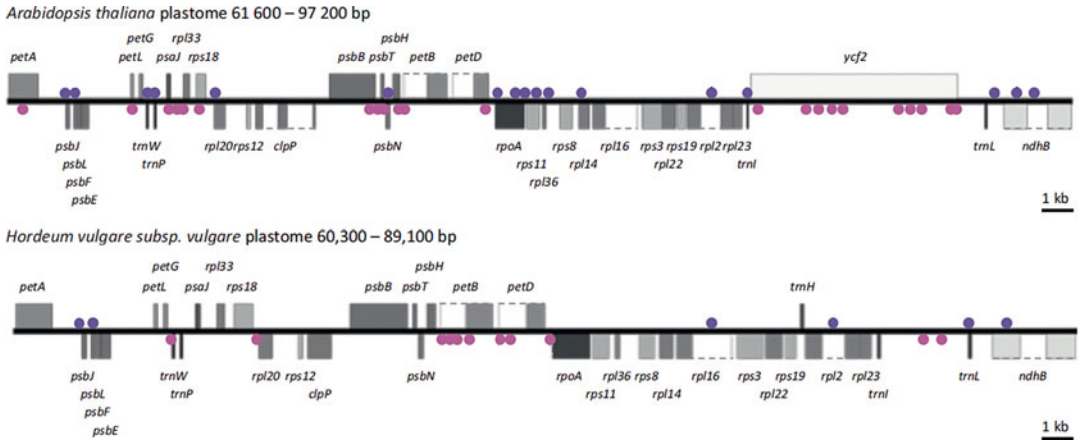


Fig. 5 **Noncoding RNA coding sites**. Sites reported as sources of ncRNAs in cpDNAs of *Arabidopsis* and barley. *Purple circles* represent ncRNAs encoded on the (+) strand and *magenta circles* represent ncRNAs encoded on the (−) strand (Adapted from Hotto et al. 2012)

be generated by the protection against exonuclease activity exerted by the PPR-like proteins involved in polycistronic processing (Ruwe and Schmitz-Linneweber 2012).

There are several available methods for the analysis of transcriptomes of plants under different stress conditions, such as DNA microarrays, SAGE (serial analysis of gene expression), and RNA-seq. Very recently, the chloroplast transcriptome of *A. thaliana* has been explored during seed development, stratification, germination, and early seedling development. It was found that abundance of most plastid mRNAs increase during maturation and decrease in dry seeds, while the abundance of mRNAs coding for the plastid gene expression machinery do not follow this pattern (Allorent et al. 2013). Another study shows that several mRNAs (including PEP subunits) are present in dry seed plastids and those mRNAs increase their abundance after stratification, at the same time antisense RNAs start accumulating during stratification and become easily detectable upon cold release. Furthermore, it was found that the expression of several housekeeping genes (PEP subunits, ribosomal proteins, and *matK*) starts during stratification, while the expression of genes related to PSI, PSII, and electron transport proteins start only upon cold release and light incidence (Demarsy et al. 2012). It was also found that little changes in mRNA expression occur once the chloroplasts have achieved photosynthetic competence, regardless of the developmental stage of the plant (Demarsy et al. 2012).

Several studies have given insight to the genes relevant to stress processes like tolerance to salt stress, drought, herbicides, and pathogens; the identified genes are perfect candidates for application in crop genetic engineering. For instance, the plant responses to pathogen infection have been explored through transcriptomics approaches in several different crop species such as canola, peanut, soybean, barley, tomato, rice, potato, grape, and wheat (Zhao et al. 2007; Luo et al. 2005; Moy et al. 2004; Zierold et al. 2005; Gibly et al. 2004; Zhou et al. 2010; Restrepo et al. 2005; Bruggmann et al. 2005; Figueiredo et al. 2008). Some studies in *A. thaliana* showed that infection with *Pseudomonas syringe* produce expression changes of several metabolic genes related with carbon metabolism, such expression changes seem to be coupled to the deviation of energy resources from biomass production to pathogen eradication (Scheideler et al. 2002); this data may be used in the near future to generate crop lines capable of efficient pathogen fighting at lower energy cost. Furthermore, in 2003 a study was conducted to investigate the responses

of rice plants to cold, salt, and drought stress; it was found that 73 genes were induced by plant exposure to stress (36 cold-induced genes, 62 drought-induced genes, 57 salt-induced genes). Among the stress-induced genes are several transcriptional factors, carbohydrate, and amino acid metabolism proteins, all enzymes whose expression may be engineered in order to improve rice yield under the mentioned stress conditions (Rabbani et al. 2003). Table 2 shows a list of the 73 stress upregulated genes found in the cited study.

Currently, a plethora of examples of transcriptomics-based studies dealing with common crop problems are available, and extensive reviews on the topic are published on a regular basis. However, the knowledge of chloroplast transcriptomes (being relatively simple and mainly focused on photosynthesis) largely lead to the emergence of a very interesting field directly related to the improvement of crops: photosynthesis/RuBisCO engineering.

Despite the existence of different alternative metabolic pathways to assimilate CO_2 to generate biomass, the carboxylating activity of RuBisCO sits at the core of the photosynthetic activity of all plant species. In spite of its great importance for plant survival, RuBisCO has some traits that make it an extremely inefficient enzyme. For instance, RuBisCO is able to catalyze a side reaction with oxygen (known as photorespiration) that leads to the formation of 2-phosphoglycolate, a metabolite that has to go through a long series of chemical reactions comprised in different

Table 2 Cold stress upregulated genes in rice (Rabbani et al. 2003)

Functional category	Number of genes	Description
Transcription factor	6	bZip DNA-binding protein, C_2H_2-type zinc finger DNA-binding protein, C_3HC_4-type RING finger protein, Myb-type DNA-binding protein, NAC-type DNA-binding protein
Receptor-like protein kinase	1	Receptor-like protein kinase
Protein phosphatase	1	Protein phosphatase 2C
Compatible solutes	6	LEA protein, dehydrin, lectin
Detoxification	3	Catalase, *O*-methyltransferase, aldehyde dehydrogenase
Photosynthesis	1	Chlorophyll *a*-/*b*-binding protein
Membrane protein	1	Chloroplast membrane protein
Carbohydrate metabolism	7	Glycoside hydrolase, glycosyl transferase, phosphoglycerate kinase, pyruvate dehydrogenase kinase 1, trehalose-6-phosphate phosphatase, UDP-Glc-4-epimease, carboxyphosphoenolpyruvate mutase
Electron transport system	1	Thioredoxin
Amino acid metabolism	2	4-Hydroxyphenylpyruvate dioxygenase, *S*-adenosylmethionine decarboxylase
Fatty acid metabolism	3	Choline kinase, lipase, and lipoxygenase
Nucleotide metabolism	1	Adenylate kinase
Hormone biosynthesis	1	Zeaxanthin epoxidase
F-box protein	1	F-box protein
Protease inhibitor	1	Protease inhibitor
Protease	1	Papain Cys protease
Dehydrogenase	3	3-Hydroxyacyl-CoA dehydrogenase, dihydroorotate dehydrogenase, glutamate dehydrogenase
Iron homeostasis	2	Ferritin metallothionein-like type 2
Cytoskeleton	2	Actin, actin-depolymerizing factor
Transporter	1	Sugar transporter
Unknown protein	28	Unknown protein

plant cell compartments (chloroplast, mitochondrion, peroxisome, and cytosol) to get its carbon atoms back into the Calvin cycle, at the cost of $NADH_2$, ATP, fixed ammonia, and CO_2 (Foyer et al. 2009). Besides photorespiration, RuBisCO is a slow enzyme that requires posttranslational modifications and conformational remodeling to stay active; hence, a great amount of RuBisCO enzyme is necessary to maintain adequate photosynthesis rates and to support plant growth; actually, around 50 % of the total soluble protein extracted from leaves is RuBisCO (Parry et al. 2012). Taken together, these observations and the pivotal role of RuBisCO for plant yield pose this enzyme or its modifying enzymes as straightforward candidates for engineering in order to achieve the creation of crop-specific RuBisCO enzymes efficient enough to greatly improve plant yield in specific environmental conditions. Several efforts have been made in this matter and will be discussed here.

As it was previously stated, RuBisCO makes up to 50 % of the total amount of leaf protein and contains around 25 % of the total leaf nitrogen; being nitrogen an essential and expensive plant nutrient, in 1994 some experiments were conducted with tobacco antisense lines with around 20 % less RuBisCO content than wild-type plants, reducing nitrogen demand in around 11 % without negatively affecting CO_2 fixation (Stitt and Schulze 1994). Besides lowering plant nitrogen demand by modulating RuBisCO levels, efforts have been made for the improvement of some of its enzymatic traits such as CO_2 affinity, CO_2/O_2 specificity, and reaction speed by amino acid substitutions in the large subunit of tobacco RuBisCO (Zhu et al. 2010). Furthermore, there are several reports that link heat stress to low rates of CO_2 assimilation and decreased levels of RuBisCO activity caused by heat-dependent RuBisCO activase inactivation (Figs. 6 and 7). RuBisCO activase has optimal activity at temperatures below 40 °C, but this trait varies among plant species depending on the average temperatures of their natural environments (Figs. 6 and 7) and correlates with the optimal temperatures for photosynthesis (Carmo-Silva and Salvucci 2011; Carmo-Silva et al. 2012). Given this data, RuBisCO activase provides opportunities for genetic engineering, in the sense that creating more thermostable variants of this enzyme would help photosynthesis to be less sensitive to heat stress; despite the apparent advantages that RuBisCO activase-engineered crops would pres-

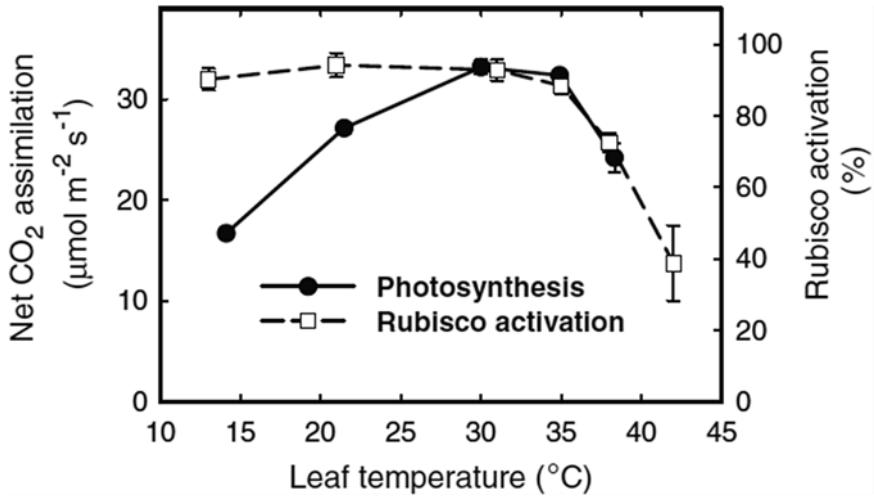


Fig. 6 Effect of heat stress on RuBisCO activation and photosynthesis. The effect of heat stress on RuBisCO activation (*squares*) and photosynthesis (*filled circles*) was evaluated by measuring the net CO_2 assimilation rate and RuBisCO in vitro activity of plants under non-photorespiratory conditions (2 % O_2) (Adapted from Carmo-Silva et al. 2012)

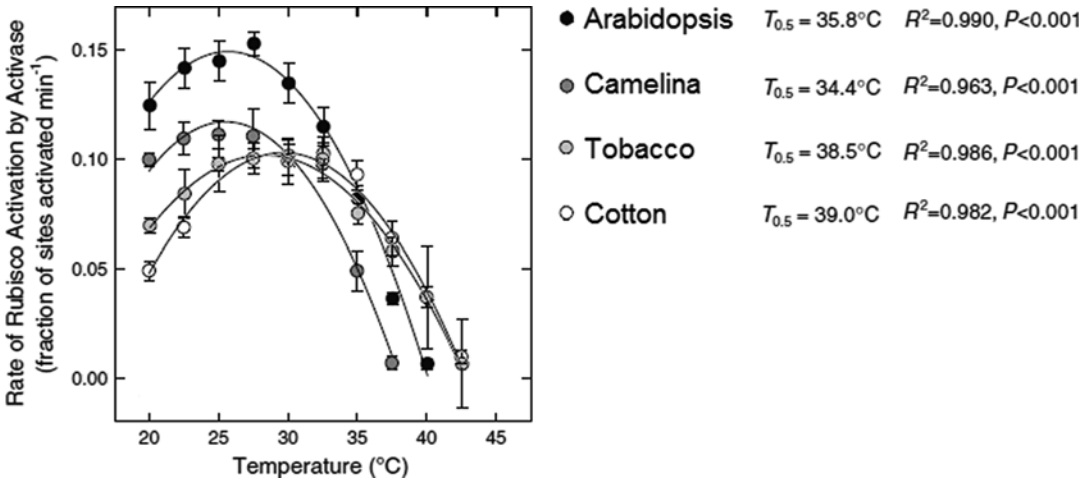


Fig. 7 Effect of heat stress on RuBisCO activation by RuBisCO activase. The effect of heat stress on RuBisCO activase was monitored at different temperatures in different plant species, Arabidopsis (*black filled circles*), Camelina (*dark-gray filled circles*), tobacco (*light-gray filled circles*), and cotton (*void circles*) (Adapted from Carmo-Silva and Salvucci 2011)

ent in an over-warming planet, such plant lines are still in the very early stages of development (Yamori et al. 2012).

Transcriptomics has proved to be a very useful tool for the identification of candidate genes to engineer in order to improve the performance of crops under diverse stressing conditions; RuBisCO itself provides a great example of the gap existing between transcript accumulation, protein expression, and protein activity; by extrapolating this problem to other genes of interest, it becomes clear that the lack of knowledge poses challenges that cannot be overcome by genomics or transcriptomics alone.

Chloroplast Proteomes

The proteome refers to the entire set of proteins expressed by a given group of cells under specific conditions or developmental stage; the term was coined in 1994 by Marc Wilkins. On the other hand, the term proteomics was coined in 1997 by P. James by blending together the terms protein and genome; this term refers to the large-scale study of proteins. Nowadays, proteomics has proved to be a very valuable tool for the progress of basic and applied sciences. Thanks to the development of different mass-spectrometry systems, researchers have gotten to explore qualitatively and quantitatively the global protein expression profiles at a level of detail never reached before. Research in proteomics of higher plant chloroplasts has achieved considerable success, and the chloroplast proteomes of different plant species are known, including specific subsets like envelope, stroma, and thylakoid proteomes (Pineda et al. 2010; Friso et al. 2004); thanks to particular protein extraction and sample fractionation protocols that enable enrichment of the samples with proteins of specific physicochemical traits and low abundance, such as membrane integral proteins whose hydrophobicity was for a long time a challenge when trying to analyze thylakoid membrane proteomes. The methods to enrich samples with membrane proteins often include steps of fractionation with organic solvents, 1-D or 2D SDS-PAGE (sodium dodecyl sulfate-polyacrylamide gel electrophoresis) or reverse-phase HPLC (high-performance liquid chromatography) (Friso et al. 2004).

Chloroplast proteomes have been reported for several plant species under different stressing conditions like *A. thaliana*, pea, barley, zucchini, sugar beet, rice and wheat (Andaluz et al. 2006; Cui et al. 2005; Zhou et al. 2006; Curto et al.

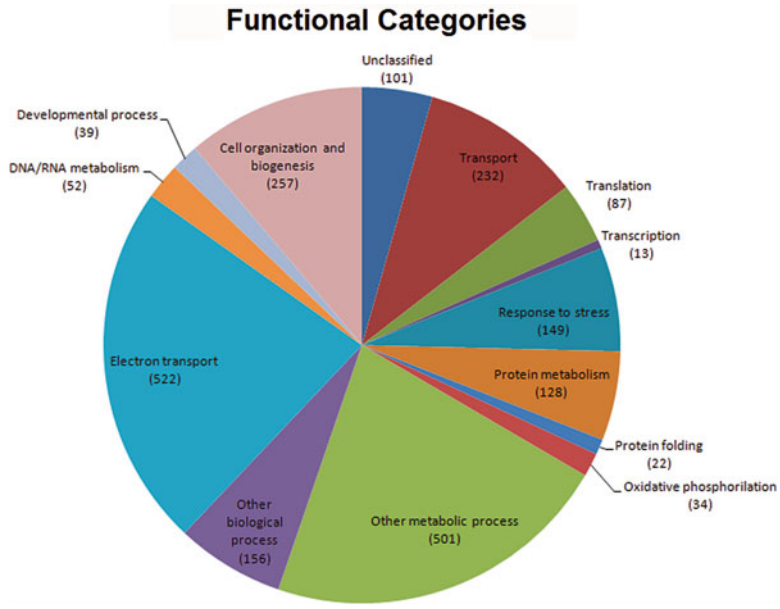


Fig. 8 Probable functions of wheat plastid proteins. Plastid protein extracts from wheat plants were analyzed and functions predicted for 767 different proteins (Data extracted from Kamal et al.2012)

2006; Ciambella et al. 2005; Aro et al. 2005); some examples will be discussed below.

In 2012 Kamal et al. reported a set of 767 unique proteins present in extracts from purified Korean winter wheat chloroplasts and sorted the identified proteins in 14 functional categories (Fig. 8), the majority of proteins were found to be involved in electron transport, cell organization and biogenesis, and metabolic processes. Furthermore, in 2007 it was reported a proteomic analysis dealing with rice chloroplast biogenesis (Kleffmann et al. 2007); the study used shotgun and 2D-PAGE-based proteomics to ensemble a rice etioplast proteome containing 477 etioplast-specific proteins, it was also found that the transition from proplastid to chloroplast induced by light is marked by a shift in protein expression that impacts the metabolic capabilities of the plastid, making it change from heterotrophic to autotrophic metabolism. In the dark, plastid protein expression was found to be focused on carbohydrate and amino acid metabolism, but 2 h after illumination proteins with functions related to carbohydrate metabolism, photosynthesis, and plastid gene expression increased their abundance, while proteins with functions like amino acid and lipid metabolism lowered their abundance, and proteins involved in nucleotide metabolism, redox regulation, and tetrapyrrole synthesis remained unchanged. Kleffmann et al. also found that proteins involved in the plastid translation machinery accumulated during this transition, the elongation factors P and Tu accumulate along with the proteins that make up the ClpP system for protein turnover. It is believed that the accumulation of translation and protein degradation proteins in the developing plastids responds to the need of replacing damaged proteins by the high levels of photooxidative stress that chloroplast proteins are subjected to. Furthermore, proteins involved in the stabilization of mRNAs increase their abundance along this transition stage, while proteins related to mRNA turn over are less abundant in mature plastids of rice. As it was discussed in the previous section, the knowledge of the genomic elements and protein effectors in RNA stability is crucial for the development of chloroplast genetic engineering methodologies, and the data available on this matter may provide opportunities for improving steps in chloroplast biogenesis that might result in enhancements of plant yield.

The analysis of chloroplast subproteomes has resulted in the better understanding of previously known functions of specific organelle compartments. In the past, it was uncertain at what extent

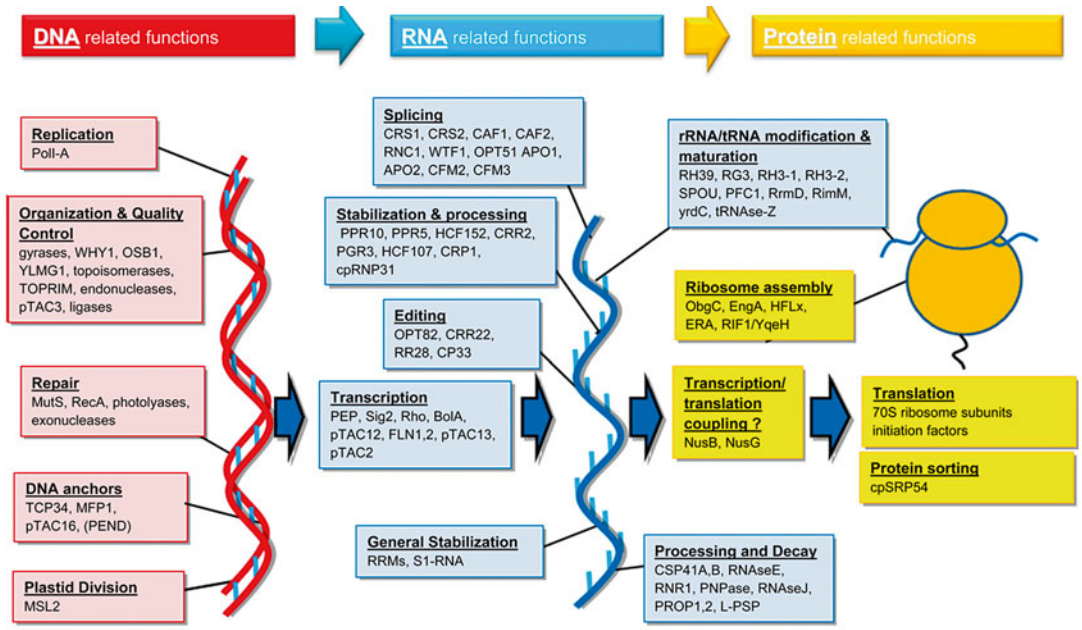


Fig. 9 Model of the nucleoid function. Proteins related to transcription, translation, RNA processing, and DNA maintenance were found to be enriched in chloroplast nucleoids, extending the previous model of nucleoid function. DNA-related processes (shown in *red*; replication, organization, repair, and plastid division), RNA-related processes (shown in *blue*; transcription, processing, stabilization, splicing, editing), protein-related processes (shown in *yellow*; translation, ribosome assembly, protein sorting) (Taken from Majeran et al. 2012)

the plastid nucleoids were the site of posttranscriptional processes of RNA maturation and ribosome assembly. However, in 2012, Majeran et al. reported a proteomic analysis of isolated maize plastid nucleoids; in the study, it was possible to identify a group of 127 nucleoid-specific proteins. Most of the identified proteins mapped to processes that are classified as DNA related (DNA anchoring, organization, repair, and replication) and RNA related (transcription, editing, splicing, stabilization, and maturation), making clear that plastid nucleoids are the site of chloroplast gene expression and RNA metabolism; besides DNA- and RNA-related processes, nucleoids were also found to be enriched in ribosomal and translation-related proteins, these proteins represented about 20 % of the total protein mass of nucleoids. These findings lead to the proposal of a novel model in which ribosome assembly and protein translation takes place in the nucleoid, in a process that may be couple to transcription (Fig. 9; Majeran et al. 2012).

Also, the effects of stress on chloroplast proteomes have been explored. For instance, in 2006 the thylakoid proteome of sugar beet was explored under iron deficiency conditions, showing that the abundance of proteins participating in photosynthetic electron transport is largely affected by iron starvation, while carbon metabolism-related proteins (RuBisCO small and large subunits, RuBisCO activase, carbonic anhydrase, phosphoglycerate kinase, aldolase, phosphoribulokinase, transketolase, and ribulose-1,7-bisphosphatase) displayed greater abundance under the stress conditions (Andaluz et al. 2006). Furthermore, the proteome of rice leaves under cold stress was reported to contain 60 unique proteins, whose functions are related to protein synthesis and folding, cell wall synthesis, protein degradation, energy production, and signal transduction. Interestingly more than 40 % of the detected proteins were identified as chloroplast proteins by bioinformatic tools of transit peptide prediction, immediately posing plastids as the

most affected organelles by cold stress and possibly as mediators of plant responses to cold stress (Cui et al. 2005).

Other aspects of plastid biology have been analyzed using proteomic approaches. For instance, chloroplast envelopes of pea and maize were isolated and their proteomes were analyzed and compared. The result was a comparison of the plastid proteomes of C_3 and C_4 plants; it was found that metabolite transporter proteins (specifically the triosephosphate/phosphate translocator and the phosphoenolpyruvate/phosphate translocator) are enriched in C_4 chloroplast proteomes probably due the higher metabolic fluxes that occur during the C_4 type of photosynthesis (Bräutigam et al. 2008). This data provides opportunities for the improvement of projects dealing with the engineering of carbon assimilation in C_4 plants.

Yet another interesting aspect of proteomics is the analysis of the phosphoproteome, a particular subproteome that consists of all the proteins that are phosphorylated under certain conditions. To this date, very little research has focused on the analysis of phosphorylated proteins from isolated plant organelles, due to low protein abundance and the low stability that phosphoproteins display under the several fractionation steps needed for organelle isolation and protein extraction; hence, rapid protocols are used for the extraction of phosphorylated proteins from whole tissues or organs, ensuring a high phosphoproteome coverage (Espina et al. 2008). Given the technical limitations that phosphoproteomic approaches face, the data of phosphoprotein abundance profiles for different organelles must be extracted from genome-scale phosphoproteomic approaches; in order to assign phosphoproteins to specific cell locations, name it the chloroplast, the whole phosphoproteome must be compared to information coming from several different sources such as organellar subproteomes and targeting experiments with GFP (green fluorescent protein)-fused proteins. Using this approach, in 2009, Reiland et al. reconstructed a phosphoproteome map for *A. thaliana* chloroplasts from shoots and rosette leaves; it was found that the most frequently phosphorylated amino acids in plastids are serine and threonine (accounting for around 80 % and 20 % of the total phosphorylated amino acids, respectively) and no evidence of tyrosine phosphorylation was found among the 174 analyzed plastid-localized phosphoproteins. Furthermore, it was found that most of the phosphoproteins have functions related to photosynthesis (26), metabolism (30), and gene expression (16); several chloroplast phosphoproteins were found in this study that support older observations and hypothesis. For instance, it was suggested that phosphorylation of the RNA-binding proteins RNP29 and RNP33 is key for RNA stabilization and both proteins were detected in this study (Reiland et al. 2009). Taken together, it is clear that phosphoproteomics is a valuable tool for the confirmation and discovery of phosphoproteins for the many functions performed by the chloroplasts; however, no further efforts have been made recently to explore in detail the phosphoproteome of plastids of other plant species.

The analysis of proteomes, particularly plastid proteomes and the recent systematic efforts to make all proteomics information publicly available through the creation of new and extensive databases, provides opportunities to explore how different processes and stress responses are regulated at the protein level, hence providing the protein candidates necessary for the generation of novel engineered research tools or crops best fit for specific environmental conditions.

Chloroplast Metabolomes

Extending the concept from transcriptomics and proteomics, metabolomics is a field that seeks the profiling of all the small-molecule metabolites present in a biological sample under a fixed condition or developmental stage, that way giving a comprehensive set of data that represents a detailed overview of the precise metabolic and physiological state of the sample analyzed. Several analytical methods are often applied in metabolomics-based studies; sample fractionation is achieved through gas chromatography, HPLC, and capillary electrophoresis (Schauer et al. 2005; Gika et al. 2007; Soga et al. 2003),

whereas metabolite detection is usually performed using mass-spectrometry technologies or nuclear magnetic resonance. However, metabolomics faces a great challenge when it comes to the analysis of its molecules of interest; in contrast to genomics, transcriptomics and proteomics, metabolomics has to deal with a very heterogeneous set of molecules with different properties, while the other omics analyze only polymers consisting of a very limited number of monomers with a predictable range of physicochemical properties. For this reason, several different sample fractionation and metabolite extraction methods must be used when trying to profile the entire metabolome of a tissue or organ. As a result, there are no reported standard methods for the analysis of whole metabolomes from plant samples, and extensive identification databases similar to those used for proteomics are nonexistent for metabolites, and the existing efforts are mainly focused on animal cell models. Other common problems related to metabolite determination are the high turnover rates displayed by some metabolites and the elevated rate of translocation that occurs between different cell compartments (Stitt et al. 1983; Weber and Fischer 2007).

Despite the many disadvantages associated with metabolomic approaches, this technologies have been applied several times with different purposes in plant biology, like the safety assessment of GMOs (genetically modified organisms) (Baker et al. 2006; Kogel et al. 2010; Kusano et al. 2011), the discovery of stress-related compounds (Leiss et al. 2009; Lawo et al. 2011; Aliferis and Jabaji 2012), and gene function discovery (Bino et al. 2004; Yonekura-Sakakibara et al. 2008). Also, several efforts have focused on the study of metabolite content of plastids using nonaqueous fractionation (NAF) methods and classical biochemical assays for the determination of metabolites (Gerhardt and Heldt 1984; Riens et al. 1991). However, recent studies exist that employ state-of-the-art high-throughput analytical procedures (such as mass spectrometry) for the profiling of the metabolite content of specific cell compartments isolated by NAF; those studies have reached a point in which it is possible to describe the metabolome of some plant cell compartments, among the studied cell compartments is of course the chloroplast (Geigenberger et al. 2011). In 2011, it was possible to reconstruct a metabolite map of the compartmentalized metabolome of *A. thaliana* leaves under regular growth conditions; from that study, it is clear that three subcellular compartments can be very well defined through nonaqueous fractionation coupled to mass spectrometry, the chloroplast, the vacuole, and the cytosol. In this study, it was found that 344 analytes are plastid-specific, while 334 metabolites are shared with the cytosol and 24 with the vacuole; the analytes were sorted into different categories including primary metabolism (7 chloroplastic, 12 cytosolic, 9 vacuolar specific metabolites), secondary metabolism (10 chloroplastic, 161 cytosolic, 158 vacuolar), lipophilic compounds (326 chloroplastic, 288 cytosolic, 0 vacuolar), and other functions (1 chloroplastic, 0 cytosolic, 0 vacuolar) (Krueger et al. 2011). A computational analysis of this data revealed that the chloroplast is enriched in amino acids and galactolipids, and some amino acids are shared with the cytosol while little overlap exists between the metabolites present at the plastid and those at the vacuole. On the other hand, a plethora of metabolites was mapped to the cytosol, stating this compartment as a gateway for metabolite flux (Klie et al. 2011). In spite of the great amount of data generated by large-scale metabolic profiling studies, only few biologically relevant hypotheses can be generated without information about the specific identities of the detected metabolites, and as this information is still out of reach for the average research project, the impact of chloroplast metabolome determination on the understanding of plant biology will be evaluated in the future.

However, one growing field within metabolomics is the identification of alterations in lipid metabolism under different conditions, a research area commonly known as lipidomics. Being the lipidome part of the metabolome, it suffers from many of the limitations usually associated to metabolomic studies, such as the lack of standard protocols for sample fractionation and metabolite isolation and identification; as a result, no specific large-scale lipidomics experiments have

been driven to characterize the lipid accumulation profile of intact chloroplasts. However, the study described above sheds light on the matter, suggesting that the lipidome of chloroplasts under normal growth conditions is composed mainly of galactolipids and only few species of phospholipids (Klie et al. 2011), an observation that is consistent with data generated using classical biochemical strategies for lipid content determination, which points out that the outer and inner membranes of rye, pea, and spinach chloroplasts are composed by a high proportion of monogalactosyldiacylglycerols (47.9 % inner membrane; 20.1 % outer membrane) and digalactosyldiacylglycerols (30 % for both membranes) with no sterols or cerebrosides present (Uemura and Steponkus 1997; Block et al. 1983; Cline et al. 1981).

Plant metabolomics is a field yet in the very early stages of technical and analytical development; hence, full chloroplastic metabolite profiles are not available yet for any plant species, but given the great correlations that can be made between metabolomics, proteomics, transcriptomics, and genomics, it is clear that plastid metabolomics will develop faster in the years to come.

Chloroplast System Biology

Systems biology is a multidisciplinary field of science that features a holistic perspective of biological phenomena instead of the classical reductionist way of addressing questions in biology. Therefore, systems biology makes use of mathematical models to analyze large datasets generated through different approaches such as genomics, transcriptomics, proteomics, and metabolomics to generate hypotheses about dynamic biological systems. The origins of this field of biological research can be tracked down in history to 1952 when a numerical simulation was published by Alan Hodgkin and Andrew Huxley, they constructed a mathematical model that explained the action potential propagation along the axon of a neuronal cell (Hodgkin and Huxley 1952). Hodgkin and Huxley model reconstructed a cellular phenomenon from the known interaction between two molecular components of the cell, a potassium and sodium channel. Nowadays and thanks to the boost that omic technologies have received over the past decade and the increasing capabilities of data processing that computing equipment has attained, systems biology is a field of research on its own, and several institutions dedicated to systems biology have been created in the world.

To this date, much effort has been put in the understanding of anterograde and retrograde signaling pathways between chloroplasts and nucleus; this efforts have been merged together to generate hypotheses about the identity of the key molecular mediators of these signaling events, which is perhaps one of the most enigmatic questions in plant biology. In order to better understand anterograde signaling, different microarray and comparative proteomics experiments have been conducted. For instance, in 2005, a large and comprehensive microarray experiment was designed to evaluate the expression profiles of 3,292 nuclear genes that code for chloroplastic proteins under 101 experimental conditions or genetic backgrounds, the analysis of the data clustered 1,590 genes in 23 independent regulons; furthermore, analysis of individual regulons made clear that photosynthesis and ribosome assembly are highly linked, a possibility never explored before (Biehl et al. 2005).

Microarray-based experiments have also been successfully performed in order to shed light on the events regulating retrograde signaling during the early stages of chloroplast development. In 2007, Koussevitzky et al. reported an approach to compare wild-type plants and two *gun* (*genomes uncoupled*) mutants (*gun1* and *gun5*) at the transcriptional level; it was found that 330 genes are regulated during plastid development and further genomics analysis of the dataset led to the identification of a shared motif in the promoter regions of the studied genes that represent the core sequence of the ABA (abscisic acid) response elements, strongly suggesting that transcription factors associated with ABA signaling might be involved in the regulation of retrograde signaling (Koussevitzky et al. 2007).

Yet another example of systems biology applied to chloroplast research is the generation of computational interactomes using proteins predicted to be translocated to the chloroplast. Yu et al. assembled a chloroplastic proteome composed of 7,592 proteins, by comparing datasets from 9 different resources, further computational analysis determined a set of 1,808 proteins as the core proteome of *A. thaliana chloroplasts*, including the 88 proteins encoded by the chloroplast genome. The rest of the proteins were regarded as putative chloroplastic proteins. By comparison with genome-wide *A. thaliana* interactomes and data derived from gene chips, Yu et al. described 22,925 interaction pairs involving 2,214 proteins; 1,043 of these proteins were from the chloroplast core proteome and the remaining 1,171 from the putative chloroplastic protein group. Further analysis revealed the existence of a large network consisting of 3,109 protein interactions between 309 proteins and 84 minor independent networks consisting of more than two interacting partners. The validation of the interactome map generated consisted in the reconstruction from core protein networks of the entire protein complexes that make up the photosystems I and II, along with their respective light harvesting complexes, plus the ATP synthase and cytochrome b_6f protein complexes, this way reassuring the reliability of the assembled interactome. In a further validation effort, 12 interactions were randomly chosen to be experimentally assessed by yeast two-hybrid assays, and the physical interactions were confirmed (Yu et al. 2008).

Even though systems biology is a not-so-young field of research and many integrative efforts have been made to try to explore chloroplast function and development with a global perspective, most of these state-of-the-art studies have been conducted using *A. thaliana* as model organism; hence, comprehensive information coming from different plant models such as maize or rice is needed for taking the development of analytical tools one step further and assemble a core proteome or interactome of the flowering plants, instead of species-specific maps. Even further, with the addition of information from more plant species, it would be possible to generate maps that depict the core transcriptomes, proteomes, interactomes, and metabolomes of the entire Plantae kingdom, which is of course the ulterior goal of systems biology.

Concluding Remarks

Plastids are the site of thousands of different chemical reactions that provide essential metabolites for plant growth and development. Given the pivotal role of chloroplasts in plant biology, the knowledge of the molecular mechanisms regulating their biogenesis, development, and functioning lays at the core of every crop improvement initiative or any other project that pretends to shed light on fundamental processes for plant life, since the most important molecular regulators (or their ancillary molecules) are usually prime targets for engineering and the information obtained about their regulation can be often extrapolated to several different species.

Nowadays, computational biology is usually very well blended with the tools to analyze omic-generated data, to the point that researchers no longer understand omics without the computational tools to analyze the raw data; take, for instance, proteomics, without image-analysis software (such as Melanie or PD-Quest; http://www.genebio.com/products/melanie/, http://www.bio-rad.com) to process and compare gel images, 2D-PAGE gels would be nothing different to a canvas full of stains, or without the search engines available online (like Mascot; http://www.matrixscience.com/), the raw mass-spectrometry data would yield very little information about the protein content of a given sample and no data about the identity of the detected proteins would be available. In spite of the plethora of analytical tools existing and given the overwhelming amount of data that is being generated on a daily basis through genomics, transcriptomics, proteomics, metabolomics, or any other novel omic approach, one of the most important challenges of modern science is to develop tools that allow researchers to integrate all the information available from different

sources, to generate new biologically relevant hypothesis that impact on the global understanding of life and the hundreds of processes that make the development and physiology of an organism by itself, and when interacting with other individuals of the same or different species.

Just as virtually all fields of biology, the study of chloroplasts has been boosted by the emergence of global analysis systems to explore genomes, transcriptomes, proteomes, metabolomes, etc., being transcriptomics and proteomics the main source of engineering targets, while genomics provides the tools for the application of the engineered proteins or transcripts. To date, the greatest problem for the application of "omic-derived" knowledge to crop engineering is the lack of data generated from enough species to create universal improvement systems instead of the species-specific systems available today.

References

Aliferis K, Jabaji S (2012) FT-ICR/MS and CG-EI/MS Metabolomics networking unravels global potato sprout's responses to *Rhizoctonia solani* infection. PLoS One 7:e42576

Allison L, Simon L, Maliga P (1996) Deletion of *rpoB* reveals a second distinct transcription system in plastids of higher plants. EMBO J 15:2802–2809

Allorent G, Courtois F, Chevalier F, Lerbs-Mache S (2013) Plastid gene expression during chloroplast differentiation and dedifferentiation into non-photosynthetic plastids during seed formation. Plant Mol Biol 82:59–70

Andaluz S, López-Millán A, De Las Rivas J, Aro E, Abadía J, Abadía A (2006) Proteomic profiles of thylakoid membranes and changes in response to iron deficiency. Photosynth Res 89:141–155

Apel W, Bock R (2009) Enhancement of carotenoid biosynthesis in transplastomic tomatoes by induced lycopene-to-provitamin A conversion. Plant Physiol 151:59–66

Aro E, Soursa M, Rokka A, Allahverdiyeva Y, Paakkarinen V, Saleem A, Battchikova N, Rintamaki E (2005) Dynamics of photosystem II a proteomic approach to thylakoid protein complexes. J Exp Bot 56:347–356

Asakura Y, Bayraktar O, Barkan A (2008) Two CRM protein subfamilies cooperate in the splicing of group IIB introns in chloroplasts. RNA 14:2319–2332

Baginsky S, Link G (2005) Redox regulation of chloroplast gene expression. In: Demmig-Adams B, Adams W, Mattoo A (eds) Photoprotection, photoinhibition, gene regulation, and environment. Springer, Dordrecht, pp 269–287

Baker J, Hawkins N, Ward J, Lovegrove A, Napier J, Shewry P, Beale M (2006) A metabolomic study of substantial equivalence of field-grown genetically modified wheat. Plant Biotechnol J 4:381–392

Barkan A (2011) Expression of plastid genes: organelle-specific elaborations on a prokaryotic scaffold. Plant Physiol 155:1520–1532

Bendich A (1991) Moving pictures of DNA released upon lysis from bacteria, chloroplasts, and mitochondria. Protoplasma 160:121–130

Bendich A (1987) Why do chloroplasts and mitochondria contain so many copies of their genome? Bioessays 6:279–282

Bendich A (2004) Circular chloroplast chromosomes: the grand illusion. Plant Cell Online 16:1661–1666

Biehl A, Richly E, Noutsos C, Salamini F, Leister D (2005) Analysis of 101 nuclear transcriptomes reveals 23 distinct regulons and their relationship to metabolism, chromosomal gene distribution and co-ordination of nuclear and plastid gene expression. Gene 344:33–41

Bino R, Hall R, Fiehn O, Kopka J, Saito K, Draper J, Nikolau B, Mendes P, Roessner-Tunaliu U, Beale M, Trethewey R, Lange B, Wurtele E, Sumner L (2004) Potential of metabolomics as a functional genomics tool. Trends Plant Sci 9:418–425

Blankenship R (2002) Molecular mechanisms of photosynthesis. Blackwell Science, Oxford

Block M, Dorne A, Joyard J, Douce R (1983) Preparation and characterization of membrane fractions enriched in outer and inner envelope membranes from spinach chloroplasts. J Biol Chem 258:13281–13286

Bock R, Kossel H, Maliga P (1994) Introduction of a heterologous editing site into the tobacco plastid genome: the lack of RNA editing leads to a mutant phenotype. EMBO J 13:4623–4628

Bohmert-Tatarev K, McAvoy S, Daughtry S, Peoples O, Snell K (2011) High levels of bioplastic are produced in fertile transplastomic tobacco plants engineered with a synthetic operon for the production of polyhydroxybutyrate. Plant Physiol 155:1690–1708

Boynton J, Gillham N, Harris E, Hosler J, Johnson A, Jones A, Randolph-Anderson B, Robertson D, Klein T, Shark K et al (1988) Chloroplast transformation in Chlamydomonas with high velocity microprojectiles. Science 240:1534–1538

Bräutigam A, Hoffmann-Benning S, Weber A (2008) Comparative proteomics of chloroplast envelopes from C_3 and C_4 plants reveals specific adaptations of the plastid envelope to C_4 photosynthesis and candidate proteins required for maintaining C_4 metabolite fluxes. Plant Physiol 148:568–579

Bruggmann R, Abderhalden O, Reymond P, Dudler R (2005) Analysis of epidermis- and mesophyll-specific transcript accumulation in powdery mildew-inoculated wheat leave. Plant Mol Biol 58:247–267

Bult C, White O, Olsen G, Zhou L, Fleischmann R, Sutton G, Blake J, FitzGerald L, Clayton R, Gocayne J, Kerlavage A, Dougherty B, Tomb J, Adams M, Reich C, Overbeek R, Kirkness E, Weinstock K, Merrick J, Glodek A, Scott J, Geoghagen N, Weidman J, Fuhrmann J, Nguyen D, Utterback T, Kelley J, Peterson J, Sadow P, Hanna M, Cotton M, Roberts K, Hurst M, Kaine B, Borodovsky M, Klenk H, Fraser C, Smith H, Woese C, Venter J (1996) Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii. Science 273:1058–1073

Campbell W, Ogren W (1990) Electron transport through photosystem I stimulates light activation of ribulose bisphosphate carboxylase/oxygenase (Rubisco) by Rubisco activase. Plant Physiol 94:479–484

Carmo-Silva A, Salvucci M (2011) The activity of Rubisco's molecular chaperone, Rubisco activase, in leaf extracts. Photosynth Res 108:143–155

Carmo-Silva A, Gore M, Andrade-Sanchez P, French A, Hunsaker D, Salvucci M (2012) Decreased CO_2 availability and inactivation of Rubisco limit photosynthesis in cotton plants under heat and drought stress in the field. Environ Exp Bot 83:1–11

Chakrabarti S, Lutz K, Lertwiriyawong B, Svab Z, Maliga P (2006) Expression of the cry9Aa2 B.t. gene in tobacco chloroplasts confers resistance to potato tuber moth. Transgenic Res 15:481–488

Chateigner-Boutin A, Small I (2010) Plant RNA editing. RNA Biol 7:213–219

Chen Z, Zhang J, Kong J, Li S, Fu Y, Li S, Zhang H, Li Y, Zhu Y (2006) Diversity of endogenous small noncoding RNAs in *Oryza sativa*. Genetica 128:21–31

Chen Y, Chen J, Wang P, Mi H, Chen G, Xu D (2010) Reversible association of ribulose-1, 5-bisphosphate carboxylase/oxygenase activase with the thylakoid membrane depends upon the ATP level and pH in rice without heat stress. J Exp Bot 61:2939–2950

Ciambella C, Roepstorff P, Aro E, Zolla L (2005) A proteomics approach for investigation of photosynthetic apparatus in plants. Proteomics 5:746–757

Cline K, Andrews J, Mersey B, Newcomb E, Keegstra K (1981) Separation and characterization of inner and outer envelope membranes of pea chloroplasts. Proc Natl Acad Sci USA 78:3595–3599

Courtois F, Meredino L, Demarsy E, Mache R, Lerbs-Mache S (2007) Phage-type RNA polymerase RPOTmp transcribes the *rrn* operon from the PC promoter at early developmental stages in *Arabidopsis*. Plant Physiol 145:712–721

Cui S, Huang F, Wang J, Ma X, Cheng Y, Liu J (2005) A proteomic analysis of cold stress responses in rice seedlings. Proteomics 5:3162–3172

Curto M, Camafeita E, López J, Maldonado A, Rubiales D, Jorrín J (2006) A proteomic approach to study pea (*Pisum sativum*) responses to powdery mildew (*Erysiphe pisi*). Proteomics 6:163–174

Danielle H, Datta R, Varma S, Gray S, Lee S (1998) Containment of herbicide resistance through genetic engineering of the chloroplast genome. Nat Biotechnol 16:345–348

De Cosa B, Moar W, Lee S, Miller M, Daniell H (2001) Overexpression of the Bt cry2Aa2 operon in chloroplasts leads to formation of insecticidal crystals. Nat Biotechnol 19:71–74

Demarsy E, Buhr F, Lambert E, Lerbs-Mache S (2012) Characterization of the plastid-specific germination and seedling establishment transcriptional programme. J Exp Bot 63:925–939

Dufourmantel N, Tissot G, Goutorbe F, Garcxon F, Muhr C, Jansens S, Pelissier B, Peltier G, Dubald M (2005) Generation and analysis of soybean plastid transformants expressing Bacillus thuringiensis Cry1Ab protoxin. Plant Mol Biol 58:659–668

Espina V, Edmiston K, Heiby M, Pierobon M, Sciro M, Merritt B, Banks S, Deng J, VanMeter A, Geho D, Pastore L, Sennesh J, Petricoin E, Liotta L (2008) A portrait of tissue phosphoprotein stability in the clinical tissue procurement process. Mol Cell Proteomics 7:1998–2018

Felder S, Meierhoff K, Sane A, Meurer J, Driemel C, Plucken H, Klaff P, Stein B, Bechtold N, Westhoff P (2001) The nucleus encoded HCF107 gene of Arabidopsis provides a link between intercistronic RNA processing and the accumulation of translation-competent psbH transcripts in chloroplasts. Plant Cell 13:2127–2141

Fiers W, Contreras R, Duerinck F, Haegeman G, Iserentant D, Merregaert J, Min Jou W, Molemans F, Raeymaekers A, Van den Berghe A, Volckaert G, Ysebaert M (1976) Complete nucleotide sequence of bacteriophage MS2 RNA: primary and secondary structure of the replicase gene. Nature 260:500–507

Figueiredo A, Fortes A, Ferreira S, Sebastiana M, Choi Y, Sousa L, Acioli-Santos B, Pessoa F, Verpoorte R, Pais M (2008) Transcriptional and metabolic profiling of grape (*Vitis vinifera* L.) leaves unravel possible innate resistance against pathogenic fungi. J Exp Bot 59:3371–3381

Fleischmann R, Adams M, White O, Clayton R, Kirkness E, Kerlavage A, Bult C, Tomb J, Dougherty B, Merrick J (1995) Whole-genome random sequencing and assembly of Haemophilus influenzae Rd. Science 269:496–512

Foyer C, Bloom A, Queval G, Noctor G (2009) Photorespiratory metabolism: genes, mutants, energetics, and redox signaling. Annu Rev Plant Biol 60:455–484

Friso G, Giacomelli L, Ytterberg A, Peltier J, Rudella A, Sun Q, Wijk K (2004) In-depth analysis of the thylakoid membrane proteome of *Arabidopsis thaliana* chloroplasts: new proteins, new functions, and a plastid proteome database. Plant Cell Online 16:478–499

Geigenberger P, Tiessen A, Meurer J (2011) Use of non-aqueous fractionation and metabolomics to study chloroplast function in *Arabidopsis*. Methods Mol Biol 775:135–160

Gerhardt R, Heldt H (1984) Measurement of subcellular metabolite levels in leaves by fractionation of freeze-stopped material in nonaqueous media. Plant Physiol 75:542–547

Gibly A, Bonshtien A, Balaji V, Debbie P, Martin G, Sessa G (2004) Identification and expression profiling of tomato genes differentially regulated during a resistance response to *Xanthomonas campestris pv. vesicatoria*. Mol Plant Microbe Interact 17:1212–1222

Gika H, Theodoridis G, Wingate J, Wilson I (2007) Within-day reproducibility of an LC-MS-based method for metabonomic analysis: application to human urine. J Proteome Res 6:3291–3303

Goffeau A, Barrell B, Bussey H, Davis R, Dujon B, Feldmann H, Galibert F, Hoheisel J, Jacq C, Johnston M, Louis E, Mewes H, Murakami Y, Philippsen P, Tettelin H, Oliver S (1996) Life with 6000 genes. Science 274:546, 563–567

Goremykin V, Nikiforova S, Biggs P, Zhong B, Delange P, Martin W, Woetzel S, Atherton R, McLenachan P, Lockhart P (2013) The evolutionary root of flowering plants. Syst Biol 62:50–61

Gruissem W, Zurawski G (1985) Identification and mutational analysis of the promoter for a spinach chloroplast transfer RNA gene. EMBO J 4:1637–1644

Hajdukiewicz P, Allison L, Maliga P (1997) The two RNA polymerases encoded by the nuclear and the plastid compartments transcribe distinct groups of genes in tobacco plastids. EMBO J 16:4041–4048

Hammani K, Cook W, Barkan A (2012) RNA binding and RNA remodeling activities of the half-a-tetratricopeptide (HAT) protein HCF107 underlie its effects on gene expression. Proc Natl Acad Sci USA 109:5651–5656

Hanson M, Gray B, Ahner B (2012) Chloroplast transformation for engineering of photosynthesis. J Exp Bot 64:731. doi:10.1093/jxb/ers325

Hodgkin A, Huxley A (1952) A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117:500–544

Hotto A, Schmitz R, Fei Z, Ecker J, Stern D (2011) Unexpected diversity of chloroplast noncoding RNAs as revealed by deep sequencing of the *Arabidopsis* transcriptome. G3 1:559–570

Hotto A, Germain A, Stern D (2012) Plastid non-coding RNAs: emerging candidates for gene regulation. Trends Plant Sci 17:737–744

Hou B, Zhou Y, Wan L, Zhang Z, Shen G, Chen Z, Hu Z (2003) Chloroplast transformation in oilseed rape. Transgenic Res 12:111–114

Ishizaki Y, Tsunoyama Y, Hatano K, Ando K, Kato K, Shinmyo A, Kobori M, Takeba G, Nakahira Y, Shiina T (2005) A nuclear-encoded sigma factor Arabidopsis SIG6, recognizes sigma-70 type chloroplast promoters and regulates early chloroplast development in cotyledons. Plant J 42:133–144

Jin S, Zhang X, Daniell H (2012) Pinellia ternata agglutinin expression in chloroplasts confers broad spectrum resistance against aphid, whitefly, Lepidopteran insects, bacterial and viral pathogens. Plant Biotechnol J 10:313–327

Kamal A, Cho K, Komatsu S, Uozumi N, Choi J, Woo S (2012) Towards an understanding of wheat chloroplasts: a methodical investigation of thylakoid proteome. Mol Biol Rep 39:5069–5083

Kannangara C, Henningsen K, Stumpf P, Appelqvist L, von Wettstein D (1971) Lipid biosynthesis by isolated barley chloroplasts in relation to plastid development. Plant Physiol 48:526–531

Keeling P (2010) The endosymbiotic origin, diversification and fate of plastids. Philos Trans R Soc Biol Sci 365:729–748

Kirk P, Leech R (1972) Amino acid biosynthesis by isolated chloroplasts during photosynthesis. Plant Physiol 50:228–234

Kleffmann T, von Zychlinski A, Russenberger D, Hirsch-Hoffmann M, Gehrig P, Gruissem W, Baginsky S (2007) Proteome dynamics during plastid differentiation in Rice. Plant Physiol 143:912–923

Klie S, Krueger S, Krall L, Giavalisco P, Flügge U, Willmitzer L, Steinhauser D (2011) Analysis of the compartmentalized metabolome – a validation of the non-aqueous fractionation technique. Front Plant Sci 2:55

Kogel K, Voll L, Schäfer P, Jansen C, Wu Y, Langen G, Imani J, Hofmann J, Schimiedl A, Sonnewald S, Wettstein D, Cook R, Soonewald U (2010) Transcriptome and metabolome profiling of field-grown transgenic barley lack induced differences but show cultivar-specific variances. Proc Natl Acad Sci USA 107:6198–6203

Kolodner R, Tewari K (1972) Molecular size and conformation of chloroplast deoxyribonucleic acid from pea leaves. J Biol Chem 247:6355–6364

Kota M, Daniell H, Varma S, Garczynski S, Gould F, Moar W (1999) Overexpression of the Bacillus thuringiensis (Bt) Cry2Aa2 protein in chloroplasts confers resistance to plants against susceptible and Bt-resistant insects. Proc Natl Acad Sci USA 96:1840–1845

Koussevitzky S, Nott A, Mockler T, Hong F, Sachetto-Martins G, Surpin M, Lim J, Mittler R, Chory J (2007) Signals from chloroplasts converge to regulate nuclear gene expression. Science 316:715–719

Krueger S, Giavalisco P, Krall L, Steinhauser M, Bussis D, Usadel B, Flugge U, Fernie A, Willmitzer L, Steinhauser D (2011) A topological map of the compartmentalized *Arabidopsis thaliana* leaf metabolome. PLoS One 6:e17806

Kumar S, Dhingra A, Daniell H (2004) Plastid-expressed betaine aldehyde dehydrogenase gene in carrot cultured cells, roots, and leaves confers enhanced salt tolerance. Plant Physiol 136:2843–2854

Kusano M, Redestig H, Hirai T, Oikawa A, Matsuda F, Fukushima A, Arita M, Watanabe S, Yano M, Hiwasa-Tanase K, Ezura H, Saito K (2011) Covering chemical diversity of genetically-modified tomatoes using metabolomics for objective substantial equivalence assessment. PLoS One 6:e16989

Lawo N, Weingart G, Schuhmacher R, Forneck A (2011) The volatile metabolome of grapevine roots: first insights into the metabolic response upon *Phylloxera* attack. Plant Physiol Biochem 49:1059–1063

Lee S, Kwon H, Kwon S, Park S, Jeong M, Han S, Byun M, Daniell H (2003) Accumulation of trehalose within transgenic chloroplasts confers drought tolerance. Mol Breed 11:1–13

Leiss K, Maltese F, Choi Y, Verpoorte R, Klinkhamer P (2009) Identification of chlorogenic acid as a resistance factor for thrips in *Chrysanthemum*. Plant Physiol 150:1567–1575

Lichtenthaler H (1999) The 1-deoxy-d-xylulose-5-phosphate pathway of isoprenoid biosynthesis in plants. Plant Mol Biol 50:47–65

Link G (2003) Redox regulation of chloroplast transcription. Antioxid Redox Signal 5:79–88

Loschelder H, Schweer J, Link B, Link G (2006) Dual temporal role of plastid sigma factor 6 in Arabidopsis development. Plant Physiol 142:642–650

Luo M, Dang P, Bausher M, Holbrook C, Lee R, Lynch R, Guo B (2005) Identification of transcripts involved in resistance responses to leaf spot disease caused by *Cercosporidium personatum* in peanut (*Arachis hypogea*). Phytopathology 95:381–387

Majeran W, Friso G, Asakura Y, Qu X, Huang M, Ponnala L, Watkins K, Barkan A, van Wijk K (2012) Nucleoid-enriched proteomes in developing plastids and chloroplasts from maize leaves: a new conceptual framework for nucleoid functions. Plant Physiol 158:156–189

McBride K, Svab Z, Schaaf D, Hogan P, Stalker D, Maliga P (1995) Amplification of a chimeric Bacillus gene in chloroplasts leads to an extraordinary level of an insecticidal protein in tobacco. Biotechnology (N Y) 13:362–365

Moy P, Qutob D, Chapman B, Atkinson I, Gijzen M (2004) Patterns of gene expression upon infection of soybean plants by Phytophthora sojae. Mol Plant Microbe Interact 17:1051–1062

Ohyama K, Fukuzawa H, Kohchi T, Shirai H, Sano T, Sano S, Umesono K, Shiki Y, Takeuchi M, Chang Z, Aota S, Inokuchi H, Ozeki H (1986) Chloroplast gene organization deduced from complete sequence of liverwort *Marchantia polymorpha* chloroplast DNA. Nature 322:572–574

Okuda K, Nakamura T, Sugita M, Shimizu T, Shikanai T (2006) A pentatricopeptide repeat protein is a site recognition factor in chloroplast RNA editing. J Biol Chem 281:37661–37667

Oldenburg D, Bendich A (2004) Most chloroplast DNA of maize seedlings in linear molecules with defined ends and branched forms. J Mol Biol 335:953–970

Palmer J (1991) Plastid chromosomes: structure and evolution. In: Hermann RG (ed) The molecular biology of plastids. Cell culture and somatic cell genetics of plants vil7A. Springer, Vienna, pp 5–53

Parry M, Andralojc P, Scales J, Salvucci M, Carmo-Silva E, Alonso H, Whitney S (2012) Rubisco activity and regulation as targets for crop improvement. J Exp Bot 64:717. doi:10.1093/jxb/ers336

Paul M, Pellny T (2003) Carbon metabolite feedback regulation of leaf photosynthesis and development. J Exp Bot 54:539–547

Pineda M, Sajnani C, Barón M (2010) Changes induced by the *Pepper mild mottle tobamovirus* on the chloroplast proteome of *Nicotiana benthamiana*. Photosynth Res 103:31–45

Quesada-Vargas T, Ruiz O, Daniell H (2005) Characterization of heterologous multigene operons in transgenic chloroplasts. Transcription, processing and translation. Plant Physiol 138:1746–1762

Rabbani M, Muruyama K, Abe H, Khan M, Katsura K, Ito Y, Yoshiwara K, Seki M, Shinozaki K, Yamaguchi-Shinozaki K (2003) Monitoring expression profiles of rice genes under cold, drought, and high-salinity stress and Abscisic acid application using cDNA microarray and RNA gel-blot analyses. Plant Physiol 133:1755–1767

Reiland S, Messerli G, Baerenfaller K, Gerrits B, Endler A, Grossmann J, Gruissem W, Baginsky S (2009) Large-scale Arabidopsis phosphoproteome profiling reveals novel chloroplast kinase substrates and phosphorylation networks. Plant Physiol 150:889–903

Restrepo S, Myers K, Pozo O, Martin G, Hart A, Buell C, Fry W, Smart C (2005) Gene profiling of a compatible interaction between Phytophthora infestans and *Solanum tuberosum* suggests a role for carbonic anhydrase. Mol Plant Microbe Interact 18:913–922

Riens B, Lohaus G, Heineke D, Heldt H (1991) Amino acid and sucrose content determined in the cytosolic, chloroplastic, and vacuolar compartments and in the phloem sap of spinach leaves. Plant Physiol 97:227–233

Rowan B, Oldenburg D, Bendich A (2004). Curr Genet 46(3):176–181

Ruf S, Hermann M, Berger I, Carrer H, Bock R (2001) Stable genetic transformation of tomato plastids and expression of a foreign protein in fruit. Nat Biotechnol 19:870–875

Ruwe H, Schmitz-Linneweber C (2012) Short non-coding RNA fragments accumulating in chloroplasts: footprints of RNA binding proteins? Nucleic Acids Res 40:3106–3116

Ruwe H, Castandet B, Schmitz-Linneweber C, Stern D (2013) Arabidopsis chloroplast quantitative editotype. FEBS Lett 587:1429–1433

Sanger F, Air G, Barrell B, Brown N, Coulson A, Fiddes C, Hutchison C, Slocombe P, Smith M (1977) Nucleotide sequence of bacteriophage phi X174 DNA. Nature 265:687–695

Saski C, Lee S, Daniell H, Wood T, Tomkins J, Kim H, Jansen R (2005) Complete chloroplast genome sequence of Gycine max and comparative analyses with other legume genomes. Plant Mol Biol 59:309–322

Saski C, Lee S, Fjellheim S, Guda C, Jansen R, Luo H, Tomkins J, Rognli O, Daniell H, Clarke J (2007) Complete chloroplast genome sequences of *Hordeum vulgare*, *Sorghum bicolor* and *Agrostis stolonifera*, and comparative analyses with other grass genomes. Theor Appl Genet 115:571–590

Schauer N, Steinhauser D, Strelkov S, Schomburg D, Allison G, Moritz T, Lundgren K, Roessner-Tunali U, Forbes M, Willmitzer L, Fernie A, Kopka J (2005) GC-MS libraries for the rapid identification of metabolites in complex biological samples. FEBS Lett 579:1332–1337

Scheideler M, Schlaich N, Fellenberg K, Beissbarth T, Hauser N, Vingron M, Slusarenko A, Hoheisel J (2002) Monitoring the switch from housekeeping to pathogen defence metabolism in *Arabidopsis thaliana* using cDNA Arrays. J Biol Chem 277:10555–10561

Schmitz-Linneweber C, Barkan A (2007) RNA splicing and RNA editing in chloroplasts. In: Bock R (ed) Cell and molecular biology of plastids. Springer, Berlin/Heidelberg, pp 213–248

Schweer J, Loschelder H, Link G (2006) A promoter switch that can rescue a plant sigma factor mutant. FEBS Lett 580:6617–6622

Schweer J, Türkeri H, Kolpack A, Link G (2010) Role and regulation of plastid sigma factors and their functional interactors during chloroplast transcription – recent lessons from *Arabidopsis thaliana*. Eur J Cell Biol 89:940–946

Shaw J, Lickey E, Schilling E, Small R (2007) Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. Am J Bot 94:275–288

Shinozaki K, Ohme M, Tanaka M, Wakasugi T, Hayashida N, Matsubayashi T, Zaita N, Chunwongse J, Obokata J, Yamaguchi-Shinozaki K, Ohto C, Torazawa K, Meng B, Sugita M, Deno H, Kamogashira T, Yamada K, Kusuda J, Takaiwa F, Kato A, Tohdoh N, Shimada H, Sugiura M (1986) The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. EMBO J 5:2043–2049

Siddell S, Ellis R (1975) Characteristics and products of protein synthesis in vitro in etioplasts and developing chloroplasts from pea leaves. Biochem J 146:675–685

Sidorov V, Kasten D, Pang S, Hajdukiewicz P, Staub J, Nehra N (1999) Stable chloroplast transformation in potato: use of green fluorescent protein as a plastid marker. Plant J 19:209–216

Soga T, Ohashi Y, Ueno Y (2003) Quantitative metabolome analysis using capillary electrophoresis mass spectrometry. J Proteome Res 2:488–494

Stitt M, Schulze E (1994) Does Rubisco control the rate of photosynthesis and plant growth? An exercise in molecular ecophysiology. Plant Cell Environ 17:465–487

Stitt M, Wirtz W, Heldt H (1983) Regulation of sucrose synthesis by cytoplasmic fructose bisphosphatase and sucrose phosphate synthase during photosynthesis in varying light and carbon dioxide. Plant Physiol 72:767–774

Sugita M, Sugiura M (1996) Regulation of gene expression in chloroplasts of higher plants. Plant Mol Biol 32:315–326

Svab Z, Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. Proc Natl Acad Sci USA 90:913–917

Svab Z, Hajdukiewicz P, Maliga P (1990) Stable transformation of plastids in higher plants. Proc Natl Acad Sci USA 87:8526–8530

Swiatecka-Hagenbruch M, Liere K, Börner T (2007) High diversity of plastidial promoters in *Arabidopsis thaliana*. Mol Genet Genomics 277:725–734

Swiatecka-Hagenbruch M, Emanuel C, Hedtke B, Liere K, Börner T (2008) Impaired function of the phage-type RNA polymerase RpoTp in transcription of chloroplasts genes is compensated by a second phage-type RNA polymerase. Nucleic Acids Res 36:785–792

Udy D, Belcher S, Williams-Carrier R, Gualberto J, Barkan A (2012) Effects of reduced chloroplast gene copy number on chloroplast gene expression in maize. Plant Physiol 160:1420–1431

Uemura M, Steponkus P (1997) Effect of cold acclimation on the lipid composition of the inner and outer membrane of the chloroplast envelope isolated from rye leaves. Plant Physiol 114:1493–1500

Vaistij F, Goldschmidt-Clermont M, Wostrikoff K, Rochaix J (2000) Stability determinants in the chloroplast psbB/T/H mRNAs of *Chlamydomonas reinhardtii*. Plant J 21:469–482

Verma D, Daniell H (2007) Chloroplast vector systems for biotechnology applications. Plant Physiol 145:1129–1143

Wang L, Yu X, Wang H, Lu Y, Ruiter M, Prins M, He Y (2011) A novel class of heat-responsive small RNAs derive from the chloroplast genome of Chinese cabbage (*Brassica rapa*). BMC Genomics 12:289–303

Watkins K, Rojas M, Friso G, van Wijk K, Meurer J, Barkan A (2011) APO1 promotes the splicing of chloroplast group II introns and harbors a plant-specific zinc-dependent RNA binding domain. Plant Cell 23:1082–1092

Weber A, Fischer K (2007) Making the connections – the crucial role of metabolite transporters at the interface between chloroplast and cytosol. FEBS Lett 581:2215–2222

Wu F, Kan P, Lee B, Daniell H, Lee W, Lin C, Lin S, Lin S (2009) Complete nucleotide sequence of Dendrocalamus latiflorus and Bambusa oldhamii chloroplast genomes. Tree Physiol 29:847–856

Wu C, Chaw S, Huang Y (2013) Chloroplast phylogenomics indicates that Ginkgo biloba is sister to cycads. Genome Biol Evol 5:243–254

Yabuta Y, Tanaka H, Yoshimura S, Suzuki A, Tamoi M, Maruta T, Shigeoka S (2012) Improvement of vitamin E quality and quantity in tobacco and lettuce by chloroplast genetic engineering. Transgenic Res. doi:10.1007/s11248-012-9656-5

Yamori W, Masumoto C, Fukayama H, Makino A (2012) Rubisco activase is a key regulator of non-steady-state photosynthesis at any leaf temperature and, to a lesser extent, of steady-state photosynthesis at high temperature. Plant J 71:870–880

Yonekura-Sakakibara K, Tohge T, Matsuda F, Nakabayashi R, Takayama H, Niida R, Watanabe-Takahashi A, Inoue E, Saito K (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlation in *Arabidopsis*. Plant Cell 20:2160–2176

Yoon H, Hackett J, Ciniglia C, Pinto G, Bhattacharya D (2004) A molecular timeline for the origin of photosynthetic eukaryotes. Mol Biol Evol 21:809–818

Yu Q, Li G, Wang G, Sun J, Wang P, Wang C, Yang Z (2008) Construction of a chloroplast protein interaction network and functional mining of photosynthetic proteins in *Arabidopsis thaliana*. Cell Res 18:1007–1019

Zhao J, Wang J, An L, Doerge R, Chen Z, Grau C, Meng J, Osborn T (2007) Analysis of gene expression profiles in response to Sclerotinia sclerotiorum in Brassica napus. Planta 227:13–24

Zhelyazkova P, Hammani K, Rojas M, Voelker R, Vargas-Suarez M, Borner T, Barkan A (2012a) Protein-mediated protection as the predominant mechanism for defining processed mRNA termini in land plant chloroplasts. Nucleic Acids Res 40:3092–3105

Zhelyazkova P, Sharma C, Förstner K, Liere K, Vogel J, Börner T (2012b) The primary transcriptome of barley chloroplasts: numerous noncoding RNAs and the dominating role of the Plastid-Encoded RNA polymerase. Plant Cell 24:123–136

Zierold U, Scholz U, Schweizer P (2005) Transcriptome analysis of mlo-mediated resistance in the epidermis of barley. Mol Plant Pathol 6:139–151

Zhou W, Eudes F, Laroche A (2006) Identification of differentially regulated proteins in response to a compatible interaction between the pathogen *Fusarium graminearum* and its host, *Triticum aestivum*. Proteomics 6:4599–4609

Zhou Y, Xu M, Zhao M, Xie X, Zhu L, Fu B, Li Z (2010) Genome-wide gene responses in a transgenic rice line carrying the maize resistance gene Rxo1 to the rice bacterial streak pathogen, Xanthomonas oryzae pv. oryzicola. BMC Genomics 11:78–88

Zhu G, Kurek I, Liu L (2010) Engineering photosynthetic enzymes involved in CO2-assimilation by gene shuffling. In: Rebeiz CA, Benning C, Bohnert HJ, Daniell H, Hoober JK, Lichtenthaler HK, Portis AR Jr, Tripathy BC (eds) The chloroplast: basics and application, Advances in photosynthesis and respiration. Springer, Dordrecht, pp 307–322

frontiers in Plant Science

ORIGINAL RESEARCH
published: 07 April 2017
doi: 10.3389/fpls.2017.00497

OPEN ACCESS

Edited by:
Dominique Job,
Centre National de la Recherche Scientifique, France

Reviewed by:
Julia Buitink,
National Institute for Agricultural Research, France
Tiago Santana Balbuena,
São Paulo State University, Brazil

***Correspondence:**
Ana P. Barba de la Rosa
apbarba@ipicyt.edu.mx

†Present address:
Alma L. Saucedo, Consejo Nacional de Ciencia y Tecnología, Laboratorio de RMN, Departamento de Química Analítica Facultad de Medicina, Universidad Autónoma de Nuevo León, Av. Madero y Av. Gonzalitos s/n, Colonia Mitras Centro, Monterrey, México; Eric E. Hernández-Domínguez, Consejo Nacional de Ciencia y Tecnología-Instituto de Ecología, A.C. Carretera antigua a Coatepec 351, El Haya, Xalapa de Enríquez, México

Specialty section:
This article was submitted to Plant Proteomics, a section of the journal Frontiers in Plant Science

Received: *22 December 2016*
Accepted: *22 March 2017*
Published: *07 April 2017*

Citation:
Saucedo AL, Hernández-Domínguez EE, de Luna-Valdez LA, Guevara-García AA, Escobedo-Moratilla A, Bojorquéz-Velázquez E, del Río-Portilla F, Fernández-Velasco DA and Barba de la Rosa AP (2017) Insights on Structure and Function of a Late Embryogenesis Abundant Protein from Amaranthus cruentus: An Intrinsically Disordered Protein Involved in Protection against Desiccation, Oxidant Conditions, and Osmotic Stress. Front. Plant Sci. 8:497. doi: 10.3389/fpls.2017.00497

Insights on Structure and Function of a Late Embryogenesis Abundant Protein from *Amaranthus cruentus*: An Intrinsically Disordered Protein Involved in Protection against Desiccation, Oxidant Conditions, and Osmotic Stress

Alma L. Saucedo[1†], Eric E. Hernández-Domínguez[1†], Luis A. de Luna-Valdez[2], Angel A. Guevara-García[2], Abraham Escobedo-Moratilla[1], Esaú Bojorquéz-Velázquez[1], Federico del Río-Portilla[3], Daniel A. Fernández-Velasco[4] and Ana P. Barba de la Rosa[1]*

[1] Department of Molecular Biology, Instituto Potosino de Investigación Científica y Tecnológica, A.C., San Luis Potosí, México, [2] Instituto de Biotecnología, Universidad Nacional Autónoma de México, Cuernavaca, México, [3] Instituto de Química, Universidad Nacional Autónoma de México, Ciudad de México, México, [4] Laboratorio de Fisicoquímica e Ingeniería de Proteínas, Departamento de Bioquímica, Facultad de Medicina, Universidad Nacional Autónoma de México, Ciudad de México, México

Late embryogenesis abundant (LEA) proteins are part of a large protein family that protect other proteins from aggregation due to desiccation or osmotic stresses. Recently, the *Amaranthus cruentus* seed proteome was characterized by 2D-PAGE and one highly accumulated protein spot was identified as a LEA protein and was named AcLEA. In this work, *AcLEA* cDNA was cloned into an expression vector and the recombinant protein was purified and characterized. *AcLEA* encodes a 172 amino acid polypeptide with a predicted molecular mass of 18.34 kDa and estimated p*I* of 8.58. Phylogenetic analysis revealed that AcLEA is evolutionarily close to the LEA3 group. Structural characteristics were revealed by nuclear magnetic resonance and circular dichroism methods. We have shown that recombinant AcLEA is an intrinsically disordered protein in solution even at high salinity and osmotic pressures, but it has a strong tendency to take a secondary structure, mainly folded as α-helix, when an inductive additive is present. Recombinant AcLEA function was evaluated using *Escherichia coli* as *in vivo* model showing the important protection role against desiccation, oxidant conditions, and osmotic stress. AcLEA recombinant protein was localized in cytoplasm of *Nicotiana benthamiana* protoplasts and orthologs were detected in seeds of wild and domesticated amaranth species. Interestingly AcLEA was detected in leaves, stems, and roots but only in plants subjected to salt stress. This fact could indicate the important role of AcLEA protection during plant stress in all amaranth species studied.

Keywords: amaranth seeds, circular dichroism, intrinsically disordered proteins (IDP), late embryogenesis abundant (LEA) proteins, nuclear magnetic resonance, Western blot

INTRODUCTION

Seeds can withstand the loss of cellular water during the maturation phase of their development by the accumulation of high levels of ubiquitous proteins named late embryogenesis abundant (LEA) proteins (Ali-Benali et al., 2005; Dalal et al., 2009; Liu et al., 2013; Avelange-Macherel et al., 2015). LEA proteins were originally discovered in cotton (*Gossypium hirsutum*) seeds (Dure, 1989), but their accumulation is not only related to the development of desiccation tolerance in orthodox seeds (desiccation-tolerant seeds). LEA proteins are also induced upon water-related stress in plant vegetative tissues and in other anhydrobiotic organisms such as eubacteria, rotifers, nematodes, tardigrades, arthropods (Ingram and Bartels, 1996; Browne et al., 2002; Hundertmark and Hincha, 2008; Campos et al., 2013; Hatanaka et al., 2014; van Leeuwen et al., 2016). In some microorganisms, LEA proteins are reported in response to water limitation, which suggests that they have an important role in desiccation tolerance (Tunnacliffe and Wise, 2007; Tunnacliffe et al., 2010; Hand et al., 2011). In spite of their widely recognized importance for desiccation tolerance, the molecular function of LEA proteins is only starting to emerge, with a variety of functions in agreement with their diversity (Battaglia and Covarrubias, 2013).

The distinctive features of LEA proteins are their high hydrophilicity due to a high percentage of charged amino acids such as alanine, serine/threonine and the absence or very low content of non-polar amino acids (tryptophan and cysteine). The presence of repeated motifs, which tend to form secondary structures, has detected in LEA proteins (Dure, 1989; Garay-Arroyo et al., 2000; Tunnacliffe and Wise, 2007). Although LEA proteins are intrinsically disordered proteins (IDP) in aqueous solutions (Wolkers et al., 2001; Goyal et al., 2003; Boucher et al., 2010; Tompa and Kovacs, 2010; Popova et al., 2011), they may acquire some structure folding into α-helical conformations during partial or complete dehydration (Shih et al., 2004; Tolleter et al., 2007; Hincha and Thalhammer, 2012).

Several hundreds of LEA protein sequences have been gathered in a dedicated database[1] and bioinformatics analyses have shown that each LEA class can be clearly characterized by a unique set of physico-chemical properties. This has led to the classification of LEA proteins into 12 non-overlapping classes with distinct properties (Battaglia et al., 2008; Hunault and Jaspard, 2010; Jaspard et al., 2012).

Although quite a few LEAs have been characterized, the functions of most members of the LEA family remain unknown (Cao and Li, 2015). Transgenic *Arabidopsis thaliana* plants overexpressing the *Nicotiana tabacum NtLEA7-3* gene are much more resistant to cold, drought, and salt stresses (Gai et al., 2011). Tomato LEA25 increases the salt and chilling stress tolerance when overexpressed in yeast (Imai et al., 1996). Wheat and rice over-expressing *HVA*1 gene (encoding an LEA protein from barley) are more tolerant to drought and salt stress (Xu et al., 1996; Sivamani et al., 2000). Olvera-Carrillo et al. (2010) reported that in *A. thaliana*, the accumulation of AtLEA4 protein leads to a drought tolerant phenotype. The overexpression of *BnLEA*4-1 from *Brassica napus* in *Escherichia coli* can enhance bacterial cellular tolerance to temperature and salt stresses (Dalal et al., 2009).

On the other hand, LEA proteins have a broad subcellular distribution; they are present in cytosol, mitochondria, chloroplasts, endoplasmic reticulum, and nucleus (Candat et al., 2014) and the specific modes of their action could be related to their intracellular location. The biological activity of these proteins seems to be associated with the stabilization of membranes during cell drying (Tolleter et al., 2010), and assistance of the transport of proteins during stress conditions (Chakrabortee et al., 2010).

Amaranth, a member of *Amaranthaceae* family, is a plant that has been cultivated and used since ancient times by Mexican and Central American civilizations. In the last decades, the nutritional role of amaranth seeds from different species has been revalued, particularly for *A. hypochondriacus* and *A. cruentus,* not only because of their high protein content and their contribution of essential amino acids, like lysine and methionine (compared to other grains), but also for their antioxidant compounds (Becker et al., 1981; Rastogi and Shukla, 2013), and bioactive peptides (Silva-Sánchez et al., 2008). Current interest in amaranth plants is also related to their extraordinary adaptability to grow in adverse weather conditions (Brenner et al., 2000). Amaranth is resistant against several types of stresses such as pest (Valdes-Rodríguez et al., 2007), heat (Maughan et al., 2009), drought (Huerta-Ocampo et al., 2011), and salinity (Aguilar-Hernández et al., 2011; Huerta-Ocampo et al., 2014). The recent report of *Amaranthus cruentus* seed proteome by 2D-PAGE revealed the over-accumulation of one spot identified as a LEA protein (Maldonado-Cervantes et al., 2014). In the present study, we have cloned the corresponding *LEA* cDNA from *A. cruentus* (*AcLEA*, GenBank accession no. KX852451), and the recombinant protein was expressed in *E. coli*. Nuclear Magnetic Resonance (NMR) and Circular Dichroism (CD) were used as tools to study the structural characteristics of this particular AcLEA protein. Its functional activity was evaluated *in vivo* using *E. coli* as model.

According to its amino acid sequence, AcLEA protein belongs to the Group 3, its hydrophilic nature and spectroscopic characteristics being *ad hoc* with IDP molecules, but exhibiting a high content of α-helix in the presence of trifluoroethanol (TFE). Overexpression of *AcLEA* in *E. coli* conferred resistance to desiccation, osmotic and oxidative stress to the bacterial cells. When accumulated in a heterologous system (*Nicotiana benthamiana* protoplasts) the amaranth protein was found to be distributed in the cytoplasm of protoplasts. Western blot analyses disclosed that AcLEA protein accumulated in seeds of wild and domesticated amaranth species. Accumulation of AcLEA in leaves, stems, and roots was observed only in plants subjected to salinity stress.

[1] http://forge.info.univ-angers.fr/~gh/Leadb

MATERIALS AND METHODS

RNA Extraction and Cloning of the cDNA Encoding *AcLEA*

Immature seeds (15 days after anthesis) of *Amaranthus cruentus* were used to extract total RNA with TRIzol Reagent (Invitrogen, Carlsbad, CA, USA) and cDNA was synthesized as previously reported (Maldonado-Cervantes et al., 2014). *AcLEA* cDNA was amplified using specific primers containing *Nde*I (5′-CATATGGCATCACATGGTCAGAGT-3′) and *Xho*I (5′-CTCGAGCTAGGGCCTAGTAGTCTTAATTGGATC-3′) restriction sites. cDNA amplification was performed using Platinum Taq DNA polymerase (Invitrogen), under standard reaction conditions. The amplified PCR product was cloned into the plasmid pGEM-T-Easy (Promega Corp., Madison, WI, USA). *AcLEA* cDNA was excised from pGEM using *Nde*I and *Xho*I (New England Biolabs, Ipswich, MA, USA) restriction enzymes. Digested fragments were purified and subcloned into pET28 expression vector restricted with *Nde*I-*Xho*I (Novagen-Merck, Darmstadt, Germany) containing the His-Tag at N-terminal. Vector was modified to have a recognition cleavage site within the amino acid sequence LeuGluValLeuPheGln/GlyPro specific for human rhinovirus 3C protease as PreScission Protease (PSP), and ending with the pET28mod vector. The resulting plasmid pET28mod-*AcLEA* was sequenced in both directions to confirm the *AcLEA* cDNA identity.

Alternatively, the *AcLEA* cDNA was PCR flanked with attB1 and attB2 recombination sites for generation of an entry clone using the gateway system entry vector pDONR-Zeo (Karimi et al., 2007), which was later used to generate the expression vector pEarlyGate 103-*AcLEA*.

Physicochemical Properties and Phylogenetic Analyses

Protein hydrophilicity analysis was performed to obtain the hydropathy plots with the Kyte and Doolittle (1982) values from the Expasy ProtScale Tool[2] (Gasteiger et al., 2005). Grand average of hydropathicity (GRAVY) and instability index were calculated using the ProtParam software[3]. Sequence similarities were determined using the BLAST program and the GenBank database on the NCBI web server. MUSCLE 3.8.31 (Edgar, 2004) was used to perform multiple sequence alignments of full-length protein sequences. The phylogenetic analyses of the LEA proteins based on amino acid sequences were carried out using the neighbor-joining method (Saitou and Nei, 1987). AcLEA protein classification was done comparing its sequence to those available in the LEA Proteins Data Base[4] (Hunault and Jaspard, 2010) and the Pfam server[5] (Finn et al., 2015).

AcLEA related protein sequences were retrieved from the recently reported genome of *Amaranthus hypochondriacus* (Clouse et al., 2016) deposited at Phytozome *v*12.0[6].

[2]http://web.expasy.org/protscale/

[3]http://web.expasy.org/protparam/

[4]http://forge.info.univ-angers.fr/~gh/Leadb/index.php

[5]http://pfam.xfam.org/

[6]http://www.phytozome.net

Expression and Purification of the Recombinant AcLEA Protein

Recombinant AcLEA protein (rAcLEA) was up-accumulated in BL21 (DE3) *E. coli* cells (Novagen) transformed with the expression vector pET28mod-*AcLEA*. LB media supplemented with kanamycin was used to grow cells at 37°C. Overnight cultures were diluted 100-fold using fresh LB medium, and incubation was continued until optical density (OD_{600}) reached 0.5–0.6. At this point, 0.1 mM isopropyl thio-β-D-galacto-pyranoside (IPTG, Sigma–Aldrich, St. Louis, MO, USA) was added to induce the protein expression. After further 4 h of incubation at 28°C, cells were harvested by centrifugation at 3,000 × *g* for 15 min at 4°C.

For structural studies cell pellets were resuspended in native buffer (150 mM NaCl, 50 mM Tris-HCl, pH 8) and for antibodies production cells pellets were resuspended in denaturing lysis buffer (500 mM NaCl, 6 M guanidine hydrochloride, 20 mM sodium phosphate, pH 7.8). Resuspended pellets were sonicated for 45 s (Misonix Sonicator 3000, Cole-Parmer, Vernon Hills, IL, USA) in ice bath. Antibodies were obtained as described in Supplementary Information. The soluble fraction was separated by centrifugation at 20,000 × *g* for 30 min at 4°C. Recombinant six-His-tagged AcLEA (rHis-AcLEA, 20.7 kDa) was purified by metal-chelate affinity chromatography (IMAC) using the Ni-NTA agarose purification system (Novex, Thermo Fischer Scientific Inc., Waltham, MA, USA), and eluted with five volumes of native (150 mM NaCl, 50 mM Tris-HCl, pH 8.0) or denaturing elution buffer (500 mM NaCl, 8 M urea, 20 mM sodium phosphate, pH 4.0). In both native and denaturing purifications, buffer exchange to 150 mM NaCl, 50 mM Tris-HCl, pH 8.0, was performed by dialysis using a 5 kDa cut-off membrane (Merck Millipore, Billerica, MA, USA), and cleavage of His-Tag was carried out overnight at 4°C. After cleavage, a second step of IMAC purification was carried out under native conditions (150 mM NaCl, 50 mM Tris-HCl, pH 8.0) in order to obtain native rAcLEA. Since rAcLEA was found to be weakly bounded to the resin, a native buffer containing 20 mM imidazole was used for protein elution. Finally, PD10 desalting columns (GE Healthcare, Piscataway, NJ, USA) were used to remove buffer components. For NMR spectroscopic and CD analyses, an additional purification step of rAcLEA was done using FPLC chromatography with a Sephacryl S-100 column (GE Healthcare) with a mobile phase of 10 mM sodium phosphate pH 7.0 (Sigma–Aldrich). All rAcLEA purification steps were followed by 12% SDS-PAGE gels stained with Coomassie Blue.

Recombinant proteins, excised from gel and/or in solution after chromatography purification were reduced with 10 mM DTT followed by protein alkylation with 55 mM iodoacetamide, and finally digested with trypsin (Promega, Madison, WI, USA) in an overnight reaction at 37°C. MS was carried out with a SYNAPT-HDMS (Waters Corp.) coupled to a nano-ACQUITY-UPLC system as described in Supplementary Information.

Structural and Functional Characteristics of AcLEA

NMR and CD Analyses

Lyophilized rAcLEA purified under native conditions was dissolved in H_2O/D_2O (95:5) to prepare a solution at final concentration of 1 mM, and transferred to a 3 mm tube. For ^{1}H-NMR water suppression signal was performed using the double-pulsed field gradient spin echo sequence (DPFGSE). Fourier transformation was applied to FID file and data were analyzed with the NUTS Data Processing Software (Acorn NMR Inc., Livermore, CA, USA). Proton nuclear magnetic resonance spectra (^{1}H-NMR) were acquired on a 500 MHz Varian Innova spectrometer (Varian, Palo Alto, CA, USA) at 298 K.

Circular dichroism spectra were recorded on a Chirascan Circular Dichroism Spectrometer (Applied Photophysics, Leatherhead, UK), equipped with a Peltier cell holder for control of temperature. A stock solution of rAcLEA protein (0.4 mg/ml) was prepared in 10 mM phosphate buffer pH 8.0. Far UV CD spectra were obtained using a quartz cell with a light path of 1 mm in the 200-260 nm range with a bandwidth of 1.0 nm and a digital resolution of 0.5 s per point. Temperature-induced conformational changes were simultaneously recorded at 210, 222 and 230 nm using a heating rate of 1°C/min in the 20 to 70°C range. After the heating ramp, the sample was cooled to 20°C then far UV CD spectra was taken to determine the reversibility of the conformational changes. CD spectra in the near UV region covering the 250-350 nm range were recorded using a quartz cell with a 10 mm path length, bandwidth of 2.0 nm and 1.0 s time per point. Molar ellipticity values, [θ], were calculated from measured θ using the equation:

$$[\theta]\,\mathrm{deg\,cm^2 dmol^{-1}} = \theta \cdot \mathrm{M} \cdot 100/\mathrm{c} \cdot l$$

where θ is the measured ellipticity in degrees, M is the protein molecular weight, c is the protein concentration in mg/ml, and *l* is the path length. Estimation of secondary structure was performed using the CDNN algorithm (Bohm et al., 1992) using a spectral window data from 200 to 260 nm. Five spectra were recorded for each experimental condition.

Assay of Protective Role of AcLEA in *E. coli*

Transformed *E. coli* BL21 cells carrying the plasmid pET28mod-*AcLEA* and the empty plasmid pET28mod (control) were grown in LB liquid medium supplemented with 37 μg/ml kanamycin overnight at 37°C. For both bacterial cultures, an aliquot was diluted 100-fold using fresh liquid LB with antibiotic and allowed to grow for 2–3 h at 37°C. When OD_{600} reached 0.5–0.6, IPTG was added to a final concentration of 0.1 mM and cultures were kept at 28°C for 2 h, for rAcLEA protein induction. At this point stress treatments were analyzed. To test the function of AcLEA protein to prevent desiccation stress, *E. coli* cells were dried at 40°C for 2 h in a flat plates under the laminar flow hood. After drying, cells were rehydrated in 200 μl of liquid LB media. Re-suspended cells were spread over Petri dishes containing LB, antibiotic, and IPTG, then were incubated overnight at 37°C. The number of colony former units (CFU) was used to compare viability (Dalal et al., 2009; He et al., 2012). Salinity stress was assessed with different concentrations of NaCl (0.4, 0.6, and 0.8 M). Sorbitol (0.6, 0.8, and 1.0 M) and PEG 4000 (5, 10, and 20% w/v) were used to decrease osmotic potential and mimic dehydration, and H_2O_2 (0.1, 0.5, and 1.0 M) was tested to promote oxidant conditions. In all experiments, absorbance at 600 nm (OD_{600}) was used to measure the bacterial growth in liquid media (Liu and Zheng, 2005; Wu et al., 2014; Hu et al., 2016). All experiments were carried out in three biological replicates each replicate was done at least three times.

Localization *In vivo* Using *Nicotiana benthamiana* Protoplasts

The expression vector pEarlyGate103-*AcLEA* was transferred to *Agrobacterium tumefaciens* C58C1 by electroporation in 0.1 cm gap cuvettes. A single colony was used to inoculate LB broth supplemented with ampicillin (100 μg/ml), rifampicin (100 μg/ml), and kanamycin (50 μg/ml). The inoculated broth was cultivated at 30°C overnight. To ensure high expression levels of the recombinant protein, *A. tumefaciens* cells containing the expression clone were used along the helper strain p19 (Voinnet et al., 2003). *A. tumefaciens* cells were harvested by centrifugation at 1,400 × *g* at room temperature and resuspended in an aqueous solution of 10 mM $MgCl_2$. Dilutions were made to adjust a final OD_{600} of 1.0 in the infiltration solution of both the p19 helper strain and the experimental strain (carry on the pEarlyGate103-AcLEA expression vector). Then acetosyringone (50 μg/ml) was added to the infiltration solution and incubated at room temperature for 3 h. This bacterial solution was used to infiltrate *N. benthamiana* leaves and the treated plants were incubated for 96 h in regular growth conditions (26°C and 16/8 h light/dark cycle) prior the protein extraction or protoplasts preparation.

Total protein was extracted from infiltrated leaves by 10 min incubation in extraction buffer (70 mM Tris-HCl, pH 8.0, 1 mM $MgCl_2$, 25 mM KCl, 5 mM NaEDTA·$2H_2O$, 0.25 mM sucrose, 7.5 mM DTT, 0.1% v/v Triton X-100) followed by centrifugation at 16,000 × *g* for 10 min at 4°C. The protein extracts were analyzed by Western blot using anti-GFP (Invitrogen) and anti-AcLEA specific antibodies. Protoplasts were released from the leaf tissue by incubation in enzyme solution composed of 0.5 M mannitol, 1% w/v cellulase R10 (KARLAN Research Products Corp., Cottonwood, AZ, USA) and 0.05% w/v macerozyme R10 (KARLAN Research) and leaf tissue was incubated in this solution for 3 h at constant agitation (1,400 × *g*). Confocal microscopy images were obtained with an Olympus FV1000 microscope (Olympus, Center Valley, PA, USA) using excitation lasers of 633 nm for chlorophyll and 514 nm for GFP.

Detection of AcLEA in Seeds, Leaves, Stems, and Roots from Different Amaranth Species

Proteins from seeds, leaves, stems, and roots were extracted from wild (*A. hybridus* and *A. powellii*) and domesticated (*A. cruentus* and *A. hypochondriacus*) amaranth species.

Seeds were milled under liquid nitrogen in order to obtain a fine powder and proteins were extracted according to their solubility properties. Aqueous soluble proteins were extracted with buffer containing 10% glycerol, 0.1 M Tris-HCl, pH 8.0 in a relation 1:20 (flour/buffer). Suspension was mixed with vortex for 15 min at 4°C and centrifuged at 17,000 × *g* at 4°C, supernatant was recovered and named as hydrophilic fraction. Resulting pellet was resuspended in 7 M urea, 2 M thiourea, 2% CHAPS (w/v), 2% Triton X-100, 0.05 M DTT and mixed as indicated above. The solubilized proteins (hydrophobic fraction) were recovered by centrifugation for 15 min at 17,000 × *g* at 4°C.

Proteins from leaves, stems, and roots were extracted from plants growing under normal and salt stress conditions. Seeds were germinated and seedlings were transferred to plastic pots with soil (Peat Moss Tourbe, Premier Horticulture, Rivière-du-Loup, QC, Canada). Amaranth seedlings were divided in two groups; the control and the salt-stressed groups, which were watered with water and water containing 150 mM NaCl (EC 16.9–17.2 ds/m), respectively. Samples from control and salt-stressed plants were collected next day after salt-stress imposition. Tissues were collected from three biological replicates containing three plants for each replicate. Samples were collected and immediately frozen in liquid nitrogen and milled to a fine powder as reported before (Huerta-Ocampo et al., 2014). The powder was suspended in extraction buffer (1:10 w/v) containing 7 M urea, 2 M thiourea, 2% Triton X-100, and 0.1 M 2-mercaptoethanol. The mixture was sonicated (GE-505, Ultrasonic Processor, Sonics & Materials, Inc., Newtown, CT, USA) for 15 min at 4°C and centrifuged as above.

Proteins (10 μg) were separated in a 12% SDS-PAGE and resolved at 75/150 V for 90 min and then transferred to a PVDF membrane using a Trans-Blot SD semi-dry transfer cell (Bio-Rad, Hercules, CA, USA) for 45 min at 15 V in transfer buffer (25 mM Tris, 192 mM glycine). Membranes were blocked for 2 h with 5% defatted milk in TBS containing 0.1% Tween-20 (TBST), washed three times for 10 min with TBST and incubated with anti-AcLEA IgG rabbit polyclonal antibody for 2 h at 1:1,000 dilution in TBST. Membranes were washed three times for 10 min each with TBST, incubated with anti-rabbit IgG-alkaline phosphatase antibody (Sigma–Aldrich) for 90 min at 1:10,000 dilution in TBST. After membranes were washed three times for 10 min with TBST. Western blots were revealed with alkaline phosphatase buffer (0.1 M Tris, pH 9.5, 0.1 M NaCl, 5 mM $MgCl_2$) and 0.5 mM BCIP, 0.4 mM NBT for 10–20 min at 37°C.

RESULTS AND DISCUSSION

AcLEA Cloning and Recombinant Protein Expression in *E. coli* System

Bioinformatics analyses, using LC-MS/MS information (Maldonado-Cervantes et al., 2014) and the *A. hypochondriacus* transcriptome (Delano-Frier et al., 2011), allowed us to design specific primers for cloning the full-length *AcLEA* cDNA. Amplified *AcLEA* fragment was ligated into pET28mod vector (Supplementary Figures S1A,B).

AcLEA cDNA contains an ORF of 516 bp that codifies for a 172 amino acids protein with a molecular mass calculated of 18.34 kDa and a theoretical p*I* of 8.58, values that corresponded to experimental data previously reported (Maldonado-Cervantes et al., 2014). The sequence (Supplementary Figure S2A) was deposited in the GenBank with access code KX852451. In order to identify AcLEA similar proteins and consensus sequences, a search was performed using protein BLAST algorithm and multiple alignment was carried out with the sequences of the highest similarity matches (**Figure 1A**). A search of related sequences in the LEAPdb database (Hunault and Jaspard, 2010) confirmed that all these sequences are grouped in the LEA_4 Pfam (PF02987). According to the classification proposed by Battaglia et al. (2008), in this family are included LEA proteins from Group 3, such as the cotton protein D-7 (Dure, 1993). Group 3 LEA proteins are characterized by a repetitive motif of 11 amino acids TAQAAKDKTSE (motif 3) in the middle of the sequence that is preceded or followed by ATEAAKQKASE (motif 5); in the N-terminal region is usually conserved the SYKAGETKGRKT (motif 4); meanwhile GGVLQQTGEQV (motif 1) and AADAVKHTLGM (motif 2) are frequently observed in the C-terminal. In many proteins motifs 3 and 5 are present more than once. Motifs 1 to 5 were detected in AcLEA wherein the motif arrangement is M4-M5-M3-M1-M2 with only one complete motif of each type (**Figure 1A**). On the other hand, the motifs arrangement for LEA group 6 is in the order M3-M1-M2-M4 (Rivera-Najera et al., 2014).

With these sequences was constructed the phylogenetic tree in which was also included commercial crops such as: *Zea mays* (NP_001150813.1), *Phaseolus vulgaris* (PvLEA_18 and PVLEA4-25), *Triticum aestivum* (AHZ35571.1), *Oryza sativa* (LEA_3 ABS44867.1 and CAA92106.1), and other crops such a *Vitis amurensis (*ADY17817.1), *Vitis vinifera* (XP_002285360.1), *Nicotiana sylvestris* (XP_009770536.1), *Nicotiana tabacum* (XP_016459037.1), *Capsicum annuum* (XP_016562822.1), *Catharanthus roseus* (AAY84145.1), *Camelina sativa* (010487398.1), *Solanum tuberosum* (XP_006364193.1), *Arabidopsis lyrata* subsp. lyrata (ARALYDRAFT_474395), among others. The phylogenetic tree shows that LEA denominated Dc3 from *B. vulgaris* (sugar beet, XP_010691209.1) shared the highest homology with AcLEA (**Figure 1B**), while the LEA-14 from *A. thaliana* (1X08, At1g01470) which structure has been reported (Singh et al., 2005) showed low similarity with AcLEA as well as for LEA proteins from the commercial cereals and legumes.

AcLEA shares a similar amino acid composition as other LEA proteins, being rich in alanine (19.2%), lysine (14.0%), glutamic acid (9.9%), glutamine (9.3%), threonine (9.3%), and glycine (8.1%) (Battaglia et al., 2008; Denekamp et al., 2010). The total number of negatively charged residues (Asp and Glu) is 27, meanwhile positive charged residues (Arg and Lys) is 29. Another characteristic of AcLEA is the lack of Trp and Cys residues. AcLEA has an aliphatic index of 29.36 with a grand average of hydropathicity (GRAVY) computed of −1.23, indicating a higher abundance of hydrophilic amino acids. Based on the AcLEA amino acid sequence, the hydropathic profile was calculated using the Kyte and Doolittle (1982) values, results showing that

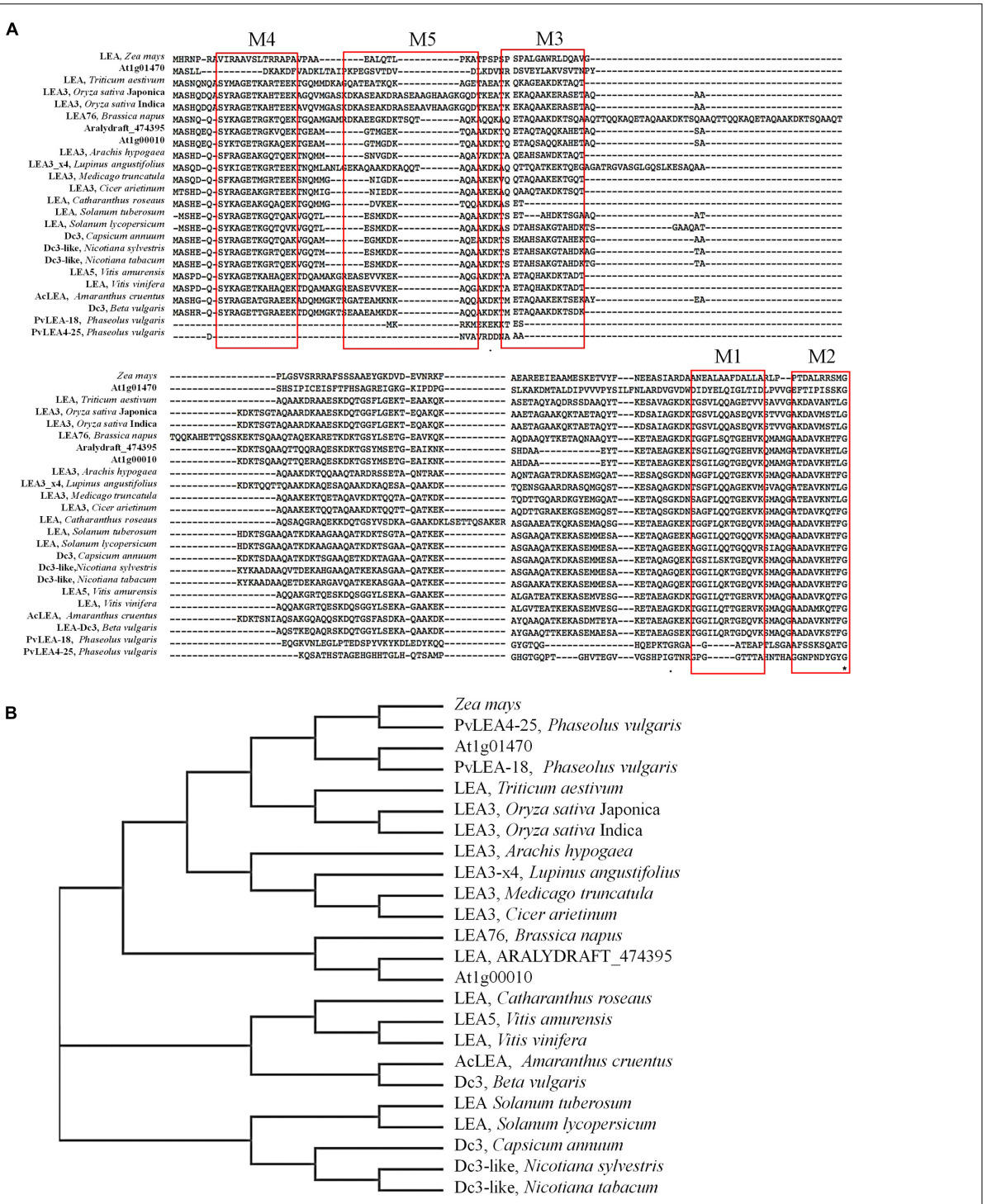


FIGURE 1 | (A) MUSCLE multiple sequence alignment of AcLEA isolated from immature seeds of *Amaranthus cruentus*. The red boxes show the category and position of the conserved motifs. **(B)** Phylogenetic tree constructed using the neighbor-joining method based on the multiple sequences alignment. Accessions numbers of published sequences in the GenBank are as follows: *Amaranthus cruentus* (AcLEA, KX852451), *Beta vulgaris* LEA_Dc3 (XP_010691209.1), *Vitis amurensis* LEA5 (ADY17817.1), *Brassica napus* LEA76, *Vitis vinifera* (XP_002285360.1), *Arabidopsis thaliana* LEA7 (AT1G00010), *Arabidopsis thaliana* LEA 1X08_A (AT1G01470), *Arabidopsis lyrata* subsp. lyrata (ARALDYDRAFT_47395), *Cicer arietinum* LEA3 (XP_004506901.1), *Medicago truncatula* LEA3 (XP_013454682.1), *Lupinus angustifolius* LEA3 (XP_019454903.1), *Arachis hypogaea* LEA3 (ADQ91835.1), *Oryza sativa* Indica LEA3 (CAA92106.1), *Oryza sativa* Japonica LEA3 (ABS44867.1), *Triticum aestivum* (AHZ35571.1), PVLEA4-25 and PvLEA18 from *Phaseolus vulgaris* (AAC49862.1 and AAC49859.1, respectively), *Zea mays* (NP_001150813.1), *Catharanthus roseus* (AAY84145.1), *Solanum tuberosum* LEA 2-like (XP_006364193.1), *Solanum lycopersicum* (NP_001238798.1), *Capsicum annuum* DC3 (XP_016562822.1), *Nicotiana sylvestris* Dc3-like (XP_009770536.1), *Nicotiana tabacum* Dc3-like (XP_016459037.1).

the hydrophilic character of this protein is clearly exhibited (Supplementary Figure S2B), as well AcLEA was predicted as disordered structure (Supplementary Figure S2C). The term hydrophilins was coined to the group of proteins with an average hydrophilicity index >1 and at least 6% Gly. Since hydrophilicity index is 1.23 and the Gly content is 8.1%, AcLEA protein fits in the definition of hydrophilins (Garay-Arroyo et al., 2000).

Protein Expression and Purification

Two distinctive bands putatively corresponding to the recombinant AcLEA were detected in SDS-PAGE, one of them was located at 21.4 kDa, correlating with the molecular weight expected for recombinant protein linked to His-tag, the second band was located at 16.0 kDa (**Figure 2**). The identities of these two bands were successfully identified by LC-MS/MS and bioinformatics analysis using *A. hypochondriacus* database (Supplementary Figure S3). Sequences of the matched peptides as well MASCOT scores are shown in **Table 1**. Data confirm that both the 21.4 and 16.0 kDa bands corresponded to AcLEA. Nevertheless peptides in the N-terminal region were not detected in the 16.0 kDa product, indicating that this short protein is a truncated fragment lacking the N-terminal region.

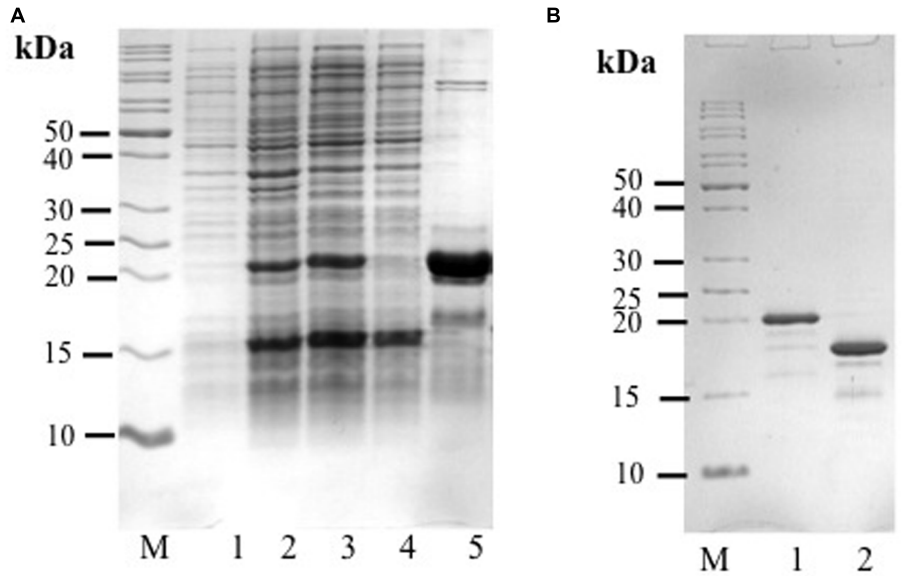


FIGURE 2 | (A) Analysis of His-rAcLEA expression and purification. Lane M = molecular weight marker, Lane 1 = total proteins from non-induced BL21 cells, Lane 2 = total proteins from BL21 cells after induction with IPTG, Lane 3 = soluble proteins from Lane 2, Lane 4 = fraction not retained from Ni^{2+}-column, Lane 5 = fraction retained in Ni^{2+}-column and eluted with 300 mM imidazole. **(B)** His-tag cleavage with PreScission Protease (PSP), Lane M = molecular weight marker, Lane 1 = His-Tag rAcLEA, Lane 2 = rAcLEA.

TABLE 1 | Recombinant AcLEA protein identities by LC-MS/MS.

Band[a]	Protein name	Homology[b]	Accession number[c]	Exp Mr[d]	Theor Mr[e]	Peptides[f]	Score[g]	pm[h]	sc[i]
10	LEA	LEA DC3 *Beta vulgaris*	AHYPO_005092-RA	16.0	15.9	K.TGGILQR.T	33	4	34
						K.ASDMTEYAK.E	54		
						K.SMAQGAADAVK.N	54		
						K.NTFGMGEPEEDDPIK.T	64		
20	LEA	LEA DC3 *Beta vulgaris*	AHYPO_005092-RA	21.4	15.9	K.DKTMETAQAA.E	11	4	27
						K.SMAQGAADAV.N	61		
						K.NTFGMGEPEEDDPIKP.	18		
						K.NTFGMGEPEEDDPIK.T	87		

[a] *10 and 20 kDa bands from **Figure 2A** and Supplementary Figure S3.*
[b] *Best homology as Blast and Muscle-Clustal analysis (**Figure 1**).*
[c] *Accession number according to Amaranthus hypochondriacus transcriptome (phytozome.jgi.doe.gov) and similar to Late embryogenesis abundant protein 76 (Brassica napus).*
[d] *Experimental mass (kDa) of identified proteins.*
[e] *Theoretical mass (kDa) of identified proteins retrieved from the databases.*
[f] *Identified peptide sequences.*
[g] *MASCOT score for each of identified peptides.*
[h] *Number of unique peptides matched with late embryogenesis abundant (LEA) sequence.*
[i] *Sequence coverage (%).*

The 21.4 kDa His-rAcLEA was retained on the Ni^{2+} column and was eluted continuously with successive low concentrations of imidazole (50 mM) washes, but high imidazole concentration (300 mM) was required to completely recover the rAcLEA (**Figure 2A** and Supplementary Figure S4). The 15.4 kDa rAcLEA was not retained by Ni^{2+} column, confirming that this protein is a truncated fragment lacking of N-terminal His-Tag, which was confirmed by MS/MS analysis (**Table 1** and Supplementary Figure S3). After exchange buffer by dialysis, the His-Tag was removed by PSP protease cleavage and rAcLEA purification was carried out again using Ni^{2+}-NTA resin (**Figure 2B**). Retention of cleaved rAcLEA in this stationary phase can be explained because after the proteolysis cleavage residues added to the N-terminal include Gly-Pro-His, and since AcLEA possesses a His in position 4 (Met-Ala-Ser-His), this combination of two histidine residues in relative positions 1–4 seems to be responsible for the rAcLEA binding to Ni^{2+}-NTA resin.

For spectroscopic analysis it is desirable to have a protein purity greater than 98% (Acton et al., 2005). To ensure this experimental condition it was necessary to use a final chromatographic purification step based on molecular exclusion. Both rAcLEA proteins purified by native and denaturing conditions were eluted in a Sephacryl S-100 column with 10 mM sodium phosphates buffer at pH 7.0 as mobile phase; no changes were detected in retention time between them. Typical chromatographic profile shows only one well-defined peak and rAcLEA showed higher purity (Supplementary Figures S5A,B).

Nuclear Magnetic Resonance Spectroscopy

rAcLEA obtained under native conditions was used to evaluate the structural conformation of the recombinant protein by proton nuclear magnetic resonance. Uni-dimensional ^{1}H-NMR spectrum provides general overviews of protein structure because chemical shifts values are strongly related with the presence of different elements of secondary structure (Wishart et al., 1991; Mielke and Krishnan, 2009). Particularly, H_N amide protons are widespread from 6 to 11 ppm in proteins with a well-defined tri-dimensional folding with a high content of α-helix and β-strand. In contrast, H_N resonances of unfolded proteins with a random coil conformation collapse in a narrow region around 7–8 ppm (Singh et al., 2005). **Figure 3** shows the ^{1}H-NMR spectrum of native rAcLEA, as can be observed amide and aromatic protons are distributed between 6.8 and 8.6 ppm, suggesting a random conformation. Moreover, Hα resonances around 4.1 ppm have also a compact distribution, which is consistent with random coil as well the absence of splitting due to coupling in aliphatic signals in the 0.8–2.0 ppm range. This spectroscopic patterns indicate that methyl and methylene groups present in aliphatic amino acid lateral chains have free rotation without limitations due to steric impediment, suggesting that rAcLEA in the experimental conditions tested lacks secondary and tertiary structure. In fact, rAcLEA possesses the typical NMR profile for IDP previously observed in a LEA protein of *T. aestivum* (Sasaki et al., 2014) and a dehydrin of *A. thaliana* (Agoston et al., 2011).

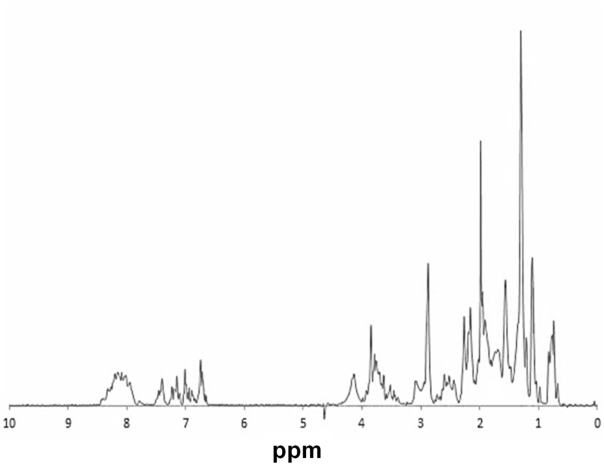


FIGURE 3 | NMR. ^{1}H-NMR spectrum of HisAcLEA purified from native conditions. Narrow signal distribution in amide region, between 6.5 and 8.5 ppm, strongly suggest the lack of a well-defined tridimentional structure distintive of intrinsically disordered proteins.

Circular Dichroism Spectroscopy

The amino acid composition of AcLEA is rich in α-helix promoters such as Ala (19.0%), Met (5.2%), Glu (9.8%), Gln (9.2%), Thr (9.2%), and Lys (13.8%), nevertheless the Gly content is high (8.1%) this amino acid does not have a high propensity for secondary structure formation (Serrano et al., 1992; Creighton, 1993). As observed in other LEA proteins, secondary structure prediction indicates the formation of vast segments of helical structures reaching up to 80% α-helix content. Interestingly, NMR data (**Figure 3**) showed that in the experimental tested conditions rAcLEA has the spectral profile of an IDP. Therefore, in order to further explore the conformational properties of rAcLEA, CD spectra were recorded in the far UV region. rAcLEA was dissolved in 10 mM phosphate buffer pH 8.0 at different NaCl or sorbitol concentrations (Furuki et al., 2011; Wu et al., 2014; Warner et al., 2016). As shown in **Figure 4A**, the AcLEA spectra were not modified by NaCl nor sorbitol presence. All these CD spectra show a negative signal near 200 nm and weak bands in the 210–220 region, suggesting a low secondary structure content. In agreement, the deconvolution of the spectra using the CDNN program (Bohm et al., 1992) indicates a limited content of secondary structure (Supplementary Table S1). Because it is well known that the temperature-induced conformational changes (Soulages et al., 2002), then the curves as a function of temperature at different wavelengths (210, 222, and 230 nm) were followed. For all samples at all the wavelengths tested, the ellipticity signal barely changed with temperature (**Figure 4B**). In agreement, the spectra obtained at 20°C before and after the heating cycle as well as that obtained at 75°C were very similar (Supplementary Figure S6). The lack of a temperature-induced transition strongly suggests that if secondary structure segments are formed, those segments are fluctuating and do not participate in the compact core structure.

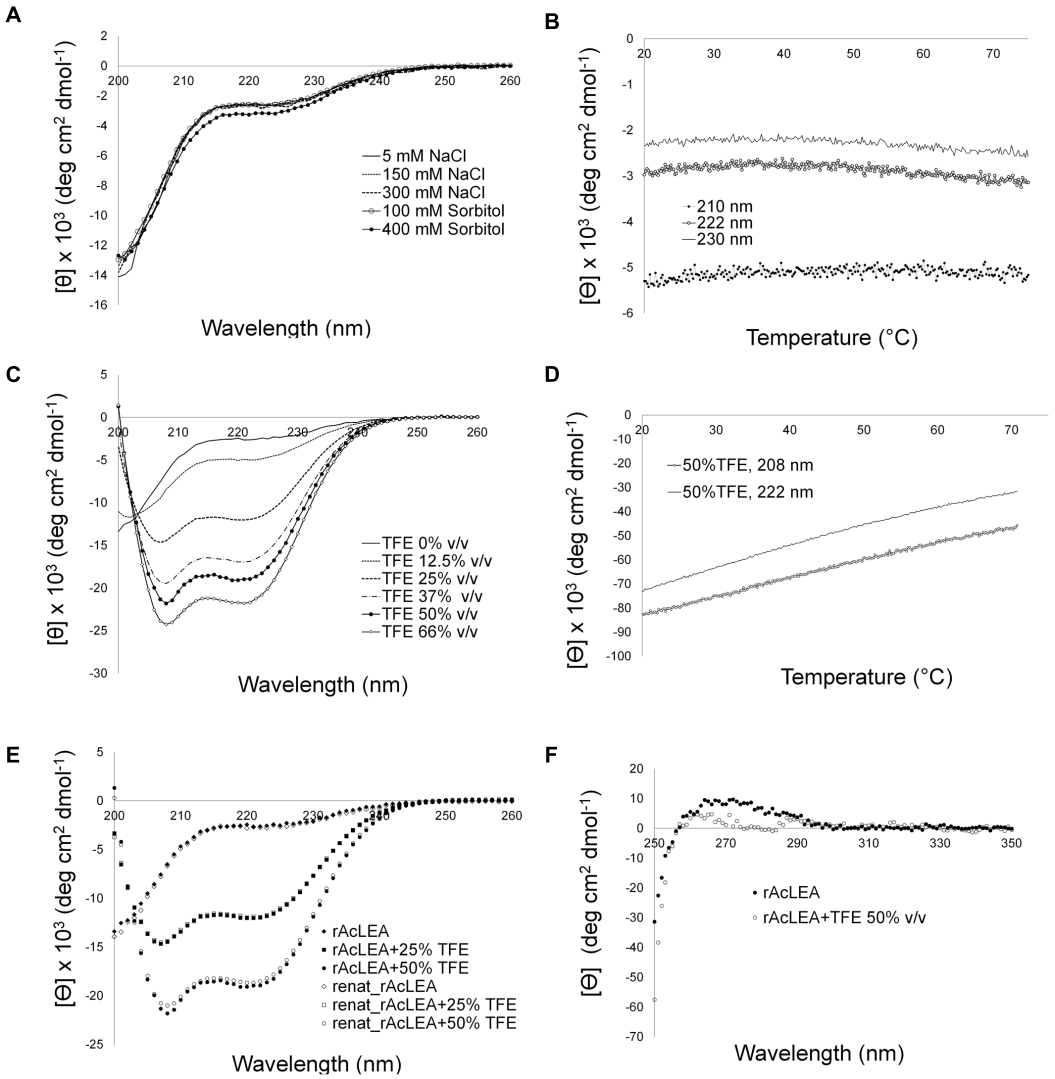


FIGURE 4 | Circular dichroism spectroscopy of rAcLEA under different environment conditions. (A) Far circular dichroism (CD) spectra of rAcLEA in presence of NaCl and sorbitol. **(B)** Recombinant AcLEA (5 mM NaCl, 10 mM phosphates, pH 8, melting monitoring at 210, 222, and 230 nm. **(C)** CD spectra of rAcLEA at different concentrations of trifluoroethanol (TFE), an additive that induce α-helix folding. **(D)** Melting denaturing of α-helix rAcLEA folded with 50% v/v TFE. **(E)** Folding recovery after melting. **(F)** Near CD spectra of rAcLEA in buffer solution with 50% v/v TFE.

It is well established that TFE can induce α-helix folding in peptides (Buck, 1998; Boswell et al., 2014), as well as in unstructured proteins with a predisposition to form secondary structure such as LEA proteins (Shih et al., 2004; Rivera-Najera et al., 2014). Therefore the effect of TFE was evaluated on the rAcLEA conformation. Far UV CD spectra clearly show the tendency of rAcLEA to adopt helical structure as TFE concentration increases (**Figure 4C**). At TFE concentrations higher than 25%, the CD spectra of rAcLEA show the distinctive minima at 208 and 222 nm characteristic of α-helix structures (Muller et al., 2008). As TFE concentration increased up to 66% a gain of helical structure up 70.7% and a decreased in all the other types of secondary structure were observed (**Figure 5** and Supplementary Table S1), this result being quantitatively confirmed using the CDNN program (Supplementary Table S1). In order to determine if this increase in helical content was accompanied with the formation of a structured core, the effect of temperature on rAcLEA dissolved in 50% TFE was assayed. It was found that the ellipticity signal at 208 and 222 nm was lost in a non-cooperative way (**Figure 4D**) and changes in CD signal were fully reversible at 25 and 50% TFE (**Figure 4E**). This strongly suggests that the helical segments induced by the addition of TFE are not arranged in a well-folded tertiary structure. To further explore the formation of tertiary structure, the CD spectra of rAcLEA in the aromatic region were also determined. In the absence of TFE, rAcLEA showed a weak signal in the region corresponding to Tyr and Phe residues, the intensity at 270 nm band

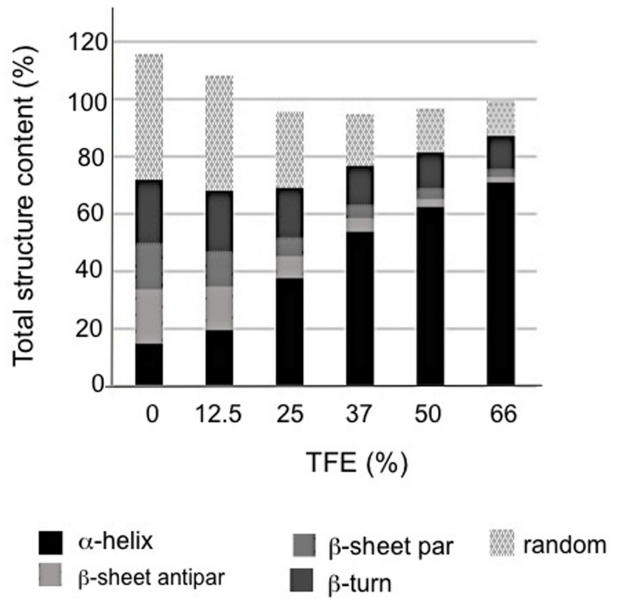


FIGURE 5 | Structural composition vs. TFE concentration in rAcLEA solution as calculated from respective CD data using CDNN software.

being further decreased in the presence of TFE (**Figure 4F**), thus confirming the absence of TFE-induced tertiary structure formation.

Biological Properties of AcLEA *In vivo* Using *E. coli* as a Model

It has been demonstrated that the expression system of *E. coli* is a simple, convenient, and effective model to determine the function of recombinant proteins (Liu and Zheng, 2005). So we used the transformed *E. coli* DE3 cells to evaluate their tolerance to diverse types of abiotic stress conditions (desiccation, NaCl, H_2O_2, sorbitol, and PEG-4000).

Figure 6A shows the growth kinetics of control *E. coli* cells transformed with empty plasmid (control) and pET28mod-AcLEA plasmid. It has been reported that expression of LEA (group 1) genes from plants has no effect on the growth kinetics of transformed *E. coli* or yeast cells (Lan et al., 2005; Campos et al., 2006; Dang et al., 2014). These results are in agreement with our results; however, Warner et al. (2016) reported that induction of *AfrLEA-1* (*Artemia franciscana* LEA group 1) was associated with inhibition of Top10F′ *E. coli* on account of basic p*I* of AfrLEA-1. Curiously AcLEA has also a basic p*I* but we did not observe such cell growth inhibition.

It has been suggested that hydrophilic and heat-stable proteins may modify the structure of other proteins and bind water directly to attenuate the damage caused by desiccation (Houde et al., 1992). **Figure 6B** shows a clear difference in the number of *E. coli* viable cells before and after desiccation stress. Before drying process very similar CFU (expressed in $\times 10^6$ units) were obtained for control and AcLEA expressing cells, but after desiccation, although only a very small fraction of cells survived, the number of CFU expressing AcLEA were three times higher than in control cells. This result suggests that AcLEA expression *E. coli* improved its survival capacity after desiccation. On the other hand, it is well known that the *E. coli* growth rate is strongly influenced by the salt content present in the growth medium (Gowrishankar, 1985). Lan et al. (2005) and Reddy et al. (2012) showed that overexpression of LEA group 1 from plants in *E. coli* provides an increased tolerance to the harmful effects of high salinity environments. Liu and Zheng (2005) indicated that the expression of PM2, a LEA group 3 from soybean, enhances salt tolerance of *E. coli* cells and that the 22-mer repeat region is an important functional region in this protein. As shown in **Figure 6C**, the *E. coli* growth was inhibited by addition of NaCl and contrarily to other reports, the expression of AcLEA did not change this behavior. Because AcLEA has been classified as LEA Group 3 it was expected that it would participated in the protection of cells against salt stress, however, the differences in the amino acid sequences detected in AcLEA (**Figure 1A**) could be responsible for this observed difference.

Low ROS concentrations can act as messengers to regulate biological process, while high ROS concentrations can have very harmful effects and dehydration will disrupt the metabolism of seeds leading to high ROS production (Bailly et al., 2008). In **Figure 6D** is shown that even at high H_2O_2 concentration, AcLEA conferred a significant tolerance to *E. coli* cells. On the other hand, Warner et al. (2016) reported that in general *E. coli* strains tolerate low sorbitol concentrations. Our results showed that AcLEA was able to overcome the negative sorbitol effect on *E. coli* growth even at 1 M concentration (**Figure 6E**). *E. coli* growth was also tested in the presence of PEG as a compound that decreases the osmotic potential of the cells. As shown in **Figure 6F**, the accumulation of AcLEA improved the growth cell supporting an osmoprotection function.

In situ Localization of AcLEA

To decipher the subcellular localization of AcLEA protein, the corresponding coding sequence was fused with green fluorescent protein (AcLEA-GFP) in vectors designed for transient transgene expression in *N. benthamiana* leaf protoplasts. Confocal microscopy images (**Figure 7A**) of protoplasts from leaves infiltrated with the expression vector pEarlyGate103-*AcLEA* clearly show that AcLEA protein exhibits a cytosolic localization in these conditions. The accumulation of AcLEA and GFP proteins in infiltrated leaves was confirmed by immunodetection analysis (**Figure 7B**). It is noted that cytosolic LEA proteins could be involved in stress protection not only within the cytosol itself but also at the level of membranes delimiting the organelles such as mitochondria, chloroplasts, endoplasmic reticulum, and nucleus (Candat et al., 2014).

AcLEA Localization in Amaranth Seeds and Plant Tissues

Anti-AcLEA antibodies were sensitive to detect the corresponding polypeptides in extracts from seed proteins from different amaranth wild and domesticated species. Among

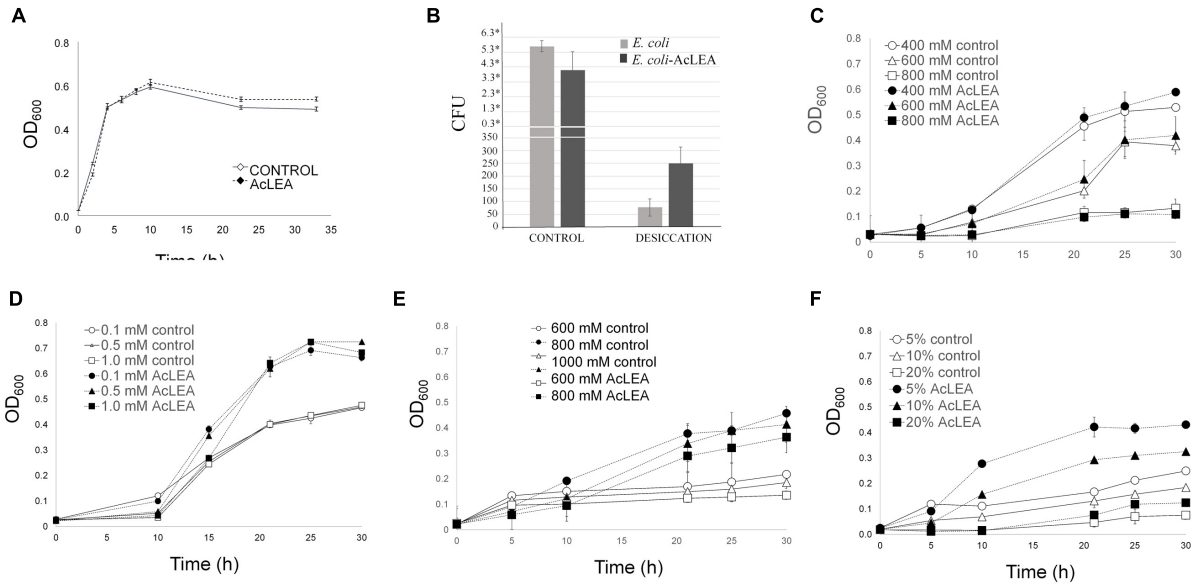


FIGURE 6 | *Escherichia coli* growth in different stress conditons. Cells were grown overnight in LB medium (with antibiotic) and their concentration was determined. An equal number of cells were added to flask with LB (control) and LB containing: NaCl, H_2O_2, sorbitol or PEG as indicated. **(A)** Growth kinetics of *E. coli* transformed with pET28mod (control) and pET28mod-AcLEA. Cells were grown until optical density at 600 nm reached 0.6 then 0.1 mM IPTG was and incubated for 2 h at 28°C. **(B)** Cell viability related to colony former units (CFU) before (control) and after desiccation. * $\times 10^6$ cells. **(C)** Effect of salinity stress induced by NaCl. **(D)** Oxidant stress caused by additon of H_2O_2. **(E,F)** Effect of dehydration stress and osmotic potential stress simulated by the additon of sorbitol and PEG-4000, respectively. Each curve/column represents an average of three biological experiments with three replicates each. Bars in each figure represent the standard deviation.

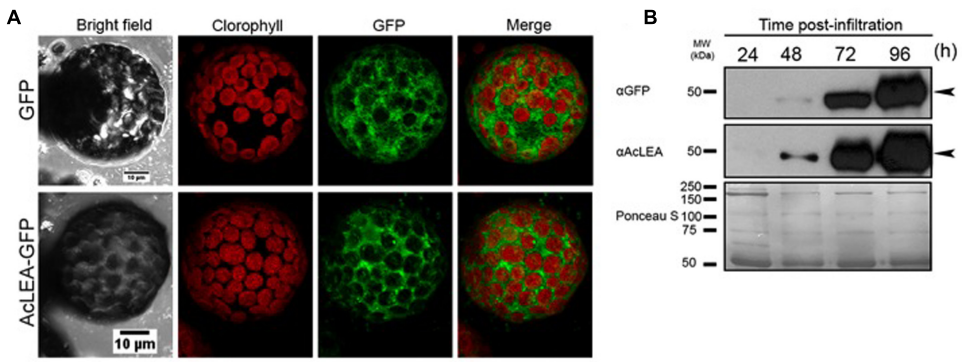


FIGURE 7 | AcLEA subcellular localization using the *Nicotiana benthamiana* system. (A) Confocal microscopy images of protoplasts from *N. benthamiana* leaves infiltrated with the expression vector pEarlyGate103 (GFP) and pEarlyGate10-*AcLEA*-GFP. Images were captured 96 h after infiltration of indicated constructs in bright field and in fluorescence using an excitation laser of 488 and emission filters BA655-755 for chlorophyll and BA505-605 for GFP. A merged image is also shown. **(B)** Immunoblots of kinetics of expression and accumulation of proteins infiltrated at the indicated times. Below immunodetection autoradiographs, the Ponceau S stained membrane is shown as loading control.

all species analyzed, no differences in abundance were observed in seeds (**Figures 8A,B** and Supplementary Figure S7). This could indicate that AcLEA plays an important function most likely during seed drying process. To identify all sequences related with LEA proteins, we carried out a search in phytozome database[7]. Sixty matches were retrieved but only one of those sequences (AHYPO_005092) was identical to AcLEA (Supplementary Figure S8), which correlates with the Western blot analysis where only one reactive band was observed (**Figure 8B**).

The abundance of AcLEA was tested also on leaves, stems, and roots of wild and domesticated amaranth species. Under normal conditions of plant growth of watering, AcLEA was not detected (**Figures 8C,D**). But very interestingly, when plants were subjected to salinity stress, we observed the accumulation of AcLEA (**Figures 8E,F**). As shown in **Figure 8F**, AcLEA accumulation was observed in *A. hypochondriacus* leaves in the

[7]https://phytozome.jgi.doe.gov/pz/portal.html

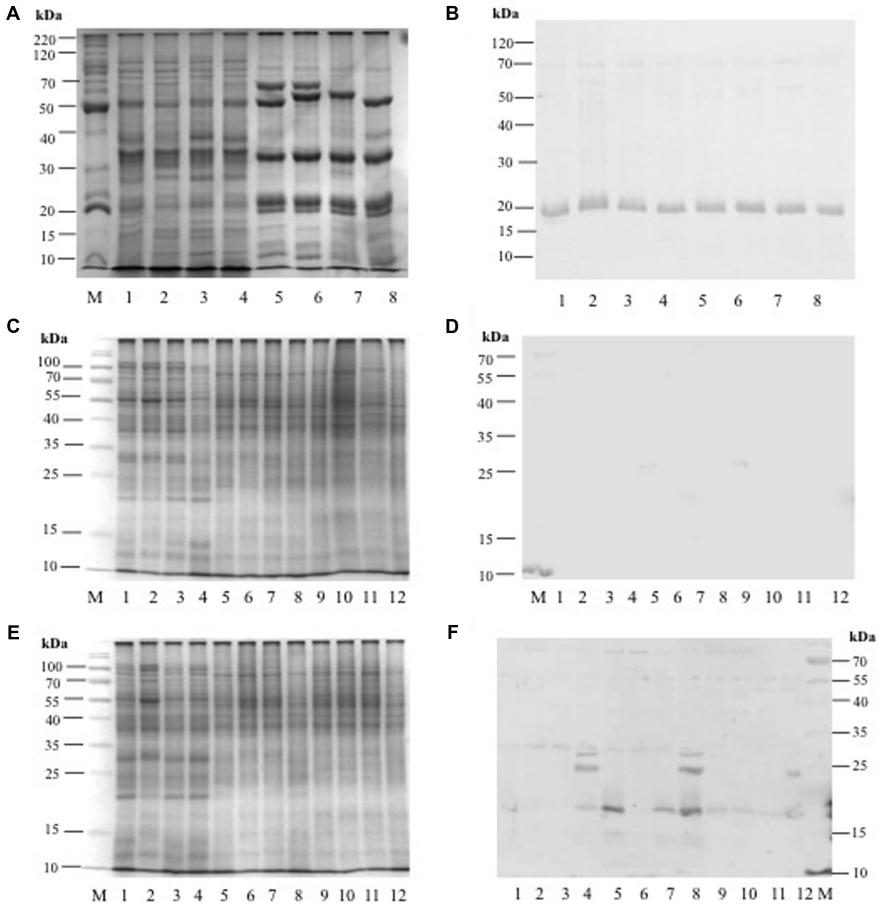


FIGURE 8 | (A) SDS-PAGE profile from amaranth seed storage proteins: Lane M = molecular weigh marker, Lanes 1–4 = hydrophilic proteins from: *A. hybridus*, *A. powellii*, *A. cruentus*, and *A. hypochondriacus*, respectively. Lanes 5–8 = hydrophobic proteins from: *A. hybridus*, *A. powellii*, *A. cruentus,* and *A. hypochondriacus*, respectively. **(B)** Western Blot analysis against anti-AcLEA. **(C)** SDS-PAGE profile from amaranth leaves, stems, and roots from plants growing under normal conditions: Lane M = molecular weight marker; Lanes 1–4 = leaf proteins from: *A. hybridus*, *A. powellii*, *A. cruentus*, and *A. hypochondriacus*, respectively; Lanes 5–8 = stem proteins from: *A. hybridus*, *A. powellii*, *A. cruentus*, and *A. hypochondriacus*, respectively; Lanes 9–12 = root proteins from: *A. hybridus*, *A. powellii*, *A. cruentus*, and *A. hypochondriacus*, respectively. **(D)** Western blot analysis against AcLEA. **(E)** SDS-PAGE profile from amaranth leaves, stems, and roots from plants subjected to salinity stress: Lane M = molecular weight marker; Lanes 1–4 = leaf proteins from of *A. hybridus*, *A. powellii*, *A. cruentus*, and *A. hypochondriacus*, respectively; Lanes 5–8 = stem proteins from of *A. hybridus*, *A. powellii*, *A. cruentus*, and *A. hypochondriacus*, respectively; Lanes 9–12 = root proteins from of *A. hybridus*, *A. powellii*, *A. cruentus*, and *A. hypochondriacus*, respectively. **(F)** Western blot analysis against AcLEA.

expected size (19 kDa, Supplementary Figure S7) but also two more bands around 25 and 30 kDa were observed. In leaves of wild species, the 19 kDa band was barely observed, but in stems a strong band was observed in the wild species *A. hybridus* and the domesticated *A. cruentus* and *A. hypochondriacus*. Meanwhile in roots the 19 kDa band was detected in all species, but at much lower accumulation. These results have shown that AcLEA is conserved in seeds among amaranth species, but that AcLEA plays an important function in response to plant stress and its tissue-specific accumulation was observed.

CONCLUSION

We present the isolation, cloning, structural and functional characterization of the first LEA from *Amaranthus species* (AcLEA). The deduced amino acid sequence of this gene showed that AcLEA belongs to the LEA proteins group 3 and structural analysis in solution has shown that it belongs to IDPs lacking of a well-defined secondary or tertiary structure, but has a strong tendency to adopt a helical conformation. Using *E. coli* as *in vivo* model to evaluate the AcLEA function it was shown that this protein displayed a protective effect against desiccation, osmotic, and oxidative stresses. In *N. benthamiana* leaf protoplasts AcLEA was observed as being localized to the cytosol. Moreover, AcLEA was detected in different tissues from wild and domesticated amaranth species suggesting the important function of AcLEA protein as osmoprotectant during seed desiccation. But interestingly, AcLEA was accumulated in leaves and stems in response to salt stress. These results highlighted the importance of AcLEA as an important protein for stress protection in amaranth species.

AUTHOR CONTRIBUTIONS

AS and APBR conceived and designed the work. AS and EEHD cloned the *AcLEA* gene in *E. coli* system. LdL-V and AG-G, carried out *N. benthamiana* transient transformation. AS and DF-V carried out CD analysis and AS and FdR-P carried out the NMR analysis. AS and AE-M prepared the anti-AcLEA antibodies in rabbits, EB-V conducted Western blot analysis. AS and APBR drafted the manuscript and all authors reviewed and approval the final manuscript.

ACKNOWLEDGMENTS

AS and EH give thanks to CONACYT-Mexico postdoctoral fellow 232286 and 290754, respectively. LdL-V thanks CONACYT-Mexico by the doctoral fellowship 240088. We also thank to CONACYT-Mexico Grants 56787 (Laboratory for Nanoscience and Nanotechnology Research-LINAN) and 204373 (Infrastructure project, INFRA-2013-01), and 251848 (CB-2015). We thank to Alberto Barrera-Pacheco for his technical assistance in MS/MS analysis. LdL-V and AG-G thank to LNMA for confocal microscopy images. We thank to Fabiola Veana Hernández for her technical assistance. We thank to Project "Problemas Nacionales-Amaranto en la soberanía alimentaria" No. 248415.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2017.00497/full#supplementary-material

REFERENCES

Acton, T. B., Gunsalus, K. C., Xiao, R., Ma, L. C., Aramini, J., Baran, M. C., et al. (2005). Robotic cloning and protein production platform of the northeast structural genomics consortium. *Methods Enzymol.* 394, 210–243. doi: 10.1016/S0076-6879(05)94008-1

Agoston, B. S., Kovacs, D., Tompa, P., and Perczel, A. (2011). Full backbone assignment and dynamics of the intrinsically disordered dehydrin ERD14. *Biomol. NMR Assign.* 5, 189–193. doi: 10.1007/s12104-011-9297-2

Aguilar-Hernández, H. S., Santos, L., Leon-Galvan, F., Barrera-Pacheco, A., Espitia-Rangel, E., De Leon-Rodriguez, A., et al. (2011). Identification of calcium stress induced genes in amaranth leaves through suppression subtractive hybridization. *J. Plant Physiol.* 168, 2102–2109. doi: 10.1016/j.jplph.2011.06.006

Ali-Benali, M. A., Alary, R., Joudrier, P., and Gautier, M. F. (2005). Comparative expression of five Lea Genes during wheat seed development and in response to abiotic stresses by real-time quantitative RT-PCR. *Biochim. Biophys. Acta* 1730, 56–65. doi: 10.1016/j.bbaexp.2005.05.011

Avelange-Macherel, M. H., Payet, N., Lalanne, D., Neveu, M., Tolleter, D., Burstin, J., et al. (2015). Variability within a pea core collection of LEAM and HSP22, two mitochondrial seed proteins involved in stress tolerance. *Plant Cell Environ.* 38, 1299–1311. doi: 10.1111/pce.12480

Bailly, C., El-Maarouf-Bouteau, H., and Corbineau, F. (2008). From intracellular signaling networks to cell death: the dual role of reactive oxygen species in seed physiology. *C R Biol.* 331, 806–814. doi: 10.1016/j.crvi.2008.07.022

Battaglia, M., and Covarrubias, A. A. (2013). Late embryogenesis abundant (LEA) proteins in legumes. *Front. Plant Sci.* 4:190. doi: 10.3389/fpls.2013.00190

Battaglia, M., Olvera-Carrillo, Y., Garciarrubio, A., Campos, F., and Covarrubias, A. A. (2008). The enigmatic LEA proteins and other hydrophilins. *Plant Physiol.* 148, 6–24. doi: 10.1104/pp.108.120725

Becker, R., Wheeler, E. L., Lorenz, K. A., Staffor, E., Grosjean, O. K., Betschart, A. A., et al. (1981). A compositional study of amaranth grain. *J. Food Sci.* 46, 1175–1180. doi: 10.1111/j.1365-2621.1981.tb03018.x

Bohm, G., Muhr, R., and Jaenicke, R. (1992). Quantitative analysis of protien far UV ciruclar dichroism spectra by neural networks. *Protein Eng.* 5, 191–195. doi: 10.1093/protein/5.3.191

Boswell, L. C., Menze, M. A., and Hand, S. C. (2014). Group 3 late embryogenesis abundant proteins from embryos of *Artemia franciscana*: structural properties and protective abilities during desiccation. *Physiol. Biochem. Zool.* 87, 640–651. doi: 10.1086/676936

Boucher, V., Buitink, J., Lin, X., Boudet, J., Hoekstra, F. A., Hundertmark, M., et al. (2010). MtM25 is an atypical hydrophobic late embryogenesis-abundant protein that dissociates cold and desiccation-aggregated proteins. *Plant Cell Environ.* 33, 418–430. doi: 10.1111/j.1365-3040.2009.02093.x

Brenner, D. M., Baltensperger, D. D., Kulakow, P. A., Lehmann, J. W., Myers, R. L., Slabbert, M. M., et al. (2000). "Genetic resources and breeding of amaranthus," in *Plant Breeding Reviews*, Vol. 19, ed. J. Janick (Oxford: John Wiley & Sons, Inc).

Browne, J., Tunnacliffe, A., and Burnell, A. (2002). Anhydrobiosis: plant desiccation gene found in a nematode. *Nature* 416:38. doi: 10.1038/416038a

Buck, M. (1998). Trifluoroethanol and colleagues: cosolvents come of age. Recent studies with peptides and proteins. *Q. Rev. Biophys.* 31, 297–355. doi: 10.1017/S003358359800345X

Campos, F., Cuevas-Velazquez, C., Fares, M. A., Reyes, J. L., and Covarrubias, A. A. (2013). Group 1 LEA proteins, an ancestral plant protein group, are also present in other eukaryotes, and in the archeae and bacteria domains. *Mol. Genet. Genomics* 288, 503–517. doi: 10.1007/s00438-013-0768-2

Campos, F., Zamudio, F., and Covarrubias, A. A. (2006). Two different late embryogenesis abundant proteins from *Arabidopsis thaliana* contain specific domains that inhibit *Escherichia coli* growth. *Biochem. Biophys. Res. Commun.* 342, 406–413. doi: 10.1016/j.bbrc.2006.01.151

Candat, A., Paszkiewicz, G., Neveu, M., Gautier, R., Logan, D. C., Avelange-Macherel, M.-H., et al. (2014). The ubiquitous distribution of late embryogenesis abundant proteins across cell compartments in *Arabidopsis* offers tailored protection against abiotic stress. *Plant Cell* 26, 3148–3166. doi: 10.1105/tpc.114.127316

Cao, J., and Li, X. (2015). Identification and phylogenetic analysis of late embryogenesis abundant proteins family in tomato (*Solanum lycopersicum*). *Planta* 241, 757–772. doi: 10.1007/s00425-014-2215-y

Chakrabortee, S., Meersman, F., Schierle, G. S. K., Bertoncini, C. W., McGee, B., Kaminski, C. F., et al. (2010). Catalytic and chaperone-like functions in an intrinsically disordered protein associated with desiccation tolerance. *Proc. Natl. Acad. Sci. U.S.A.* 107, 16084–16089. doi: 10.1073/pnas.1006276107

Clouse, J. W., Adhikary, D., Page, J. T., Ramaraj, T., Deholos, M. K., Udall, J. A., et al. (2016). The amaranth genome: genome, transcriptome, and physical map assembly. *Plant Genome* 9, 1–14. doi: 10.3835/plantgenome2015.07.0062

Creighton, T. (1993). *Proteins. Structure and Molecular Properties*, Second Edn. New York, NY: W. H. Freeman.

Dalal, M., Tayal, D., Chinnusamy, V., and Bansal, K. C. (2009). Abiotic stress and ABA-inducible Group 4 LEA from *Brassica napus* plays a key role in salt and drought tolerance. *J. Biotechnol.* 139, 137–145. doi: 10.1016/j.jbiotec.2008.09.014

Dang, N. X., Popova, A. V., Hundertmark, M., and Hincha, D. K. (2014). Functional characterization of selected LEA proteins from *Arabidopsis thaliana* in yeast and in vitro. *Planta* 240, 325–336. doi: 10.1007/s00425-014-2089-z

Delano-Frier, J. P., Aviles-Arnaut, H., Casarrubias-Castillo, K., Casique-Arroyo, G., Castrillon-Arbelaez, P. A., Herrera-Estrella, L., et al. (2011). Transcriptomic analysis of grain amaranth (*Amaranthus hypochondriacus*) using 454 pyrosequencing: comparison with A. tuberculatus, expression profiling in stems and in response to biotic and abiotic stress. *BMC Genomics* 12:363. doi: 10.1186/1471-2164-12-363

Denekamp, N. Y., Reinhardt, R., Kube, M., and Lubzens, E. (2010). Late embryogenesis abundant (LEA) proteins in nondesiccated, encysted, and diapausing embryos of rotifers. *Biol. Reprod.* 82, 714–724. doi: 10.1095/biolreprod.109.081091

Dure, L. III (1989). Common amino acid sequence domains among the LEA proteins of higher plants. *Plant Mol. Biol.* 12, 475–486. doi: 10.1007/BF00036962

Dure, L. (1993). A repeating 11-mer amino acid motif and plant desiccation. *Plant J.* 3, 363–369. doi: 10.1046/j.1365-313X.1993.t01-19-00999.x

Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340

Finn, R. D., Coggill, P., Eberhardt, R. Y., Eddy, S. R., Mistry, J., Mitchell, A. L., et al. (2015). The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res.* 44, D279–D285. doi: 10.1093/nar/gkv1344

Furuki, T., Shimizu, T., Kikawada, T., Okuda, T., and Sakurai, M. (2011). Salt effects on the structural and thermodynamic properties of a group 3 LEA protein model peptide. *Biochemistry* 50, 7093–7103. doi: 10.1021/bi200719s

Gai, Y. P., Ji, X. L., Lu, W., Han, X. J., Yang, G. D., and Zheng, C. C. (2011). A novel late embryogenesis abundant like protein associated with chilling stress in *Nicotiana tabacum* cv. bright yellow-2 cell suspension culture. *Mol. Cell. Proteom.* 10, M111.010363. doi: 10.1074/mcp.M111.010363

Garay-Arroyo, A., Colmenero-Flores, J. M., Garciarrubio, A., and Covarrubias, A. A. (2000). Highly hydrophilic proteins in prokaryotes and eukaryotes are common during conditions of water deficit. *J. Biol. Chem.* 275, 5668–5674. doi: 10.1074/jbc.275.8.5668

Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). "Protein identification and analysis tools on the ExPASy server," in *The Proteomics Protocols Handbook*, eds J. M. Walker and N. J. Totowa (New York City, NY: Humana Press Inc), 571–607. doi: 10.1385/1-59259-890-0:571

Gowrishankar, J. (1985). Identification of osmoresponsive genes in *Escherichia coli*: evidence for participation of potassium and proline transport systems in osmoregulation. *J. Bacteriol.* 164, 434–445.

Goyal, K., Tisi, L., Basran, A., Browne, J., Burnell, A., Zurdo, J., et al. (2003). Transition from natively unfolded to folded state induced by desiccation in an anhydrobiotic nematode protein. *J. Biol. Chem.* 278, 12977–12984. doi: 10.1074/jbc.M212007200

Hand, S. C., Menze, M. A., Toner, M., Boswell, L., and Moore, D. (2011). LEA proteins during water stress: not just for plants anymore. *Annu. Rev. Physiol.* 73, 115–134. doi: 10.1146/annurev-physiol-012110-142203

Hatanaka, R., Furuki, T., Zhimizu, T., Takezawa, D., Kikawada, T., Sakurai, M., et al. (2014). Biochemical and structural characterization of an endoplasmic reticulum-localized late embryogenesis abundant (LEA) protein from the liverwort *Marchantia polymorpha*. *Biochem. Biophys. Res. Commun.* 454, 588–593. doi: 10.1016/j.bbrc.2014.10.130

He, S., Tan, L., Hu, Z., Chen, G., Wang, G., and Hu, T. (2012). Molecular characterization and functional analysis by heterologous expression in *E. coli* under diverse abiotic stresses for OsLEA5, the atypical hydrophobic LEA protein from *Oryza sativa* L. *Mol. Genet. Genomics* 287, 39–54. doi: 10.1007/s00438-011-0660-x

Hincha, D. K., and Thalhammer, A. (2012). LEA proteins: IDPs with versatile functions in cellular dehydration tolerance. *Biochem. Soc. Trans.* 40, 1000–1003. doi: 10.1042/BST20120109

Houde, M., Danyluk, J., Laliberte, J. F., Rassart, E., Dhindsa, R. S., and Sarhan, F. (1992). Cloning, characterization, and expression of a cDNA encoding a 50-kilodalton protein specifically induced by cold acclimation in wheat. *Plant Physiol.* 99, 1381–1387. doi: 10.1104/pp.99.4.1381

Hu, T., Zhou, N., Fu, M., Qin, J., and Huang, X. (2016). Characterization of OsLEA1a and its inhibitory effect on the resistance of *E. coli* to diverse abiotic stresses. *Int. J. Biol. Macromol.* 91, 1010–1017. doi: 10.1016/j.ijbiomac.2016.06.056

Huerta-Ocampo, J. A., Barrera-Pacheco, A., Mendoza-Hernández, C. S., Espitia-Rangel, E., Mock, H. P., and Barba de la Rosa, A. P. (2014). Salt stress-induced alterations in the root proteome of *Amaranthus cruentus* L. *J. Proteome Res.* 13, 3607–3627. doi: 10.1021/pr500153m

Huerta-Ocampo, J. A., Leon-Galvan, M. F., Ortega-Cruz, L. B., Barrera-Pacheco, A., De Leon-Rodriguez, A., Mendoza-Hernandez, G., et al. (2011). Water stress induces up-regulation of DOF1 and MIF1 transcription factors and down-regulation of proteins involved in secondary metabolism in amaranth roots (*Amaranthus hypochondriacus* L.). *Plant Biol.* 13, 472–482. doi: 10.1111/j.1438-8677.2010.00391.x

Hunault, G., and Jaspard, E. (2010). LEAPdb: a database for the late embryogenesis abundant proteins. *BMC Genomics* 11:221. doi: 10.1186/1471-2164-11-221

Hundertmark, M., and Hincha, D. K. (2008). LEA (late embryogenesis abundant) proteins and their encoding genes in *Arabidopsis thaliana*. *BMC Genomics* 9:118. doi: 10.1186/1471-2164-9-118

Imai, R., Chang, L., Ohta, A., Bray, E. A., and Takagi, M. (1996). A lea-class gene of tomato confers salt and freezing tolerance when expressed in *Saccharomyces cerevisiae*. *Gene* 170, 243–248. doi: 10.1016/0378-1119(95)00868-3

Ingram, J., and Bartels, D. (1996). The molecular basis of dehydration tolerance in plants. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 47, 377–403. doi: 10.1146/annurev.arplant.47.1.377

Jaspard, E., Macherel, D., and Hunault, G. (2012). Computational and statistical analyses of amino acid usage and physic-chemical properties of the twelve late embryogenesis abundant protein classes. *PLoS ONE* 7:e36868. doi: 10.1371/journal.pone.0036968

Karimi, M., Depicker, A., and Hilson, P. (2007). Recombinational cloning with plant gateway vectors. *Plant Physiol.* 145, 1144–1154. doi: 10.1104/pp.107.106989

Kyte, J., and Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132. doi: 10.1016/0022-2836(82)90515-0

Lan, Y., Cai, D., and Zheng, Y.-Z. (2005). Expression in *Escherichia coli* of three different soybean late embryogenesis abundant (LEA) genes to investigate enhanced stress tolerance. *J. Integr. Plant Biol.* 47, 613–621. doi: 10.1111/j.1744-7909.2005.00025.x

Liu, Y., Wang, L., Xing, X., Sun, L., Pan, J., Kong, X., et al. (2013). ZmLEA3, a multifunctional Group 3 LEA protein from maize (*Zea mays* L.), is involved in biotic and abiotic stresses. *Plant Cell Physiol.* 54, 944–959. doi: 10.1093/pcp/pct047

Liu, Y.-L., and Zheng, Y.-Z. (2005). PM2, a group 3 LEA protein from soybean, and its 22-mer repeating region confer salt tolerance in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* 331, 325–332. doi: 10.1016/j.bbrc.2005.03.165

Maldonado-Cervantes, E., Huerta-Ocampo, J. A., Montero-Moran, G. M., Barrera-Pacheco, A., Espitia-Rangel, E., and Barba de la Rosa, A. P. (2014). Characterization of *Amaranthus cruentus* L. seed proteins by 2-DE and LC/MS-MS: identification and cloning of a novel late embryogenesis-abundant protein. *J. Cereal Sci.* 60, 172–178. doi: 10.1016/j.jcs.2014.02.008

Maughan, P. J., Yourstone, S. M., Jellen, E. N., and Udall, J. A. (2009). SNP discovery via genomic reduction, barcoding, and 454-Pyrosequencing in amaranth. *Plant Genome* 2, 260–270. doi: 10.3835/plantgenome2009.08.0022

Mielke, S. P., and Krishnan, V. V. (2009). Characterization of protein secondary structure from NMR chemical shifts. *Prog. Nucl. Magn. Reson. Spectrosc.* 54, 141–165. doi: 10.1016/j.pnmrs.2008.06.002

Muller, I., Sarramégna, V., Renault, M., Lafaquière, V., Sebai, S., Milon, A., et al. (2008). The full-length Mu-opioid receptor: a conformational study by circular dichorism in trifluoroethanol and membrane-mimetic environments. *J. Membrane Biol.* 223, 49–57. doi: 10.1007/s00232-008-9112-x

Olvera-Carrillo, Y., Campos, F., Reyes, J. L., Garciarrubio, A., and Covarrubias, A. A. (2010). Functional analysis of the Group 4 late embryogenesis abundant proteins reveals their relevance in the adaptive response during water deficit in *Arabidopsis*. *Plant Physiol.* 154, 373–390. doi: 10.1104/pp.110.158964

Popova, A. V., Hundertmark, M., Seckler, R., and Hincha, D. K. (2011). Structural transitions in the intrinsically disordered plant dehydration stress protein LEA7 upon drying are modulated by the presence of membranes. *Biochim. Biophys. Acta* 1808, 1879–1887. doi: 10.1016/j.bbamem.2011.03.009

Rastogi, A., and Shukla, S. (2013). Amaranth: a new millennium crop of nutraceutical values. *Crit. Rev. Food Sci. Nutr.* 53, 109–125. doi: 10.1080/10408398.2010.517876

Reddy, P. S., Reddy, G. M., Pandey, P., Chandrasekhar, K., and Reddy, M. K. (2012). Cloning and molecular characterization of a gene encoding late embryogenesis abundant protein from *Pennisetum glaucum*: protection against abiotic stresses. *Mol. Biol. Rep.* 39, 7163–7174. doi: 10.1007/s11033-012-1548-5

Rivera-Najera, L. Y., Saab-Rincon, G., Battaglia, M., Amero, C., Pulido, N. O., Garcia-Hernandez, E., et al. (2014). A group 6 late embryogenesis abundant

protein from common bean Is a disordered protein with extended helical structure and oligomer-forming properties. *J. Biol. Chem.* 289, 31995–32009. doi: 10.1074/jbc.M114.583369

Saitou, N., and Nei, M. (1987). The neighbor-joining method: a new method for reconstruction phylogenetic trees. *Mol. Biol. Evol.* 4, 406–425.

Sasaki, K., Christov, N. K., Tsuda, S., and Imai, R. (2014). Identification of a novel LEA protein involved in freezing tolerance in wheat. *Plant Cell Physiol.* 55, 136–147. doi: 10.1093/pcp/pct164

Serrano, L., Sancho, J., Hirshberg, M., and Fersht, A. R. (1992). Alpha-helix stability in proteins. I. Empirical correlations concerning substitution of side-chains at the N and C-caps and the replacement of alanine by glycine or serine at solvent-exposed surfaces. *J. Mol. Biol.* 227, 544–559. doi: 10.1016/0022-2836(92)90906-Z

Shih, M. D., Lin, S. D., Hsieh, J. S., Tsou, C. H., Chow, T. Y., Lin, T. P., et al. (2004). Gene cloning and characterization of a soybean (*Glycine max* L.) LEA protein, GmPM16. *Plant Mol. Biol.* 56, 689–703. doi: 10.1007/s11103-004-4680-3

Silva-Sánchez, C., Barba de la Rosa, A. P., León-Galván, M. F., de Lumen, B. O., De León-Rodríguez, A., and González de Mejía, E. (2008). Bioactive peptides in amaranth (*Amaranthus hypochondriacus*) seed. *J. Agric. Food Chem.* 56, 1233–1240. doi: 10.1021/jf072911z

Singh, S., Cornilescu, C. C., Tyler, R. C., Cornilescu, G., Tonelli, M., Lee, M. S., et al. (2005). Solution structure of a late embryogenesis abundant protein (LEA14) from *Arabidopsis thaliana*, a cellular stress-related protein. *Protein Sci.* 2005, 2601–2609. doi: 10.1110/ps.051579205

Sivamani, E., Bahieldin, A., Wraith, J. M., Al-Niemi, T., Dyer, W. E., Ho, T., et al. (2000). Improved biomass productivity and water use efficiency under water deficit conditions in transgenic wheat constitutively expressing the barley HVA1 gene. *Plant Sci.* 155, 1–9. doi: 10.1016/S0168-9452(99)00247-2

Soulages, J. L., Kim, K., Walters, C., and Cushman, J. C. (2002). Temperature-induced extended helix/random coil transitions in a group 1 late embryogenesis-abundant protein from soybean. *Plant. Physiol.* 128, 822–832. doi: 10.1104/pp.010521

Tolleter, D., Hincha, D. K., and Macherel, D. (2010). A mitochondrial late embryogenesis abundant protein stabilizes model membranes in the dry state. *Biochim. Biophys. Acta* 1798, 1926–1933. doi: 10.1016/j.bbamem.2010.06.029

Tolleter, D., Jaquinod, M., Mangavel, C., Passirani, C., Saulnier, P., Manon, S., et al. (2007). Structure and function of a mitochondrial late embryogenesis abundant protein are revealed by desiccation. *Plant Cell* 19, 1580–1589. doi: 10.1105/tpc.107.050104

Tompa, P., and Kovacs, D. (2010). Intrinsically disordered chaperones in plants and animals. *Biochem. Cell Biol.* 88, 167–174. doi: 10.1139/o09-163

Tunnacliffe, A., Hincha, D. K., Leprince, O., and Macherel, D. (2010). "LEA proteins: versa-tility of form and function," in *Sleeping Beauties: Dormancy and Resistance in Harsh Environments*, eds E. Lubzens, J. Cerda, and M. Clark (Berlin: Springer), 91–108. doi: 10.1007/978-3-642-12422-8_6

Tunnacliffe, A., and Wise, M. J. (2007). The continuing conundrum of the LEA proteins. *Naturwissenschaften* 94, 791–812. doi: 10.1007/s00114-007-0254-y

Valdes-Rodríguez, S., Guerrero-Rangel, A., Melgoza-Villagomez, C., Chagolla-Lopez, A., Delgado-Vargas, F., Martinez-Gallardo, N., et al. (2007). Cloning of a cDNA encoding a cystatin from grain amaranth (*Amaranthus hypochondriacus*) showing a tissue-specific expression that is modified by germination and abiotic stress. *Plant Physiol. Biochem.* 45, 790–798. doi: 10.1016/j.plaphy.2007.07.007

van Leeuwen, M. R., Wyatt, T. T., van Doorn, T. M., Lugones, L. G., Wösten, H. A. B., and Dijksterhuis, J. (2016). Hydrophilins in the filamentous fungus Neosartorya fischeri (*Aspergillus fischeri*) have protective activity against several types of microbial water stress. *Environ. Microbiol. Rep.* 8, 45–52. doi: 10.1111/1758-2229.12349

Voinnet, O., Rivas, S., Mestre, P., and Baulcombe, D. (2003). An enhanced transient expression system in plants based on suppression of gene silencing by the p19 protein of tomato bushy stunt virus. *Plant J.* 33, 949–956. doi: 10.1046/j.1365-313X.2003.01676.x

Warner, A. H., Guo, Z.-H., Moshi, S., Hudson, J. W., and Kozarova, A. (2016). Study of model systems to test the potential function of Artemia group 1 late embryogenesis abundant (LEA) proteins. *Cell Stress Chaperones* 21, 139–154. doi: 10.1007/s12192-015-0647-3

Wishart, D. S., Sykes, B. D., and Richards, F. M. (1991). Simple techniques for the quantification of protein secondary structure by H-1 NMR spectroscopy. *FEBS Lett.* 293, 1–2. doi: 10.1016/0014-5793(91)81155-2

Wolkers, W. F., McCready, S., Brandt, W. F., Lindsey, G. G., and Hoekstra, F. A. (2001). Isolation and characterization of a D-7 LEA protein from pollen that stabilizes glasses in vitro. *Biochim. Biophys. Acta* 1544, 196–206. doi: 10.1016/S0167-4838(00)00220-X

Wu, Y., Liu, C., Kuang, J., Ge, Q., Zhang, Y., and Wang, Z. (2014). Overexpression of SmLEA enhances salt and drought tolerance in *Escherichia coli* and *Salvia miltiorrhiza*. *Protoplasma* 251, 1191–1199. doi: 10.1007/s00709-014-0626-z

Xu, D., Duan, X., Wang, B., Hong, B., Ho, T., and Wu, R. (1996). Expression of a late embryogenesis abundant protein gene, HVA1, from barley confers tolerance to water deficit and salt stress in transgenic rice. *Plant Physiol.* 110, 249–257. doi: 10.1104/pp.110.1.249

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Plant Physiology and Biochemistry 139 (2019) 293–306

Contents lists available at ScienceDirect

Plant Physiology and Biochemistry

journal homepage: www.elsevier.com/locate/plaphy

Research article

Functional analysis of the Chloroplast GrpE (CGE) proteins from *Arabidopsis thaliana*

L.A. de Luna-Valdez, C.I. Villaseñor-Salmerón, E. Cordoba, R. Vera-Estrella, P. León-Mejía, A.A. Guevara-García*

Departamento de Biología Molecular de Plantas, Instituto de Biotecnología, Universidad Nacional Autónoma de México, Av. Universidad 2001, Col. Chamilpa, Cuernavaca, Morelos, CP 62210, Mexico

ARTICLE INFO

Keywords:
Chloroplast
Chaperones
Cochaperones
Arabidopsis
GrpE
Hsp70

ABSTRACT

The function of proteins depends on specific partners that regulate protein folding, degradation and protein-protein interactions, such partners are the chaperones and cochaperones. In chloroplasts, proteins belonging to several families of chaperones have been identified: chaperonins (Cpn60s), Hsp90s (Hsp90-5/Hsp90C), Hsp100s (Hsp93/ClpC) and Hsp70s (cpHsc70s). Several lines of evidence have demonstrated that cpHsc70 chaperones are involved in molecular processes like protein import, protein folding and oligomer formation that impact important physiological aspects in plants such as thermotolerance and thylakoid biogenesis. Despite the vast amount of data existing around the function of cpHcp70s chaperones, very little attention has been paid to the roles of DnaJ and GrpE cochaperones in the chloroplast. In this study, we performed a phylogenetic analysis of the chloroplastic GrpE (CGE) proteins from 71 species. Based on their phylogenetic relationships and on a motif enrichment analysis, we propose a classification system for land plants' CGEs, which include two independent groups with specific primary structure traits. Furthermore, using *in vivo* assays we determined that the two CGEs from *A. thaliana* (AtCGEs) complement the mutant phenotype displayed by a knockout *E. coli* strain defective in the bacterial grpE gene. Moreover, we determined *in planta* that the two AtCGEs are *bona fide* chloroplastic proteins, which form the essential homodimers needed to establish direct physical interactions with the cpHsc70-1 chaperone. Finally, we found evidence suggesting that AtCGE1 is involved in specific physiological phenomena in *A. thaliana*, such as the chloroplastic response to heat stress, and the correct oligomerization of the photosynthesis-related LHCII complex.

1. Introduction

Chloroplasts are photosynthetic organelles of endosymbiotic origin that contain their own genome (Martin et al., 1998), which encodes 130 genes in average (RefSeq:https://www.ncbi.nlm.nih.gov/genome/organelle/). Estimations indicate that functional chloroplasts require 2000 to 3500 proteins (van Wijk and Baginsky, 2011), however the chloroplast genome codes for an average of only 83 proteins (RefSeq:https://www.ncbi.nlm.nih.gov/genome/organelle/). The rest of the chloroplastic proteome is encoded in the nuclear genome; using the cellular transcription and translation apparatus such proteins are synthesized in the cytoplasm and post-translationally imported into the organelle (Villarejo et al., 2005; Jarvis and Kessler, 2014; Paila et al., 2015). From their synthesis, chloroplast proteins are subjected to regulatory events that impact the performance of their functions; targeting, folding, post-translational stabilization, and degradation are processes that affect the homeostasis of chloroplast proteins (Trösch et al., 2015); these processes are regulated by molecular chaperones. In the cytoplasm, precursor proteins interact with the Hsp70, 14-3-3 and Hsp90 chaperones that keep them in a proper folding state and facilitate their interaction with the translocon receptor proteins Toc159 and Toc34 (Rial et al., 2000, May and Soll, 2000; Qbadou et al., 2006; Flores-Pérez and Jarvis, 2013).

In the chloroplast stroma, several chaperone systems have been identified. The chaperonins (Cpn60α and Cpn60β) are involved in the folding of RbcL (Ribulose Bisphosphate Carboxylase/Oxygenase large chain) and NdhH (NAD(P)H-quinone oxidoreductase subunit H) (Barraclough and Ellis, 1980; Peng et al., 2011), Hsp90s mediate protein import, protein maturation and thylakoid formation (Pratt and Toft, 2003; Heide et al., 2009; Inoue et al., 2013), Hsp100s prevent

* Corresponding author.
E-mail addresses: ldeluna@ibt.unam.mx (L.A. de Luna-Valdez), carlos.i.villasenor@gmail.com (C.I. Villaseñor-Salmerón), eliza@ibt.unam.mx (E. Cordoba), rosario@ibt.unam.m (R. Vera-Estrella), patricia@ibt.unam.mx (P. León-Mejía), aguevara@ibt.unam.mx (A.A. Guevara-García).

https://doi.org/10.1016/j.plaphy.2019.03.027
Received 8 December 2018; Received in revised form 26 February 2019; Accepted 17 March 2019
Available online 20 March 2019
0981-9428/

Nomenclature

A. thaliana CGEs	AtCGEs
BiFC	Bimolecular Fluorescence Complementation
RbcL	Ribulose Bisphosphate Carboxylase/Oxygenase large chain
CGEs	chloroplast GrpE proteins
CrCGE1	C. reinhardtii CGE1
cYFP	Yellow Fluorescent Protein C-terminus
DXS	1-deoxy-D-xylulose-5-phosphate synthase)
Hip	Hsp70 interacting protein
Hop	Hsp70-Hsp90-organizing protein
NdhH	NAD(P)H-quinone oxidoreductase subunit H
nYFP	Yellow Fluorescent Protein N-terminus
pCE	pSPYCE vector
pD14	pDEST14 vector
pD22	pDEST22 vector
pD32	pDEST32 vector
pEG	pEarleyGate103 vector
pNE	pSPYNE vector
TCA	Trichloroacetic acid
VIPP1	Vesicle-inducing Protein in Plastids 1

protein aggregation and facilitate protein import and degradation (Kovacheva et al., 2007; Doyle and Wickner, 2009; Olinares et al., 2011; Clarke, 2012; Sjogren et al., 2014), and Hsp70s (cpHsc70-1 and cpHsc70-2) participate in thylakoid formation, protein import and photosystem assembly (Yalovsky et al., 1992; Schroda et al., 1999; Liu et al., 2007; Su and Li, 2008; Shi and Theg, 2010; Su and Li, 2010).

Mutant plants defective in chloroplast chaperones display pigment-defective phenotypes and contain aberrant chloroplasts. For example, *Arabidopsis thaliana* mutants in Cpn60α display albino phenotypes with impaired growth (Apuya et al., 2001), and point mutations in the Hsp90-5 gene also generate albino seedlings while null mutants are embryo-lethal (Cao et al., 2003; Inoue et al., 2013). The knockout of the major isoform of Hps93/ClpC causes chloroplast development and protein import defects (Constan et al., 2004; Kovacheva et al., 2005, 2007). Finally, mutation of cpHsc70-1 gene causes cotyledon and leaf variegation with chloroplast protein import defects, and double mutants for the two plastidic Hsc70s are embryo lethal (Su and Li, 2008).

To function efficiently, chaperones interact with other proteins known as cochaperones (Fink, 1999; Mayer, 2010). In chloroplasts, the cochaperonins Cpn10 and Cpn20 form hetero-oligomeric complexes to form the lid of the Cpn60 barrel-shaped oligomer (Tsai et al., 2012). Also, the C-terminus of Tic40 has homology to the proteins Hip (Hsp70 interacting protein) and Hop (Hsp70-Hsp90-organizing protein), which are known cochaperones that bind and organize Hsp70 and Hsp90 in the cytoplasm (Stahl et al., 1999; Bédard et al., 2007). In accordance, Tic40 has been found to directly interact with Hsp93/ClpC to enhance ATP hydrolysis (Chou et al., 2003).

Additionally, other cochaperones have been identified in chloroplasts, including DnaJ and GrpE domain-containing proteins. In *Escherichia coli,* such cochaperones directly interact with DnaK (Hsp70) to mediate protein folding. DnaJ is known to recruit target proteins and enhance the ATP hydrolysis of DnaK, while GrpE is a nucleotide exchange factor that regulates the exchange of ADP to ATP in the nucleotide-binding site of DnaK, thus regulating the affinity of the chaperone for the unfolded polypeptides (Fink, 1999; Mayer, 2010). In plastids, 19 DnaJ domain-containing proteins have been identified (Chiu et al., 2013), recent studies have found that mutant plants defective in individual DnaJ proteins (AtJ8, AtJ11, and AtJ20) display defects in PSII (Photosystem II) homeostasis in *A. thaliana* (Chen et al., 2010). Also, AtJ20 was found involved in the turnover of DXS (1-deoxy-D-xylulose-5-phosphate synthase) enzyme in *A. thaliana* (Pulido et al., 2013), whereas the DnaJ-containing protein CDJ2 from *Chlamydomonas reinhardtii* is required for the formation of the essential VIPP1 (Vesicle-inducing Protein in Plastids 1) oligomers during thylakoid biogenesis (Liu et al., 2005, 2007).

In contrast, the information regarding the function of the chloroplast GrpE proteins (CGEs) is fragmentary, with a comprehensive description available only for the *C. reinhardtii* CGE1. It is known that the CGE1 from *C. reinhardtii* is localized in the chloroplast stroma, where it interacts with HSP70B (the chloroplastic Hsp70 from *C. reinhardtii*), forms homodimers and participates in both protein import and VIPP1 oligomerization (Schroda et al., 2001; Liu et al., 2007; Willmund et al., 2007). Additionally, in *Physcomitrella patens* CGE proteins are localized in the chloroplast, where they interact with the stromal Hsp70-2

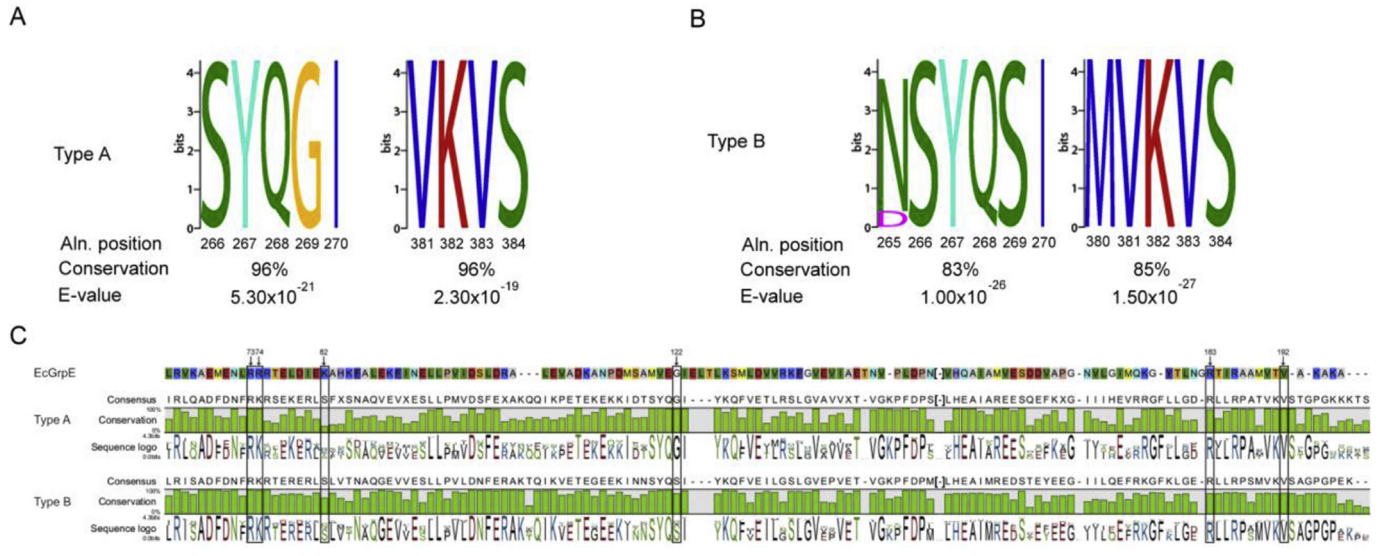


Fig. 1. Conserved motifs in CGE proteins. Sequence motifs found enriched in Type A (A) and Type B (B) CGEs based on the analysis with DREME tool (http://meme-suite.org/doc/dreme.html?man_type=web). The input sequences were randomized to generate a control set, the E-value threshold for motif discovery was set to 0.05. Alignment between the *E. coli* GrpE and the consensus sequences of Type A and Type B CGEs is shown (C). The numbers, arrows and boxes over positions 73, 74, 82, 122, 183, and 192 point to the amino acids that have been shown to be important for GrpE activity in *E. coli*.

protein to regulate protein import into the organelle (Shi and Theg, 2010). However, no experimental evidence exists that analyzes CGE function from land plants. This study presents data that gives insight into the functional traits of the *A. thaliana* CGEs, it is focused on the *in vivo* determination of several CGE protein traits that are relevant to its proposed function as a cochaperones for cpHsc70-1 protein, such as activity as nucleotide exchange factor, subcellular localization, physical interaction with cpHsc70 and dimer formation, it also presents information regarding the biological processes mediated by CGEs in *A. thaliana*.

2. Results

2.1. CGE proteins share functionally-relevant amino acid residues

To gain understanding of the evolution of CGEs, we performed a phylogenetic analysis of 136 CGE protein sequences from 71 species (61 embryophytes: 60 tracheophytes and 1 bryophyte, 7 chlorophytes, 1 charophyte, and 1 cryptophyte) (Fig. S1). A full alignment between all the analyzed CGEs and the *E. coli* GrpE is shown in the supplemental file CGEs full alignment. This analysis showed that the CGE protein sequences from embryophytes form a monophyletic group that further dives into two subgroups, named Type A and Type B. Analysis of the phylogenetic tree showed that 57% of the plants contain both Type A and Type B CGEs, while 18% contain only Type A and 24% only Type B CGEs. The two subgroups of CGEs can be distinguished from each other by variations in short motifs conserved in the analyzed CGE protein sequences. The SYQGI and VKVS motifs are present in Type A CGEs, and the (N/D)SYQSI and MVKVS are characteristic of Type B sequences (Fig. 1A and B). Furthermore, sequence comparison with the *E. coli* GrpE showed that two of the six amino acid positions essential for GrpE activity (Harrison et al., 1997; Gelinas et al., 2003, 2004) have been modified in the plant sequences. In embryophytes, amino acid position R74 was substituted by a K residue, and position K82 by an S residue (Fig. 1C and Table S2). Moreover, the G residue at position 122 of the *E. coli* GrpE is only conserved in Type A CGEs, while it has been substituted by S in 86% of the Type B CGEs (Fig. 1C and Table S2). The remaining functionally-relevant amino acids (R73, R183, and V192) from the bacterial GrpE are fully conserved among all the CGEs analyzed (Fig. 1C and Table S2). Finally, we observed that the two *CGE* genes present in the genome of *A. thaliana* belong to each of the two subgroups; *CGE2* to Type A and *CGE1/EMB1241* to Type B.

2.2. A. thaliana type A and type B CGEs are functional nucleotide exchange factors

To assess the functionality of the *A. thaliana* CGEs (AtCGEs), *AtCGE1* and *AtCGE2* genes were used to complement an *E. coli* knockout line (OD212) that is defective in the *grpE* gene. This mutant carries an additional mutation the Hsp70 gene (*dnaK332*) that suppresses the lethal phenotype conferred by the knockout of the endogenous *grpE*. Thus, OD212 line can grow normally at 25 °C but has growth defects when cultured at higher temperatures (Deloche et al., 1997). The predicted transit peptides were deleted from the CDSs of *AtCGE1* and *AtCGE2* genes, and the Shine-Dalgarno sequence was added to the genetic constructs to facilitate bacterial translation; the chimeric genes were cloned into the pDEST14 bacterial expression vector, generating the constructs *pD14::ΔCGE1* and *pD14::ΔCGE2* (Fig. S2). These constructs were transformed into OD212 cells and their phenotype was evaluated under permissive (25 °C), mild (37 °C), and restrictive (43 °C) temperature conditions (Fig. 2). The empty *pD14::ΔccdB* vector was used as negative control and the full lenght CDS of the *E. coli grpE* gene was cloned into pDEST14 and used as positive control (*pD14::EcgrpE*). Under permissive conditions, no detectable differences in growth were observed between the OD212 strain transformed with the empty vector and those carrying *pD14::ΔCGE1*, *pD14::ΔCGE2* and *pD14::EcgrpE* plasmids (Fig. 2). However, under mild temperature conditions (37 °C) improved growth was observed in the cells containing *pD14::ΔCGE1*, *pD14::ΔCGE2* and *pD14::EcgrpE* plasmids compared with the empty vector (Fig. 2). At the restrictive conditions (43 °C), the growth defects displayed by all the strains were more severe; however, cells transformed with *pD14::ΔCGE1*, *pD14::ΔCGE2* and *pD14::EcgrpE* plasmids performed better than the empty vector-transformed cells (Fig. 2). Furthermore, *pD14::CGE2*-carrying cells displayed improved growth compared with those containing *pD14::CGE1* (Fig. 2), indicating a better complementation capacity for *AtCGE2* compared to *AtCGE1*. These results demonstrate that the AtCGEs alleviate the defective phenotype of the OD212 mutant; supporting the idea that both AtCGEs retain key molecular traits that allow interaction with DnaK, to promote the essential nucleotide exchange needed by this chaperone.

2.3. A. thaliana CGEs are chloroplast localized

To investigate the role of CGEs as cochaperones for chloroplastic Hsp70s, their subcellular localization was determined *in vivo* and contrasted to the localization of the chloroplastic Hsp70, the cpHsc70-1 protein. Translational fusions of AtCGE1, AtCGE2, and cpHsc70-1 to

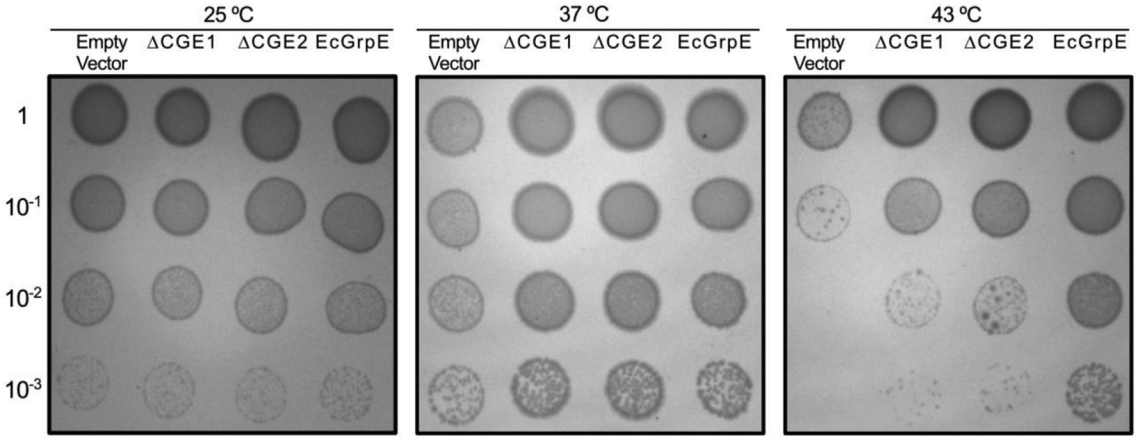


Fig. 2. Functional complementation assay in bacteria. The growth of the thermo-sensitive *E. coli* OD212 strain, transformed with the pD14:Δ*ccdB* (empty vector) or the expression vectors pD14:Δ*CGE1* (Δ*CGE1*), pD14:Δ*CGE2* (Δ*CGE2*) or pD14:*EcgrpE* (EcGrpE), was analyzed at 25 °C (A), 37 °C (B), and 43 °C (C). Five μL drops from the indicated bacterial samples were cultured for 20 h in LB-agar plates at the corresponding dilutions (1, OD_{600nm} = 0.1; 10^{-1}, OD_{600nm} = 0.01; 10^{-2}, OD_{600nm} = 0.001; 10^{-3}, OD_{600nm} = 0.0001) and temperature conditions.

GFP and c-Myc tag were generated (Fig. S2) and used to transform *N. benthamiana* leaves. Protoplasts from transformed leaves were analyzed 96 h post-transformation under the confocal microscope. The protoplasts expressing AtCGE1-GFP (Fig. 3B) or AtCGE2-GFP (Fig. 3C) fusions displayed GFP fluorescence distributed inside the chloroplast stroma and in discrete cumuli at the periphery of plastids. The formation of peripheral fluorescence cumuli has been described as chloroplast envelope deformations caused by overaccumulation of membrane-associated proteins (Breuers et al., 2012). Interestingly, protoplasts expressing cpHsc70-1-GFP fusion displayed a distribution pattern characteristic of stromal proteins (Fig. 3D), with intense fluorescence in a few foci inside the organelle (Farmaki et al., 2006; Perello et al., 2016). Additionally, we addressed the possibility of AtCGEs being transported to mitochondria by colocalization analysis of AtCGE1-GFP and AtCGE2-GFP and a translational fusion between mCherry and the yeast mitochondrial protein COX4 (Cytochrome c oxidase subunit 4) (Nelson et al., 2007), but no co-localization was found (Fig. S5). Altogether, these data indicate that both AtCGEs are *in vivo* imported into the chloroplast, where they coexist with cpHsc70-1 chaperone.

To confirm our data and to determine the suborganellar localization of the AtCGEs and cpHsc70-1 proteins, chloroplasts from *N. benthamiana* leaves transformed with *pNE::CGE1*, *pNE::CGE2* or *pNE::CPHSC70-1* constructs were isolated and fractionated into stromal and envelope fractions. Western blot analysis of purified suborganellar fractions was carried out using the c-Myc epitope tag present in the *pNE* expression vectors (Fig. S2). Fig. 4 shows that AtCGE1 and AtCGE2 proteins accumulate in both the chloroplast envelope and in the stromal fractions (Fig. 4: lanes 3 and 5), while cpHsc70-1 protein was only detected in the stromal fractions (Fig. 4 stromal fractions: lanes 2 and 4). Western blots against RbcS (stromal protein marker) and Tic40

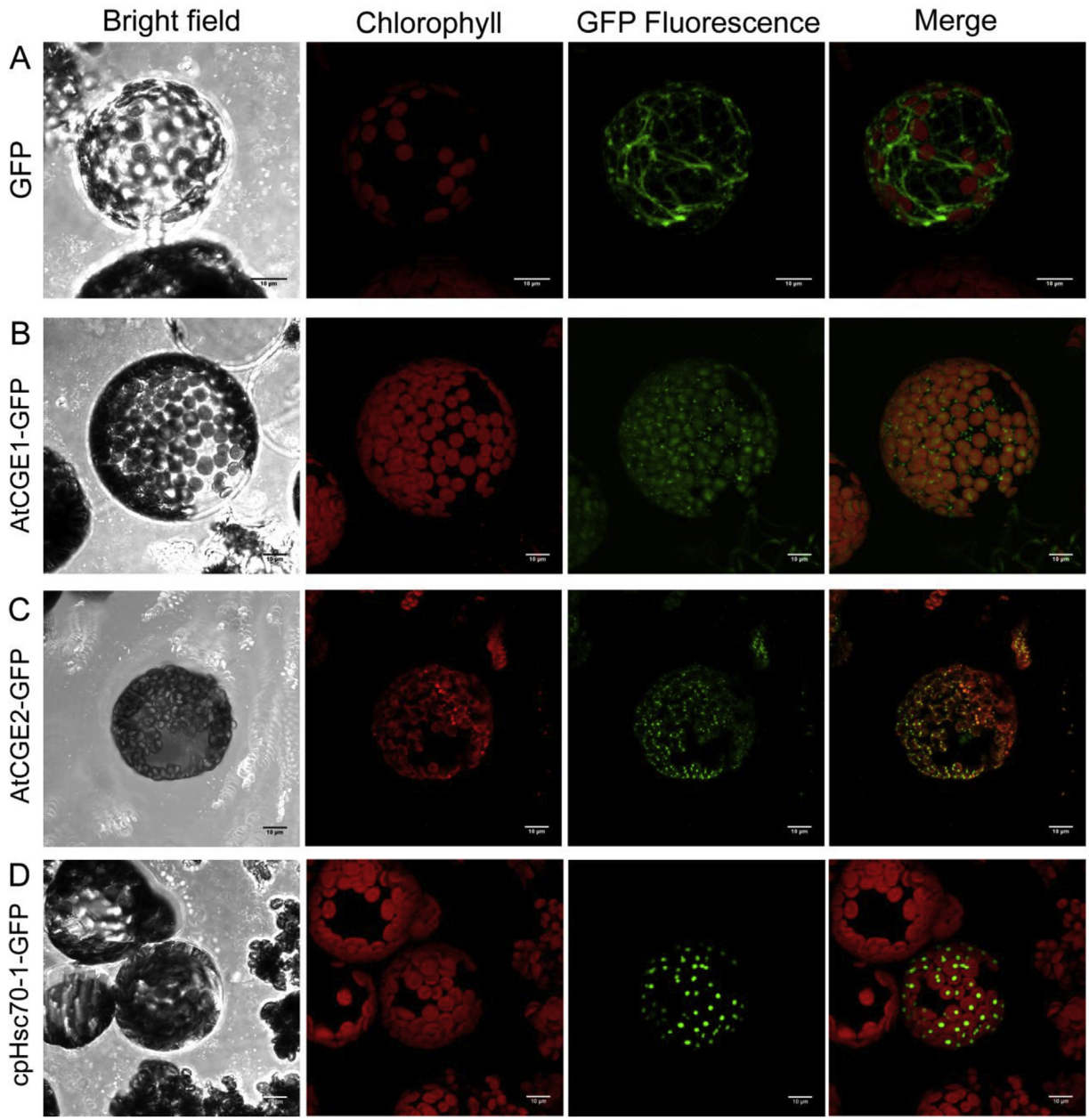


Fig. 3. Subcellular localization of AtCGE1, AtCGE2 and cpHsc70–1 proteins. Images of mesophyll protoplasts from *N. benthamiana* leaves expressing GFP (A) and the translational fusions AtCGE1-GFP (B), AtCGE2-GFP (C), and cpHsc70-1-GFP (D) are shown. Images corresponding to the bright field, chlorophyll fluorescence (Chlorophyll), GFP fluorescence, and the merge between the two fluorescence channels are shown in the indicated columns. Protoplasts were prepared 96 h after leaf agroinfiltration. Scale bars correspond to 10 μm.

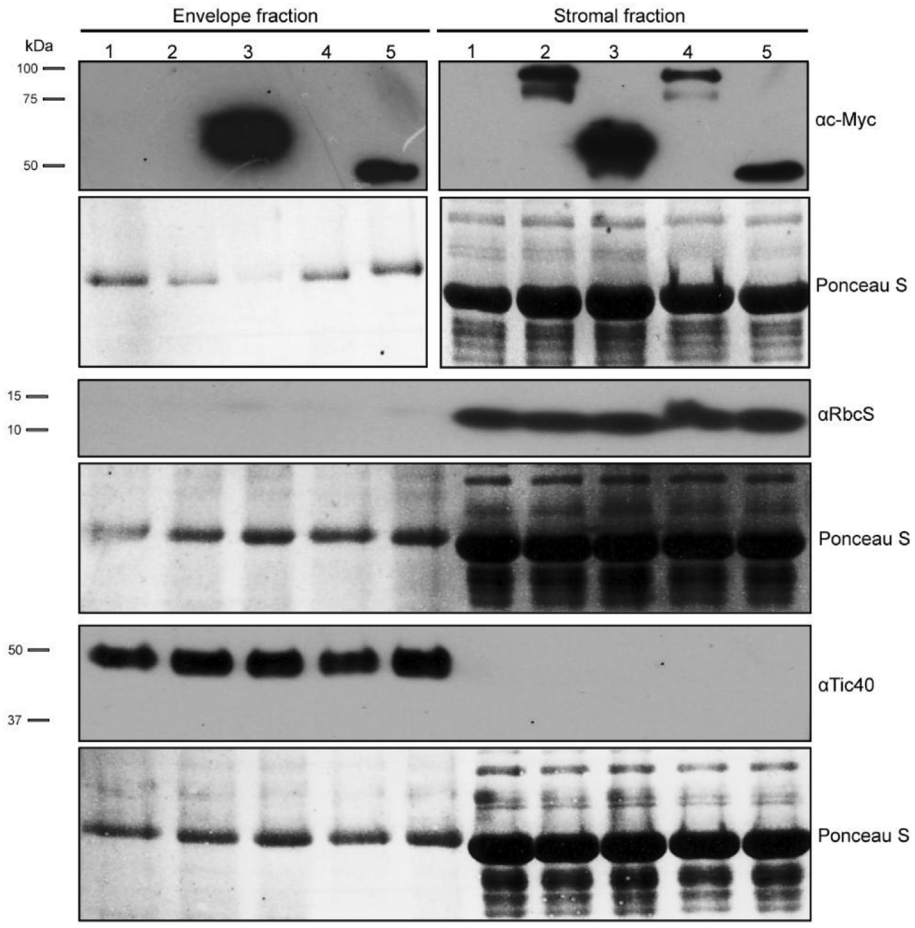


Fig. 4. Suborganellar localization of AtCGE1, AtCGE2 and cpHsc70–1 proteins. Western blot analysis of chloroplast proteins from purified chloroplast envelope and stroma from not agroinfiltrated leaves (lane 1) or leaves agroinfiltrated with *pNE::CPHSC70-1* (lanes 2 and 4), *pNE::CGE1* (lane 3), and *pNE::CGE2* (lane 5). Immunodetection of the c-Myc tagged AtCGE1, AtCGE2 and cpHsc70-1 protein fusions (αc-Myc), RbcS (αRbcS), and Tic40 (αTic40) proteins in the suborganellar samples is shown. The Ponceau S staining of the protein-containing membranes is shown as loading control. Molecular weight marker (kDa) is shown at the left. αc-Myc western-blots were performed on independent protein membranes.

(envelope protein marker) confirmed that no cross-contamination between the suborganellar fractions existed in the samples. This information corroborates the confocal microscopy data (Fig. 3) and demonstrates the co-localization between AtCGEs and cpHsc70-1 in the chloroplast stroma, supporting the hypothesis of a common function for these proteins.

2.4. A. thaliana CGEs interact with cpHsc70-1 in vivo

To analyze the existence of a physical interaction between the AtCGEs and the cpHsc70-1 chaperone *in vivo,* BiFC (Bimolecular Fluorescence Complementation) assays were performed. *N. benthamiana* leaves were transformed by agroinfiltration with the plasmids containing cpHsc70-1, AtCGE1 and AtCGE2 proteins fused to the N- (nYFP) and C-terminal (cYFP) regions of YFP protein (Fig. S2). Reciprocal co-transformation assays (i.e. cpHsc70-1-nYFP/AtCGE1-cYFP and cpHsc70-1-cYFP/AtCGE1-nYFP) were performed to corroborate the complementation of fluorescence. Fig. 5 shows that fused to nYFP or to cYFP, AtCGE1 directly interacts with cpHsc70-1 chaperone, as demonstrated by the presence of YFP fluorescence in the chloroplasts of transformed plant cells (Fig. 5A and B). Similar results were found for AtCGE2 and cpHsc70-1, the cells co-transformed with the nYFP and cYFP protein fusions displayed YFP fluorescence complementation in the chloroplast stroma (Fig. 5C and D). Furthermore, *N. benthamiana* protoplasts transformed with the individual translational fusions (i.e only AtCGE1-nYFP or AtCGE2-cYFP) were used as negative control for these experiments. Alternatively, using c-Myc and GFP-tagged proteins, we detected co-immunoprecipitation of AtCGE1 and AtCGE2 with cpHsc70-1 (Fig. S6), corroborating the interactions detected in the BiFC assays. Altogether, these observations demonstrate that the two AtCGEs form stable physical interactions with the cpHsc70-1 chaperone inside the chloroplasts *in vivo*.

2.5. A. thaliana CGEs form homo and heterodimers

It is known that bacterial GrpE interacts with DnaK as a homodimer, and this conformation is key for the regulation of the DnaK reaction cycle (Harrison et al., 1997). Dimerization has been reported for *C. reinhardtii* CGE1 (CrCGE1) (Schroda et al., 2001). To test the capacity of AtCGEs to form dimers, BiFC experiments were performed in *N. benthamiana* leaves transformed with the expression vectors containing AtCGE1 and AtCGE2 proteins fused to the cYFP and nYFP moieties of YFP protein (Fig. S2). Complementation of YFP fluorescence was detected in the chloroplasts of the cells transformed with the plasmids containing the fusions AtCGE1-cYFP and AtCGE1-nYFP (Fig. 6A), indicating that AtCGE1 protein forms homodimers *in vivo.* Similarly, YFP fluorescence was detected in the chloroplasts of the cells expressing the fusions AtCGE2-cYFP and AtCGE2-nYFP (Fig. 6B), demonstrating that AtCGE2 is capable of homodimerization *in vivo* too. Given that AtCGE1 and AtCGE2 belong to different clades of the CGE phylogenetic tree (Fig. S1), we were also interested in analyzing their ability to form heterodimers. YFP fluorescence was detected in the chloroplasts of the cells expressing the protein fusions AtCGE1-cYFP and AtCGE2-nYFP (Fig. 6C), and AtCGE2-cYFP and AtCGE1-nYFP (Fig. 6D), demonstrating that AtCGE1 and AtCGE2 can form heterodimers *in vivo*. Alternatively, using c-Myc and GFP-tagged proteins, we detected co-immunoprecipitation of AtCGE1 and AtCGE2 with AtCGE1, and AtCGE2 with AtCGE2 (Fig. S6), corroborating the interactions detected in the BiFC assays. In conclusion, these experiments support that the AtCGE1 and AtCGE2 proteins can form homodimers and heterodimers as well. Finally, as no data suggests the formation of cpHsc70-1 homodimers, protoplasts from *N. benthamiana* leaves transformed using the plasmids

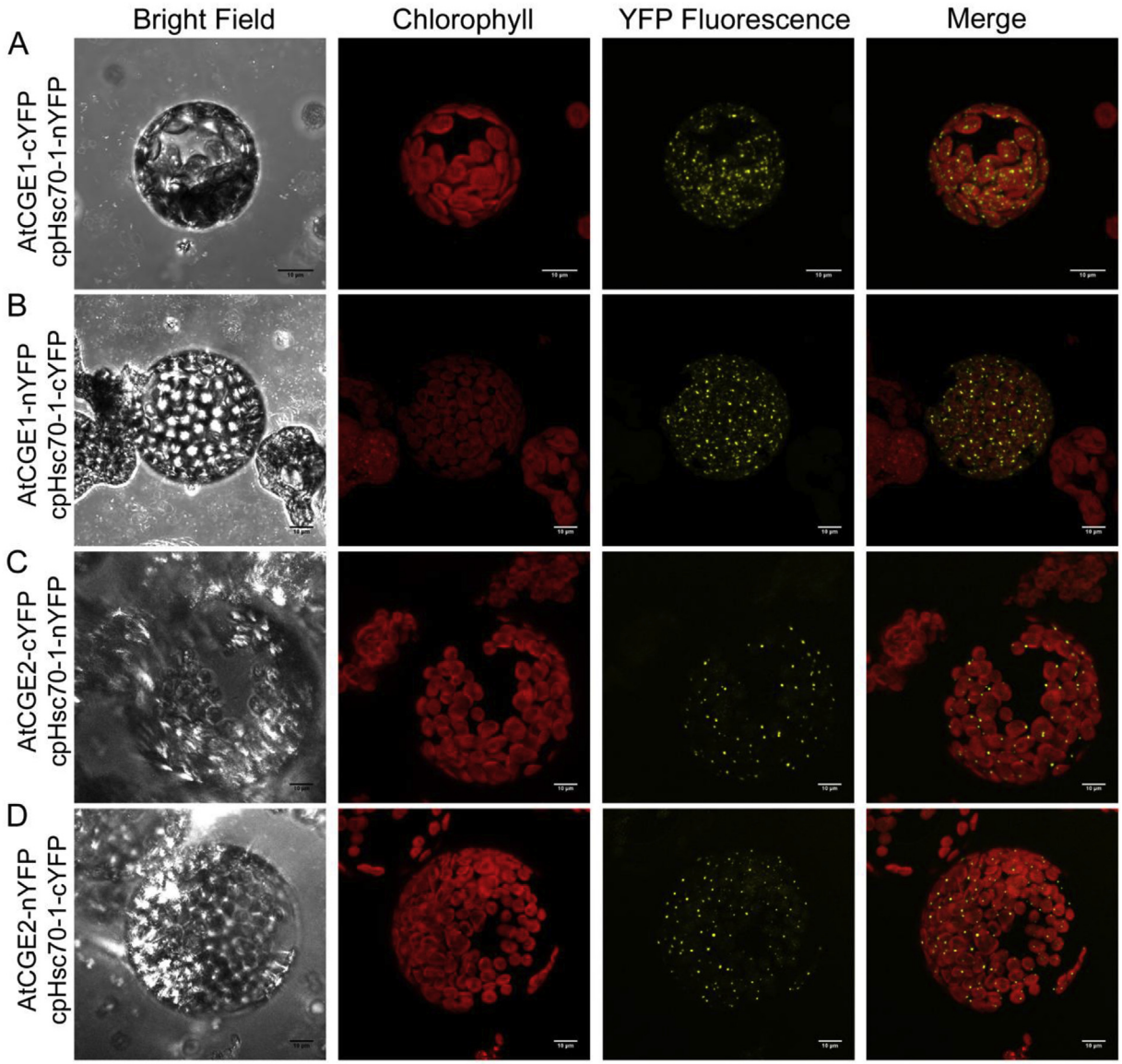


Fig. 5. *In vivo* interaction between AtCGEs and cpHsc70–1 proteins. Mesophyll protoplasts from *N. benthamiana* leaves co-expressing the translational fusions cpHsc70-1-nYFP and AtCGE1-cYFP (A), AtCGE1-nYFP and cpHsc70-1-cYFP (B), cpHsc70-1-nYFP and AtCGE2-cYFP (C), and AtCGE2-nYFP and cpHsc70-1-cYFP (D) are shown. Images corresponding to the bright field, chlorophyll fluorescence (Chlorophyll), reconstituted YFP fluorescence (YFP fluorescence), and the merge between the two fluorescence channels are shown in the indicated columns. Protoplasts were prepared 96 h after the agroinfiltration of the *N. benthamiana* leaves. Scale bars correspond to 10 μm. Negative controls are shown in Fig. S4.

containing the fusions cpHsc70-1-cYFP and cpHsc70-1-nYFP were used as a control, no fluorescence was detected in protoplasts from such leaves (Fig. 6E), indicating that the overexpression of proteins targeted to the chloroplast stroma does not result in unspecific interactions between the two moieties of YFP, validating the BiFC observations presented in Figs. 5 and 6.

2.6. CGE1 and CGE2 genes are differentially regulated in response to heat stress

In *E. coli*, *grpE* and *dnaK* genes are transcriptionally activated upon heat stress (Yura et al., 1993), a regulation that is conserved in the *CGE1* and *HSP70B* genes of *C. reinhardtii* but not in mitochondrial GrpEs from yeast or rat (Ang et al., 1986; Ikeda et al., 1994; Naylor et al., 1996; Schroda et al., 2001). To determine the transcriptional response of *AtCGEs* and *CPHSC70-1* genes to heat stress in seedlings, we first determined by Northern-blot the mRNA accumulation levels of *AtCGE1*, *AtCGE2* and *CPHSC70-1* under normal growth conditions in several developmental stages of wild-type seedlings (Fig. 7A and B). We found that *AtCGE1*, *AtCGE2* and *CPHSC70-1* genes are co-expressed at 8, 10, 12, and 14 days after germination (Fig. 7A and B). We also observed that the mRNA levels of *AtCGE2* and *CPHSC70-1* increase after 8 days of development and remain the same for the rest of the evaluated time points; furthermore, we found that the mRNA levels of *AtCGE1* are consistently downregulated 12 days after germination, but the initial levels are recovered later during development (Fig. 7A and B). Considering the downregulation of *AtCGE1* in 12-day-old seedlings, we decided to perform the heat stress experiments in 14-day-old seedlings exposed to 40 °C for 30, 60 and 90 min, and the mRNA levels of *AtCGE1*, *AtCGE2* and *CPHSC70-1* were analyzed by Northern blot. As shown in Fig. 7, *AtCGE1* transcript levels do not significantly change in response to heat treatment (Fig. 7C and D). In contrast, At*CGE2* mRNA levels decrease after 30 min of heat treatment (Fig. 7C and D). This response was like the observed in *DXS1*, a transcript that encodes a chloroplast-

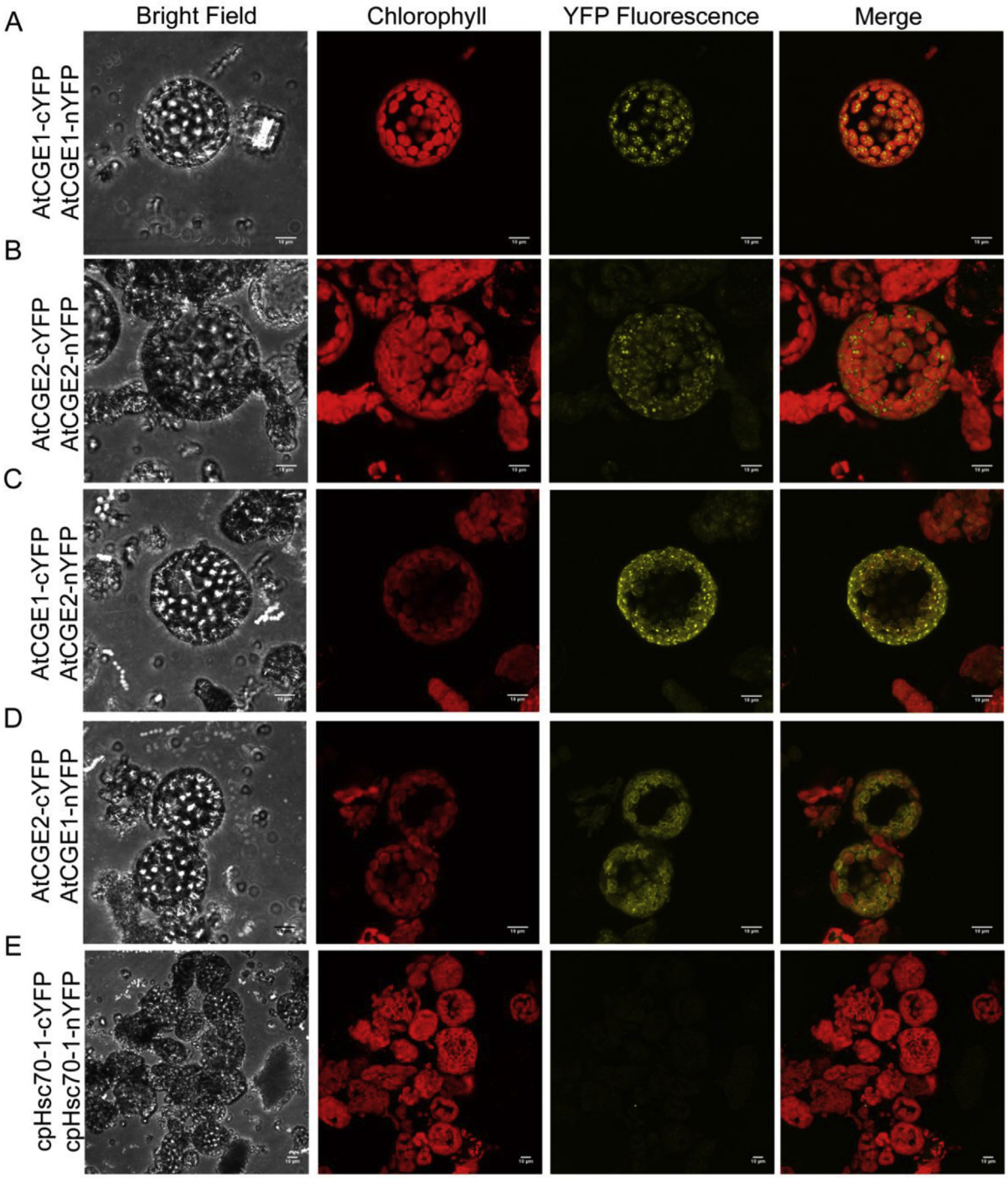


Fig. 6. *In vivo* determination of AtCGE1 and AtCGE2 dimer formation. Mesophyll protoplasts from *N. benthamiana* leaves co-expressing the translational fusions AtCGE1-nYFP and AtCGE1-cYFP (A), AtCGE2-nYFP and AtCGE2-cYFP (B), AtCGE2-nYFP and AtCGE1-cYFP (C), AtCGE1-nYFP and AtCGE2-cYFP (D), and cpHsc70-1-nYFP and cpHsc70-1-cYFP (E). Images corresponding to the bright field, chlorophyll fluorescence (Chlorophyll), reconstituted YFP fluorescence (YFP fluorescence), and the merge between the two fluorescence channels are shown in the indicated columns. Protoplasts were prepared 96 h after the agroinfiltration of the *N. benthamiana* leaves. Scale bars correspond to 10 μm.

localized protein with no chaperone function that was used as a control of the treatment (Fig. 7C and D). Finally, the abundance of *CPHSC70-1* mRNA shows downregulation after 30 and 60 min of heat stress but increases 90 min after the onset of the heat treatment (Fig. 7C and D); this response agrees with previous reports (Schroda et al., 2001). To our understanding, the fact that *AtCGE1* transcript levels do not change in response to heat stress, suggests that the AtCGE1 protein levels remain unchanged during the development of the stress response; at the same time, the downregulation of *AtCGE2* mRNA accumulation suggests that AtCGE2 protein levels decrease in response to heat stress. This hypothesis suggests that there are specific molecular mechanisms that differentially regulate the transcription or mRNA accumulation of the two AtCGEs; specifically, during the heat stress response, these mechanisms increase the AtCGE1 to AtCGE2 ratio, favoring the interactions between AtCGE1 and cpHsc70-1. These data suggest that AtCGE1 is the major AtCGE involved in the chloroplast's response to heat stress,

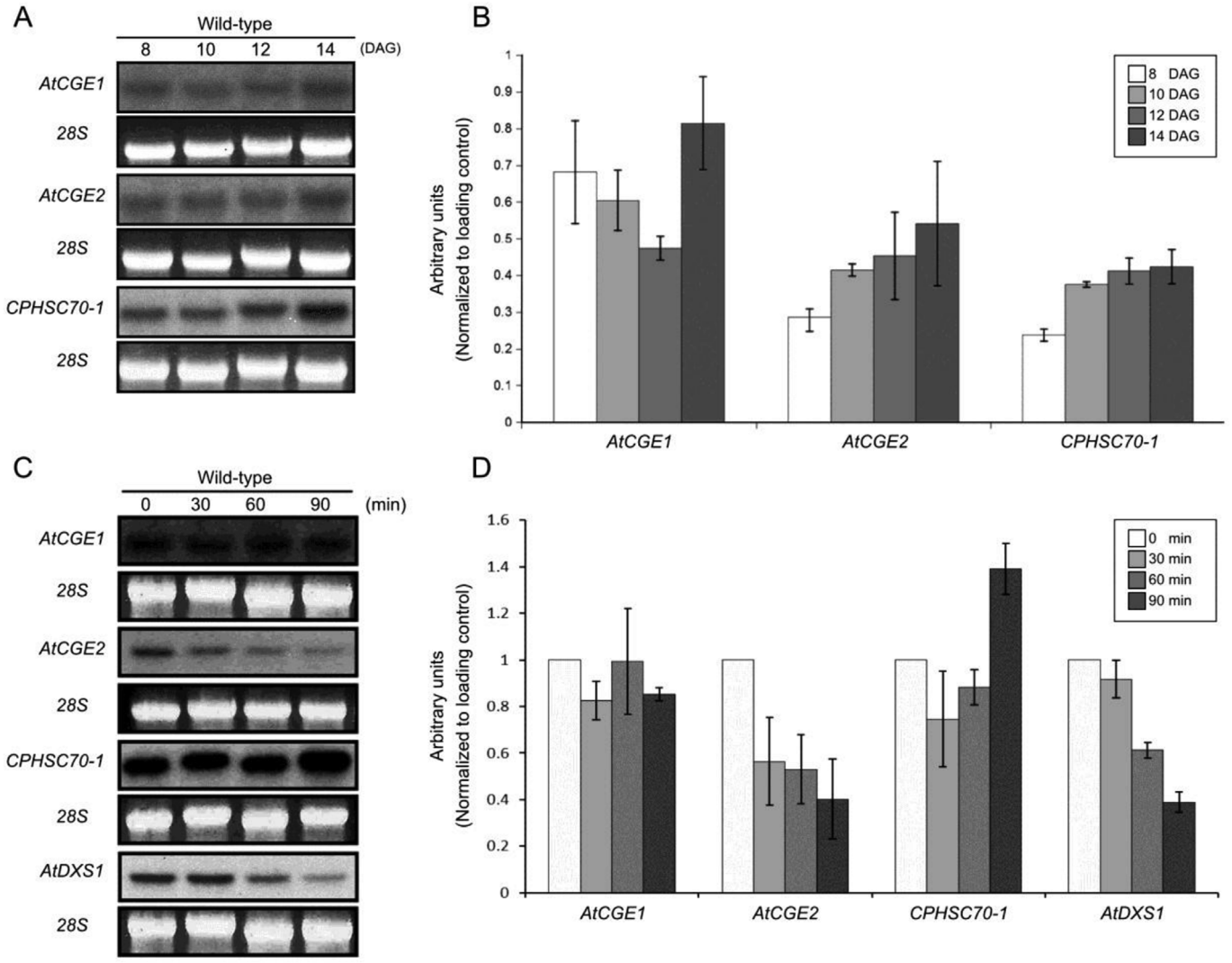


Fig. 7. Expression analysis of *AtCGE1*, *AtCGE2,* and *cpHsc70-1*. Northern blot hybridization of *AtCGE1*, *AtCGE2* and *CPHSC70-1* transcripts in four stages of seedling development (A), and the corresponding densitometry analysis is shown (B). Northern blot hybridization of *AtCGE1*, *AtCGE2*, *CPHSC70-1* and *DXS1* transcripts from wild-type seedlings after 0, 30, 60, and 90 min of heat stress(C). Densitometry analysis of the northern-blot data is shown considering 0 min as a reference (D). For each condition10 μg of total RNA were loaded and the ethidium bromide staining of the rRNA 28S is shown as a loading control. The DNA probes used are shown in Fig. S3. Error bars represent the standard error calculated from three independent experiments.

while AtCGE2 might be involved in a different biological process. However, the post-transcriptional regulation that controls the abundance of AtCGE proteins under heat stress remains to be analyzed.

2.7. *cge1 mutant plants have distinctive phenotypes*

To further explore the physiological roles of AtCGEs, we performed a phenotypic characterization of mutant plants for *AtCGE1*, and *CPHSC70-1* genes. As previously described, homozygous T-DNA insertion mutants in the *AtCGE1* gene (*emb1241-1* and *emb1241-2* mutants) are embryo-lethal (Meinke et al., 2008), preventing the phenotypic characterization of such plants. Thus, we analyzed the available heterozygote mutant lines *emb1241-1(±)*, *emb1241-2(±)*, and *cge2-1(±)*, and the homozygous mutant *Dcphsc70-1* (Fig. S3). The pigment-content analysis in these mutants showed a significant decrease in the content of chlorophyll *b* (chl *b*) in *Dcphsc70-1* and *emb1241-2(±)* mutant lines (Fig. 8A and Table S3). In contrast, no significant differences in pigment content were found in the *emb1241-1(±)* plants (Fig. 8A and Table S3). These results indicate the presence of alterations in the photosynthetic apparatus of *emb1241-2(±)* and *Dcphsc70-1* plants.

To support these findings, the accumulation of the photosystem I psaD1 and photosystem II D1 reaction center proteins was analyzed by Western blot in *emb1241-2(±)* and *Dcphsc70-1* plants. In Fig. 8B, we observed a higher accumulation of psaD1 and D1 in *emb1241-2(±)* and *Dcphsc70-1* mutants than in the wild-type plants (Fig. 8B), suggesting that the stoichiometry of the reaction centers or other parts of the photosynthetic apparatus might be altered in these plants. Using BN-PAGE, the accumulation of photosynthetic complexes was analyzed in *emb1241-2(±)* and *Dcphsc70-1* mutants. It was observed that both *emb1241-2(±)* and *Dcphsc70-1* accumulate lower levels of the functional oligomers of LHCII than the wild-type plants (Fig. 8C). Accordingly, the levels of LHCII monomer were higher in the *emb1241-2(±)* and *Dcphsc70-1* than in the wild-type plants (Fig. 8C). No significant differences in the abundance of the other major complexes (PSI and PSII) resolved in the gel were detected. These results suggest that cpHsc70-1 and AtCGE1 protein have a role in the proper oligomerization of the LHCII complex.

3. Discussion

Protein function largely depends on the protein's ability to interact

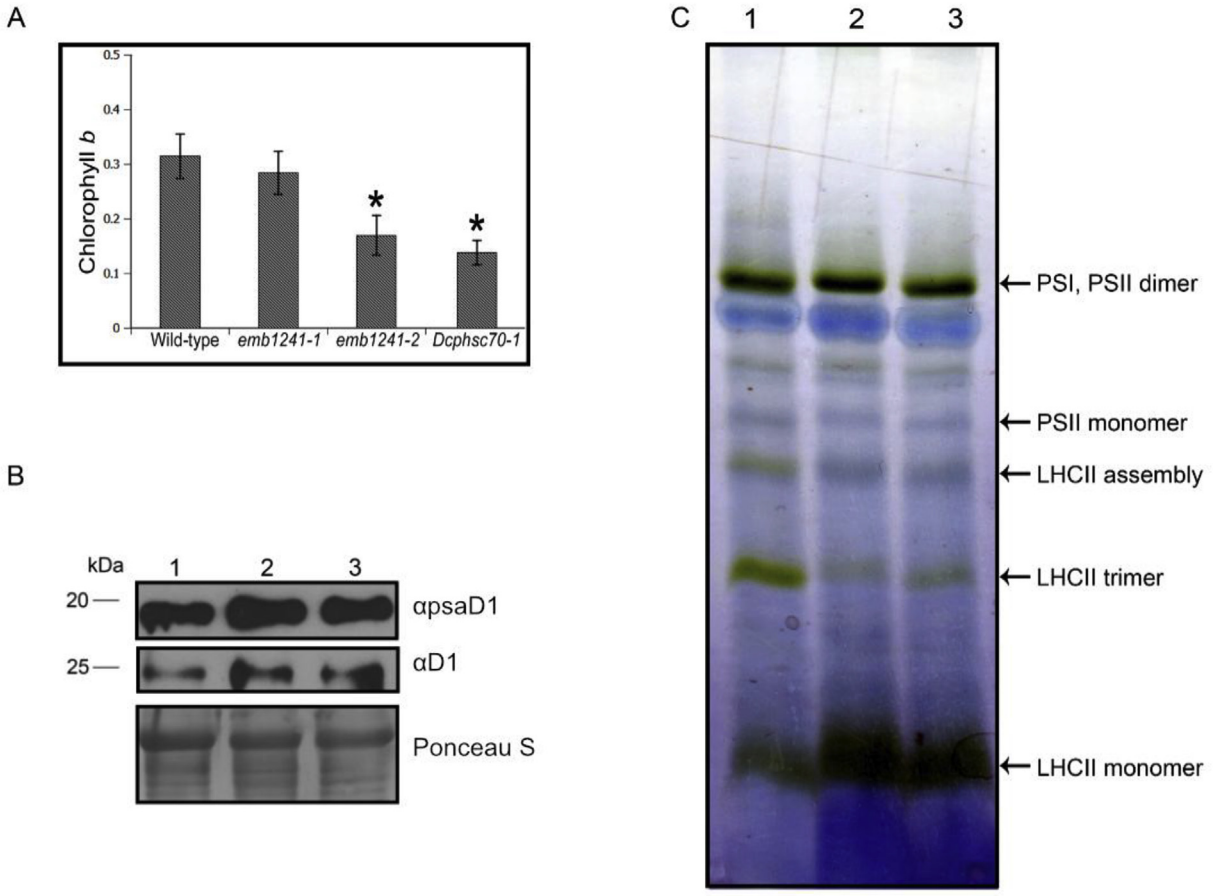


Fig. 8. Biochemical phenotype of *emb1241 (±)* and *Dcphsc70-1* mutants. The Chlorophyll *b* accumulation in wild-type, *emb1241-1 (±)*, *emb1241-2 (±)*, and *Dcphsc70-1* plants is shown (A). The error bars represent the standard error, the symbol (*) over the bars represent statistically significant differences calculated in a one-way analysis of variance with a significance level of $p^{\circ}0.05$. The Immunodetection of psaD1 and D1 proteins (B) in total protein extracted from wild-type (1), *emb1241-2 (±)* (2), *and Dcphsc70-1* (3) adult plants is shown. The Ponceau S staining of the protein-containing membranes is shown as a loading control, molecular weight marker (in kDa) is shown at the left. Blue native-PAGE of total leaf protein (C) from wild-type (1), *emb1241-2 (±)* (2), and *Dcphsc70-1* (3) plants. Arrows point to the bands corresponding to the supramolecular complexes visible in the samples. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

with specific molecular partners to regulate aspects like folding, localization, and degradation; the most well-known protein partners are the chaperones. Chloroplastic chaperones of the Hsp70 family are known to be involved in protein import into the chloroplast and the formation of protein oligomers, and to have roles in cellular and physiological traits such as thylakoid biogenesis and thermotolerance (Schroda et al., 2001; Liu et al., 2007; Su and Li, 2008; Shi and Theg, 2010; Su and Li, 2010). It is known that Hsp70-family proteins work together with DnaJ and GrpE proteins (Mayer, 2010). In the plastids of land plants, it has been demonstrated that chloroplastic DnaJ proteins have roles in the homeostasis of proteins and PSII function (Pulido et al., 2013; Kong et al., 2013). However, the role that GrpE proteins play in the chloroplasts of higher plants has not been analyzed in detail.

Through a phylogenetic analysis, we identified the chloroplastic protein homologues of the bacterial GrpE protein in 61 land plant species, in accordance to previous reports we found two CGEs in *A. thaliana* that were named AtCGE1 and AtCGE2 (Schroda et al., 2001; Shi and Theg, 2010). Using a genetic complementation assay, we established the activity of the CGEs from *A. thaliana in vivo*. Due to the knockout of the *grpE* gene, the *E. coli* mutant strain OD212 has growth defects under heat stress conditions. Here we found that transformation with the coding regions of *AtCGE1* and *AtCGE2* genes can rescue the defective-growth phenotype of the *E. coli* strain OD212 at high temperatures. However, the complementation shown by the cells transformed with the *AtCGEs* is partial, since the heat-resistant phenotype conferred by the expression of the AtCGEs is weaker than the conferred by EcGrpE at 43 °C. This result might be due to differences in the amino acid positions important for the establishment of interactions with DnaK protein of AtCGE1 and AtCGE2, in comparison to the bacterial GrpE. It is known that mutation of six amino acid positions in the bacterial GrpE protein negatively affects the interaction with DnaK (Harrison et al., 1997; Gelinas et al., 2003, 2004); three of the six mentioned positions are conserved between GrpE and CGEs (R73, R183, and V192). However, positions R74 and K82 in the GrpE sequence are substituted by K and S residues respectively in CGEs. It is possible that these amino acid substitutions generate interactions between AtCGEs and DnaK, that are not as stable at 43 °C as the interactions between DnaK and EcGrpE, suggesting that these substitutions are responsible for the partial complementation displayed by the AtCGEs. Furthermore, the amino acid residue corresponding to position 122 of GrpE is the only functionally-relevant position that is different between Type A and Type B CGEs; Type A CGEs (such as AtCGE2) conserve the G residue present in the bacterial GrpE, while Type B CGEs (such as AtCGE1) present an S residue substitution at the corresponding position. The conservation of G122 in Type A CGEs explains the improved complementation displayed by AtCGE2 at 43 °C, apparently this position makes AtCGE2 more similar to the bacterial GrpE and likely results in better DnaK binding in comparison to AtCGE1. These observations suggest that AtCGEs form stable interactions with DnaK to supplement for the lack of endogenous GrpE in *E. coli*. Similar

complementation has been shown for other eukaryotic GrpE functional orthologues, such as the *C. reinhardtii* CGE1 (CrCGE1) and the yeast mitochondrial protein Mge1p (Deloche and Georgopoulos, 1996; Schroda et al., 2001), indicating that the two AtCGEs are functional nucleotide exchange factors that can be considered functional orthologues of the bacterial GrpE.

In addition to the activity and the amino acid conservation, several aspects of the AtCGEs biology were investigated to find a correlation between CGEs and chloroplastic Hsp70s in higher plants, given the fact that there are not available specific αCGE antibodies against the CGEs from higher plants, our main experimental strategies are based on the expression of recombinant proteins in *N. benthamiana* leaves. Using GFP-tagged proteins, colocalization of the chloroplastic homolog of DnaK (cpHsc70-1) and the two AtCGEs was detected, indicating that the two AtCGEs are effectively targeted to the chloroplast. Additionally, our mRNA accumulation experiments showed that *AtCGE1*, *AtCGE2* and *CPHC70-1* genes are co-expressed during seedling development in *A. thaliana*, indicating a strong correlation between the AtCGEs and CPHSC70-1 at the transcriptional and posttranscriptional level. Accordingly, CrCGE1 and the chloroplastic Hsp70 HSP70B, have been shown to accumulate mainly in the chloroplast stroma, indicating that their activities are spatially coordinated in algae (Schroda et al., 2001). In contrast, the present work showed that c-Myc tagged AtCGE1 and AtCGE2 accumulate not only in the stroma but also at the envelope of chloroplasts. In these experiments, c-Myc tagged cpHsc70-1 was only detected in the stromal fraction of the plastids. However, AtCGEs do not display features of integral membrane proteins, thus their localization at the chloroplast envelope must be mediated by interactions with unknown membrane proteins with functions not related to chaperone activity, or that might serve as scaffold to mediate interaction with the cpHsc70-1. In chloroplasts, the stromal side of Tic40 protein has sequence homology to the Hip and Hop cochaperones that are known to bind and regulate the activities of Hsp70 and Hsp90 chaperones (Stahl et al., 1999; Chou et al., 2003; Bédard et al., 2007); this observation suggests that similar interacting proteins might work to regulate the interaction between AtCGEs and cpHsc70-1 at the chloroplast envelope. Additionally, the immunodetection of AtCGEs at the chloroplast envelope corroborates the fluorescence distribution displayed by the translational fusions AtCGEs-GFP, which (in addition to the stromal localization) show the formation of cumuli that resemble envelope deformations caused by the overaccumulation of envelope proteins (Breuers et al., 2012). Altogether, these data demonstrate that the AtCGEs have a dual suborganellar distribution that might be due to interactions with unknown membrane proteins.

BiFC and co-immunoprecipitation assays were performed to demonstrate the existence of direct interactions between CGEs and chloroplastic Hsp70s. These experiments demonstrate that AtCGE1 and AtCGE2 form stable physical interactions with cpHsc70-1 *in vivo* inside the chloroplasts, corroborating the information obtained in the GFP-tagging experiments. Previous work has demonstrated that dimerization of GrpE proteins is fundamental for the establishment of interactions with Hsp70 proteins in *E. coli* and *S. cerevisiae* (Deloche and Georgopoulos, 1996; Wu et al., 1996; Azem et al., 1997). Taking advantage of our genetic constructs for BiFC and GFP-tagged proteins, we showed that the two CGEs from *A. thaliana* can form homodimers *in vivo*, indicating that the interaction mechanism between AtCGEs and cpHsc70-1 might be like the described for *E. coli* and *S. cerevisiae* (Deloche and Georgopoulos, 1996; Wu et al., 1996; Azem et al., 1997). In addition, we found that AtCGE1 and AtCGE2 form heterodimers *in vivo*; however, the relevance to chaperone activity of the formation of such heterodimers remains to be addressed. Together, these results indicate that the chloroplastic GrpE proteins of *A. thaliana* have the specific traits needed to physically interact with cpHsc70-1 and that their biological roles might be functionally linked to the activity of cpHsc70-1.

The activity of DnaK and GrpE proteins has been linked to heat stress responses (Yura et al., 1993). In *C. reinhardtii*, increased accumulation of *HSP70B* transcript has been detected upon incubation at 40 °C (Schroda et al., 2001). However, the transcriptional response of *A. thaliana CPHSC70-1* to heat stress is not fully consistent, with some studies detecting upregulation (Suzuki et al., 2013) and others detecting downregulation (Sung et al., 2001) in response to heat stress. In this regard, the heat stress experiments performed in this study showed overaccumulation of the *CPHSC70-1* mRNA 90 min after the onset of the heat treatment, supporting the physiological data of Su and Li (2008) that suggested the involvement of cpHsc70-1 protein in the thermotolerance of *A. thaliana* seedlings. In *C. reinhardtii*, increased accumulation of *CrCGE1* transcript was found in response to heat stress (Schroda et al., 2001). In our experiments, no significant changes in the accumulation of *AtCGE1* mRNA were detected after the heat treatment, whereas downregulation of the *AtCGE2* transcript was found after the onset of the treatment. The transcriptional response of *AtCGE2* is commonly observed in genes not related to heat stress responses (Rizhsky et al., 2002; Echevarría-Zomeño et al., 2016; Jiang et al., 2017). Based on this transcriptional data, we hypothesized that AtCGE1 protein has a higher likelihood to have a function that is coordinated with cpHsc70-1 in the chloroplastic heat stress responses. Also, the data presented suggest that the two AtCGE proteins might have independent roles in chloroplast biology.

In the past, cpHsc70-1 protein activity has been linked to several biological roles, including thermotolerance, thylakoid biogenesis, protein import and photosystem II assembly (Yalovsky et al., 1992; Schroda et al., 1999; Liu et al., 2007; Su and Li, 2008; Shi and Theg, 2010; Su and Li, 2010). As stated before, Hsp70 proteins work together with DnaJ cochaperones. Accordingly, different DnaJ proteins have been found to have biological roles like those of cpHsc70-1; such as the essential oligomerization of VIPP1 protein during thylakoid biogenesis, and the formation of PSII dimers and PSII-LHCII supercomplexes (Liu et al., 2005, 2007; Chen et al., 2010). To investigate the existence of a shared role for AtCGEs and cpHsc70-1, the accumulation of photosynthetic pigments was determined in leaves of the plant lines *Dcphsc70-1*, and the two alleles of *AtCGE1 emb1241-1(±)* and *emb1241-2(±)*. Such analysis detected a decrease in the content of chlorophyll *b* in the *Dcphsc70-1*, *emb1241-1(±)*, and *emb1241-2(±)* plant lines. As the biochemical phenotype is stronger in *emb1241-2(±)* than in *emb1241-1(±)*, only *emb1241-2(±)* plants were used for further analysis. In photosynthetic organisms, chlorophyll *b* is mainly associated with the LHCII complex of PSII (Green and Durnford, 1996; Kitajima and Hogan, 2003; Mascia et al., 2017). In agreement, the BN-PAGE analysis showed that the intensity of the protein bands corresponding to the LHCII trimer and the LHCII assembly from the *emb1241-2(±)* and *Dcphsc70-1* mutants, is reduced in comparison to the intensity of the corresponding bands in wild-type plants. In contrast, the abundance of the band corresponding to the monomeric state of the LHCII is increased in both mutants compared to the wild-type. This observation indicates that similar to what is observed for cpHsc70-1, the AtCGE1 protein is necessary for the correct assembling of the supramolecular organization of the LHCII complex, further supporting that the activity of AtCGE1 is functionally linked to that of cpHsc70-1 in *A. thaliana*.

As reported before, mutation of *CGE1* causes defects that result in the stunning of embryo development at the preglobular stage (Meinke et al., 2008). The fact that homozygous mutants for *AtCGE1* are embryo-lethal indicates that AtCGE2 protein is not able to supplement the absence of AtCGE1 protein. Furthermore, the phylogenetic analysis performed here shows that 35 plant species contain both Type A and Type B CGEs, while 26 species contain only one type of CGE protein (11 species have Type A and 15 species have Type B). These observations indicate that either type of CGE can perform all the functions of the two types of CGEs in the species that have only one type of CGE protein, a phenomenon that does not happen in the plant species that have the two types of CGEs, such as *A. thaliana*. This hypothesis suggests that

CGEs from species containing only one type of CGE must have different biochemical or molecular properties compared to CGEs from two type-containing species, that enables them to carry out the entire set of biological functions associated to CGE activity. These fundamental regulatory aspects of CGE biology will be investigated in detail in future research. Finally, the fact that the two AtCGEs are functional nucleotide-exchange factors that establish direct physical interactions with cpHsc70-1, suggests that despite their different biological roles, the two AtCGEs exert their functions through the enhancement of the activity of Hsp70 chaperones.

In conclusion, AtCGE1 and AtCGE2 proteins are the chloroplastic functional orthologues of the bacterial GrpE protein, they physically interact with cpHsc70-1 and mediate independent biological processes. We have evidence that suggests the involvement of AtCGE1, but not AtCGE2, in the homeostasis of LHCII functional oligomers.

4. Materials and methods

4.1. Plant material and growth conditions

Nicotiana benthamiana plants were cultured for 6 weeks in Metro-Mix 300 (Sun Gro Horticulture, USA) substrate supplemented with a controlled-release fertilizer (Osmocote Smart-release, The Scotts Miracle-Gro Company, USA) at 26 °C in a 16h light:8h dark photoperiod. For the experiments involving aseptic culture of plants, seeds were surface sterilized by incubation in a solution of 1% (v/v) NaClO, and cultured *in vitro* on 1X GM agar plates [4.3 g/L Murashige and Skoog salts with Gamborg's B5 vitamins (Phytotechnology Laboratories, USA), 1% w/v sucrose and 0.8% (w/v) phytoagar] at 22 °C in a 16:8h light:dark photoperiod. The *A. thaliana* T-DNA insertion lines *emb1241-1* (CS16149), *emb1241-2* (CS24098), and *Dcphsc70-1* (CS860808) were obtained from the Arabidopsis Biological Resource Center (www.arabidopsis.org). Populations of 30 adult plants (30 days old) of the mutant lines, were PCR-screened (Table S1) and only T-DNA-carrying plants were used for further analysis.

4.2. In silico analysis

Chloroplast GrpE protein sequences were identified and retrieved using the UniProtKB (http://www.uniprot.org/help/uniprotkb) BLAST tool, using the *E. coli* GrpE protein sequence as a query against the Plants database. In total, 136 CGE protein sequences were sorted by the presence of putative N-terminal chloroplast transit peptide. Transit peptide prediction was performed using TargetP 1.1 (Emanuelsson et al., 2000), and the predicted transit peptides were manually removed from the protein sequences. The analysis involved 137 amino acid sequences (136 CGEs and the GrpE from *E. coli*) that were aligned using Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) (CGEs full alignment file), the alignment was processed using MEGA 7.0.18 tools (Sudhir et al., 2016). The evolutionary history was inferred using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992), a discrete Gamma distribution was used to model evolutionary rate differences among sites [5 categories (+G, parameter = 1.3736)]. Less than 30% alignment gaps, missing data, and ambiguous bases were allowed at any position. The phylogenetic tree was drawn using FigTree v1.4 (http://tree.bio.ed.ac.uk/software/figtree/). CGEs were sorted into Type A (48 sequences) and Type B (78 sequences) based on the information from the phylogenetic tree. Motif discovery was performed using the DREME tool from the MEME-suite (Bailey, 2011) fed with Type A or Type B sequences.

4.3. Gene cloning and plasmid construction

The *A. thaliana CGE1* (At5g17710), *CGE2* (At1g36390) and *CPHSC70-1* (At4g24280), and the bacterial *EcgrpE* (NC_000913.3) coding regions (CDS) were amplified by PCR using specific oligonucleotides to obtain the full CDS or to delete the transit peptides (Table S1). Entry vectors for Gateway cloning were generated by either recombination into pDONR™/Zeo donor vector or by directional cloning into pENTR/D-TOPO® entry vector. The expression plasmids (Fig. S2) were generated by LR clonase II-mediated (Invitrogen, USA) recombination of the entry vectors containing the coding regions of interest and destination vectors pDEST14 (*pD14*), pDEST22 (*pD22*), pDEST32 (*pD32*), pEarleyGate103 (*pEG*), pSPYCE (*pCE*), and pSPYNE (*pNE*). The expression plasmids constructed using *pEG*, *pCE*, and *pNE* were transferred to *Agrobacterium tumefaciens* C58C1 cells, individual clones were used for plant agroinfiltration.

4.4. E. coli complementation assay

The *E. coli* strain OD212 (*dnaK332 ΔgrpE::Ω-cam*^{R}) (Deloche et al., 1997) was transformed with *pD14::ΔCGE1, pD14::ΔCGE2, pD14::EcgrpE* and *pD14::ΔccdB* plasmids (Fig. S2), the cells were cultured overnight at 25 °C in solid lysogeny broth (LB) media. Individual colonies were cultured overnight in liquid LB medium at 25 °C, 100 µL of these cultures were used to inoculate fresh medium and allowed to grow to an OD_{600nm} of 0.1. Serial dilutions were prepared and 5 µL drops were plated on solid LB medium. The inoculated plates were incubated for 20 h at either 25 °C, 37 °C or 43 °C. All the LB media was supplemented with ampicillin to a final concentration of 100 µg/mL. The complementation experiments were replicated at least three times.

4.5. Subcellular and suborganellar localization, and Bimolecular Fluorescence Complementation (BiFC)

The subcellular localization and the BiFC experiments were performed using agroinfiltrated *Nicotiana benthamiana* leaves. Subcellular localization was determined using translational fusions between the proteins of interest and GFP, while BiFC experiments were carried out using the split YFP system (Walter et al., 2004). The agroinfiltration of *N. benthamiana* leaves was performed using *A. tumefaciens* clones individually containing the plasmids *pBN::GFP* (Voinnet et al., 2000), *pEG::CGE1*, *pEG::CGE2*, or *pEG::CPHSC70-1* for the subcellular localization experiments, and the corresponding plasmid combinations *pCE::CGE1, pCE::CGE2, pCE::CPHSC70-1*, *pNE::CGE1*, *pNE::CGE2,* and *pNE::CPHSC70-1* for BiFC assays. All the agroinfiltration experiments included the helper *A. tumefaciens* p19 strain (Leuzinger et al., 2013). The bacterial strains were cultured at 30 °C overnight, harvested by centrifugation (1, 400 × *g*, 10 min), and resuspended in a solution of 10 mM $MgCl_2$. The bacterial cultures were diluted to a final OD_{600nm} of 1 in 10 mM $MgCl_2$. The samples were incubated at room temperature for 3 h in the presence of 5 µg/mL acetosyringone prior to plant inoculation. Leaves of 6 weeks-old *N. benthamiana* plants were infiltrated with the bacterial samples and cultured at 26 °C in a 16h light:8h dark cycle for 96 h prior to analysis. Protoplasts from the infiltrated leaves were obtained by enzymatic digestion of the cell wall. Leaf strips were incubated in enzymatic solution (0.5 M mannitol, 1% cellulase R-10, and 0.05% macerozyme R-10) for 3 h in the dark, at room temperature, and mild shaking (~30 rpm). After digestion, leaf debris was removed and protoplasts were analyzed. Chloroplasts and suborganellar fractions from infiltrated leaves were isolated according to the method described by Salvi et al., (2011). All the experiments described in this section were performed by triplicate.

4.6. Confocal microscopy

Confocal microscopy images of *N. benthamiana* leaf protoplast were obtained with an Olympus FV1000 microscope (Olympus, USA) using excitation lasers of 488 nm for GFP and 515 nm for YFP, chlorophyll fluorescence was captured using a 515 nm laser and a barrier filter BA655-755. Z projections were rendered using the Fiji software maximum intensity projection type (Schindelin et al., 2012).

4.7. Plant heat stress experiments and northern blot

A. thaliana (Col-0) seedlings were cultured for 8, 10, 12 and 14 days under normal growth as described before. Fourteen days old *A. thaliana* (Col-0) seedlings were exposed to heat stress by incubation at 40 °C for 30, 60 or 90 min. Total RNA was prepared from the seedlings using TRIzol (Ambion, Life Technologies, USA) reagent, following the manufacturer's protocol. For northern blot analyses, 10 μg of total RNA were fractionated in 1.5% (w/v) agarose gels under denaturating conditions [2% (v/v) formaldehyde] and transferred to Hybond-N^+ nylon membrane (GE Healthcare Bio-Sciences, UK). Specific DNA probes for each of the genes were isolated from the CDSs and encompass the regions shown in Fig. S3. The specific DNA probes were radiolabeled with α-dCT^{32}P using the Megaprime DNA labeling system (GE Healthcare Bio-Sciences, UK), following the protocol provided by the manufacturer. Hybridization and washing of the RNA-containing membranes were performed under stringent conditions (55 °C and 0.0825 M Na^+). All the experiments described in this section were performed by triplicate.

4.8. Pigment quantitation

Photosynthetic pigments (chlorophyll *a*, chlorophyll *b*, and carotenes) of adult plants were extracted by overnight incubation of leaves in 80% acetone solution. The absorbance of the samples was measured at 663 nm, 646 nm, and 470 nm. Using the absorbance data, the concentration of pigments was calculated with the following equations (Lichtenthaler and Wellburn, 1983):

Chlorophyll a (Ca) = $(12.21*A_{663})-(2.81*A_{646})$

Chlorophyll b (Cb) = $(20.13*A_{646})-(5.03*A_{663})$

Carotenes = $((1000*A_{470})-(3.27*Ca)-(104*Cb))/229$

The statistical significance of the data was determined using a One-Way ANOVA calculator (http://www.socscistatistics.com/tests/anova/Default2.aspx), with a significance level of $p < 0.05$.

4.9. Protein analyses

The protein isolated from suborganellar fractions was precipitated by addition of Tris-EDTA [100 mM Tris, 10 mM EDTA, pH 8], 0.3% sodium deoxycholate and 72% Trichloroacetic acid (TCA) in a reagent to sample ratio of 1:5 (v/v), the samples were incubated on ice for 1 h. The protein was collected by centrifugation at 25, 000 × *g*, resuspended in 90% acetone and incubated overnight at 4 °C. The samples were centrifuged at 25, 000 × *g*, and the pellets were allowed to air-dry before resuspension in a minimal volume of 2X Laemmli buffer [65 mM Tris-HCl pH 6.8, 30% (v/v) glycerol, 2% (w/v) SDS, 0.01% (w/v) bromophenol blue, 355 mM 2-mercaptoethanol] (Laemmli, 1970). The protein samples (10–20 μL) were fractionated in denaturating 12% polyacrylamide gels, transferred to nitrocellulose blotting membrane (GE Healthcare Bio-Sciences, UK), and incubated in PBS-T buffer [137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 0.1% (v/v) Triton X-100] supplemented with 0.5% (w/v) nonfat dry milk. For Western blot analysis, monoclonal primary antibodies αc-Myc (Sigma-Aldrich, Mexico) and αHA (Santa Cruz Biotechnology Inc, USA), and secondary antibodies αMouse-HRP (Thermo Fisher Scientific, USA) and αRabbit-HRP (Thermo Fisher Scientific, USA) were used. Detection of the recombinant proteins was performed using the Amersham ECL Prime Western blotting detection reagent kit (GE Healthcare Bio-Sciences, UK), following the manufacturer's instructions.

For the BN-PAGE analysis, total protein from wild-type, *emb1241-2*, and *DcpHsc70-1* mutants was extracted by incubation of frozen-pulverized leaf tissue with BN extraction buffer [70 mM Tris-HCl pH 7.5, 1 mM $MgCl_2$, 25 mM KCl, 5 mM EDTA pH 8, 0.25 mM Sucrose, 39.1 mM n-Dodecyl β-D-maltoside], supplemented with the cOmplete™ protease inhibitor cocktail as indicated by the manufacturer (Sigma-Aldrich, Mexico). Protein samples were subjected to BN-PAGE in a 4%–14% acrylamide gradient, according to the protocol of Heinemeyer et al., (2007). Additionally, the BN-PAGE gels were destained by 10 consecutive washes of boiling-hot distilled water. All the experiments described in this section were performed at least three times.

Contribution

L. A. de Luna-Valdez: performed experiments, analyzed the results and wrote the manuscript.

C. I. Villaseñor-Salmerón: performed experiments.

A. A. Guevara-García: designed and supervised experiments, edited the manuscript and funded the research.

P. León-Mejía: edited the manuscript funded the research.

E. Cordoba: designed experiments and edited the manuscript.

R. Vera-Estrella: designed experiments and edited the manuscript.

Acknowledgements and Funding

The authors thank M.C. Andrés Saralegui and Dr. Arturo Pimentel from the Laboratorio Nacional de Microscopía Avanzada- Universidad Nacional Autónoma de México, for confocal microscopy services. Also, we thank Dr. Michael Schroda for the donation of the OD212 *E. coli* mutant strain. This work was supported by the PAPIIT-DGAPA-UNAM [grant numbers IN207214, IN210917 to AG and IN204617 to PL], also by the Consejo Nacional de Ciencia y Tecnología-Mexico [grant numbers CB2015-251848 to AG, CB2015-220534 and FC2016-96 to PL, and the student scholarship 240088 to LdL].

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.plaphy.2019.03.027.

References

Ang, D., Chandrasekhar, G.N., Zylicz, M., Georgopoulos, C., 1986. Escherichia coli grpE gene codes for heat shock protein B25.3, essential for both lambda DNA replication at all temperatures and host growth at high temperature. J. Bacteriol. 167, 25–29.

Apuya, N.R., Yadegari, R., Fischer, R.L., Harada, J.J., Zimmerman, J.L., Goldberg, R.B., 2001. The Arabidopsis embryo mutant schlepperless has a defect in the chaperonin-60alpha gene. Plant Physiol. 126, 717–730.

Azem, A., Oppliger, W., Lustig, A., Jenö, P., Feifel, B., Schatz, G., Horst, M., 1997. The mitochondrial hsp70 chaperone system: effect of adenine nucleotides, peptide substrate, and mGrpE on the oligomeric state of mhsp70. J. Biol. Chem. 272, 20901–20906.

Bailey, T.L., 2011. DREME: motif discovery in transcription factor ChIP-seq data. Bioinformatics 27, 1653–1659.

Barraclough, R., Ellis, R.J., 1980. Protein synthesis in chloroplasts IX. Assembly of newly-synthesized large subunits into ribulose bisphosphate Carboxylase in isolated intact pea chloroplasts. Biochim. Biophys. Acta 608, 19–31.

Bédard, J., Kubis, S., Bimanadham, S., Jarvis, P., 2007. Functional similarity between the chloroplast translocon component, Tic40, and the human Co-chaperone, hsp70-interacting protein (Hip). J. Biol. Chem. 282, 21404–21414.

Breuers, F.K.H., Bräutigam, A., Geimer, S., Welzel, U.Y., Stefano, G., Renna, L., Brandizzi, F., Weber, A.P.M., 2012. Dynamic remodeling of the plastid envelope membranes a tool for chloroplast envelope in vivo localizations. Front. Plant Sci. 3, 7.

Cao, D., Froehlich, J.E., Zhang, H., Cheng, C.L., 2003. The chlorate-resistant and photomorphogenesis-defective mutant cr88 encodes a chloroplast-targeted HSP90. Plant J. 33, 107–118.

Chen, K.M., Holmström, M., Raksajit, W., Suorsa, M., Piippo, M., Aro, E.M., 2010. Small chloroplast-targeted DnaJ proteins are involved in optimization of photosynthetic reactions in Arabidopsis thaliana. BMC Plant Biol. 10, 43.

Chiu, C.C., Chen, L.J., Su, P.H., Li, Hm, 2013. Evolution of chloroplast J proteins. PLoS One. https://doi.org/10.1371/journal.pone.0070384.

Chou, M.L., Fitzpatrick, L.M., Tu, S.L., Budziszewski, G., Potter-Lewis, S., Akita, M., Levin, J.Z., Keegstra, K., Li, H.M., 2003. Tic40 a membrane-anchored Co-chaperone homolog in the chloroplast protein translocon. EMBO J. 22, 2970–2980.

Clarke, A.K., 2012. The chloroplast ATP-dependent clp protease in vascular plants - new dimensions and future challenges. Physiol. Plantarum 145, 235–244.

Constan, D., Froehlich, J.E., Rangarajan, S., Keegstra, K., 2004. A stromal Hsp100 protein is required for normal chloroplast development and function in Arabidopsis. Plant Physiol. 136, 3605–3615.

Deloche, O., Georgopoulos, C., 1996. Purification and biochemical properties of *Saccharomyces cerevisiae*'s Mge1p, the mitochondrial cochaperones of Ssc1p. J. Biol. Chem. 271, 23960–23966.
Deloche, O., Kelley, W.L., Georgopoulos, C., 1997. Structure-function analyses of the Ssc1p, Mdj1p, and Mge1p Saccharomyces cerevisiae mitochondrial proteins in Escherichia coli. J. Bacteriol. 179, 6066–6075.
Doyle, S.M., Wickner, S., 2009. Hsp104 and ClpB: protein disaggregating machines. Trends Biochem. Sci. 34, 40–48.
Echevarría-Zomeño, S., Fernández-Calvino, L., Castro-Sanz, A.B., López, J.A., Vázquez, J., Castellano, M.M., 2016. Dissecting the proteome dynamics of the early heat stress response leading to plant survival or death in Arabidopsis. Plant Cell Environ. 39, 1264–1278.
Emanuelsson, O., Nielsen, H., Brunak, S., von Heijne, G., 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. J. Mol. Biol. 300, 1005–1016.
Farmaki, T., Sanmartin, M., Jimenez, P., Paneque, M., Sanz, C., Vancanneyt, G., Leon, J., Sanchez-Serrano, J.J., 2006. Differential distribution of the lipoxygenase pathway enzymes within potato chloroplasts. J. Exp. Bot. 58, 555–568.
Fink, A.L., 1999. Chaperone-mediated protein folding. Physiol. Rev. 79, 425–449.
Flores-Pérez, U., Jarvis, P., 2013. Molecular chaperone involvement in chloroplast protein import. Biochim. Biophys. Acta 1833, 332–340.
Gelinas, A.D., Toth, J., Bethoney, K.A., Langsetmo, K., Stafford, W.F., Harrison, C.J., 2003. Thermodynamic linkage in the GrpE nucleotide exchange factor, a molecular thermosensor. Biochemistry 42, 9050–9059.
Gelinas, A.D., Toth, J., Bethoney, K.A., Stafford, W.F., Harrison, C.J., 2004. Mutational analysis of the energetics of the GrpE.DnaK binding interface: equilibrium association constants by sedimentation velocity analytical ultracentrifugation. J. Mol. Biol. 33, 447–458.
Green, B.R., Durnford, D.G., 1996. The chlorophyll-carotenoid proteins of oxygenic photosynthesis. Annu. Rev. Plant Physiol. Plant Mol. Biol. 47, 685–714.
Harrison, C.J., Hayer-Hartl, M., Di, L.M., Hartl, F., Kuriyan, J., 1997. Crystal structure of the nucleotide exchange factor GrpE bound to the ATPase domain of the molecular chaperone DnaK. Science 276, 431–435.
Heide, H., Nordhues, A., Drepper, F., Nick, S., Schulz-Raffelt, M., Haehnel, W., Schroda, M., 2009. Application of quantitative immunoprecipitation combined with knock-down and cross-linking to Chlamydomonas reveals the presence of vesicle-inducing protein in plastids 1 in a common complex with chloroplast HSP90C. Proteomics 9, 3079–3089.
Heinemeyer, J., Lewejohann, D., Braun, H.P., 2007. Blue-native gel electrophoresis for the characterization of protein complexes in plants. Methods Mol. Biol. 355, 343–352.
Ikeda, E., Yoshida, S., Mitsuzawa, H., Uno, I., Toh-e, 1994. A. YGE1 is a yeast homologue of Escherichia coli GrpE and is required for maintenance of mitochondrial functions. FEBS (Fed. Eur. Biochem. Soc.) Lett. 339, 265–268.
Inoue, H., Li, M., Schnell, D.J., 2013. An essential role for chloroplast heat shock protein 90 (Hsp90C) in protein import into chloroplasts. Proc. Natl. Acad. Sci. Unit. States Am. 110, 3173–3178.
Jarvis, P., Kessler, F., 2014. Mechanisms of chloroplast protein import in plants. In: In: Theg, S., Wollman, F.A. (Eds.), Plastid Biology. Advances in Plant Biology, vol. 5. Springer, New York, pp. 241–270.
Jiang, J., Liu, X., Liu, C., Liu, G., Li, S., Wang, L., 2017. Integrating omics and alternative splicing reveals insights into grape response to high temperature. Plant Physiol. 173, 1502–1518.
Jones, D.T., Taylor, W.R., Thornton, J.M., 1992. The rapid generation of mutation data matrices from protein sequences. Bioinformatics 8, 275–282.
Kitajima, K., Hogan, K.P., 2003. Increases of chlorophyll a/b ratios during acclimation of tropical woody seedlings to nitrogen limitation and high light. Plant Cell Environ. 26, 857–865.
Kong, F., Deng, Y., Zhou, B., Wang, G., Wang, Y., Meng, Q., 2014. A chloroplast-targeted DnaJ protein contributes to maintenance of photosystem II under chilling stress. J. Exp. Bot. 65, 143–158.
Kovacheva, S., Bédard, J., Patel, R., Dudley, P., Twell, D., Ríos, G., Koncz, C., Jarvis, P., 2005. In vivo studies on the roles of Tic110, Tic40 and Hsp93 during chloroplast protein import. Plant J. 41, 412–428.
Kovacheva, S., Bédard, J., Wardle, A., Patel, R., Jarvis, P., 2007. Further in vivo studies on the role of the molecular chaperone Hsp93, in plastid protein import. Plant J. 50, 364–379.
Laemmli, U.K., 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature 227, 680–685.
Leuzinger, K., Dent, M., Hurtado, J., Stahnke, J., Lai, H., Zhou, X., Chen, Q., 2013. Efficient agroinfiltration of plants for high-level transient expression of recombinant proteins. JoVE 77, 50521.
Lichtenthaler, H.K., Wellburn, A.R., 1983. Determinations of total carotenoids and chlorophylls a and b of leaf extracts in different solvents. Biochem. Soc. Trans. 11, 591–592.
Liu, C., Willmund, F., Whitelegge, J.P., Hawat, S., Knapp, B., Lodha, M., Schroda, M., 2005. J-domain protein CDJ2 and HSP70B are a plastidic chaperone pair that interacts with vesicle-inducing protein in plastids 1. Mol. Biol. Cell 16, 1165–1177.
Liu, C., Willmund, F., Golecki, J.R., Cacace, S., Heß, B., Markert, C., Schroda, M., 2007. The chloroplast HSP70B-CDJ2-CGE1 chaperones catalyse assembly and disassembly of VIPP1 oligomers in Chlamydomonas. Plant J. 50, 265–277.
Martin, W., Stoebe, B., Goremykin, V., Hapsmann, S., Hasegawa, M., Kowallik, K.V., 1998. Gene transfer to the nucleus and the evolution of chloroplasts. Nature 393, 162–165.
Mascia, F., Girolomoni, L., Alcocer, M.J.P., Bargigia, I., Perozeni, F., Cazzaniga, S., Cerullo, G., D'Andrea, C., Ballottar, M., 2017. Functional analysis of photosynthetic pigment binding complexes in the green alga *Haematococcus pluvialis* reveals distribution of astaxanthin in Photosystems. Sci. Rep. 7, 16319.
May, T., Soll, J., 2000. 14-3-3 proteins form a guidance complex with chloroplast precursor proteins in plants. Plant Cell 12, 53–63.
Mayer, M.P., 2010. Gymnastics of molecular chaperones. Mol. Cell 39, 321–331.
Meinke, D., Muralla, R., Sweeney, C., Dickerman, A., 2008. Identifying essential genes in Arabidopsis thaliana. Trends Plant Sci. 13, 483–491.
Naylor, D.J., Hoogenraad, N.J., Høj, P.B., 1996. Isolation and characterisation of a CDNA encoding rat mitochondrial GrpE a stress-inducible nucleotide-exchange factor of ubiquitous appearance in mammalian organs. FEBS (Fed. Eur. Biochem. Soc.) Lett. 396, 181–188.
Nelson, B.K., Cai, X., Nebenführ, A., 2007. A multicolored set of in vivo organelle markers for Co-localization studies in Arabidopsis and other plants. Plant J. 51, 1126–1136.
Olinares, P.D., Kim, J., van Wijk, K.J., 2011. The clp protease system; a central component of the chloroplast protease network. Biochim. Biophys. Acta 1807, 999–1011.
Paila, Y.D., Richardson, L.G.L., Schnell, D.J., 2015. New insights into the mechanism of chloroplast protein import and its integration with protein quality control organelle biogenesis and development. J. Mol. Biol. 427, 1038–1060.
Peng, L., Fukao, Y., Myouga, F., Motohashi, R., Shinozaki, K., Shikanai, T., 2011. A chaperonin subunit with unique structures is essential for folding of a specific substrate. PLoS Biol. https://doi.org/10.1371/journal.pbio.1001040.
Perello, C., Llamas, E., Burlat, V., Ortiz-Alcaide, M., Phillips, M.A., Pulido, P., Rodriguez-Concepcion, M., 2016. Differential subplastidial localization and turnover of enzymes involved in isoprenoid biosynthesis in chloroplasts. PLoS One. https://doi.org/10.1371/journal.pone.0150539.
Pratt, W.B., Toft, D.O., 2003. Regulation of signaling protein function and trafficking by the hsp90/hsp70-Based chaperone machinery. Exp. Biol. Med. 228, 111–133.
Pulido, P., Toledo-Ortiz, G., Phillips, M.A., Wright, L.P., Rodriguez-Concepcion, M., 2013. Arabidopsis J-protein J20 delivers the first enzyme of the plastidial isoprenoid pathway to protein quality control. Plant Cell 25, 4183–4194.
Qbadou, S., Becker, T., Mirus, O., Tews, I., Soll, J., Schleiff, E., 2006. The molecular chaperone Hsp90 delivers precursor proteins to the chloroplast import receptor Toc64. EMBO J. 25, 1836–1847.
Rial, D.V., Arakaki, A.K., Ceccarelli, E.A., 2000. Interaction of the targeting sequence of chloroplast precursors with Hsp70 molecular chaperones. Eur. J. Biochem. 267, 6239–6248.
Rizhsky, L., Liang, H., Mittler, R., 2002. The combined effect of drought stress and heat shock on gene expression in tobacco. Plant Physiol. 130, 1143–1151.
Salvi, D., Moyet, L., Seigneurin-Berny, D., Ferro, M., Joyard, J., Rolland, N., 2011. Preparation of envelope membrane fractions from Arabidopsis chloroplasts for proteomic analysis and other studies. In: Jarvis, R. (Ed.), Chloroplast Research in Arabidopsis. Humana Press, pp. 189–206.
Schindelin, J., Arganda-Carreras, I., Frise, E., et al., 2012. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682.
Schroda, M., Vallon, O., Wollman, F.A., Beck, C.F., 1999. A chloroplast-targeted heat shock protein 70 (HSP70) contributes to the photoprotection and repair of photosystem II during and after photoinhibition. Plant Cell 11, 1165–1178.
Schroda, M., Vallon, O., Whitelegge, J.P., Beck, C.F., Wollman, F.A., 2001. The chloroplastic GrpE homolog of Chlamydomonas: two isoforms generated by differential splicing. Plant Cell 13, 2823–2839.
Shi, L.X., Theg, S.M., 2010. A stromal heat shock protein 70 system functions in protein import into chloroplasts in the moss Physcomitrella patens. Plant Cell 22, 205–220.
Sjogren, L.L.E., Tanabe, N., Lymperopoulos, P., Khan, N.Z., Rodermel, S.R., Aronsson, H., Clarke, A.K., 2014. Quantitative analysis of the chloroplast molecular chaperone ClpC/hsp93 in Arabidopsis reveals new insights into its localization interaction with the clp proteolytic core, and functional importance. J. Biol. Chem. 289, 11318–11330.
Stahl, T., Glockmann, C., Soll, J., Heins, L., 1999. Tic40, a new old subunit of the chloroplast protein import translocon. J. Biol. Chem. 274, 37467–37472.
Su, P.H., Li, H.M., 2008. Arabidopsis stromal 70-KD heat shock proteins are essential for plant development and important for thermotolerance of germinating seeds. Plant Physiol. 146, 1231–1241.
Su, P.H., Li, H.M., 2010. Stromal Hsp70 is important for protein translocation into pea and Arabidopsis chloroplasts. Plant Cell 22, 1516–1531.
Sudhir, K., Stecher, G., Tamura, K., 2016. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. Mol. Biol. Evol. 33, 1870–1874.
Sung, D.Y., Vierling, E., Guy, C.L., 2001. Comprehensive expression profile Analysis of the Arabidopsis Hsp70 gene family. Plant Physiol. 126, 789–800.
Suzuki, N., Miller, G., Salazar, C., et al., 2013. Temporal-spatial interaction between reactive oxygen species and abscisic acid regulates rapid systemic acclimation in plants. Plant Cell 25, 353–3569.
Trösch, R., Mühlhaus, T., Schroda, M., Willmund, F., 2015. ATP-dependent molecular chaperones in plastids–more complex than expected. Biochim. Biophys. Acta 1847, 872–888.
Tsai, Y.C., Mueller-Cajar, O., Saschenbrecker, S., Hartl, F.U., Hayer-Hartl, M., 2012. Chaperonin cofactors, Cpn10 and Cpn20, of green algae and plants function as hetero-oligomeric ring complexes. J. Biol. Chem. 287, 20471–20481.
van Wijk, K.J., Baginsky, S., 2011. Plastid proteomics in higher plants: current state and future goals. Plant Physiol. 155, 1578–1588.
Villarejo, A., Burén, S., Larsson, S., et al., 2005. Evidence for a protein transported through the secretory pathway en route to the higher plant chloroplast. Nat. Cell Biol. 7, 1224–1231.
Voinnet, O., Lederer, C., Baulcombe, D.C., 2000. A viral movement protein prevents spread of the gene silencing signal in *Nicotiana benthamiana*. Cell 103, 157–167.
Walter, M., Chaban, C., SchützenK, et al., 2004. Visualization of protein interactions in living plant cells using bimolecular fluorescence complementation. Plant J. 40,

428–438.
Willmund, F., Muhlhaus, T., Wojciechowska, M., Schroda, M., 2007. The NH2-terminal domain of the chloroplast GrpE homolog CGE1 is required for dimerization and co-chaperone function in vivo. J. Biol. Chem. 282, 11317–11328.
Wu, B., Ang, D., Snavely, M., Georgopoulos, C., 1996. Structure-function analysis of the *Escherichia coli* GrpE heat shock protein. EMBO J. 15, 4806–4816.
Yalovsky, S., Paulsen, H., Michaeli, D., Chitnis, P.R., Nechushtai, R., 1992. Involvement of a chloroplast HSP70 heat shock protein in the integration of a protein (Light-Harvesting complex protein precursor) into the thylakoid membrane. Proceedings of the National Academy of Sciences USA 89, 5616–5619.
Yura, T., Nagai, H., Mori, H., 1993. Regulation of the heat-shock response in bacteria. Annu. Rev. Microbiol. 47, 321–350.